



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Davi Duarte de Paula

Reconstrução de Tomossíntese Mamária Utilizando Redes Neurais com Aprendizado Profundo

São José do Rio Preto

2020

Davi Duarte de Paula

**Reconstrução de Tomossíntese Mamária Utilizando Redes
Neurais com Aprendizado Profundo**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientador: Prof. Dr. Denis Henrique Pinheiro Salvadeo

São José do Rio Preto
2020

P324r

Paula, Davi Duarte de

Reconstrução de Tomossíntese Mamária Utilizando Redes Neurais com Aprendizado Profundo / Davi Duarte de Paula. -- São José do Rio Preto, 2020

130 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Denis Henrique Pinheiro Salvadeo

1. Tomossíntese Mamária. 2. Aprendizado Profundo. 3. Reconstrução Tomográfica. 4. Redução de Ruído. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Davi Duarte de Paula

**Reconstrução de Tomossíntese Mamária Utilizando Redes Neurais
com Aprendizado Profundo**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Prof. Dr. Denis Henrique Pinheiro Salvadeo
UNESP – Câmpus de Rio Claro - Orientador

Prof. Dr. Fabricio Aparecido Breve
UNESP – Câmpus de Rio Claro

Prof. Dr. José Hiroki Saito
UFSCAR

Rio Claro
28 de fevereiro de 2020

Agradecimentos

Primeiramente, à Deus, pelas inúmeras bênçãos.

Aos meus pais, pelo apoio incondicional, regado a sabedoria e inteligência.

Ao meu orientador, pelo esforço, tempo e inteligência despendidos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, à qual agradeço.

Resumo

Tomossíntese Mamária Digital (DBT) é uma técnica de imageamento radiográfico, com aquisição de projeções em ângulos limitados utilizando dose reduzida de radiação. Ela tem por objetivo reconstruir fatias tomográficas do interior da mama, possibilitando o diagnóstico precoce de possíveis lesões e aumentando, conseqüentemente, a probabilidade de cura do paciente. Contudo, devido ao fato de que DBT utiliza doses baixas de radiação, a imagem gerada contém mais ruído que a mamografia digital. Embora a qualidade do exame esteja diretamente relacionada com a dose utilizada, espera-se que a dose de radiação empregada no exame seja a mais baixa possível, mas ainda com qualidade suficiente para que o diagnóstico possa ser realizado, conforme o princípio *As Low As Reasonably Achievable* (ALARA). Uma das etapas importantes para se buscar o princípio ALARA é a reconstrução tomográfica, que consiste em um software que gera as fatias do interior da mama a partir de um conjunto de projeções 2D de DBT adquiridas. Por outro lado, técnicas de Aprendizado de Máquina, especialmente redes neurais com aprendizado profundo, que recentemente tem evoluído consideravelmente o estado da arte em diversos problemas de Visão Computacional e Processamento de Imagens, tem características adequadas para serem aplicadas também na etapa de reconstrução. Deste modo, este trabalho investigou uma arquitetura básica de rede neural artificial com aprendizado profundo que seja capaz de reconstruir imagens de DBT, especialmente focada na redução de ruído. Ainda, considerando uma etapa adicional de filtragem usando o método *Denoising Convolutional Neural Networks* (DnCNN) após a reconstrução, o método proposto foi superior a uma abordagem em estado da arte de reconstrução de DBT em termos de SSIM (*Structural Similarity Index Measure*) e PSNR (*Peak Signal-to-Noise Ratio*) usando dados de corpos de prova virtuais de DBT, atingindo valores de 37,644 db em PSNR e 0,869 em SSIM. Por fim, ressalta-se que esta implementação de arquitetura básica para reconstrução de DBT é relevante para o desenvolvimento futuro de redes neurais artificiais únicas que realizam o processo completo de reconstrução de DBT (envolvendo filtragem e retroprojeção).

Palavras-chave: Tomossíntese Mamária, Aprendizado Profundo, Reconstrução Tomográfica, Aprendizado de Máquina, Redes Neurais Artificiais, Redução de Ruído

Abstract

Digital Breast Tomosynthesis (DBT) is a technique of radiographic imaging, with acquisition of projections at limited angles using reduced dose of radiation. It aims to reconstruct tomographic slices inside the breast, making possible the early diagnosis of possible lesions and, consequently, increasing the probability of cure of the patient. However, due to the fact that DBT uses low doses of radiation, the generated image contains more noise than digital mammography. Although the quality of the exam is directly related to the dose applied, the radiation dose used in the examination is expected to be as low as possible, but still keeping enough quality for the diagnosis to be made, as determined by the As Low As Reasonably Achievable (ALARA) principle. One of the important steps to achieve the ALARA principle is the tomographic reconstruction, which consists of a software that generates slices inside the breast from an acquired set of 2D DBT projections. On the other hand, Machine Learning techniques, especially neural networks with deep learning, that have recently evolved considerably the state-of-the-art in several problems in Computer Vision and Image Processing areas, it has suitable characteristics to be applied also in the reconstruction step. Thus, this work investigated a basic architecture of artificial neural network with deep learning that is capable to reconstruct DBT images, especially focused on noise reduction. Furthermore, considering an additional filtering step using the Denoising Convolutional Neural Networks (DnCNN) method after reconstruction, the proposed method was superior to a state-of-the-art approach to reconstruct DBT in terms of SSIM (Structural Similarity Index Measure) and PSNR (Peak Signal-to-Noise Ratio) using data from DBT phantoms, achieving values of 37.644 dB in PSNR and 0.869 in SSIM. Finally, it is emphasized that this implementation of a basic architecture for DBT reconstruction is relevant for the future development of a unique artificial neural networks that realizes the complete process of DBT reconstruction (involving filtering and backprojection).

Keywords: Digital Breast Tomosynthesis, Deep Learning, Tomographic Reconstruction, Machine Learning, Artificial Neural Networks, Noise Reduction

Lista de Figuras

Figura 1 – Incidência e mortalidade de vários tipos de cânceres sobre a população de ambos os sexos (A), do sexo feminino (B) e do sexo masculino (C). Os valores estão expressos em termos da porcentagem em relação ao total de casos de câncer registrados.	13
Figura 2 – Corpo cúbico sendo imageado por raios X. Ênfase deve ser dada na questão de que a seção transversal PSRQ é imageada por raios mutuamente paralelos, formando uma projeção 1D no filme receptor. O conjunto de todos os raios emitidos sobre o corpo irá formar uma projeção 2D do corpo 3D.	18
Figura 3 – Características principais de um aparelho de mamografia.	21
Figura 4 – Característica do espalhamento de raios, que causa ruído nas imagens geradas.	22
Figura 5 – Vistas usadas em exames cotidianos.	22
Figura 6 – Ilustração do processo de tomografia.	25
Figura 7 – Planos e eixos anatômicos.	26
Figura 8 – Exemplo do princípio de reconstrução tomográfica.	27
Figura 9 – Ilustração do processo de aquisição de DBT.	29
Figura 10 – Geometria dos equipamentos de tomossíntese mamária.	30
Figura 11 – Modelo do Perceptron.	38
Figura 12 – Função de ativação Limiar com $\alpha = 0$	40
Figura 13 – Função de ativação Linear com $\beta = 1$	41
Figura 14 – Função de ativação Sigmóid.	42
Figura 15 – Função de ativação Tanh.	43
Figura 16 – Função de ativação ReLu.	44
Figura 17 – Parte da curva da função de custo. Os vetores representam os gradientes para quatro diferentes valores de peso w_1 , w_2 e w_3 e w_{opt} , onde w_{opt} é o valor de mínimo da função (ponto ótimo). Baseado em Samarasinghe (2016).	46
Figura 18 – Representação da arquitetura LeNet-5.	49
Figura 19 – Representação da arquitetura AlexNet.	51
Figura 20 – Módulos Inception.	53
Figura 21 – Blocos residuais.	55

Figura 22 – Abordagens de reconstrução tomográfica.	61
Figura 23 – Processo de aquisição de imagem utilizando tomografia clássica.	62
Figura 24 – Ilustração da retroprojeção.	63
Figura 25 – Representação da transformada de Radon de uma função $f(x, y)$	65
Figura 26 – Representação do Teorema de Corte de Fourier.	66
Figura 27 – Representação do Método ART.	71
Figura 28 – Exemplificação do método de Kaczmarz.	74
Figura 29 – Rede neural que simula o algoritmo FDK.	81
Figura 30 – Rede Neural que simula os operadores Primal e Dual.	84
Figura 31 – Rede Neural aplicada no processo de reconstrução proposto por Chen et al. (2018a).	87
Figura 32 – Rede neural proposta para escolher os melhores parâmetros em abordagens de reconstrução iterativas.	89
Figura 33 – Diagrama do Gradiente Descendente Projetado. Uma CNN é utilizada para fazer o papel do projetor.	91
Figura 34 – Rede neural recorrente residual que representa a Equação 4.47.	93
Figura 35 – Arquitetura de reconstrução de DBT utilizando os pesos de Parker.	103
Figura 36 – Arquitetura de reconstrução de DBT sem a utilização dos pesos de Parker.	104
Figura 37 – Metodologia de avaliação das arquiteturas de reconstrução propostas.	106
Figura 38 – RNA para reconstrução de DBT seguida de um método de filtragem.	107
Figura 39 – Procedimento de filtragem das projeções pré-reconstrução usado nos métodos de comparação (filtros NLM ou BM3D para filtrar projeções), com a reconstrução por FBP.	107
Figura 40 – Valores de perda no conjunto de treinamento e validação. Linha azul é o erro no conjunto de treinamento, enquanto a linha laranja é o erro no conjunto de validação.	111
Figura 41 – Reconstruções de uma amostra do conjunto de teste.	113
Figura 42 – Valores de perda no conjunto de treinamento e validação. Gráfico azul é o erro no conjunto de treinamento, enquanto o gráfico laranja é o erro no conjunto de validação.	114
Figura 43 – Comparação de um corpo de prova presente no conjunto de teste.	115
Figura 44 – Comparação dos resultados em relação a uma etapa adicional para filtragem pre- ou pós-reconstrução.	116

Lista de Tabelas

Tabela 1 – Principais características de diversos equipamentos de DBT.	31
Tabela 2 – Modelos ganhadores do desafio ILSVRC.	48
Tabela 3 – Classificação e detalhes experimentais dos métodos discutidos nas seções anteriores.	99
Tabela 4 – Detalhes sobre a performance dos métodos discutidos nas seções anteriores.	99
Tabela 5 – Desempenho da Arquitetura Parker para a diminuição do ruído.	112
Tabela 6 – Resultados experimentais para os métodos de filtragem.	115

Lista de Abreviaturas

ANN	<i>Artificial Neural Network</i>
DNN	<i>Deep Neural Network</i>
CNN	<i>Convolutional Neural Network</i>
DBT	<i>Digital Breast Tomosynthesis</i>
MRI	<i>Magnetic Resonance Imaging</i>
ACRIN	<i>American College of Radiology Imaging Network</i>
DMIST	<i>Digital Mammographic Imaging Screening Trial</i>
BP	<i>Backprojection</i>
FBP	<i>Filtered Backprojection</i>
SAA	<i>Shift-And-Add</i>
ART	<i>Algebraic Reconstruction Technique</i>
SIRT	<i>Simultaneous Iterative Reconstruction Technique</i>
SART	<i>Simultaneous Algebraic Reconstruction Technique</i>
mAs	<i>Milliampere-seconds</i>
GPU	<i>Graphics Processing Unit</i>

Sumário

1	INTRODUÇÃO	12
2	MAMOGRAFIA E TOMOSSÍNTESE MAMÁRIA	17
2.1	Interação da Radiação com a Matéria	17
2.2	Mamografia	19
2.3	Tomografia	24
2.4	Tomossíntese Mamária	28
2.5	Ruído	32
2.5.1	Simulação	35
2.6	Considerações Finais do Capítulo	35
3	REDES NEURAIIS COM APRENDIZADO PROFUNDO	37
3.1	Funções de Ativação	39
3.2	Treinamento	44
3.2.1	Função de Custo	45
3.2.2	Backpropagation	45
3.3	Redes Neurais Convolucionais	47
3.3.1	LeNet-5 (1998)	48
3.3.2	AlexNet (2012)	50
3.3.3	ZFNet (2013)	50
3.3.4	VGG (2014)	51
3.3.5	Inception (2014)	52
3.3.6	ResNet (2015)	53
3.3.7	MobileNet (2017)	55
3.4	Frameworks	57
3.5	Considerações Finais do Capítulo	59
4	RECONSTRUÇÃO TOMOGRÁFICA DE DBT	60
4.1	Reconstrução Baseada em Retroprojeção	61
4.1.1	Shift-And-Add (SAA) e Backprojection (BP)	61

4.2	Reconstrução Baseada em Transformada de Fourier	64
4.2.1	Transformada de Radon	64
4.2.2	Teorema de Corte de Fourier	65
4.2.3	Filtered Backprojection (FBP)	68
4.3	Métodos de Reconstrução Algébricos	70
4.3.1	Técnica de Reconstrução Algébrica (ART)	70
4.3.2	Técnica de Reconstrução Iterativa Simultânea (SIRT) e Técnica de Reconstrução Simultânea Algébrica (SART)	73
4.4	Reconstrução Estatística Iterativa	75
4.5	Reconstrução Baseada em Sensoriamento Comprimido	76
4.6	Reconstrução Baseada em Aprendizado de Máquina	77
4.6.1	FPB Simulado via RNA	78
4.6.1.1	Experimentos e Resultados	81
4.6.2	Reconstrução de Primal-Dual	82
4.6.2.1	Experimentos e Resultados	84
4.6.3	Método Estatístico Iterativo para CBCT baseado em CNNs	85
4.6.3.1	Experimentos e Resultados	86
4.6.4	Aperfeiçoamento de parâmetros via Deep Learning	88
4.6.5	PGD baseado em Rede Neural	90
4.6.5.1	Experimentos e Resultados	91
4.6.6	Método Learn	92
4.6.6.1	Experimentos e Resultados	93
4.6.7	PWLS-ULTRA e PWLS-ST	94
4.6.7.1	Experimentos e Resultados	96
4.6.8	Discussões e avaliações	98
4.7	Considerações Finais do Capítulo	100
5	DESENVOLVIMENTO DO PROJETO	102
5.1	Metodologia realizada para reconstrução de DBT	102
5.2	Metodologia de Avaliação Experimental	105
5.3	Materiais e Métodos	107
5.4	Avaliação dos Resultados	108

6	EXPERIMENTOS E RESULTADOS	110
6.1	Arquitetura Parker	110
6.2	Arquitetura Ram-Lak	112
6.2.1	Comparação com Abordagens de filtragem	114
7	CONCLUSÕES E TRABALHOS FUTUROS	117
	REFERÊNCIAS	119

1 Introdução

As células de uma determinada região do corpo podem sofrer alterações no mecanismos de crescimento e passarem a se reproduzir em ritmo intenso e sem controle. Quando isso acontece, o resultado é chamado de tumor, massa ou neoplasma (RAMACHANDRAN; LAKSHMINARAYANAN, 1971). Embora seja possível que esses tumores não se desloquem para outras áreas do corpo, eles ainda podem causar complicações na região em que se encontram. Portanto, também é possível que o tumor invada os tecidos vizinhos ou que células defeituosas se desprendam e se desloquem para outras regiões do indivíduo, causando a metástase. Quando os tumores estão restritos às suas áreas iniciais, eles são chamados de benignos. Quando o tumor se desenvolve de tal maneira a ponto de invadir tecidos vizinhos ou acontece a metástase, ele é chamado de maligno. Por sua vez, o termo câncer é utilizado para definir tumores malignos.

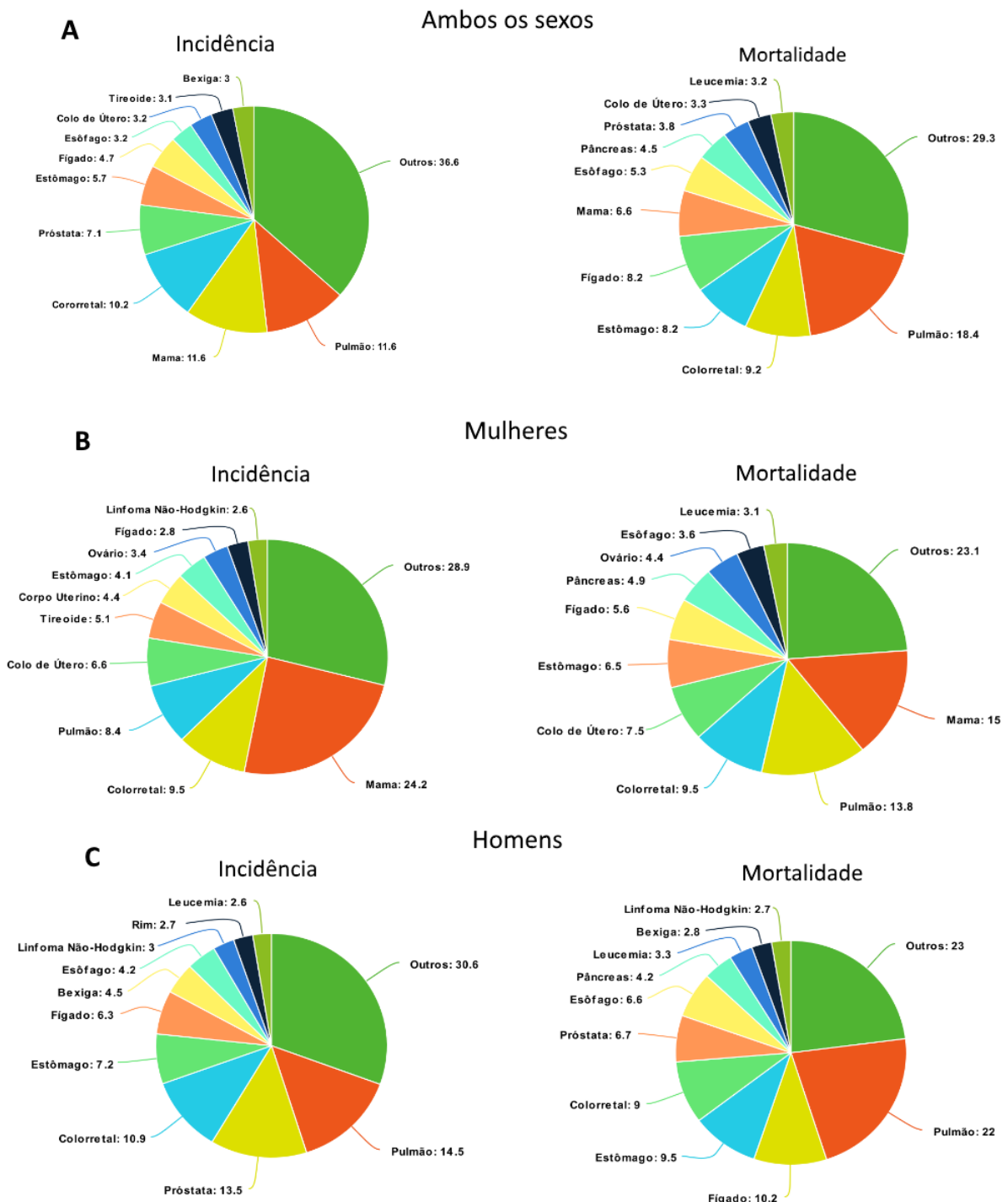
No mundo, segundo Bray et al. (2018), o câncer de mama possui a terceira maior taxa de incidência (11,6%) entre todos os tipos de câncer pesquisados, além da sexta maior taxa de mortes (6,6%)¹, levando em consideração a população mundial de ambos os sexos. Contudo, ao considerar somente o sexo feminino, o câncer de mama possui a maior incidência (24,2%) e a maior taxa de óbitos (15%), representando o perigo dessa doença para as mulheres. A Figura 1 mostra a distribuição dos casos de incidência e mortes causados pelos 10 tipos de cânceres mais comuns em 2018 no mundo.

No Brasil, no biênio de 2018 e 2019 foram estimados 59700 novos casos de câncer de mama (INCA - Instituto Nacional de Câncer José Alencar Gomes da Silva, 2018). Por outro lado, foram registrados 16254, 15593, 14786, 14388 e 13746 casos de óbito por causa de câncer de mama nos anos de 2016, 2015, 2014, 2013 e 2012, respectivamente (DATASUS,). É importante observar esse aumento no número de mortes no Brasil (SILVA et al., 2011), que pode estar relacionado com o aumento da expectativa de vida e envelhecimento da população, agravados por maior exposição da população à fatores de risco, resultado dos processos de urbanização (PORTER, 2008). Na América do Sul, os países que apresentam as maiores taxas de incidência e mortalidade são Argentina, Brasil e Uruguai (SIBIO et al., 2016).

Embora as estatísticas de incidência e óbitos sejam impactantes, há diversos trabalhos que

¹ A lista não considera o câncer de pele não-melanoma, devido à alta incidência na população.

Figura 1 – Incidência e mortalidade de vários tipos de cânceres sobre a população de ambos os sexos (A), do sexo feminino (B) e do sexo masculino (C). Os valores estão expressos em termos da porcentagem em relação ao total de casos de câncer registrados.



Fonte – Adaptado de Bray et al. (2018)

reportam que o uso de técnicas de imageamento reduziram, com o passar do tempo, a quantidade de óbitos de mulheres por causa de câncer de mama (BERRY et al., 2005; BROEDERS et al., 2012; MUNOZ et al., 2014; TABÁR et al., 2011), embora técnicas de terapia também

contribuíram para a redução desse número. Em especial, o uso da mamografia também tem se mostrado eficiente. Com o uso de tais técnicas de imageamento, o especialista em diagnóstico pode detectar anomalias nas mamas muitas vezes em estados iniciais, possibilitando o início antecipado do tratamento e aumentando a probabilidade de cura, assim, resultando em maiores chances de sobrevivência. Dessa maneira, a busca de técnicas e métodos para a melhor visualização dessas anomalias devem ser constantes, para permitir diagnósticos melhores e mais precisos, contribuindo para a maior redução das estatísticas de mortalidade.

Segundo o INCA (Instituto Nacional de Câncer), duas estratégias são utilizadas para detectar e tratar os casos de câncer de mama, que são: (i) *Diagnóstico Precoce* e (ii) *Rastreamento* (INCA, 2018). A estratégia (i), também conhecida como *Downsampling* (ORGANIZATION, 2007), busca encontrar e encaminhar para o tratamento as mulheres que já apresentam sinais ou sintomas iniciais da doença, com o objetivo de diminuir as taxas de diagnóstico em estágios avançados. Para que isso seja realizado, a estratégia adotada por diversos países é estimular as mulheres a efetuar a autopalpação/observação das mamas de forma a procurar pequenas anomalias nas mamas. O sistema de saúde deve dar suporte às mulheres que detectarem tais anomalias, fornecendo diagnóstico adequado e priorizando seu atendimento. Portanto, segundo Rangayyan (2004), os tumores que tiverem em estágio de desenvolvimento suficiente para serem detectáveis pela mulher, provavelmente já teriam realizado a metástase e formado tumores secundários; frisando a importância da detecção precoce de um tumor.

Com isso, para a detecção ainda mais antecipada de tumores, a estratégia (ii) tem por objetivo ofertar exames clínicos para um determinado grupo de risco, utilizando as seguintes abordagens: (a) *Rastreamento Oportunístico* e (b) *Rastreamento Organizado*. No primeiro, são ofertados exames às mulheres que voluntariamente chegam aos postos de saúde, enquanto que no segundo, as mulheres são explicitamente convidadas para exames clínicos periódicos. É evidente que as técnicas de Rastreamento devem ser implementadas zelando pela qualidade, com boa cobertura da população-alvo e tratamento adequado. Pode-se observar a redução na mortalidade pelo câncer de mama em países que implementam o Rastreamento, sendo isso suficiente para que a OMS recomende essa técnica como política pública para o combate de câncer de mama.

No mundo, diferentes entidades de saúde recomendam que as mulheres sejam incluídas em programas de aplicação de mamografia em diferentes faixas etárias. No Brasil, a recomendação do Ministério da Saúde (MIGOWSKI et al., 2018) é que exames mamográficos sejam efetuados em mulheres entre 50 a 69 anos, a cada dois anos, utilizando a Mamografia como método de

exame padrão, devido ao fato de que tal método tem eficiência comprovada na diminuição da mortalidade por câncer de mama, quando utilizada em estratégias de rastreamento (BRASIL; SAÚDE, 2015). Nos Estados Unidos, segundo a American Cancer Society (ACS) (KC et al., 2015), é recomendado que as mulheres comecem a realizar Mamografia anualmente com 45 anos, até atingir 54 anos. Mulheres com 55 anos ou mais devem passar a realizar o exame a cada dois anos, enquanto que mulheres entre 40 a 44 anos ou acima de 55 anos podem optar por realizar o exame anualmente.

É importante mencionar que essas recomendações são para as mulheres que compõem o grupo de risco médio, as quais não possuem agravantes conhecidos, como histórico pessoal de câncer de mama, mutação genética conhecida que aumenta o risco da doença (como a presença do gene BRCA (em inglês, *Breast Cancer*)) ou histórico familiar de câncer de mama. Para esse grupo de alto risco, exames mais frequentes são aconselhados, assim como o uso de Imagem por Ressonância Magnética (MRI) (SMITH; COKKINIDES; EYRE, 2003).

Assim como qualquer outro exame, a Mamografia possui riscos, como a ocorrência de falsos positivos, gerando ansiedade desnecessária à mulher, ou a realização de um sobretratamento de um possível tumor que não representaria nenhum dano à saúde se não fosse detectado.

Quanto maior a dose de radiação presente em um exame de mamografia, melhor a qualidade da imagem, facilitando o diagnóstico. Contudo, há grande esforço para encontrar métodos que diminuam a quantidade de radiação ionizante na Mamografia sem prejudicar a imagem resultante e a capacidade de identificação precoce de um possível tumor. O princípio ALARA (As Low As Reasonably Achievable) (PRASAD; COLE; HAASE, 2004), utilizado como base para as pesquisas científicas e a construção de equipamentos de mamografia, sugere a utilização do menor nível de radiação possível de modo a manter a boa qualidade da imagem e conseguir bons resultados. Imagens com níveis baixos de radiação ionizante tendem a ter mais imperfeições, como ruído. Um exemplo de trabalho que busca recuperar imagens adquiridas por meio de baixos níveis de radiação é o de Borges (2017), que aplica métodos de processamento de imagens para melhorar a qualidade dos resultados.

Outro exame de imagem que contribui para a detecção de câncer de mama é a Tomossíntese Mamária (DBT, do inglês, *Digital breast tomosynthesis*), que tem por objetivo gerar fatias do interior da mama, diminuindo o problema de sobreposição de tecidos presente em exames

mamográficos 2D.

Com isso, visto a importância do diagnóstico precoce e da diminuição da taxa de radiação emitida em um aparelho de DBT, o assunto abordado neste documento é a utilização de técnicas de Inteligência Artificial para o melhoramento de imagens derivadas de exames de DBT. Especialmente, este texto também irá abordar técnicas de Redes Neurais com Aprendizado Profundo para reconstruir imagens de DBT 3D a partir de projeções de DBT 2D ruidosas.

Para tal, a principal referência deste trabalho (WÜRFL et al., 2018) apresentou uma DNN que simula a operação de retroprojeção em uma camada de DNN, possibilitando a reconstrução analítica de imagens de CT. Desta maneira, um dos objetivos deste trabalho foi propor uma arquitetura básica de DNN para reconstrução de imagens de DBT. A motivação para propor tal arquitetura é que ela pode ser treinada para reconstruir imagens de DBT com qualidade maior ou equivalente aos métodos de reconstrução de estado-da-arte, mas apresentado um tempo de reconstrução (na etapa de teste) inferior. Além disso, uma arquitetura básica para a reconstrução de DNN possibilita que novas camadas possam ser utilizadas junto com a reconstrução, como camadas de filtragem pré-reconstrução ou pós-reconstrução, ou camadas para a detecção de lesões, assim, possibilitando que a etapa de treinamento e testes ocorra fim-a-fim.

Além disso, como outras contribuições, este trabalho disponibiliza como um software livre o código fonte utilizado para a reconstrução de DBT², além de ter submetido um artigo para a revista *IET Image Processing*, com o título *Image Reconstruction from Projections of Digital Breast Tomosynthesis using DeepLearning*, estando em análise neste momento.

Por fim, o restante deste texto está organizado da seguinte maneira: No Capítulo 2, serão discutidos os conceitos sobre mamografia. No Capítulo 3, serão discutidos os conceitos sobre Redes Neurais Artificiais com Aprendizado Profundo. No Capítulo 4, serão discutidas as técnicas de reconstrução tomográfica. No Capítulo 5, serão discutidas as propostas desta dissertação e a metodologia para a execução dos experimentos. No Capítulo 6, serão discutidos os resultados experimentais. Por último, no Capítulo 7, serão feitas as considerações finais desse texto.

² O link para o repositório no GitHub é: <https://github.com/daviduarte/DeepLearningDBT>

2 Mamografia e Tomossíntese Mamária

Um longo percurso em termos de pesquisa e desenvolvimento foi percorrido para ser possível a construção e a utilização segura de equipamentos de mamografia. Como já mostrado em outros trabalhos (SPRAGUE et al., 2017; SICKLES et al., 2005; PISANO et al., 2005), a mamografia analógica e digital são utilizadas como métodos seguros para a detecção precoce de câncer de mama, aumentando a taxa de sobrevivência em mulheres de todo o mundo. Portanto, há diversas pesquisas que buscam melhorar ainda mais a qualidade dos exames mamográficos e diminuir a dose de radiação aplicada, utilizando abordagens como a tomossíntese mamária digital (DBT).

Assim, neste Capítulo, são discutidos: a interação da radiação com a matéria (Seção 2.1), os conceitos sobre mamografia (Seção 2.2), o princípio de Tomografia (Seção 2.3) como base para o entendimento da tecnologia de Tomossíntese Mamária (Seção 2.4), os modelos de ruído presente nestas tecnologias (Seção 2.5) e, finalmente, as considerações finais (Seção 2.6).

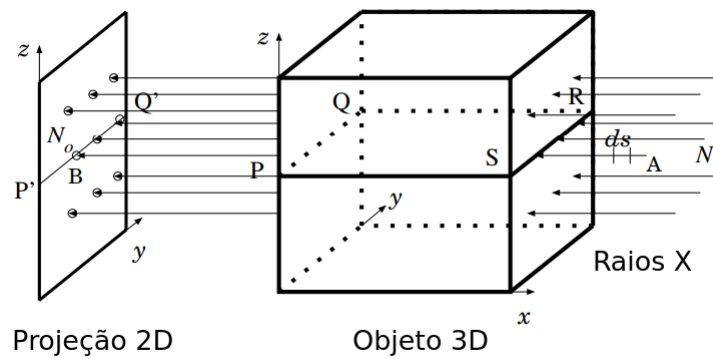
2.1 Interação da Radiação com a Matéria

Com especial interesse na área médica, o exame radiográfico (ou radiografia), é importante para a realização de exames de imagem, onde a sua forma mais simplificada se traduz nas aplicações de fótons de raios X para obter uma imagem planar (2D) do interior de um corpo 3D. Assim, todo o volume do corpo é resumido em uma única imagem que é exibida em um filme detector, filme este que recebe os feixes de raios X que atravessam o corpo examinado. É importante mencionar que, ao resumir todos os detalhes de um objeto 3D em uma imagem 2D, diversos detalhes são sobrepostos e informações importantes podem ser perdidas. Portanto, existem abordagens que buscam contornar esse problema, as quais serão discutidas neste Capítulo.

Na aplicação de raios X, os feixes de fótons emitidos são atenuados dependendo do material atravessado. Assim, a integral da atenuação linear durante todo o trajeto do feixe irá determinar a intensidade do sinal final, que irá corresponder a um nível de cinza no ponto em que o respectivo feixe colidir com o filme detector. No final, quando todos os feixes colidirem no filme, uma imagem 2D será formada, correspondendo ao interior do respectivo corpo 3D examinado.

Assim como exemplificado em Rangayyan (2004), considere o corpo cúbico mostrado na Figura 2, onde raios X mutuamente paralelos atravessam a seção transversal PSRQ. Na figura, N_i corresponde à quantidade de energia de um raios entrando no corpo e N_0 à quantidade de energia de um raios saindo do corpo. Considere também que a fonte da emissão esteja longe do corpo. Dessa maneira, as Equações 2.1 e 2.2 descrevem o processo de atenuação da passagem de cada raio pela Seção transversal PSRQ. Elas são versões modificadas da lei Beer-Lambert.

Figura 2 – Corpo cúbico sendo imageado por raios X. Ênfase deve ser dada na questão de que a seção transversal PSRQ é imageada por raios mutuamente paralelos, formando uma projeção 1D no filme receptor. O conjunto de todos os raios emitidos sobre o corpo irá formar uma projeção 2D do corpo 3D.



Fonte – Adaptado de Rangayyan (2004).

$$N_0 = N_i \exp\left[-\int \mu(x, y) ds\right] \tag{2.1}$$

$$\int_{\text{raio AB}} \mu(x, y) ds = \ln\left(\frac{N_i}{N_0}\right) \tag{2.2}$$

O valor de $\mu(x, y)$ representa o coeficiente de atenuação linear no ponto (x, y) , que depende do material no respectivo ponto e da frequência (comprimento de onda) da radiação eletromagnética utilizada.

Nas Equações 2.1 e 2.2, assumem-se raios X monocromáticos. A integral da Equação 2.2 representa o percurso do raio AB, sendo este presente no plano transversal PSRQ. Os raios que atravessam o plano PSRQ possuem um ângulo θ em relação ao plano (x, y) , e são representados pela coordenada (s, t) , sendo o eixo s paralelo aos raios. A relação entre (x, y) e (s, t) é dada

pela Equação 2.3. A variável de integração ds representa a distância elementar percorrida no raio AB. É importante observar que a Figura 2 corresponde ao caso particular de $\theta = 90$.

$$s = -x \sin \theta + y \cos \theta. \quad (2.3)$$

Alguns problemas inerentes à utilização de raios X ocorrem quando eles são utilizados para exames de imageamento, podendo resultar em degradação da imagem final. Quando um feixe de raios X é emitido, os fótons interagem com a matéria, podendo ser espalhados, causando ruído na imagem no ponto incidente do detector, sendo que o ângulo de espalhamento do fóton é aleatório. Uma técnica que diminui esse efeito é a utilização de uma grade que limita o ângulo de incidência dos fótons, discutido com mais detalhes na próxima Seção. Outro fenômeno que pode ocorrer é a absorção de fótons pela matéria, o que pode também ocasionar problemas na imagem final.

2.2 Mamografia

Entre todos os tipos de câncer incidentes nas mulheres, o de mama é o que mais as acomete. Segundo Rangayyan (2004) esse é um mal que possui uma incidência relativamente antecipada na vida das mulheres em comparação com outros tipos de câncer e doenças. Assim, é importante efetuar o diagnóstico o mais rápido possível, pois como mencionado anteriormente, muitos neoplasmas já terão formado tumores secundários quando a mulher puder detectá-lo manualmente (RANGAYYAN, 2004). Por isso, a aplicação da mamografia tem por objetivo principal detectar lesões nas mamas. Outras técnicas de imageamento podem ser usadas para o auxílio no diagnóstico de lesões, como ultrassonografia, ressonância magnética (MRI), tomossíntese mamária digital (DBT), etc. (BERRY et al., 2005; BROEDERS et al., 2012; MUNOZ et al., 2014; TABÁR et al., 2011).

Embora qualquer exame de imageamento das mamas possa ser chamado de mamografia, esse texto usará este termo para se referir à mamografia tradicional utilizando raios X. Nesse exame, é importante levar em consideração o fato de que a mama é composta por tecidos macios homogêneos, resultando em imagens de baixo contraste. Contudo, várias técnicas tem sido desenvolvidas para contornar esse problema (DROMAIN; BALLEYGUIER, 2010; JOCHELSON, 2014). Um método típico para a construção de um aparelho de mamografia está representado na Figura 3, onde é possível observar a presença de um emissor de raios X, cujos raios são

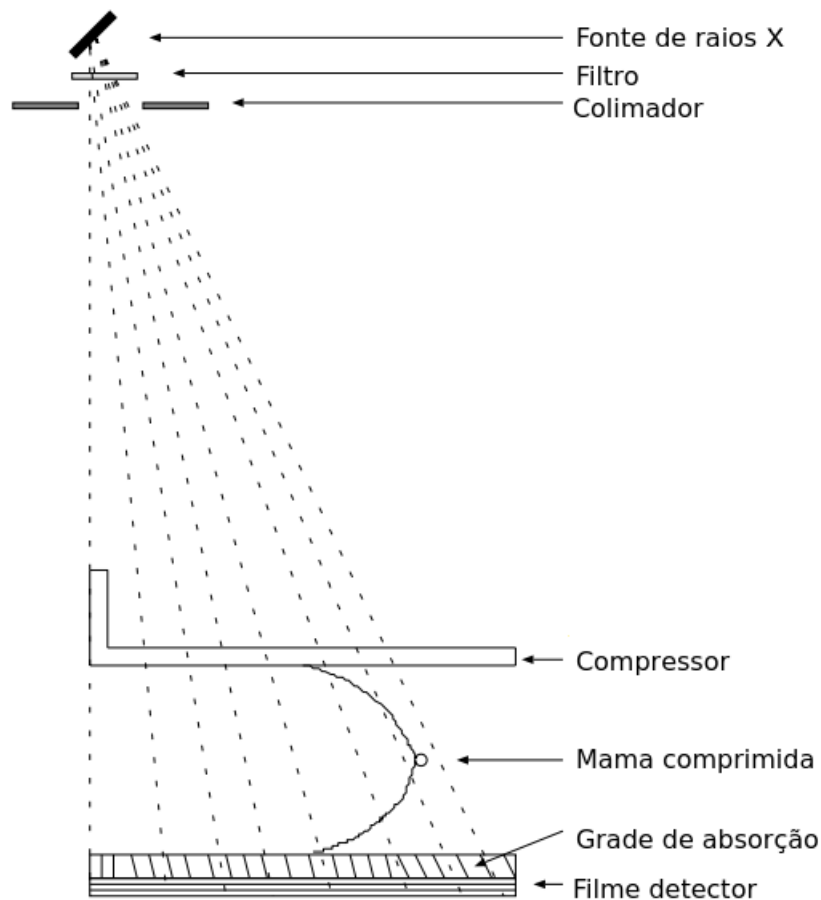
controlados por um filtro e por um colimador (material que absorve a radiação para poder direcioná-la). Também é possível notar a presença de uma haste comprimindo a mama, deixando os tecidos uniformes (aumentando a qualidade da imagem) e deixando-a imóvel.

Na Figura 3, há também a presença de uma grade de absorção desenhada especialmente para a mamografia, que é composta por um material que absorve radiação. Ela impede a passagem de fótons que tenham desviado de seu feixe original por causa do fenômeno de espalhamento (ver Seção 2.1). Parte da Figura 3 foi ampliada e editada na Figura 4 para ilustrar esse fenômeno. Nela é possível observar que os raios emitidos pela fonte (linhas tracejadas grossas) têm sua trajetória passando entre os receptáculos da grade de absorção. É possível observar também que os receptáculos da grade, na mamografia, são construídos respeitando a inclinação dos raios emitidos.

Os raios que foram perdidos de sua trajetória inicial devido ao espalhamento e absorção (linhas pontilhadas finas), podem colidir com as laterais dos receptáculos, sendo absorvidos antes de colidir com o filme detector, reduzindo ruído na imagem. Embora alguns fótons possam passar por essa grade (primeiro raio fino da esquerda), os efeitos negativos são amenizados, havendo uma melhora no contraste, especialmente importante no imageamento de mamas com estruturas densas.

O posicionamento adequado do detector em relação à mama (ou vista do detector), é um importante aspecto que deve ser levado em consideração no momento do imageamento. Características específicas do paciente devem ser analisadas para a escolha da vista adequada, como anomalias corporais, anomalias nas mamas previamente conhecidas, ou até mesmo doenças que impactam no exame, cabendo ao profissional avaliar cada caso individualmente. Na mamografia, o objetivo é obter uma imagem que mostra todas as regiões da mama, portanto, algumas vistas podem realçar determinadas regiões e dificultar a análise em outras, justamente pelo fato da mamografia produzir uma imagem 2D a partir de um corpo 3D (havendo sobreposição de detalhes e perda de informação), sendo comum o uso de mais de uma vista por exame. Segundo Andolina e Lillé (2011), é comum o uso de duas vistas (*Two-Views*), a cranial-caudal (CC) e a mediolateral-oblíqua (MLO). A vista CC tem por objetivo visualizar a mama de cima para baixo, assim como mostra a Figura 5a. O detector é posicionado com 0° em relação ao solo e os feixes de raio X são emitidos de cima para baixo, com os raios, por sua vez, perpendiculares ao solo. Na vista MLO (Figura 5b), o detector é posicionado em um eixo oblíquo em relação à mama, com inclinação aproximada de 45° com o solo (o ângulo pode variar de 30° a 60° dependendo

Figura 3 – Características principais de um aparelho de mamografia.

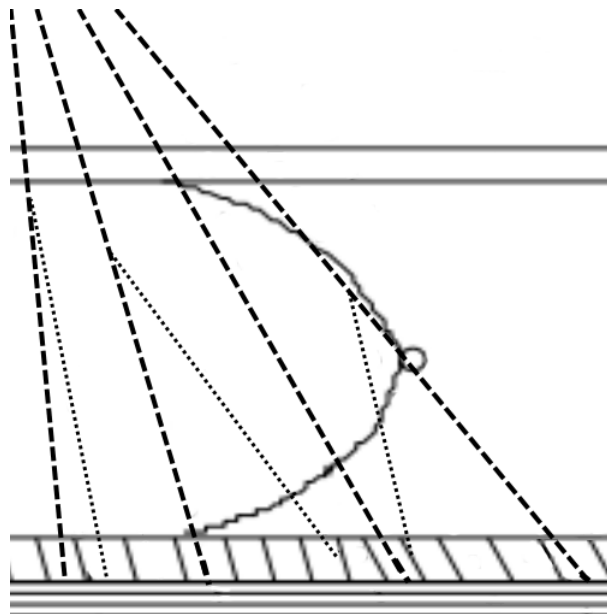


Fonte – Adaptado de Rangayyan (2004).

das características físicas do paciente).

Segundo Rangayyan (2004), há relatos da utilização de raios X para a obtenção de imagens das mamas desde 1930. Porém, somente em 1960 essa modalidade apresentou maiores progressos. Nos anos de 1980, diversas modalidades de imageamento se beneficiaram do advento das tecnologias digitais, mas na mamografia o primeiro protótipo de equipamento digital foi introduzido em 1995 na Sociedade de Radiologia da América do Norte (RSNA, do inglês *Radiological Society of North America*). Contudo, a agência de Administração de Alimentos e Medicamentos dos Estados Unidos (FDA, do inglês *Food and Drug Administration*) classificou o uso de equipamentos de mamografia digital como de risco significativo (mesmo depois do longo e demorado processo de avaliação), embora equipamentos digitais de outras modalidades de raios X tivessem sido classificadas com riscos mais baixos. Naquele momento, a questão sobre qual tinha a melhor performance (equipamentos de mamografia digital ou mamografia analógica) estava em alta, e alguns trabalhos começaram a surgir discutindo sobre o tema. O trabalho

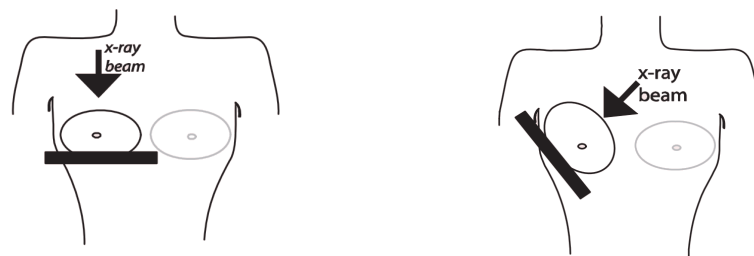
Figura 4 – Característica do espalhamento de raios, que causa ruído nas imagens geradas.



Fonte – Adaptado de Rangayyan (2004).

Figura 5 – Vistas usadas em exames cotidianos.

- (a) Representação da posição da mama e do detector na vista cranial-caudal. (b) Representação da posição da mama e do detector na vista mediolateral-oblíqua.



Fonte – Adaptado de Andolina e Lillé (2011).

de Lewin et al. (2002) foi um dos primeiros estudos que tentaram responder a essa pergunta, concluindo que a mamografia digital teve um *recall*¹ de 11,8% e a mamografia analógica teve um *recall* de 14,9%. No final de 2005, um estudo importante e longo foi realizado pela ACRIN (do inglês, *American College of Radiology Imaging Network*), denominado DMIST (do inglês, *Digital Mammographic Imaging Screening Trial*), com 49,528 mulheres que não apresentavam sintomas de doenças nas mamas, tornando possível a análise de pequenas diferenças clínicas e a obtenção de resultados comparativos da mamografia digital e mamografia analógica (PISANO et al., 2005). Os resultados mostraram que ambos os equipamentos mamográficos obtiveram

¹ Proporção dos exames que são identificados como suspeitos e solicitados ao paciente novos exames (seja por imagem ou por biópsia).

resultados semelhantes. Porém, a mamografia digital teve resultados significativamente superiores para mulheres abaixo dos 50 anos, mulheres com mamas heterogeneamente ou extremamente densas e mulheres na pré-menopausa² ou perimenopausa³. Esses resultados foram suficientes para aquecer as vendas de mamógrafos digitais logo na metade de 2006.

No exame utilizando mamografia analógica, uma imagem estática é gerada. Essa imagem, exibida em um filme especial, pode ser comparada, assim como feito em Andolina e Lillé (2011), a pinturas, desenhos, fotografias, etc, não podendo ser modificada, ampliada ou explorada com maiores detalhes. A imagem digital, gerada no exame de mamografia digital, é dinâmica, onde em vez do exame “impresso” em um filme especial, a imagem é armazenada em forma de bits, podendo ser modificada e processada de várias maneiras para servir a diversos propósitos. Essa liberdade é um dos principais interesses em torno da mamografia digital. Nesse formato, diversas técnicas e algoritmos de pós-processamento podem ser aplicados para melhorar o contraste e diminuir possíveis ruídos da imagem. Contudo, outro fator importante é que a imagem digital possui maior capacidade de representar as estruturas dos tecidos e suas relações.

O trabalho de Berns et al. (2006) analisou o tempo de aquisição e avaliação da imagem produzida por exames de mamografia analógica e digital. Os resultados mostraram que o uso da mamografia digital diminuiu o tempo de aquisição da imagem, pois a tecnologia digital permite que a imagem esteja disponível quase que instantaneamente, enquanto que o filme deveria passar por um outro processo antes de ficar disponível para análise. Em contrapartida, a necessidade de manipulação das imagens digitais ocasionaram problemas técnicos que aumentam o tempo de avaliação dos resultados, .

Muitas das características dos dois tipos de mamografia são compartilhadas, como o processo de aquisição da imagem e as características que distinguem tecidos de doenças. A transmissão de raios X através da mama, nos quais os feixes de fótons são atenuados em diferentes níveis pelas estruturas internas, gerando uma imagem única, segue o mesmo processo. O tecido adiposo atenua muito pouco a radiação, portanto ele aparece na imagem com tons mais escuros. O tecido glandular atenua mais a radiação, aparecendo com tons mais claros. Câncer de mama possui a característica de atenuar muito a radiação, aparecendo com tons mais claros. Portanto, se o câncer estiver em uma região de tecido adiposo, seu diagnóstico é fácil, mas se ele tiver em uma região de tecido glandular, seu diagnóstico é mais desafiador. Além disso, os dois métodos

² Transição do período reprodutivo para o não-reprodutivo. Ocorre por volta de 10 anos antes da menopausa.

³ Período que marca o fim da vida reprodutiva da mulher, marcando a aproximação da última menstruação.

tem por objetivo produzir uma imagem 2D a partir de um corpo 3D. Algumas diferenças entre os dois tipos de mamografia são (ANDOLINA; LILLÉ, 2011):

- A mamografia analógica utiliza um filme detector, enquanto que a mamografia digital utiliza uma matriz de detectores digitais.
- A imagem na mamografia analógica é latente⁴, enquanto que na mamografia digital é composta por sinais eletrônicos.
- É necessária a utilização de uma sala especial para a revelação da imagem obtida na mamografia analógica, enquanto que isso não é necessário na mamografia digital.
- Na mamografia analógica, tem-se um conjunto de filmes originais, e eles podem ser perdidos ou extraviados. Na mamografia digital, como a imagem é armazenada de forma virtual, sua perda ou extravio é difícil de ocorrer. Além disso, há a possibilidade de ter inúmeras cópias, assim como é possível ver a imagem em vários lugares ao mesmo tempo.

A mamografia se confirmou como o método mais indicado para a realização de exames mamográficos, sendo comprovado que seu uso diminuiu a taxa de mortalidade para o câncer de mama. Com o aperfeiçoamento da tecnologia digital, os mamógrafos digitais se tornaram o tipo mais comum de equipamento disponível atualmente. Contudo, ambos os tipos de equipamentos sofrem da mesma dificuldade na geração da imagem final, que é a sobreposição de tecidos, característica inerente quando se deseja criar uma projeção 2D a partir de um objeto 3D, dificultando diagnósticos de possíveis doenças.

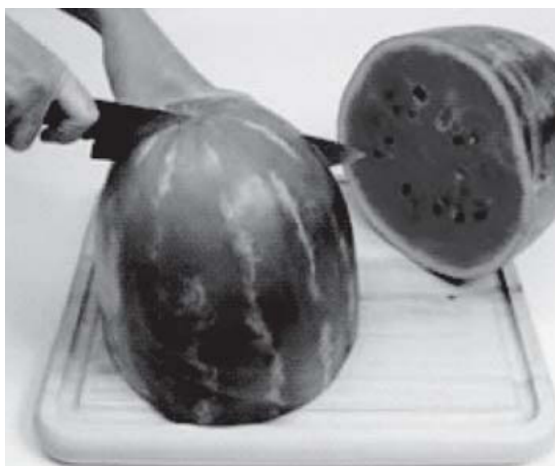
2.3 Tomografia

A palavra tomografia se refere à obtenção de imagem de uma fatia, corte ou seção de algum objeto. A origem da palavra deriva do grego, *tomos*, que significa seção ou fatia. Zeng (2010) mostra um interessante exemplo no qual utiliza uma melancia para ilustrar o processo de tomografia em exames médicos. Se alguém tivesse por objetivo analisar uma fatia do interior da melancia, o melhor método seria cortá-la (ver Figura 6). Contudo, no contexto médico, cortar um paciente para analisar seu interior está fora de cogitação. Essa necessidade de avaliar planos internos distintos de forma não invasiva fez com que experimentos utilizando raios X começassem

⁴ Imagem formada no processo de revelação por meio de reações químicas com o filme detector.

a ser realizados poucos anos após o seu descobrimento. Como resultado desse esforço foram desenvolvidas abordagens de imageamento que tem por objetivo focar em um plano no interior do corpo que está sendo imageado e borrar os demais planos, resultando em uma imagem focada apenas na seção de interesse. Para isso, uma abordagem clássica é movimentar a fonte de raio X e do filme detector de forma síncrona, como será discutida mais adiante.

Figura 6 – Ilustração do processo de tomografia.

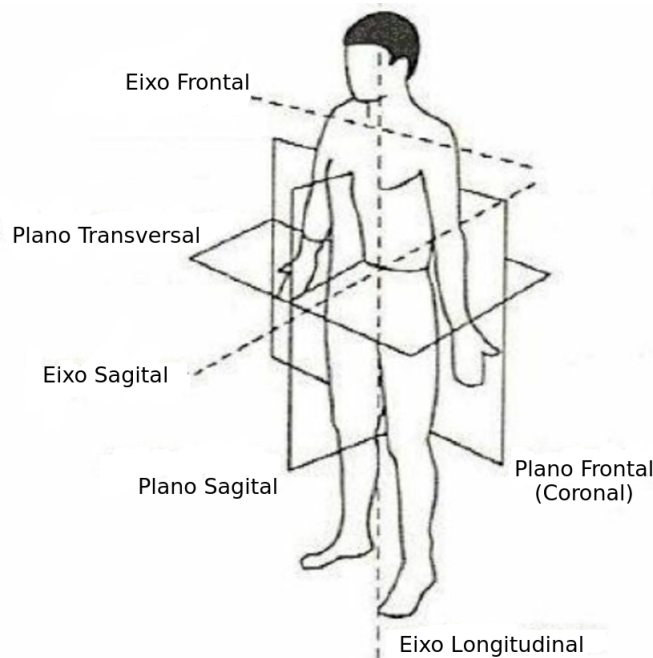


Fonte – Extraído de Zeng (2010).

Portanto, quando se fala em obter imagens de planos no interior do corpo humano, é necessário dividi-lo em partes e nomear as estruturas em relação às dimensões espaciais. É possível dividir o corpo em 3 partes, que são representadas por planos ortogonais: (i) plano Sagital, (ii) plano Coronal e (iii) plano Transversal. Esta divisão está ilustrada na Figura 7. O plano Sagital corta o corpo humano no sentido antero-posterior (que passa entre os olhos). O plano Coronal corta o corpo humano de forma lateral, que passa de uma orelha à outra. O plano Transversal corta o corpo humano em fatias transversais. A intersecção dos planos Sagital e Transversal forma o eixo Sagital. A intersecção dos planos Frontal e Transversal forma o eixo Frontal. Por fim, a intersecção dos planos Frontal e Sagital forma o eixo Longitudinal.

Com os planos anatômicos definidos, é possível discutir as técnicas de tomografia. Como mencionado anteriormente, tais técnicas já começavam a serem propostas pouco tempo após a descoberta dos raios X, quando o notório trabalho de Radon (1917) descreveu os fundamentos matemáticos que permitiram que projeções adquiridas por feixes de raios passando por um determinado corpo pudessem ser utilizadas para reconstruir a estrutura interna desse corpo. Após isso, o trabalho de Plantes (1932) utilizou esses princípios para descrever o conceito de imagens tomográficas, que logo possibilitou o desenvolvimento dos equipamentos de tomografia clássica.

Figura 7 – Planos e eixos anatômicos.



Fonte – Adaptado de RAMACHANDRAN et al. (2016).

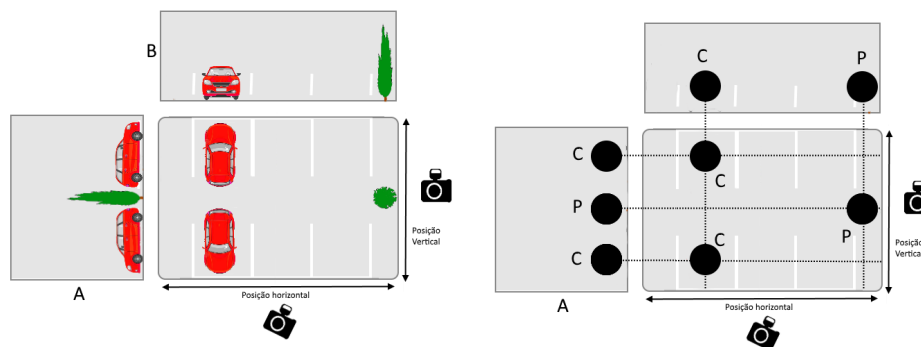
Tais equipamentos utilizam apenas o movimento da fonte de radiação e do filme detector para formar uma imagem no interior do paciente.

Segundo Rangayyan (2004), a tomografia computadorizada (CT) teve seu início no final dos anos de 1960 e começo dos anos de 1970, possibilitando a obtenção de imagens transversais (Figura 7) da cabeça humana. Equipamentos de CT modernos geram uma imagem em 3D do interior do paciente. Para isso, a região do paciente é irradiada repetidas vezes variando o ângulo de emissão de radiação, obtendo-se várias projeções do mesmo local. Ao final, algoritmos computacionais são usados para reconstruir a imagem desta área utilizando as projeções adquiridas de todos os ângulos. Tais algoritmos são chamados de algoritmos de reconstrução tomográfica e serão discutidas com mais detalhes no Capítulo 4. A utilização destes algoritmos é fundamental para o funcionamento da Tomografia Computadorizada, em contraste com a tomografia clássica, que utiliza apenas princípios mecânicos e de ótica para formar a imagem. Desta maneira, o computador se torna fundamental nesse novo método, justificando o seu nome. A escolha adequada de um algoritmo de reconstrução é necessária para se atingir um nível de qualidade adequado da imagem final.

Para exemplificar o funcionamento de um algoritmo de reconstrução analítico, considere a Figura 8. Nela, tem-se um estacionamento com dois carros e um pinheiro, posicionados em lugares

Figura 8 – Exemplo do princípio de reconstrução tomográfica.

- (a) Um estacionamento com 2 carros e 1 pinheiro. (b) Posições dos elementos conhecidos a partir das 2 projeções do estacionamento, sendo mapeados em suas posições originais.



Fonte – Autoria própria.

distintos (Figura 8a). Se uma pessoa se posicionasse na lateral direita desse estacionamento (representado pela imagem da câmera fotográfica) e tirar uma fotografia, a foto resultante (A) seria uma projeção lateral do estacionamento e de todos os objetos que nele estão. Se essa pessoa agora se posicionasse na parte inferior do estacionamento e tirasse outra fotografia, a foto resultante (B) seria uma projeção do estacionamento vista de baixo. Veja, portanto, que na foto tirada de baixo (B), não é possível distinguir que há 2 carros no estacionamento, pois eles estão sobrepostos. Já na foto tirada da lateral (A), é possível distinguir que há dois carros, mas não é possível inferir em qual vaga (posição horizontal) eles estão. Da mesma maneira, olhando apenas para a foto tirada de baixo (B) ou para a foto tirada na lateral (A), não é possível saber a posição exata do pinheiro. Portanto, para saber a posição exata de todos os elementos, pode-se desenhar uma reta partindo da posição de cada elemento nas duas fotos, sendo a intersecção das retas a posição exata em que cada elemento se encontra. A Figura 8b exemplifica o método explicado, sendo que os carros e o pinheiro foram representados como círculos (para simplificar a visualização), cada um distinguido por uma letra ao lado do respectivo círculo, sendo a letra C para distinguir os Carros e a letra P para distinguir o Pinheiro. Nesse exemplo, foram utilizadas 2 fotos (projeções), mas o mesmo princípio poderia ser aplicado para mais fotos tiradas de diferentes ângulos.

Embora a tomografia possa contribuir para a visualização de detalhes internos, é necessário o imageamento de todo o intervalo angular em torno de uma seção transversal no paciente, ou seja, a obtenção do conjunto completo de projeções. De maneira similar, a técnica chamada

de *Tomossíntese Digital* consiste em obter imagens de quaisquer fatias interna, mas utilizando apenas algumas projeções obtidas em um conjunto limitado de ângulos, diminuindo a exposição do paciente à radiação, conforme detalhado na Seção 2.4.

2.4 Tomossíntese Mamária

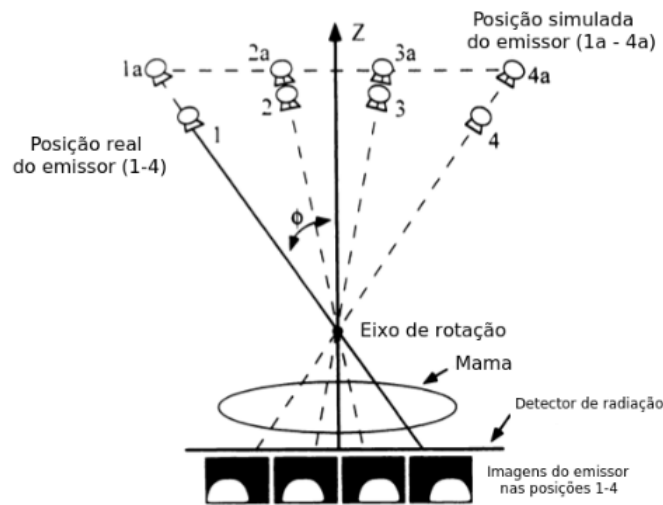
A Tomossíntese Digital é semelhante à Tomografia Computadorizada pois utiliza o mesmo mecanismo de obtenção das imagens digitais, utilizando um único movimento da fonte e do detector de radiação. Ambas as técnicas são baseadas na transformada inversa de Radon (RADON, 1917). Portanto, como a tomossíntese funciona por meio de uma amostragem parcial de projeções, técnicas de aproximação devem ser utilizadas.

A Tomossíntese Digital pode ser aplicada na detecção de câncer de mama, sendo chamada de Tomossíntese Mamária Digital (DBT) (NIKLASON et al., 1997). Os trabalhos de Friedewald et al. (2014) e Rafferty et al. (2013) sugerem que a aplicação de DBT junto com a mamografia digital aumenta a acurácia e diminui a taxa de *recall* para os casos em que não há presença de câncer de mama.

É possível que os equipamentos de mamografia digital sejam modificados para que se tornem capazes de realizar exames de Tomossíntese Mamária. Portanto, a fonte de radiação deve ter uma trajetória ao longo de um arco sobre o paciente, para que projeções de diversos ângulos possam ser adquiridas, como mostra a Figura 9. Assim, uma modificação que deve ser feita nos equipamentos de mamografia digital é permitir a movimentação da fonte de radiação (assim como o filtro e o colimador na Figura 3). Considerações de projeto devem ser feitas levando em conta o movimento do emissor de raios X e o posicionamento do paciente, garantindo que o emissor não interfira no seu posicionamento. Além disso, considerações a respeito da vibração das partes móveis, além do alinhamento das peças, devem ser feitas, pois, segundo Mainprize et al. (2011), isso influencia criticamente no contraste das imagens geradas. Devido ao movimento angular do emissor de raios X, o colimador (veja Figura 3) deve ser variável, certificando que os raios X não incidam em outros lugares a não ser na área delimitada pelo detector digital. Ainda, por definição, o processo de DBT obtém projeções utilizando baixas doses de radiação.

Os equipamentos de DBT podem emitir raios X de duas maneiras: (i) continuamente ao longo de seu movimento ou (ii) passo a passo, no qual o emissor para em uma determinada posição, emite o feixe de raio X e, em seguida, vai para a próxima posição, repetindo o processo

Figura 9 – Ilustração do processo de aquisição de DBT.



Fonte – Adaptado de Niklason et al. (1997).

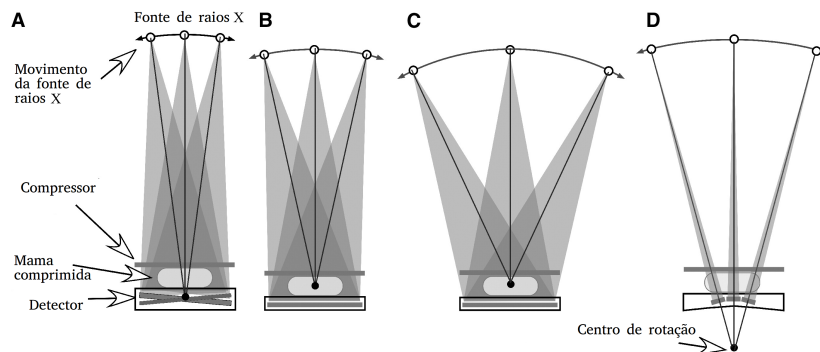
ao longo de sua trajetória. A desvantagem da abordagem (i) é a ocorrência de borramento nas projeções, devido ao movimento do emissor, contudo, sua implementação é mais simples, devido à não necessidade de um mecanismo de parada e controle de emissão. A abordagem (ii) evita o efeito de borramento nas projeções, mas sua implementação se torna mais complexa, em virtude da necessidade de um mecanismo que controle a emissão dos raios X passo a passo. A ocorrência de borramento nas projeções da abordagem (i) pode ser, segundo Mainprize et al. (2011), parcialmente evitado utilizando raios X pulsantes (ao invés de emitir continuamente).

Para produzir uma imagem de maior qualidade, é possível aumentar a faixa angular que a mama é imageada. Tal aumento, conseqüentemente, produz um incremento nas informações tomográficas usadas na reconstrução, resultando em uma melhor separação das seções transversais nas imagens finais. Contudo, isso deve ser acompanhado de uma quantidade de projeções também maior, o que deve ser analisado, pois a radiação empregada em cada projeção deve ser baixa, uma fração da quantidade total empregada no exame. Além disso, para diminuir a probabilidade de movimentação do paciente (e evitar possíveis danos na imagem), o processo de aquisição deve ser o mais rápido possível, fazendo com que o detector digital tenha que ter uma rápida taxa de leitura da radiação incidente. Em todos os casos, o detector digital deve ser especialmente configurado para DBT.

Os equipamentos de DBT podem conter várias peculiaridades em suas geometrias. A Figura 10 mostra algumas dessas características. Em A, a posição do detector é alterada em relação ao centro de rotação, e pode adquirir ângulos que variam em $\pm 2, 1^\circ$, enquanto que a fonte

de raios X pode variar em $\pm 7,5^\circ$. Em tais geometrias, é dito que o detector possui movimento isocêntrico. Em B, o detector se mantém parado enquanto a fonte de raios X realiza a sua trajetória. Em C, a abordagem é similar à mostrada em B, porém o emissor de raios X percorre uma trajetória maior. Em D, uma geometria de varredura de fenda (*slot-scan*) é mostrada, cujo centro de rotação se encontra abaixo do detector.

Figura 10 – Geometria dos equipamentos de tomossíntese mamária.



Fonte – Adaptado de Vedantham et al. (2015).

Cada geometria possui suas características, que podem ser positivas ou negativas. Na geometria ilustrada em A, o compartimento que abriga o detector deve ser grande o suficiente para comportar o seu movimento, algo que pode influenciar negativamente na facilidade de posicionamento do paciente. Isso não acontece nas geometrias em que o detector fica estacionário (geometrias B e C). Portanto, o detector estacionário implica na incidência oblíqua dos raios X, quando o emissor estiver em posições angulares altas, causando diminuição da resolução espacial das projeções (MAINPRIZE et al., 2006). Existem técnicas que tentam contornar esse problema, mas não é evidente que tais correções sejam implementadas nos equipamentos de DBT, como sugerido por Vedantham et al. (2015). Além disso, os detectores estacionários devem ser mais largos para diminuir a área truncada da mama na projeção resultante. Segundo Vedantham et al. (2015), a geometria de varredura de fenda (D) possui excelente desempenho em impedir os efeitos de espalhamento de raio. A Tabela 2.4 mostra as características de alguns equipamentos de DBT encontrados comercialmente (SECHOPOULOS, 2013).

Tabela 1 – Principais características de diversos equipamentos de DBT.

Sistema	GE Essential	Hologic Selenia Dimensions	IMS Giotto TOMO	Philips MicroDose	Planned Nuance Excel DBT	Siemens MAMMOMAT Inspiration
Tipo do Detector	Full field—indireto	Full field—direto (a- <i>Se</i>)	Full field—direct (a- <i>Se</i>)	Linear slit scan—contagem de fótons espectral (Si)	Full field—direto (a- <i>Se</i>)	Full field—direto (a- <i>Se</i>)
Tamanho do detector (cm)	24 x 30	24 x 29	24 x 30	detectores com 21 linhas, cada uma com 24 cm	24 x 30	24 x 30
Tamanho do pixel do detector (μm)	100	70 (binned 2x2)	85	50 (perpendicular ao movimento)	85	85
Movimento do detector	Estático	Rotacional	Estático	Escaneamento de fenda contínua	Rotacional durante exposição	Estático
Intervalo angular (graus)	25	15	40	11	30	50
Número de projeções	9	15	13	21	15	25
Tempo de escaneamento (s)	7	3.7	12	3-10	20	25
Distância da fonte de raios X ao detector (cm)	66	70	68	66	65	65.5
Método de reconstrução	Iterativo	FBP	Iterativo com variação total	Iterativo	Iterativo	FBP

Fonte – Adaptado de (SECHOPOULOS, 2013).

2.5 Ruído

Qualquer procedimento que tenha por objetivo a recepção de sinais digitais pode estar sujeito a interferências, podendo resultar em uma interpretação de menor qualidade em comparação ao mesmo sinal se não houvesse tal interferência. Sua origem pode ser inerente aos instrumentos usados, ou mesmo ao ambiente no qual determinado experimento está sendo realizado. No domínio das imagens biomédicas, tal interferência resulta em dados corrompidos nas imagens, podendo gerar ruído ou objetos que não estão realmente no paciente imageado (artefatos). Ela pode gerar perda de qualidade no exame final, impactando negativamente na qualidade do diagnóstico. A seguir será feita uma caracterização dos diversos tipos de ruído encontrado na aquisição e reconstrução de imagens tomográficas.

Em exames radiológicos a própria natureza estatística da emissão e detecção dos feixes de raios X apresenta uma certa natureza aleatória. Assim, essa componente aleatória $N_{\theta_i}(t)$ é descrita pela distribuição de probabilidade Poisson (KAK; SLANEY, 1988; SALVADEO et al., 2013)

$$p\{N_{\theta_i}(k\tau)\} = \frac{\bar{N}_{\theta_i}(k\tau)^{N_{\theta_i}(k\tau)}}{N_{\theta_i}(k\tau)!} e^{-\bar{N}_{\theta_i}(k\tau)}, \quad (2.4)$$

onde $N_{\theta_i}(k\tau)$ é o número de fótons detectados do k -ésimo raio emitido no ângulo θ_i , p é a probabilidade e $\bar{N}_{\theta_i}(k\tau)$ é o valor esperado (E) de fótons, sendo expresso pela Equação 2.5 (KAK; SLANEY, 1988; SALVADEO et al., 2013):

$$\bar{N}_{\theta_i}(k\tau) = E\{N_{\theta_i}(k\tau)\}. \quad (2.5)$$

Como esta é uma distribuição de Poisson, é possível definir a variância como na Equação 2.6:

$$Var\{N_{\theta_i}(k\tau)\} = \bar{N}_{\theta_i}(k\tau). \quad (2.6)$$

Assim, é possível concluir que o ruído presente em projeções tomográficas é governado pela distribuição de Poisson, sendo chamado de ruído quântico. A variância do ruído presente em uma determinada projeção é inversamente proporcional ao número de fótons medido no detector (EPSTEIN, 2008), sendo que quanto menor a contagem de fótons, maior o ruído.

Para eliminar o ruído de Poisson, que é inerente à etapa de projeção, há diversos trabalhos que buscam melhorar os receptores digitais para que sejam mais eficientes e menos suscetíveis ao ruído (MAINPRIZE; WANG; YAFFE, 2009, 2009; ZHAO; ZHAO, 2008; REN et al., 2010; CHEUNG et al., 2005).

Na imagem reconstruída, é possível caracterizar o ruído como Gaussiano (distribuição normal) (SALVADEO et al., 2013). É possível avaliar essa afirmação considerando o Teorema Central do Limite (EPSTEIN, 2008), pois os métodos de reconstrução tratam a imagem como um grande número de variáveis aleatórias, e esse teorema afirma que quando se tem n variáveis aleatórias, e n aumenta muito, a distribuição amostral da média desse conjunto aproxima-se de uma distribuição normal, podendo concluir que os pixels da imagem reconstruída são governados aproximadamente por uma distribuição gaussiana. Entretanto, Scarparo et al. (2019) mostram evidências que o ruído predominante na imagem reconstruída obedece a distribuição de Burr (BURR, 1942).

O processamento de sinais por circuitos eletrônicos inevitavelmente gera ruído, sendo chamado de ruído eletrônico. Contudo, há detectores que conseguem reduzir o ruído eletrônico de modo que ele represente apenas uma fração do ruído total (HANSON, 1981; COHEN; DIBIANCA, 1979). A adição de ruído a partir de circuitos eletrônicos acontece quando se deseja processar a quantidade de energia dos raios X. Em computadores pessoais, os circuitos que processam sinais discretos são resistentes a tais interferências (COHEN; DIBIANCA, 1979). No caso de DBT, em geral, esse ruído é modelado nas projeções por uma distribuição Gaussiana. Desse modo, combinado ao ruído quântico, o ruído nas projeções de DBT pode ser modelado como uma distribuição Poisson-Gaussiana, expresso pela Equação 2.7 (AZZARI; BORGES; FOI, 2018)

$$z(x) = ap(x) + n(x), \quad (2.7)$$

onde

$$p(x) \sim \mathcal{P}(a^{-1}y(x)) \text{ e } n(x) \sim \mathcal{N}(0, b), \quad (2.8)$$

onde $a > 0$ é o fator de escala para o Poisson e $b \geq 0$ é a variância da Gaussiana.

Por sua vez, a média e variância condicional de z são (AZZARI; BORGES; FOI, 2018),

respectivamente,

$$E\{z(x)|y(x)\} = y(x) \quad (2.9)$$

e

$$\text{var}\{z(x)|y(x)\} = ay(x) + b, \quad (2.10)$$

sendo que o ruído Poisson-Gaussiano é definido por $z(x) - y(x)$, possuindo média zero e variância afim (AZZARI; BORGES; FOI, 2018).

Além disso, existem também os artefatos, que são estruturas presentes na imagem que não possuem sua contraparte no objeto real. Tais estruturas podem prejudicar a análise de um exame de CT ou DBT (KALENDER, 2000).

Ainda, outro tipo de ruído que pode estar presente é o ruído estrutural, que está relacionado com a sobreposição de estruturas do objeto imageado. Um corpo que possui regiões com diferentes densidades pode sobrepor informações na imagem resultante. Embora DBT tenha eliminado parte considerável deste tipo de ruído (HANSON, 1981), interferências em fatias de DBT podem ser causadas por fatias adjacentes.

Em teoria, aumentar o número de projeções e do intervalo angular imageado resultaria em um aumento da qualidade da imagem. Contudo, para manter os níveis de radiação equivalentes aos níveis utilizados na mamografia digital, cada projeção teria uma parcela ainda menor de radiação utilizada, aumentando a quantidade de ruído e artefatos na imagem. Portanto, encontrar a melhor combinação da quantidade de projeções, o intervalo angular e a distribuição das projeções pode se tornar um problema de otimização, cujo resultado pode ser diferente para objetivos clínicos específicos.

Levando em consideração a natureza Poisson-Gaussiana das imagens de CT, Borges et al. (2017) descrevem um pipeline para processar imagens de FFDM (do inglês *Full-Field Digital Mammography*) e DBT utilizando uma etapa de calibração e transformação de estabilização de variância.

2.5.1 Simulação

Como mencionado nas seções anteriores, a qualidade da imagem em um exame é dependente de diversos fatores, entre eles, a quantidade de radiação empregada. Contudo, para o estudo de métodos de processamento de imagens biomédicas, uma ampla quantidade de imagens para testes, derivadas do mesmo paciente, com diferentes níveis de radiação, deve estar disponível. Mas na prática isso não é possível devido ao risco biológico. Uma alternativa é a utilização de *phantoms* (físicos ou virtuais) para simular uma mama real, mas tais objetos não são capazes de simular certas características, podendo impactar negativamente nos resultados dos estudos. Outra alternativa é manipular uma imagem obtida com uma certa dose de radiação para aparentar ter sido obtida com uma dose menor.

Na mamografia digital, Borges et al. (2016b) propôs um método para simular redução de dose em exames, cujo mesmo método foi adaptado mais tarde para DBT (BORGES et al., 2016a). No caso de DBT, a abordagem proposta tem 3 conjuntos de projeções como entrada e, como saída, as fatias reconstruídas com a dose reduzida de forma simulada. O primeiro conjunto de entrada deve conter as projeções com dose de radiação padrão. Os outros conjuntos devem conter projeções uniformes, sendo necessário que um conjunto tenha exposição (mAs) igual ao conjunto padrão e o outro tenha exposição inferior àquela que se deseja simular. Esses três conjuntos são utilizados na simulação das projeções, que por sua vez são utilizadas na reconstrução. O processo para simular as fatias finais é dividido em 3 etapas: (i) linearizar e escalar todas as imagens de entrada; (ii) modular uma máscara com ruído simulado; (iii) introduzir essa máscara nas imagens no domínio de Anscombe. A transformação de Anscombe leva a imagem com o ruído governado pela distribuição de Poisson, para uma imagem com o ruído governado aproximadamente pela distribuição Gaussiana.

2.6 Considerações Finais do Capítulo

Diversas abordagens surgiram no decorrer dos anos com o objetivo de obter uma imagem do interior dos pacientes, possibilitando o diagnóstico de diversas enfermidades. Em especial, a mamografia, que utiliza raios X para obter imagens das mamas, possibilitou uma diminuição na taxa de mortes de mulheres em todo o mundo, pois lesões nas mamas puderam ser detectadas e tratadas muito antes de se tornarem um risco alto para a saúde. Logo após, surgiu a mamografia digital (FFDM), que uniu os avanços da tecnologia digital com a mamografia tradicional e abriu

caminho para o advento da Tomossíntese Mamária (DBT), ou mamografia 3D. Essa abordagem elimina a sobreposição de tecidos na imagem causados pela mamografia 2D. Entretanto, seu uso ainda é recomendado apenas como complemento à mamografia 2D, mas a tendência é que ela irá tomar o papel principal.

A área da pesquisa correspondente aos exames mamográficos está muito ativa, buscando constantemente novas técnicas para aumentar a taxa de detecção de doenças e diminuir os riscos dos exames, como a ocorrência de falsos positivos e os possíveis danos causados pela radiação (princípio ALARA).

Dessa maneira, este Capítulo teve por objetivo apresentar as principais referências sobre exames mamários e discutir os principais problemas, principalmente sobre DBT. No próximo Capítulo, os fundamentos sobre Redes Neurais e Aprendizado Profundo serão discutidos.

3 Redes Neurais com Aprendizado Profundo

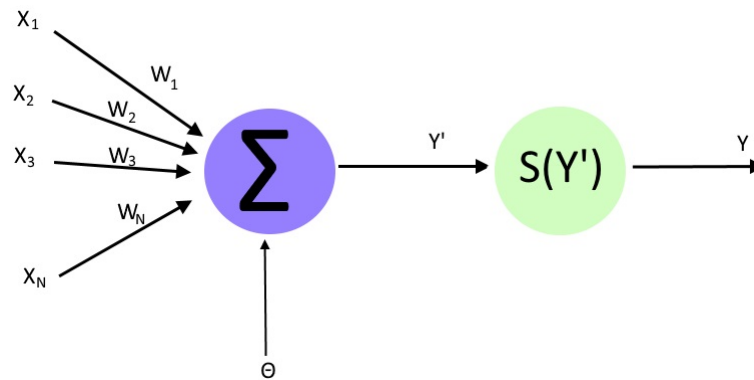
Redes Neurais Artificiais (RNAs) (BENGIO et al., 2009; DENG; YU et al., 2014; LECUN; BENGIO; HINTON, 2015; SCHMIDHUBER, 2015) são modelos matemáticos que tem por objetivo simular o comportamento do sistema nervoso de mamíferos. As RNAs existem como uma subárea do Aprendizado de Máquina, estudando algoritmos e técnicas com o objetivo de aprender padrões presentes em conjuntos de dados. As unidades básicas de sistemas nervosos biológicos são os neurônios, que formam grandes redes interconectadas capazes de armazenar informações e efetuar processamento. Uma clara definição de RNA é apresentada por Samarasinghe (2016, tradução nossa):

“... é uma coleção de neurônios interconectados que incrementalmente aprendem a partir de seu ambiente (dados) para capturar tendências essenciais lineares e não lineares em dados complexos, assim, possibilitando previsões confiáveis para novas situações ruidosas ou com informações faltando.”

Um neurônio artificial com a capacidade de encontrar superfícies de separação para padrões linearmente separáveis foi proposto por (ROSENBLATT, 1958), chamado de Perceptron, que foi desenvolvido para simular o comportamento do neurônio biológico, implementando suas principais estruturas. O modelo de um Perceptron é mostrado na Figura 11. Ele pode se conectar com outras unidades para formar redes mais complexas, assim, possibilitando o aprendizado de padrões também mais complexos. Essa conexão corresponde às sinapses nos neurônios biológicos. Na Figura 11, as entradas ($X_1, X_2, X_3, \dots, X_N$) representam os dendritos, que armazenam as saídas de outros neurônios. O produto escalar é então efetuado com os pesos ($W_1, W_2, W_3, \dots, W_N$) e um bias θ é somado, resultando em um valor Y' . Por último, o valor de Y' é submetido a uma função de ativação $S(Y')$ (correspondendo ao axônio no neurônio biológico), resultando no valor Y , que é a saída do neurônio. Esse valor pode ser submetido a outros neurônios subjacentes ou ser apresentado como uma resposta final. A Equação 3.1 representa o modelo matemático de um Perceptron, onde N é quantidade de sinapses (ROSENBLATT, 1958),

$$Y = S\left(\theta + \sum_{i=1}^N W_i X_i\right). \quad (3.1)$$

Figura 11 – Modelo do Perceptron.



Fonte – Autoria própria.

Uma RNA pode ser composta por vários Perceptrons divididos em três tipos de camadas: (i) camada de entrada, que recebe os estímulos (dados) externos; (ii) camadas ocultas, que recebem os estímulos das camadas anteriores e generalizam as informações do ambiente, armazenando informações e efetuando processamento; e (iii) camada de saída, que emite o estímulo-resposta para os dados presentes na camada de entrada. O agrupamento de vários Perceptrons em camadas torna possível a aproximação de funções matemáticas arbitrárias (HORNIK, 1991). A variação do número de camadas, quantidade de neurônios em cada camada, função de ativação utilizada, entre outras características, resultam em diferentes modelos ou arquiteturas de RNAs.

Atualmente, vários modelos foram propostos com uma grande quantidade de camadas (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; SIMONYAN; ZISSERMAN, 2014; SZEGEDY et al., 2015; HE et al., 2015), sendo esses modelos chamados de Redes Neurais Profundas (DNNs, do inglês *Deep Neural Networks*). Em geral, tais modelos tem conseguido bons resultados em suas áreas de aplicação, em parte por causa da capacidade de processamento disponibilizada atualmente. As DNNs necessitam de grande capacidade de processamento, devido à sua quantidade de camadas. O surgimento de grandes conjuntos de dados também contribuiu com a obtenção de bons resultados, principalmente em tarefas que são consideradas ambíguas, como a classificação de imagens do dia-a-dia (DENG et al., 2009). Um grande conjunto de dados deve estar disponível para que a rede neural aprenda com sucesso as relações presentes. Cada dado individual do conjunto, no decorrer desta dissertação, será chamado de amostra.

A utilização de uma RNA é efetuada em duas etapas: treinamento e inferência. Na etapa de treinamento, a RNA irá percorrer todo o conjunto de dados e codificar seu aprendizado na forma dos pesos W_i , representados na Figura 11. Esses valores são chamados de parâmetros, ao

passo que os valores que devem ser definidos manualmente pelo usuário, antes da execução da RNA, e nunca são alterados, são chamados de hiperparâmetros. Na etapa de inferência, a RNA utiliza o aprendizado adquirido na etapa de treinamento para processar corretamente novos dados apresentados em sua entrada. O sucesso de um modelo de rede neural depende basicamente da capacidade dela aprender as representações dos dados apresentados na etapa de treinamento. Um método de treinamento é o *Backpropagation*, que corrige os parâmetros da RNA propagando o erro para as camadas anteriores. Este método será explicado com mais detalhes na Seção 3.2.

Por sua vez, as funções de ativação apresentam uma importante função nas redes neurais, estando relacionadas com o sucesso das DNN atuais (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A próxima Seção irá discutir as principais funções de ativação e suas características.

Na Seção 3.1, algumas funções de ativação serão discutidas. Na Seção 3.2, o treinamento das RNA será explicado. Na Seção 3.3, as Redes Neurais Convolucionais serão apresentadas. Na Seção 3.4, os *Frameworks* disponíveis para trabalhar com RNAs serão discutidos.

3.1 Funções de Ativação

O valor de Y' , mostrado na Figura 11 e expresso pela Equação 3.1, é resultado do produto escalar entre os vetores $(X_1, X_2, X_3, \dots, X_N)$ e $(W_1, W_2, W_3, \dots, W_N)$ somado com o bias θ . Dessa maneira, o valor de Y' está relacionado de forma linear com os dados de entrada, não apresentando as variações não-lineares que permitem ao neurônio aprender padrões complexos. Por isso, o Perceptron possui a função de ativação não-linear $S(Y')$, permitindo que o valor de Y' seja modificado para se adaptar aos padrões não-lineares. Consequentemente, uma RNA com mais de uma camada que não utiliza uma função de ativação irá se comportar como um Perceptron simples, independente da quantidade de camadas, pois a composição de várias funções lineares resulta em outra função linear.

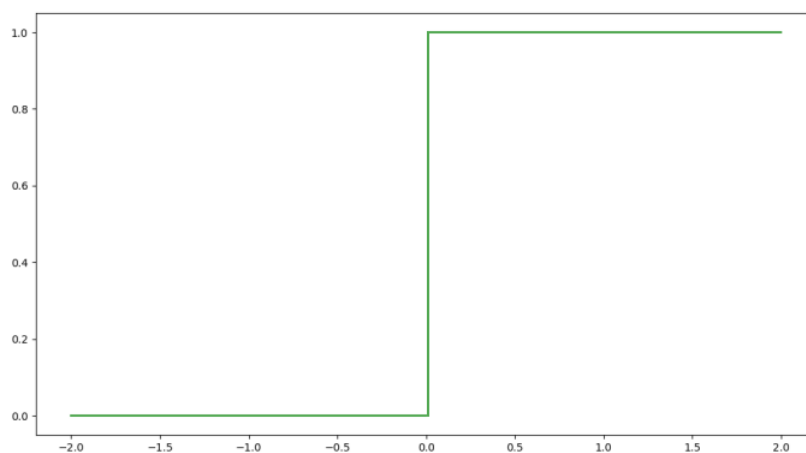
A função de ativação possui grande importância no desempenho da RNA, sendo que sua escolha impacta nos resultados obtidos. A seguir é feita uma discussão de algumas funções de ativação:

- **Função Limiar:** utiliza um valor limite α para decidir se o valor de saída será ativado (1) ou desativado (0). Esse limite é um hiperparâmetro, pois o usuário deve defini-lo antes da execução da RNA. A Equação 3.2 expressa a função de ativação Limiar (MCCULLOCH;

PITTS, 1943), enquanto a Figura 12 mostra seu gráfico. Essa função é útil quando se deseja utilizar uma RNA para classificar dados em 2 classes (classificadores binários), mas apresenta problemas quando se deseja utilizar mais classes. Por exemplo, em um conjunto de dados com 3 ou mais classes, é possível utilizar uma arquitetura de RNA com 3 ou mais neurônios da camada de saída, cada um representando uma classe. Espera-se que, ao classificar um dado, apenas o neurônio correspondente à classe correta seja ativado (1), enquanto todos os outros fiquem desativados (0). Portanto, há a possibilidade que mais de um neurônio seja ativado ao mesmo tempo, tornando difícil a convergência e interpretação dos resultados. Uma solução seria a utilização de probabilidades (como na função Sigmóide, analisada mais adiante) ao invés de valores binários, pois seria muito mais difícil a ocorrência de dois ou mais neurônios emitindo o mesmo valor.

$$\text{Limiar}(Y') = \begin{cases} 0, & \text{se } Y' \leq \alpha \\ 1, & \text{caso contrário} \end{cases} \quad (3.2)$$

Figura 12 – Função de ativação Limiar com $\alpha = 0$.



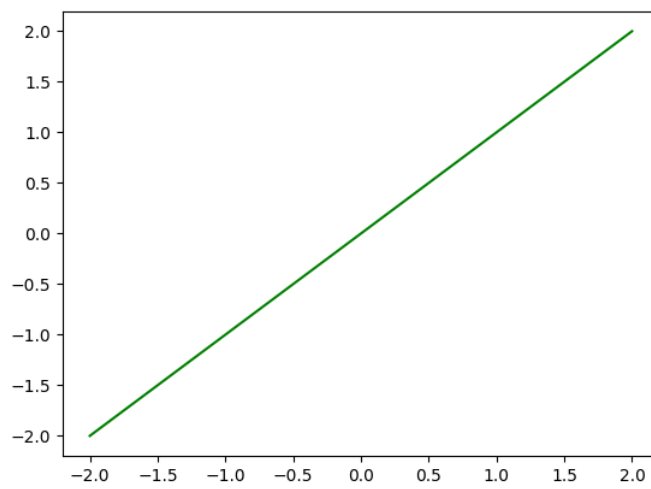
Fonte – Autoria própria.

- **Função Linear:** Como dito anteriormente, é interessante que os neurônios na camada de saída emitam valores contínuos, facilitando a interpretação e o treinamento da rede neural. Uma alternativa para isso é a utilização de uma função de ativação linear, representada pela Equação 3.3 (ROSENBLATT, 1958), com β sendo um hiperparâmetro. O gráfico da função Linear é expresso pela Figura 13. O valor de saída do neurônio é proporcional ao valor de entrada, mas se mais de um neurônio for ativado ao mesmo tempo na camada de

saída, pode-se obter aquele que apresenta o maior valor. Entretanto, a utilização de uma função de ativação linear faz com que, como discutido anteriormente, a RNA não seja capaz de aprender padrões não-lineares.

$$\text{Linear}(Y') = \beta Y' \quad (3.3)$$

Figura 13 – Função de ativação Linear com $\beta = 1$.



Fonte – Autoria própria.

- **Função Sigmóide (Logística):** Essa função de ativação, expressa pela Equação 3.4 (GLOROT; BORDES; BENGIO, 2011), ao contrário da função Linear, consegue aprender os padrões complexos presentes nos dados de entrada. Conseqüentemente, é possível criar RNAs com diversas camadas, pois combinações de diversas funções Sigmóide resulta em uma função não-linear (ao contrário da função Linear, onde combinações resultavam em outra função linear). Além disso, o valor resultante está presente no intervalo $[0, 1]$, possibilitando a interpretação como probabilidades, em contraste da função Linear, cujo resultado está contido em $[-\infty, \infty]$.

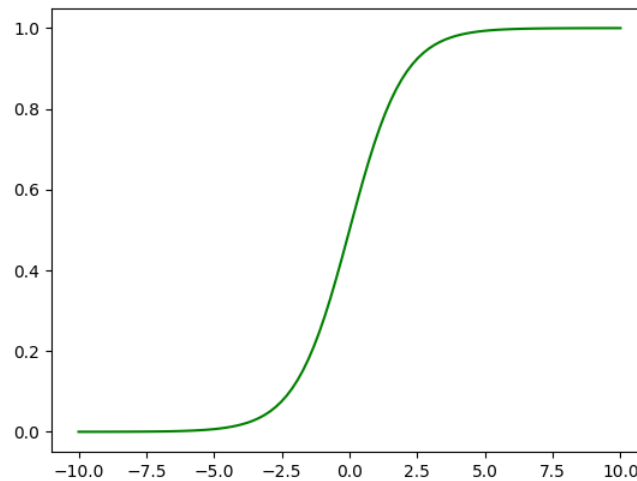
Como pode ser visto na Figura 14, a função Sigmóide no intervalo $[-2, 2]$ possui grande inclinação, sendo que pequenas variações no eixo horizontal resultam em grandes variações no eixo vertical, resultando na tendência de levar o resultado para os finais da curva.

Portanto, devido à pouca variação presente na função quando $Y' \geq 2$ e $Y' \leq -2$, a etapa de aprendizado pode sofrer do problema conhecido como “Desaparecimento dos gradientes”

(*Vanishing Gradient*), quando os valores são tão pequenos que a rede demora muito ou se torna incapaz de aprender mais.

$$\text{Sigmoide}(Y') = \frac{1}{1 + e^{-Y'}} \quad (3.4)$$

Figura 14 – Função de ativação Sigmóid.



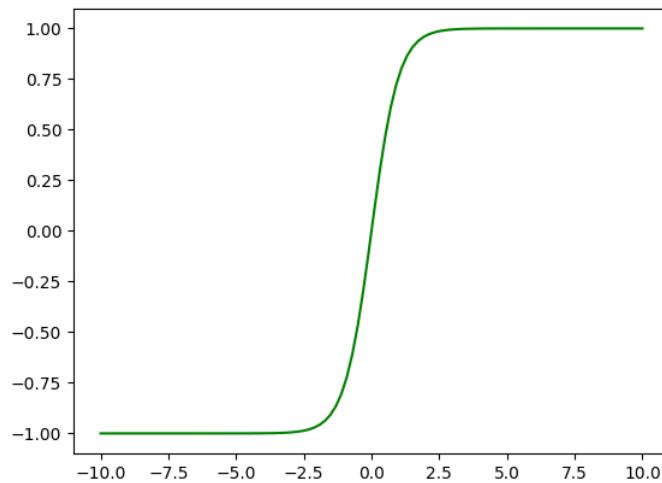
Fonte – Autoria própria.

- **Função Tanh (Tangente Hiperbólica:)** A função de ativação Tanh é uma versão escalada da Sigmóide, como expressa pela Equação 3.5 e pela Figura 15. Sendo uma função não-linear, é possível utilizá-la em uma RNA mais complexa. Além disso, seu contradomínio está contido no intervalo de $[-1, 1]$, evitando que o valor de Y seja muito grande (como acontece na função Linear).

Portanto, a principal diferença em relação à função Sigmóide está no fato de que a derivada da função Tanh é mais impactante no processo de treinamento. A Figura 15 mostra que a inclinação da curva na região central é maior. Mesmo assim, a escolha entre Sigmóide e Tanh depende dos requerimentos do problema a ser resolvido.

$$S(Y') = 2 \text{Sigmoide}(2Y') - 1 = \frac{2}{1 + e^{-2Y'}} - 1 \quad (3.5)$$

Figura 15 – Função de ativação Tanh.



Fonte – Autoria própria.

- **Função ReLU:** A função ReLU, como expressa pela Equação 3.6 (HAHNLOSER et al., 2000) e mostrada na Figura 16, apresenta uma reta nos valores positivos de Y' e é nula nos valores negativos de Y' . Mesmo assim ela é uma função não-linear.

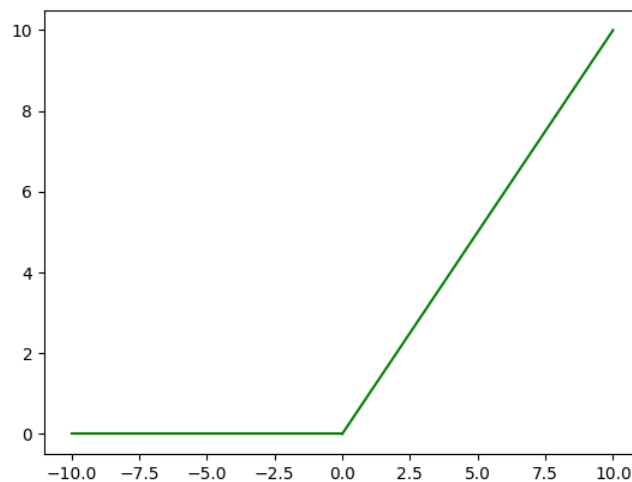
Pelo fato de que $ReLU(Y' < 0) = 0$, muitos neurônios em uma RNA estariam naturalmente desativados, economizando processamento. Por exemplo, em uma RNA com parâmetros $(W_1, W_2, W_3, \dots, W_n)$ inicializados aleatoriamente e com a média dos valores em 0, aproximadamente metade dos neurônios estariam desativados, efeito esse chamado de ativação esparsa. No caso da Sigmóide e da Tanh, todos os neurônios estariam ativados e teriam que ser levados em consideração no cálculo da saída, comportamento este chamado de ativação densa.

Um ponto negativo da ReLU é que quando $Y' > 0$ o resultado está contido em $[0, \infty)$, possibilitando que os valores resultantes sejam muito grandes, prejudicando a interpretação da RNA. Quando $Y' \leq 0$, o valor resultante é 0, formando uma linha horizontal, prejudicando o processo de treinamento. Esse é um problema conhecido como *Dying ReLU*. Existe uma variação da função ReLU que consiste em simplesmente mudar a componente horizontal para uma reta um pouco inclinada, como a ReLU Paramétrica, mostrada na Equação 3.7.

$$ReLU(Y') = \max(0, Y') \quad (3.6)$$

$$\text{ReLUParamétrica}(Y') = \begin{cases} Y', & \text{se } Y' > 0 \\ 0.01Y', & \text{otherwise} \end{cases} \quad (3.7)$$

Figura 16 – Função de ativação ReLu.



Fonte – Autoria própria.

Entretanto, não existe uma função de ativação definitiva, cada uma pode ser melhor adaptada em certos tipos de problemas.

3.2 Treinamento

Como já mencionado anteriormente, a fase de treinamento é extremamente importante para o sucesso de uma rede neural, pois é ela que determinará se a rede será capaz de generalizar o conjunto de treinamento de modo a efetuar inferências com boa precisão.

Quando uma RNA começa a ser treinada, seus parâmetros são atribuídos aleatoriamente, fazendo com que, provavelmente, qualquer exemplo colocado em sua entrada tenha a classificação incorreta. Entretanto, utiliza-se uma *função de custo* para avaliar quanto a rede neural está desviando das respostas corretas e, em seguida, efetuar a correção dos parâmetros por meio do algoritmo *Backpropagation*. Mais detalhes serão expostos nas Seções 3.2.1 e 3.2.2.

3.2.1 Função de Custo

Para que uma rede neural possa ajustar os seus parâmetros de modo a representar os dados corretamente, é necessário uma maneira de comparar a saída com os verdadeiros rótulos, calculando um erro. A escolha da função de custo é importante para o problema que se deseja resolver, havendo vários trabalhos na literatura que propõem funções para aplicações específicas, como é o caso da Perda Tripla (SCHROFF; KALENICHENKO; PHILBIN, 2015), utilizada em reconhecimento facial. Entretanto, há diversas funções de custo que podem ser usadas em tarefas mais genéricas, como é o caso do Erro Quadrático Médio, exibido na Equação 3.8. Os exemplos nesta Seção utilizarão essa função,

$$E_{total} = \frac{1}{n} \sum_{i=1}^n (Y_i^p - Y_i^l)^2. \quad (3.8)$$

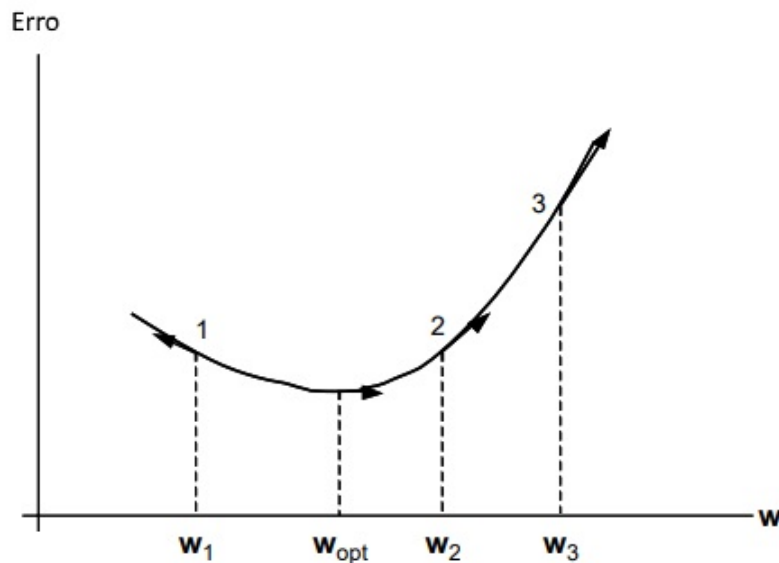
A variável n representa a quantidade de amostras no conjunto de treinamento, Y_i^p e Y_i^l representam, respectivamente, o valor predito pela RNA e o rótulo do i -ésimo exemplo do conjunto de treinamento. Essa função, assim como todas as funções de custo, tem por objetivo calcular quanto o resultado gerado está longe do resultado correto. O resultando é alto quando o modelo apresenta valores muito distantes do correto, e pequeno quando o modelo apresenta valores próximos ao correto.

3.2.2 Backpropagation

Em uma rede neural com várias camadas, os parâmetros presentes em cada neurônio devem ser modificados continuamente para que a rede apresente resultados mais precisos e, portanto, deixar a função de custo próxima a zero. Esse processo é iterativo, de tal modo que a cada etapa, um ou mais exemplos são colocados na entrada da rede, a função de erro é calculada e os parâmetros são corrigidos. A técnica de Gradiente Descendente é utilizada para a correção dos parâmetros em outras abordagens de aprendizado supervisionado (KIEFER; WOLFOWITZ et al., 1952). Em redes neurais artificiais, o mesmo conceito é utilizado, portanto, o erro é propagado de volta para cada neurônio da rede, passando a se chamar *Backpropagation* (WERBOS, 1974). Mais detalhes desta técnica serão fornecidos a seguir.

Gradiente Descendente, como o próprio nome diz, utiliza o gradiente da função de custo para descer na superfície do erro. Lembre-se que o gradiente de uma função, no cálculo vetorial, é um vetor que mostra a direção e o sentido do maior incremento, considerando um ponto no

Figura 17 – Parte da curva da função de custo. Os vetores representam os gradientes para quatro diferentes valores de peso w_1 , w_2 e w_3 e w_{opt} , onde w_{opt} é o valor de mínimo da função (ponto ótimo). Baseado em Samarasinghe (2016).



Fonte – Autoria própria.

espaço. Portanto, pode-se usar o sentido contrário do vetor gradiente para guiar as alterações nos parâmetros da rede neural até atingir um ponto de mínimo. O gradiente de uma função f é expresso pela Equação 3.9, sendo x , y e z coordenadas espaciais:

$$\nabla f = \frac{\partial f}{\partial x} \hat{e}_x + \frac{\partial f}{\partial y} \hat{e}_y + \frac{\partial f}{\partial z} \hat{e}_z. \quad (3.9)$$

Levando em consideração apenas a variável x , o gradiente de f ficaria conforme a Equação 3.10:

$$\nabla f = \frac{df}{dx}. \quad (3.10)$$

A Figura 17, mostra a curva de uma função de custo com apenas uma dimensão e quatro configurações de parâmetros. Observe que os valores de w_1 , w_2 e w_3 não são ótimos, portanto, por meio da direção inversa do gradiente (representado pelas setas), é possível encontrar o ponto ótimo w_{opt} , que é o menor valor da função.

Na Figura 17, o método Gradiente Descendente foi mostrado para uma função de custo de apenas uma variável, entretanto, ele pode ser generalizado para mais variáveis. Em uma rede neural, cada parâmetro é uma variável otimizável por meio da técnica *Backpropagation*. Cada

parâmetro em uma RNA de múltiplas camadas se relaciona com a função de custo por meio da regra da cadeia, mostrada na Equação 3.11:

$$\frac{\partial E_{total}}{\partial w_x} = \frac{\partial E_{total}}{\partial w_g} \cdot \frac{\partial w_g}{\partial w_x}, \quad (3.11)$$

sendo w_g um parâmetro que está entre a função de custo E e o parâmetro w_x . É importante mencionar que a função de custo deve ser diferenciável.

A regra da cadeia afirma que a derivada de uma função pode ser calculada por meio da derivada de funções que a compõem. Em redes neurais artificiais, ela permite que a derivada parcial da função de custo em relação a um parâmetro específico seja encontrada de uma forma computacionalmente eficiente. Como mostrado na Equação 3.9, a derivada parcial permite encontrar o vetor gradiente, que por sua vez mostrará qual o incremento que o determinado parâmetro terá que ter para minimizar a função de custo.

Por meio da regra da cadeia a derivada parcial da função de custo E_{total} pode ser calculada em relação a qualquer parâmetro w_x . Assim, todos os parâmetros da rede podem ser ajustados para minimizar o erro total da rede.

3.3 Redes Neurais Convolucionais

O surgimento da Rede Neural Convolucional (CNN, do inglês *Convolutional Neural Network*) representou um grande avanço para o reconhecimento e manipulação de imagens em geral (Gatys; Ecker; Bethge, 2016; CAO et al., 2016). Uma das primeiras CNNs foi apresentada por LeCun et al. (1998), que descreveu como utilizar uma rede neural para aprender e detectar formas 2D em imagens sem utilizar outros módulos externos, como extrator de características, segmentação, etc. Contudo, devido aos poucos recursos computacionais existentes na época e à baixa disponibilidade de conjuntos de dados de imagens para treinamento, seu trabalho foi pouco explorado pela comunidade científica. Mas com o rápido desenvolvimento da indústria dos computadores e, em especial, das placas gráficas (GPUs), os pesquisadores se voltaram novamente para essa abordagem e encontraram uma alternativa capaz de superar os métodos tradicionais de processamento de imagens existentes até então.

No ano de 2010, uma importante competição anual de reconhecimento visual iniciou-se, a *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC), onde pesquisadores puderam

Tabela 2 – Modelos ganhadores do desafio ILSVRC.

ILSVRC Ano	Erro Top-5	Camadas
2010	28,2	-
2011	25,8	-
2012 (AlexNet) (KRIZHEVSKY; SUTSKEVER; HINTON, 2012)	16,4	8
2013 (ZFNet) (ZEILER; FERGUS, 2014)	11,7	8
2014 (VGG) (SIMONYAN; ZISSERMAN, 2014)	7,3	19
2014 (GoogleNet) (SZEGEDY et al., 2015)	6,7	22
2015 (ResNet) (HE et al., 2015)	3,57	152

treinar e avaliar seus modelos de classificação de imagens em um grande conjunto de dados (DENG et al., 2009). Nas duas primeiras edições, abordagens que não eram baseadas em redes neurais conseguiram, respectivamente, um erro Top-5 de 28,2% e 25,8%. O Erro Top-5 é avaliado como a quantidade de vezes que a classe correta da imagem esteve entre as 5 classes mais prováveis preditas pelo modelo, dividido pela quantidade total de imagens.

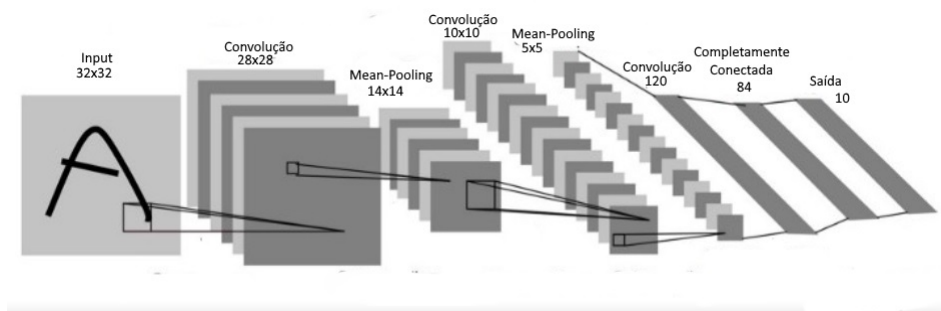
Na terceira edição, em 2012, Krizhevsky, Sutskever e Hinton (2012) apresentaram o primeiro modelo de CNN do desafio, conseguindo um erro Top-5 de 16,4%. Nos anos seguintes, os modelos propostos foram todos baseados em CNNs, diminuindo a taxa do erro Top-5 progressivamente. A Tabela 2 mostra o modelo (primeira coluna), a taxa de erro alcançada (segunda coluna) e a quantidade de camadas utilizadas (terceira coluna) dos modelos ganhadores do desafio. Observe que a maior quantidade de camadas está diretamente relacionada com uma menor taxa de erro, enfatizando o potencial das RNAs com grande quantidade de camadas, ou Redes Neurais Profundas. As linhas com fundo mais escuro mostram as abordagens que utilizam CNN. Essas arquiteturas serão analisadas com mais detalhes a seguir.¹

3.3.1 LeNet-5 (1998)

O modelo de rede neural LeNet-5 foi proposto por LeCun et al. (1998) e foi um dos primeiros trabalhos a utilizar redes neurais convolucionais no reconhecimento de imagens. O autor propôs a extração de características das imagens por meio da utilização de filtros (*kernels*), que podem ser representados por neurônios com entradas conectadas em um pequeno conjunto de pixels próximos uns aos outros na camada anterior. Essa ideia é chamada de campos receptores locais e é justificada pela descoberta da sensibilidade local no sistema visual de gatos por Hubel e Wiesel (1962). É importante mencionar que o trabalho de Fukushima (1980) é anterior ao de LeCun et al. (1998), e também apresenta uma abordagem para reconhecimento de

¹ Mais detalhes dos modelos de 2010 e 2011 podem ser encontrados no site oficial do ILSVRC: <http://image-net.org/>.

Figura 18 – Representação da arquitetura LeNet-5.



Fonte – Adaptado de (LECUN et al., 1998).

imagens baseada na ideia dos campos receptores locais. Ela apresenta os conceitos de células-S, células-C e planos-de-células, dispostos em uma estrutura hierárquica, utilizando aprendizado supervisionado. Entretanto, foi o trabalho de LeCun et al. (1998) que popularizou as RNAs para o reconhecimento de imagens, o qual será discutido com mais detalhes a seguir.

Na etapa de treinamento, os filtros aprendem as características presentes nas imagens de uma determinada classe. Após o treinamento, quando uma nova imagem é apresentada para a rede, os filtros são utilizados para verificar se essas características existem na nova imagem. Esse processo se chama convolução, no qual os filtros percorrem toda a imagem montando um mapa de ativação, estrutura que resume a localização de todas as ocorrências do filtro na imagem.

O neurônio convolucional possui a mesma estrutura daquele mostrado na Figura 11. É importante destacar que o mapa de ativação resultante já está submetido à função de ativação. A junção do neurônio convolucional com o mapa de ativação é chamada de camada convolucional. É possível unir várias camadas convolucionais, sendo a saída da anterior (mapa de ativação), a entrada da outra. O nível de abstração das características presentes no conjunto de dados aumenta conforme mais camadas são inseridas. Ou seja, em uma CNN já treinada, os filtros de camadas mais profundas contém abstrações muito mais complexas do que os filtros das primeiras camadas.

A variação do número de camadas e de outras configurações resultam em diferentes arquiteturas de CNNs. No caso da arquitetura LeNet-5, mostrada na Figura 18, foram utilizadas 3 camadas convolucionais, 2 camadas de *pooling* (subamostragem) e 2 camadas completamente conectadas. A camada de *pooling* tem por objetivo diminuir o tamanho do mapa de ativação anterior, assim, economizando custo computacional. As camadas convolucionais e de *pooling* foram colocadas em sequência. No final da rede, duas camadas completamente conectadas foram inseridas em sequência para classificar os mapas de ativação resultantes.

3.3.2 AlexNet (2012)

As CNNs como propostas por LeCun et al. (1998) foram pouco exploradas nos anos seguintes, até que Krizhevsky, Sutskever e Hinton (2012) apresentaram uma arquitetura com mais camadas, representada na Figura 19, para atingir o então estado da arte em classificação de imagens no desafio ILSVRC (RUSSAKOVSKY et al., 2015). A rede neural proposta possui 5 camadas convolucionais e 2 camadas completamente conectadas. Devido ao alto custo computacional, os autores dividiram o processamento em 2 GPUs GTX 580 3GB. Esse modelo é notável por usar a função de ativação ReLU (Equação 3.6), que apresenta uma velocidade de treinamento muito maior que a Tanh (Equação 3.5).

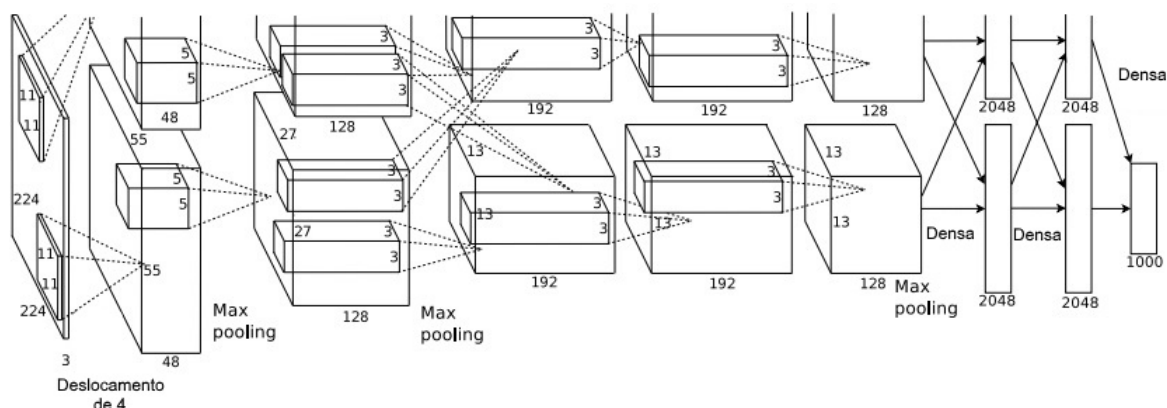
O aumento na quantidade de camadas em redes neurais eleva também a tendência de sobreajuste (*overfitting*), quando a rede neural passa a diminuir o erro no conjunto de treinamento e, ao mesmo tempo, aumentar o erro no conjunto de validação (a rede passa a decorar o conjunto de dados, perdendo a capacidade de classificar novos exemplos). Para diminuir esse efeito, técnicas de regularização são utilizadas. A arquitetura AlexNet utilizou a técnica chamada de *Dropout* (HINTON et al., 2012), que tem por objetivo desativar conexões aleatórias entre os neurônios de camadas adjacentes em cada iteração na fase de treinamento, fazendo com que os neurônios que ficaram ativos sejam forçados a aprender características extras que melhoram o desempenho do modelo final.

Além disso, Krizhevsky, Sutskever e Hinton (2012) utilizaram *max-pooling* (diminuição das dimensões de um mapa de ativação obtendo o maior valor em um conjunto de valores), aumentando a acurácia da AlexNet, enquanto que na arquitetura LeNet-5 foi utilizado *mean-pooling* (diminuição das dimensões de um mapa de ativação por meio da média dos valores). O trabalho mostrou ainda que a remoção de uma única camada convolucional do modelo degrada consideravelmente a acurácia final, devido ao fato que as características aprendidas pelos filtros são hierárquicas.

3.3.3 ZFNet (2013)

Por meio de uma análise da arquitetura AlexNet, o trabalho de Zeiler e Fergus (2014) propôs alterações nos hiperparâmetros utilizados, conseguindo um resultado superior ao apresentado por Krizhevsky, Sutskever e Hinton (2012). Uma das mudanças foi o tamanho dos filtros.

Figura 19 – Representação da arquitetura AlexNet.



Fonte – Extraído de Russakovsky et al. (2015).

Enquanto que na arquitetura AlexNet foram utilizados filtros de tamanho 11x11, a ZFNet utilizou filtros de tamanho 7x7. Para efetuar essa mudança, o autor utilizou o método de deconvolução (ZEILER; TAYLOR; FERGUS, 2011) para projetar as características dos filtros de volta para o espaço de pixels, possibilitando que formas identificáveis pudessem ser visualizadas. Assim, foi possível identificar alterações nos hiperparâmetros que resultaram em uma melhora no desempenho da arquitetura.

Além disso, o trabalho ainda efetuou uma análise de sensibilidade escondendo partes da imagem para verificar quais são importantes para a classificação da mesma. Com isso, foi possível verificar se as CNNs possuem um bom desempenho identificando os principais objetos presentes na imagem, ou se outras características menos importantes (como plano de fundo, objetos secundários, etc) influenciam nos resultados. Os experimentos mostraram que os objetos principais das figuras analisadas desempenham papel fundamental para a classificação na categoria correta.

3.3.4 VGG (2014)

A arquitetura proposta por Simonyan e Zisserman (2014), chamada de VGG, obteve o segundo melhor resultado na classificação de imagens no desafio ILSVRC 2014 (ver Tabela 2), com Erro Top-5 de 7.3%, perdendo apenas para a arquitetura Inception (descrita com mais detalhes na Seção 3.3.5), que por sua vez obteve um Erro Top-5 de 6.7%.

A arquitetura VGG está disponível com quantidades diferentes de camadas, entre elas, as mais conhecidas são a VGG 16 e VGG 19, respectivamente com 13 e 16 camadas convolucionais,

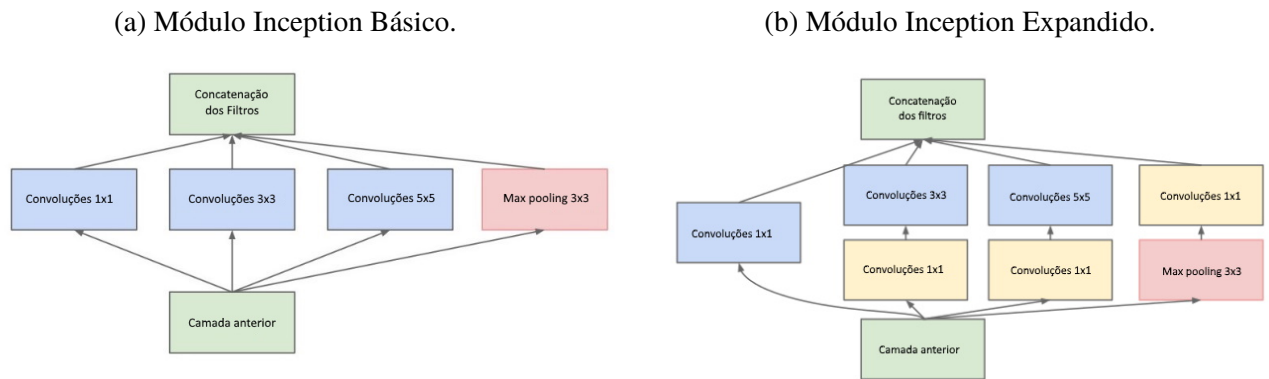
ou 16 e 19 camadas, se levar em consideração as 3 camadas completamente conectadas no final. Elas tem um aspecto muito uniforme, cuja principal diferença da arquitetura AlexNet (Seção 3.3.2) é a quantidade de camadas e o tamanho do filtro nas camadas convolucionais. Nas arquiteturas VGG, o filtro utilizado tem tamanho reduzido para 3x3. Isso é justificado pelo fato de que os grandes filtros são aqui simulados pela sequência de várias camadas convolucionais com filtros 3x3 em sequência.

3.3.5 Inception (2014)

Pesquisadores do Google apresentaram a primeira versão da arquitetura conhecida como GoogLeNet (SZEGEDY et al., 2015), em homenagem à Yann LeCun, autor do trabalho precursor das CNNs (Seção 3.3.1). A arquitetura apresentou 22 camadas, contando apenas as que possuem parâmetros (ou 27 contando as camadas de *pooling*), entretanto, com uma quantidade 12 vezes menor de parâmetros que a AlexNet (Seção 3.3.2), sendo muito menos custosa computacionalmente. Além de ser mais rápida para ser processada, ela conseguiu uma acurácia maior que suas predecessoras. Isso foi possível graças aos Módulos *Inception*, que são operações realizadas em paralelo entre duas camadas, como pode ser vista na Figura 20a. Os Módulos Inception são as unidades básicas da arquitetura, podendo ser empilhados para formar a rede neural final. É possível observar que esses módulos possuem operações de convolução utilizando filtros de tamanho 1x1, 3x3 e 5x5 em paralelo, para que características de diversos tamanhos possam ser detectadas. O trabalho argumenta que as informações visuais devem ser processadas em várias escalas no mesmo momento e, depois, concatenadas para que a próxima camada possa processar tais informações simultaneamente.

O problema é que a utilização de vários filtros torna a rede muito lenta para ser processada, devido à grande quantidade de mapas de ativação fornecidos para as camadas seguintes. Portanto, para resolver esse problema, os autores colocaram operações adicionais de convolução 1x1 antes das operações de convolução 3x3 e 5x5, para diminuir a dimensionalidade dos mapas de ativação, controlando a complexidade da rede e o tempo de processamento. O resultado pode ser visto na Figura 20b.

Figura 20 – Módulos Inception.



Fonte – Adaptado de Szegedy et al. (2015).

Para alcançar uma quantidade de parâmetros 12 vezes menor que a AlexNet, os autores substituíram a camada completamente conectada no final da rede por uma camada de *mean-pooling*. Assim como na convolução, os neurônios utilizados em uma camada de *mean-pooling* possuem entradas conectadas a uma pequena quantidade de pixels vizinhos na camada anterior, tornando o processamento muito mais rápido em comparação a uma camada completamente conectada.

Cada vez que as arquiteturas se tornam mais profundas, o problema de *overfitting* fica mais presente. Para contornar esse problema, os autores da GoogLeNet colocaram classificadores auxiliares no meio da rede, que não influenciavam diretamente no resultado final mas auxiliava na contenção do *overfitting*. Entretanto, no início de 2015, o trabalho de Ioffe e Szegedy (2015) apresentou o conceito de Normalização em Lote (*Batch Normalization*), que foi inserido na arquitetura Inception, passando a se chamar Inception V2. Basicamente, esse método calcula a média e o desvio padrão de todos os mapas de ativação e, em seguida, normaliza-os para que seus valores estejam no mesmo intervalo. Essa técnica possibilitou o controle eficiente do *overfitting* em redes profundas, sendo que até então isso era feito com métodos pouco elegantes.

3.3.6 ResNet (2015)

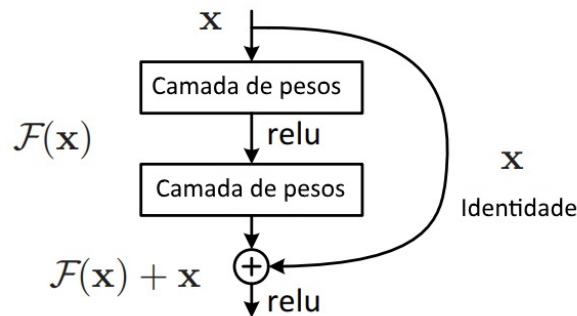
Como visto nas Seções anteriores, diversas arquiteturas tem conseguido bons resultados na classificação de imagens aumentando a quantidade de camadas. Portanto, uma questão naturalmente levantada é se existe algum limite para a adição de mais camadas em DNNs, de modo que a acurácia não apresente mais melhora. Alguns obstáculos conhecidos são o *overfitting*, o desaparecimento de gradiente (discutido brevemente na função de ativação Sigmóide, na Seção

3.1) e a explosão de gradientes, que é semelhante ao desaparecimento de gradiente, porém ocorre quando os valores dos gradientes são muito altos, dificultando o treinamento da RNA (BENGIO; SIMARD; FRASCONI, 1994; GLOROT; BENGIO, 2010). Esses problemas podem ser amenizados com a utilização da técnica Normalização em Lote (IOFFE; SZEGEDY, 2015). Mesmo assim, He e Sun (2014), Srivastava, Greff e Schmidhuber (2015) observaram que um modelo de RNA com muitas camadas é impedido de convergir devido ao problema da degradação, que é caracterizado pela saturação da acurácia à medida que se coloca mais camadas. Esse problema é diferente do *overfitting*, uma vez que a degradação aumenta o erro no conjunto de treinamento.

Por exemplo, considere que uma RNA com 10 camadas possibilite uma determinada acurácia. Se mais 5 camadas fossem adicionadas ao final dessa rede, espera-se que, ao menos, a mesma acurácia seja atingida. Isso deveria acontecer pois a nova rede pode copiar os mesmos parâmetros nas 10 primeiras camadas e depois copiar as saídas das últimas 5 camadas. Entretanto, não é isso o que acontece, uma rede com mais camadas sofre dificuldade na fase de treinamento sem nenhum motivo aparente. Dessa maneira, He et al. (2015) apresenta a abordagem de aprendizado residual para contornar esse problema.

Os autores argumentam que redes neurais com muitas camadas tem dificuldade em chegar às funções ótimas a partir do zero. Deste modo, a abordagem proposta tem por objetivo auxiliar o aprendizado fornecendo às camadas um ponto de partida para encontrar a função que melhor aproxima as características do conjunto de dados. Para isso, o resultado de uma camada anterior é copiado para a camada atual, fornecendo um auxílio para uma função mais complexa ser aproximada. Essa “ligação direta” não necessita de nenhum parâmetro adicional e é chamada de aprendizado residual profundo. Na Figura 21 é mostrado o bloco básico de formação das redes neurais residuais (ResNets), o bloco residual. Nele, a saída de uma camada convolucional (x) é copiada para a saída da camada convolucional duas posições à frente ($\mathcal{F}(x) + x$). É importante mencionar que a função de ativação é aplicada somente depois dessa soma.

Figura 21 – Blocos residuais.



Fonte – Adaptado de He et al. (2015).

Dessa maneira, os autores conseguiram aumentar a quantidade de camadas desviando do problema de degradação. Para o desafio ILSVRC, uma Rede Neural Residual com 152 camadas foi treinada, conseguindo um Erro Top-5 de 3,57%.

3.3.7 MobileNet (2017)

Nos últimos anos surgiu a demanda por arquiteturas que pudessem ser utilizadas em tempo real e que mantivessem sua qualidade, podendo ser aplicadas em situações que necessitam de baixa latência, como em *streaming* de vídeo. Além disso, houve também a necessidade de arquiteturas mais compactas, que pudessem ser utilizadas em sistemas embarcados e em smartphones com capacidade de processamento reduzida. Pensando nisso, o trabalho de Howard et al. (2017) propôs uma abordagem chamada de convolução separada por profundidade, na qual é uma maneira mais eficiente de se efetuar convoluções em CNNs, ao custo de uma piora aceitável na acurácia. Contudo, os autores também propuseram dois hiperparâmetros que servem para controlar o balanço entre a complexidade do modelo e a acurácia final, cabendo ao usuário fazer a melhor escolha se baseando na situação em que esse modelo seria utilizado. Esses dois parâmetros são: multiplicador de largura, que reduz o número de canais dos mapas de ativação, e multiplicador de resolução, que diminuiu a resolução da imagem de entrada e dos mapas de ativação nas camadas intermediárias.

Para iniciar a discussão do método proposto por Howard et al. (2017), primeiramente é necessário formalizar o modelo tradicional de convolução. Uma CNN comum tem como entrada um mapa de ativação F com dimensões $D_f \times D_f \times M$, um filtro K de dimensões

$D_k \times D_k \times M \times N$ e resulta em um mapa de ativação G com dimensões² $D_f \times D_f \times N$. Os valores D_f e M representam, respectivamente, a dimensão espacial (largura e altura) e a quantidade de canais do mapa de ativação de entrada, enquanto que N representa o número de canais do mapa de ativação de saída. O filtro K é convolvido sobre o mapa de ativação F de acordo com a Equação 3.12 (HOWARD et al., 2017):

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} F_{k+i-1,l+j-1,m}. \quad (3.12)$$

Observando a Equação acima, é possível notar que a complexidade da convolução tradicional está relacionada com as dimensões espaciais do mapa de ativação de entrada e com as dimensões do filtro. Ou seja, a complexidade algorítmica pode ser descrita por $O(D_K \times D_K \times M \times N \times D_F \times D_F)$, assim como feito por Howard et al. (2017).

Na arquitetura MobileNet, o processo de convolução é dividido em 2 etapas. A primeira, chamada de convolução por profundidade (*depthwise convolutions*), recebe na entrada um mapa de ativação de dimensões $D_f \times D_f \times M$ e efetua uma convolução com M filtros K' de dimensões $D_k \times D_k \times 1$, resultando em um mapa de ativação G' de dimensões $D_f \times D_f \times M$. Essa primeira etapa é expressa pela Equação 3.13 (HOWARD et al., 2017):

$$G'_{k,l,m} = \sum_{i,j,m} K'_{i,j,m} F_{k+i-1,l+j-1,m}. \quad (3.13)$$

Observe que a complexidade algorítmica dessa primeira etapa pode ser representada por $O(D_k \times D_k \times M \times D_f \times D_f)$, assim como feito por Howard et al. (2017).

A segunda etapa, chamada de convolução por pontos (*pointwise convolution*), recebe como entrada o mapa de ativação resultante da etapa anterior, de dimensões $D_f \times D_f \times M$, e efetua uma convolução com N filtros de dimensões 1×1 , resultando em uma saída de $D_f \times D_f \times N$. A complexidade dessa etapa é, portanto, $O(M \times N \times D_f \times D_f)$.

Somando as complexidades das duas etapas, tem-se que a complexidade total de uma convolução na arquitetura MobileNet é de $O(D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f)$. Dividindo a complexidade da convolução tradicional por aquela encontrada na convolução da

² Considere que o mapa de ativação resultante tenha as mesmas dimensões espaciais que o mapa de ativação de entrada.

MobileNet, tem-se a redução de custo, que é representado pela Equação 3.14 (HOWARD et al., 2017):

$$R = \frac{D_K \times D_K \times M \times N \times D_F \times D_F}{D_k \times D_K \times M \times D_f \times D_f + M \times N \times D_f \times D_f} = \frac{1}{N} + \frac{1}{D_k^2}. \quad (3.14)$$

Portanto, em termos de complexidade algorítmica, a arquitetura MobileNet é melhor que a convolução empregada pelas arquiteturas analisadas em seções anteriores. Além disso, em termos de qualidade de classificação de imagens, os autores mostram que se tem apenas uma perda pequena de qualidade. Se algum usuário necessitar de maior velocidade ou menor quantidade de parâmetros, ele pode alcançar isso utilizando os parâmetros multiplicador de largura e multiplicador de resolução, ao custo de qualidade de classificação.

3.4 Frameworks

Existem diversas ferramentas e *frameworks* que possibilitam ao usuário trabalhar com redes neurais artificiais facilmente. Tais *frameworks* também oferecem a vantagem de utilizar todo o poder de processamento disponível em vários núcleos e em placas de processamento gráfico. As redes neurais artificiais, assim como as biológicas, possuem a vantagem de serem altamente paralelizáveis, tornando essas ferramentas aliadas no desenvolvimento. Abaixo seguem informações sobre os principais *frameworks* disponíveis atualmente:

- **Tensorflow (ABADI et al., 2016):** Essa é uma biblioteca de código aberto desenvolvida pelo Google, que é utilizada para pesquisa e produção dos produtos Google. O seu antecessor, o DistBelief, foi criado inicialmente como uma biblioteca de Aprendizado Profundo de código fechado com grande ênfase em processamento distribuído, por meio do algoritmos de treinamento distribuído Downpour SGD e do *framework* Sandblaster (ANIL et al., 2018). Em 2015, o TensorFlow foi liberado e o seu código fonte disponibilizado para a comunidade. As operações no TensorFlow são altamente paralelizáveis, escondendo a complexidade de utilizar diversas CPUs e GPUs do usuário. As camadas das redes neurais são representadas como arranjos de dados denominados *Tensores*. Internamente, uma rede neural nada mais é do que grafos orientando o fluxo de cálculo dos tensores. Além disso, o

framework possui uma intuitiva plataforma gráfica para visualizar os modelos criados. As linguagens suportadas oficialmente são Python, C++, Java e GO.

Em 2016, o Google anunciou a Unidade de Processamento de Tensores (TPU, do inglês *Tensor Processing Unit*) (ROSS et al., 2017), um hardware de propósito específico para aumentar o desempenho da inferência em redes neurais. O Google iniciou o desenvolvimento de tal hardware após perceber que teria que dobrar a capacidade de seus servidores se a demanda de processamento de DNNs em seus produtos continuasse a crescer em tal proporção (JOUPPI et al., 2017). O TensorFlow oferece suporte para a utilização de TPU quando hospedado na *Google Cloud* (GCP).

- **Caffe (JIA et al., 2014):** Arquitetura Convolutacional para Incorporação Rápida de Recursos (Caffe, do inglês *Convolutional Architecture for Fast Feature Embedding*) é escrito em C++, mas oferece interfaces em várias linguagens, como Python, C, C++, Matlab e linha de comando. Ele foi desenvolvido pela Universidade da Califórnia, inicialmente como uma solução para o processamento eficiente de CNN, e seu código está disponível para a comunidade. Atualmente é possível executar modelos de CNN, LSTM e redes completamente conectadas. Como o TensorFlow, o Caffe possui suporte para processamento em CPU e GPU.

Baseado no Caffe, o Facebook tomou a iniciativa da segunda versão desse *framework*, o Caffe2 (CAFFE2,). Nesta nova versão, os desenvolvedores focaram no processamento distribuído e para padrões de desenvolvimento mais atuais, além da necessidade de execução em dispositivos móveis e de mais modelos voltados para outras tarefas além da visão computacional.

- **PyTorch (PASZKE et al., 2017):** PyTorch é um *framework* de Aprendizado de Máquina que utiliza tensores (assim como o TensorFlow), escrito em C e com interface disponível na linguagem Lua. PyTorch utiliza o núcleo do Torch e oferece uma interface em Python, possibilitando também a aceleração via GPU e manipulação de matrizes multidimensionais. Uma vantagem do PyTorch é a possibilidade de reutilização de vários pacotes do Python, como NumPy, SciPy e Cython. Isso atrai a atenção dos programadores pois permite a programação em um estilo que ele já está familiarizado, em vez de escrever o código especificamente no ambiente do PyTorch.

3.5 Considerações Finais do Capítulo

Este Capítulo possibilitou uma visão geral dos últimos trabalhos disponíveis na literatura relacionados a redes neurais artificiais no processamento de imagens. No próximo Capítulo serão mostrados como RNAs podem ajudar na reconstrução e processamento de imagens de Tomossíntese Mamária Digital, além de discutir as técnicas de estado da arte em reconstrução de DBT e CT.

4 Reconstrução Tomográfica de DBT

Como discutido no Capítulo 2, exames de raios X, como DBT, CT ou exames mamográficos 2D convencionais, utilizam radiação emitida através do corpo a ser imageado, que são capturados por um detector que está do lado oposto. Durante o trajeto dos feixes de raios X, eles interagem com a matéria e sofrem atenuações, que podem variar dependendo do tecido ou de possíveis lesões. Dessa maneira, os raios incidentes no detector formam uma imagem. Em exames 2D convencionais, esta imagem é utilizada pelo especialista para fazer um diagnóstico, enquanto que em exames de DBT e CT, essa imagem será uma entre várias projeções utilizadas em um processo de reconstrução que resulta em uma imagem do corpo em 3D, só então analisada pelo especialista para fornecer um diagnóstico mais preciso. Esse processo é chamado de reconstrução tomográfica, e foi discutido brevemente na Seção 2.2.

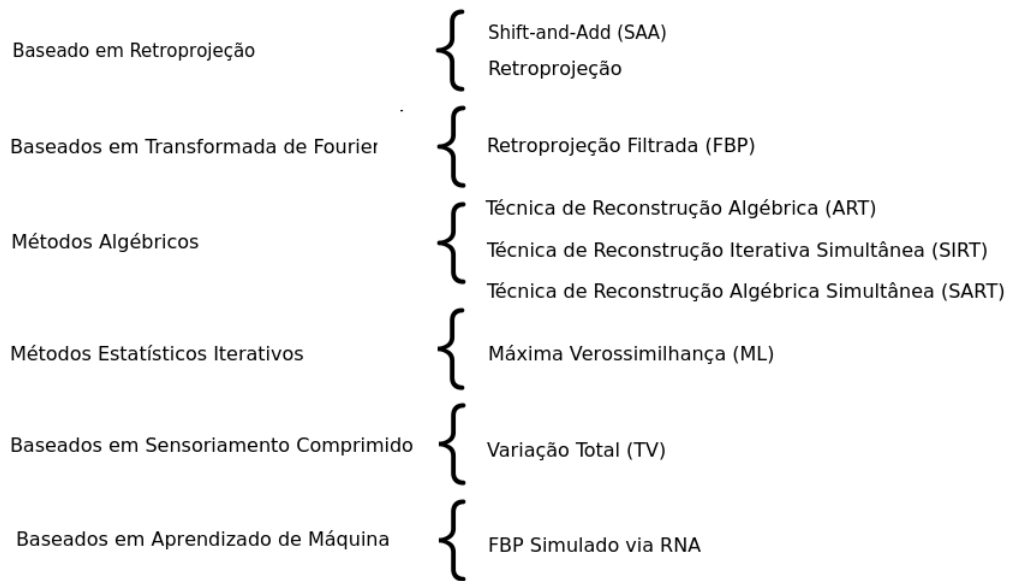
A reconstrução tomográfica, assim como outros problemas como superresolução, filtragem de ruído e retoque (*inpainting*), é caracterizada como um problema inverso, no qual se deseja saber a estrutura interna de um objeto dadas as observações externas. Matematicamente, um problema inverso pode ser caracterizado como (ULYANOV; VEDALDI; LEMPITSKY, 2018)

$$x^* = \min_x E(x, x_0) + R(x), \quad (4.1)$$

sendo $E(x, x_0)$ o termo de reconstrução, que depende diretamente da tarefa que se está executando, x_0 é a observação externa (projeções tomográficas, imagem degradada com ruído, baixa resolução, danificada, etc), $R(x)$ é um regularizador, que tem por objetivo auxiliar o processo de reconstrução inserindo um conhecimento prévio, e x é a imagem restaurada/reconstruída que se está buscando, que pode ser encontrada aplicando um processo de otimização.

Na Figura 22, é possível ver a classificação na qual diversas técnicas de reconstrução tomográfica de DBT se enquadram. Assim, neste Capítulo serão explicados os métodos Baseados em Retroprojeção (Seção 4.1), os baseados em Transformada de Fourier (Seção 4.2), os Métodos Algébricos (Seção 4.3), os métodos Estatísticos Iterativos (Seção 4.4), os baseados em Sensoriamento Comprimido (Seção 4.5) e os baseados em Aprendizado de Máquina (Seção 4.6). Esse último ganhou bastante atenção nos últimos anos devido ao sucesso das RNAs, e tem apresentado bons resultados na reconstrução tomográfica de imagens.

Figura 22 – Abordagens de reconstrução tomográfica.



Fonte – Autoria própria.

4.1 Reconstrução Baseada em Retroprojeção

Embora a Equação 4.1 estabeleça que a reconstrução de imagens tomográficas pode ser resolvida por meio de um processo de otimização, existem algumas possibilidades de encontrar a imagem reconstruída x^* de maneira analítica, as quais são explicadas com mais detalhes na próxima subseção.

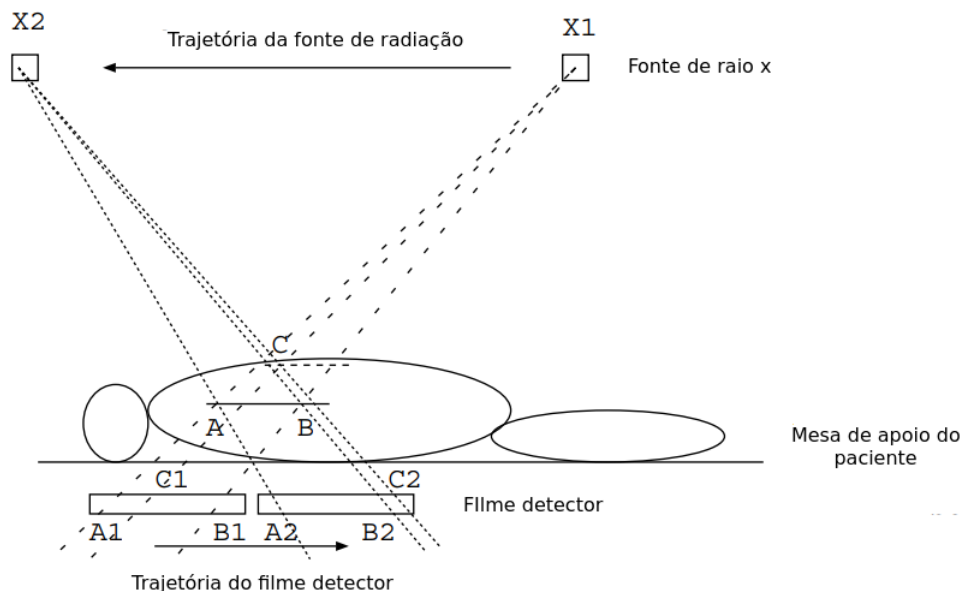
4.1.1 Shift-And-Add (SAA) e Backprojection (BP)

O método SAA leva em consideração somente o movimento do emissor e do detector de radiação para desfocar os planos que não são de interesse. Várias projeções são obtidas e depois, dependendo de qual seção se deseja focar, as projeções são deslocadas e adicionadas.

Para ilustrar esse processo, na Figura 23 tem-se a fonte de raios X em um movimento linear, cuja trajetória vai do ponto $X1$ ao ponto $X2$. O detector mantém uma trajetória síncrona com a fonte de radiação, se movimentando na mesma direção, porém em sentido oposto. No exemplo, a imagem será adquirida em relação ao plano frontal (Figura 7), indicado pelo segmento de reta AB , que estará em foco no filme detector, enquanto que as outras partes do corpo (em cima e em baixo) ficarão desfocadas. É possível observar que o feixe de fótons com início em $X1$ e trajetória passando por B ($X1-B$) e o feixe de fótons com início em $X2$ e trajetória passando por B ($X2-B$) atingirão o mesmo ponto no detector ($B1 = B2$). Da mesma maneira, o feixe de fótons

com início em X1 e trajetória passando por A (X1-A) e o feixe de fótons com início em X2 e trajetória passando por A (X2-A), atingirão o mesmo ponto no detector ($A1 = A2$). Isso ocorre pois os pontos A e B pertencem à mesma seção transversal em que se deseja obter a imagem. Por outro lado, considere o ponto C, que está fora do plano de interesse. O feixe de fótons com início em X1 e trajetória passando por C (X1-C), atingirá o ponto C1, enquanto que o feixe de fótons com início em X2 e trajetória passando por C (X2-C) atingirá o ponto C2, sendo estes pontos diferentes no filme detector ($C1 \neq C2$). Dessa maneira, a radiação incidente no plano AB faz com que os detalhes da seção transversal de interesse se mantenha em foco, enquanto que os detalhes de outros planos são borrados.

Figura 23 – Processo de aquisição de imagem utilizando tomografia clássica.



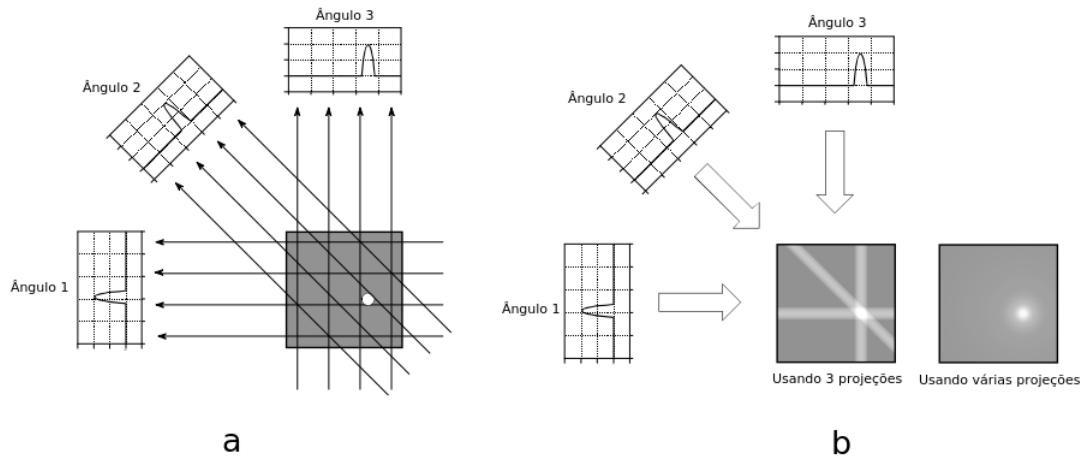
Fonte – Adaptado de Rangayyan (2004).

O método SAA é equivalente ao algoritmo Backprojection (BP), embora o BP tenha a capacidade de incorporar a geometria de movimento do emissor de radiação, enquanto que o SAA é válido apenas em movimento linear, paralelo e em distância fixa ao detector (YANG, 2012). Ambos os métodos são derivados dos princípios da Transformada de Radon (explicado com mais detalhes na Seção 4.2.1), uma vez que os valores das atenuações sofridas pelos raios no interior do corpo imageado são somadas, obtendo uma projeção, e então retroprojetado para formar a imagem reconstruída.

O conjunto de projeções adquiridas em vários ângulos diferentes é chamado de sinograma, que por sua vez pode ser usado para reconstruir a imagem. No algoritmo BP, cada projeção é

retrojetada, contribuindo igualmente para os pixels da nova imagem a ser formada. A Figura 24 ilustra o processo de retroprojeção.

Figura 24 – Ilustração da retroprojeção.



Fonte – Adaptado de Smith et al. (1997).

Na Figura 24 (a), é possível observar que a imagem original é um círculo branco em um fundo preto, no qual são obtidas 3 projeções a partir de um emissor de radiação com raios paralelos posicionado em ângulos diferentes. Na Figura 24 (b), observa-se que as projeções são retrojetadas para formar uma nova imagem, buscando igualar à imagem original. Portanto, como cada projeção contribui igualmente para os pixels da nova imagem, faixas indesejadas são formadas (utilizando 3 projeções), gerando borrão. É importante destacar também que, quanto mais projeções utilizadas, melhor o resultado, mas o borrão se torna algo inevitável, sendo necessária a utilização de outras técnicas para contornar esse problema.

Matematicamente, o processo de retroprojeção pode ser modelado como mostra a Equação 4.2 (YANG, 2012),

$$\hat{f}(x, y) \approx \int_0^\pi P_\theta(t) d\theta, \text{ sendo } t = x \cos \theta + y \sin \theta, \quad (4.2)$$

sendo $\hat{f}(x, y)$ a imagem reconstruída, $P_\theta(t)$ a projeção adquirida no ângulo θ e $t = x \cos \theta + y \sin \theta$ a equação do raio a uma distância t do centro. Esta equação diz que um determinado pixel (x, y) da imagem a ser reconstruída é a soma de todas as projeções de ângulo θ , formada pelo conjunto de raios paralelos $t = x \cos \theta + y \sin \theta$.

Na próxima Seção, será explicado o algoritmo FBP, que tem por objetivo minimizar o problema de borrão na imagem reconstruída aplicando um filtro nas projeções, antes de

retroprojetá-las.

4.2 Reconstrução Baseada em Transformada de Fourier

Nesta Seção será descrita a abordagem Retroprojeção Filtrada (FBP, do inglês *Filtered Backprojection*) (KAK; SLANEY, 1988). Contudo, alguns conceitos matemáticos introdutórios devem ser discutidos antes da Seção que introduz definitivamente tal método de reconstrução, por serem conceitos fundamentais no entendimento. Para tanto, a Seção 4.2.1 detalha o conceito da Transformada de Radon de uma função, a Seção 4.2.2 introduz o Teorema de Corte de Fourier, e, finalmente, a Seção 4.2.3 introduz o algoritmo Retroprojeção Filtrada.

4.2.1 Transformada de Radon

A Transformada de Radon (RADON, 1917), utilizada como princípio base do algoritmo Retroprojeção Filtrada, será discutida nesta Seção. Para iniciar a discussão, primeiramente considere feixes de raios X paralelos e um corpo imageado cuja distribuição é representada por $f(x, y)$, sendo (x, y) um ponto no espaço. A Figura 25 ilustra essa situação. As Equações apresentadas nesta Seção utilizam notação de espaço contínuo devido à simplicidade, embora aplicações reais sejam modeladas no espaço discreto.

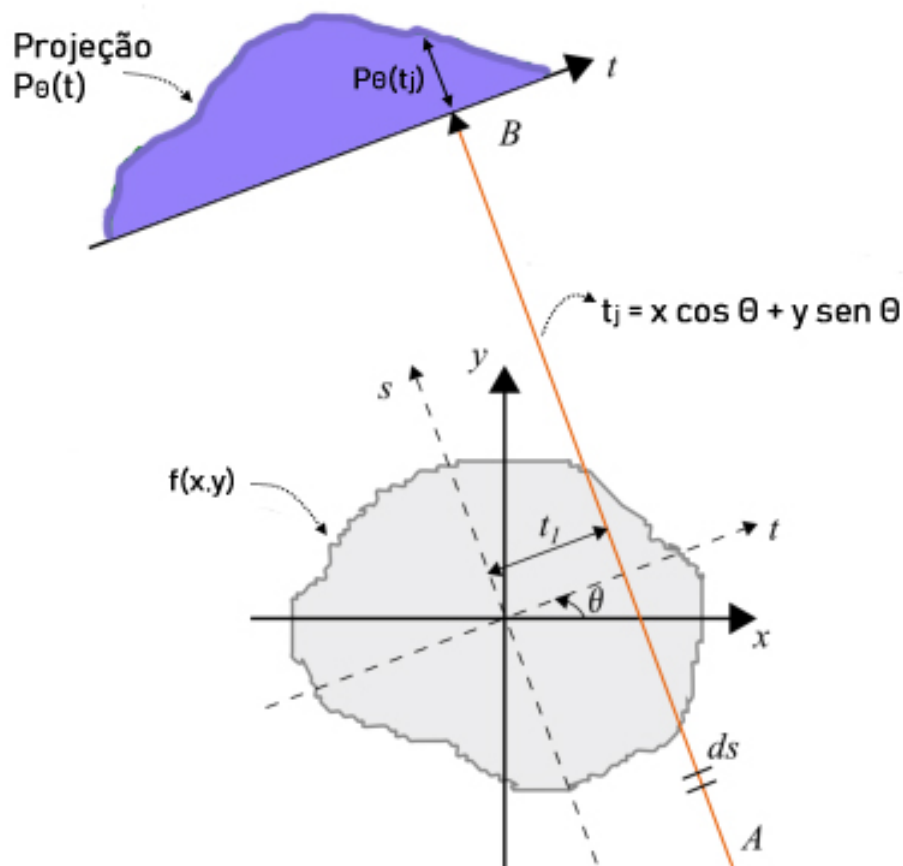
Assim como feito por Rangayyan (2004), um raio definido como um segmento \overline{AB} pode ser descrito por meio de coordenadas polares, representado pela Equação 4.3, onde t_1 é a distância do raio a um eixo central de mesma inclinação, representada por θ ,

$$x \cos \theta + y \sin \theta = t_1, \quad (4.3)$$

sendo que a atenuação que um determinado raio \overline{AB} sofre ao passar através do corpo $f(x, y)$ pode ser calculada somando todas as contribuições pontuais que $f(x, y)$ exerce ao longo de sua trajetória, conforme modelado pela Equação 4.4 (RANGAYYAN, 2004):

$$P_\theta(t_1) = \int_{AB} f(x, y) ds = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - t_1) dx dy. \quad (4.4)$$

Lembre-se que, em um exame de tomografia ou tomossíntese, várias projeções de diversos ângulos θ são obtidas. Dessa maneira, o resultado $P_\theta(t_1)$, levando em consideração

Figura 25 – Representação da transformada de Radon de uma função $f(x, y)$.

Fonte – Adaptado de Rangayyan (2004).

todos os valores de θ , é chamado de Transformada de Radon da função $f(x, y)$, e pode ser visto, opcionalmente, como uma função de 2 variáveis na Equação 4.4. É possível também que o volume reconstruído tenha 3 dimensões ($f(x, y, z)$) e, por sua vez, as projeções tenham 2 dimensões ($P_\theta(t_1, t_2)$). Os raios paralelos rotacionados em um ângulo θ em relação à base (x, y) são representados pelo sistema de coordenadas (t, s) , δ se refere ao delta de Dirac e $s = -x \sin \theta + y \cos \theta$ se refere à posição ao longo do feixe (RANGAYYAN, 2004).

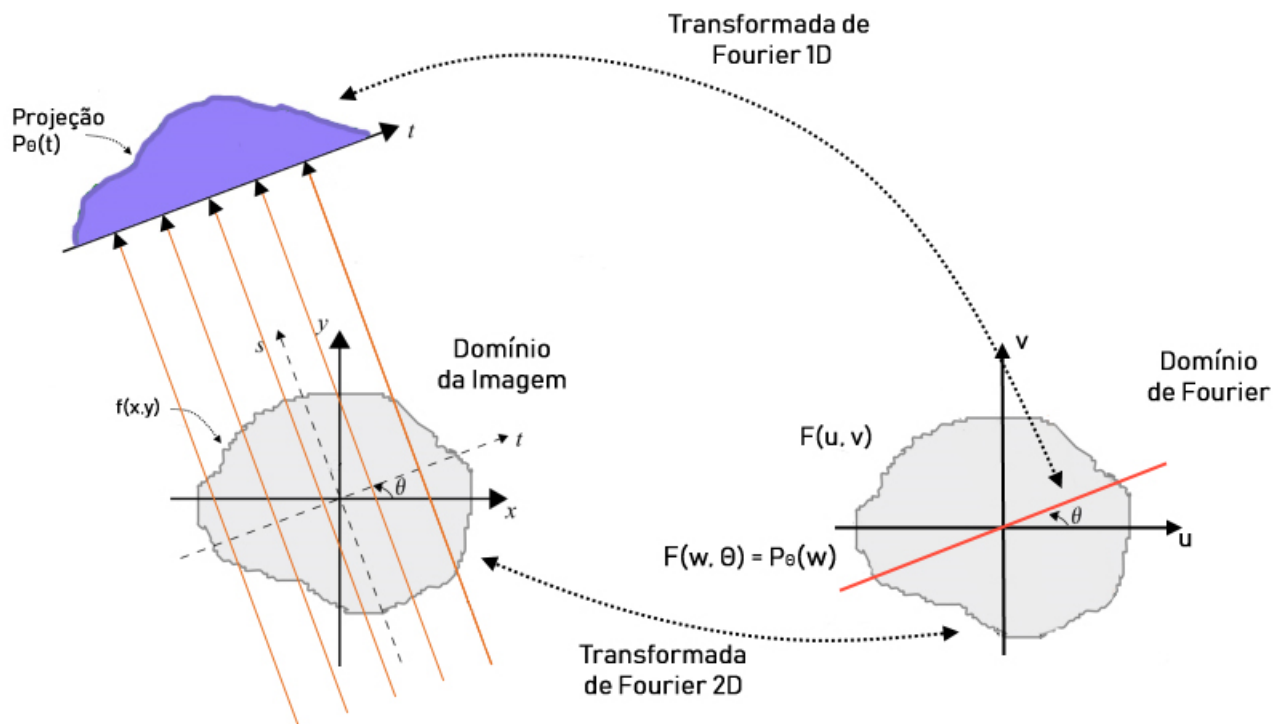
Após a definição da Transformada de Radon, é possível utilizá-la em um algoritmo de reconstrução. Dessa maneira, na próxima Seção é descrito como a Transformada de Radon se relaciona com a Transformada de Fourier da função $f(x, y)$.

4.2.2 Teorema de Corte de Fourier

Em reconstrução de exames tomográficos, as imagens podem se apresentar em diversos domínios, tais como (i) domínio da imagem, (ii) domínio de Fourier e (iii) domínio de Radon.

Os métodos de reconstrução utilizam as imagens representadas por esses domínios para facilitar o processo de reconstrução tomográfica. Portanto, é necessário que eles estejam relacionados de modo que a imagem possa ser transformada de um para outro. O Teorema de Corte de Fourier revela um importante relacionamento entre o domínio de Radon e o domínio de Fourier, utilizado no algoritmo de reconstrução FBP, descrito na Seção 4.2.3. Como posto por Rangayyan (2004), considerando uma imagem 2D, esse teorema mostra que a Transformada de Fourier 1D ($S_{\theta}(w)$) da projeção 1D ($P_{\theta}(t)$), que por sua vez foi obtida a partir da imagem 2D $f(x, y)$, é igual a uma seção radial da Transformada de Fourier 2D $F(u, v)$ da imagem 2D em um determinado ângulo de projeção θ . A Figura 26 ilustra essa definição. Na prática, como não é possível adquirir um número infinito de projeções, interpolações devem ser feitas, gerando degradação na imagem reconstruída final.

Figura 26 – Representação do Teorema de Corte de Fourier.



Fonte – Adaptado de Rangayyan (2004).

Para demonstrar tal teorema, considere a Transformada de Fourier 2D $F(u, v)$ de uma função (x, y) (RANGAYYAN, 2004; SALVADEO et al., 2013):

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy \quad (4.5)$$

e sua inversa como:

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{j2\pi(ux+vy)} du dv. \quad (4.6)$$

Considere também a transformada de Fourier 1D $S_{\theta}(w)$ de uma projeção $P_{\theta}(t)$:

$$S_{\theta}(w) = \int_{-\infty}^{\infty} P_{\theta}(t) e^{-j2\pi wt} dt. \quad (4.7)$$

Em um exame tomográfico real, o corpo mantém sua posição fixa e o emissor de radiação é deslocado em uma posição radial. Portanto, isso é equivalente a fixar o emissor de radiação e rotacionar objeto. Assim, $f(t, s)$ representa a imagem $f(x, y)$ rotacionada por um ângulo θ :

$$\begin{bmatrix} t \\ s \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (4.8)$$

Dessa maneira, uma projeção $P_{\theta}(t)$ em (t, s) é dada por:

$$P_{\theta}(t) = \int_{-\infty}^{\infty} f(t, s) ds. \quad (4.9)$$

Substituindo a Equação 4.9 em 4.7, é possível obter a Equação 4.10 (RANGAYYAN, 2004; SALVADEO et al., 2013):

$$S_{\theta}(w) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t, s) ds \right] e^{-j2\pi wt} dt. \quad (4.10)$$

Efetuada novamente a troca das coordenadas de (t, s) para (x, y) , tem-se a Equação 4.11:

$$\begin{aligned} S_{\theta}(w) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi w(x \cos \theta + y \sin \theta)} dx dy \\ &= F(u, v), \text{ com } y = w \cos \theta, v = w \sin \theta \\ &= F(w, \theta). \end{aligned} \quad (4.11)$$

Assim, a Equação 4.11 finaliza a derivação do Teorema de Corte de Fourier, relacionando a Transformada de Fourier 1D ($S_{\theta}(w)$) das projeções $P_{\theta}(t)$ à Transformada de Fourier 2D da imagem representada por $f(x, y)$ ($S_{\theta}(w) = F(w, \theta)$).

4.2.3 Filtered Backprojection (FBP)

O algoritmo de reconstrução FBP, apresentado inicialmente por Ramachandran e Lakshminarayanan (1971), tem por objetivo reconstruir funções bidimensionais a partir de suas projeções no domínio de Fourier, utilizando filtragem e convolução. Com isso, esse algoritmo consegue melhores resultados do que o BP e o SAA em reconstrução de imagens médicas, reduzindo a degradação causada pelas projeções adquiridas em ângulos limitados. Nesta Seção, será feita a derivação do FBP, assim como feita por Rangayyan (2004).

Primeiramente, a Equação 4.12 descreve a Inversa da Transformada de Fourier 2D:

$$f(x, y) = \int_0^{2\pi} \int_0^{\infty} F(w, \theta) e^{j2\pi w(x \cos \theta + y \sin \theta)} w \, dw \, d\theta, \quad (4.12)$$

onde as coordenadas do domínio da imagem foram trocadas de retangulares (u, v) para polares (w, θ) , tal que $w = \sqrt{u^2 + v^2}$ e $\theta = \tan^{-1}(\frac{v}{u})$.

Essa equação pode ser reescrita em duas partes, assim como mostra a Equação 4.13, ao considerar θ variando primeiramente de 0 a θ e depois de θ a 2θ :

$$\begin{aligned} f(x, y) = & \int_0^{\pi} \int_0^{\infty} f(w, \theta) e^{j2\pi w(x \cos \theta + y \sin \theta)} w \, dw \, d\theta + \\ & \int_0^{\pi} \int_0^{\infty} f(w, \theta + \pi) e^{j2\pi w[x \cos(\theta + \pi) + y \sin(\theta + \pi)]} w \, dw \, d\theta. \end{aligned} \quad (4.13)$$

É possível concluir que $F(w, \theta + \pi) = F(-w, \theta)$, pois a projeção gerada quando o detector estiver em uma posição angular θ , será a mesma projeção, porém invertida, quando o detector estiver em uma posição $\theta + \pi$. Usando esta conclusão junto com a propriedade $t = x \cos \theta + y \sin \theta$, é possível chegar à Equação 4.14:

$$f(x, y) = \int_0^{\pi} \left[\int_{-\infty}^{\infty} F(w, \theta) |w| e^{j2\pi wt} \, dw \right] d\theta. \quad (4.14)$$

Utilizando o teorema de Corte de Fourier ($S_{\theta}(w) = F(w, \theta)$), explicado na Seção 4.2.2, é possível chegar na Equação 4.15:

$$f(x, y) = \int_0^{\pi} \left[\int_{-\infty}^{\infty} S(w) |w| e^{j2\pi wt} \, dw \right] d\theta. \quad (4.15)$$

Dessa maneira, a função a ser reconstruída $f(x, y)$ passa a depender apenas da Transformada de Fourier 1D (S_{θ}) das projeções, e não mais da Transformada de Fourier 2D $F(w, \theta)$.

Para facilitar o entendimento, é possível reescrever a Equação acima conforme a Equação 4.16 (RANGAYYAN, 2004; SALVADEO et al., 2013):

$$f(x, y) = \int_0^\pi Q_\theta(x \cos \theta + y \sin \theta) d\theta, \quad (4.16)$$

onde $Q_\theta(t)$ é descrito como na Equação 4.17:

$$Q_\theta(t) = \int_{-\infty}^{\infty} S_{\theta(w)} |w| e^{j2\pi wt} dw. \quad (4.17)$$

Dessa maneira, diz-se que a Equação 4.16 descreve a retroprojeção da imagem, enquanto que a Equação 4.17 descreve uma operação de filtragem. Este filtro tem por finalidade diminuir os problemas de degradação originados das projeções limitadas. Assim, ele fornece uma filtragem unidimensional em cada ângulo θ , sendo depois adicionadas na Equação 4.16 para originar a reconstrução completa da imagem. Esta filtragem pode ser expressa por uma operação de convolução entre a imagem e o filtro, resultando na Equação 4.18 (RANGAYYAN, 2004):

$$Q_\theta(n\tau) = \tau \sum_{k=0}^{N-1} h(n\tau - k\tau) P_\theta(k\tau), n = 0, \dots, N - 1. \quad (4.18)$$

Um raio t_j descrito por $t_j = x \cos \theta_i + y \sin \theta_i$ passa sobre determinados pixels da imagem, contribuindo apenas com o valor $Q_{\theta_i}(t_j)$. Na etapa de retroprojeção, o valor $Q_{\theta_i}(t_j)$ irá contribuir com o mesmo valor nestes mesmos pixels na Figura 24.

Em aplicações práticas, um número limitado K de projeções são obtidas. Em CT, considera-se que uma projeção é obtida a cada π/K radianos, no intervalo de rotação $[0, \pi]$. No caso de DBT, o intervalo angular é bem maior, o que tende a gerar mais artefatos. O número de raios paralelos presente no feixe que origina a projeção é dado por N , sendo τ a distância entre eles e $h(t)$ uma equação que representa um filtro. Quando os valores de k estão fora do intervalo $[0, N - 1]$, $P_\theta(k\tau)$ é nulo. O filtro $h(\pi\tau)$ é definido conforme a Equação 4.19 (RANGAYYAN, 2004), chamado de filtro da Rampa:

$$\begin{cases} \frac{1}{4\tau^2}, & \text{se } n = 0 \\ 0, & \text{se } n \text{ é par} \\ \frac{-1}{(n\pi\tau)^2}, & \text{se } n \text{ é ímpar.} \end{cases} \quad (4.19)$$

Em implementações práticas, é possível utilizar algoritmos como a Transformada Rápida de Fourier (FFT, do inglês *Fast Fourier Transform*) e a sua inversa (IFFT, do inglês *Inverse FFT*) para reescrever a Equação 4.18, como mostra a Equação 4.20:

$$Q_{\theta}(n\tau) = \tau \times IFFT\{FFT[P_{\theta}(n\tau)] \times FFT[h(n\tau)]\}. \quad (4.20)$$

Da mesma maneira, é possível discretizar a Equação 4.16, para se obter a Equação 4.21, sendo que $\hat{f}(x, y)$ representa a imagem reconstruída:

$$\hat{f}(x, y) = \frac{\pi}{K} \sum_{i=1}^K Q_{\theta_i}(x \cos \theta_i + y \sin \theta_i). \quad (4.21)$$

4.3 Métodos de Reconstrução Algébricos

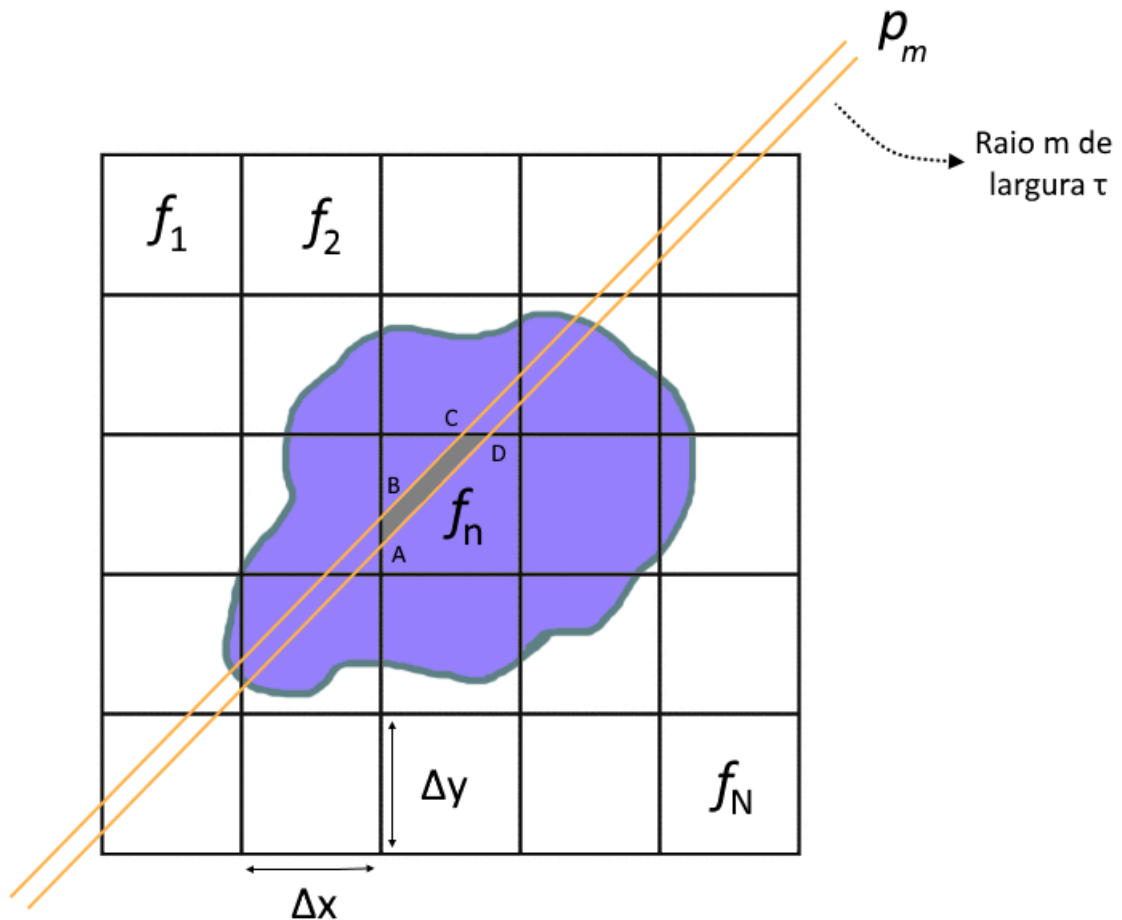
A Seção anterior descreveu o funcionamento do método FBP, que utiliza princípios algébricos para transformar as projeções do domínio de Radon para o Domínio de Fourier, fornecendo um método rápido e eficaz para reconstrução da imagem. Nesta Seção, a imagem será tratada como um sistema linear, onde os valores dos pixels reconstruídos são tratados como componentes de equações lineares que precisam ter seus valores descobertos, e para isso, são utilizadas técnicas descritas a seguir.

As próximas Seções irão descrever os métodos de reconstrução utilizando exemplos em 2D, portanto, em aplicações práticas de DBT, tem-se projeções 2D e imagens reconstruídas em 3D, sendo que tais métodos podem ser generalizados para a aplicação prática. A Seção 4.3.1 descreve o método de Reconstrução Algébrica (ART, do inglês *Algebraic Reconstruction Technique*), a Seção 4.3.2 descreve a Técnica de Reconstrução Iterativa Simultânea (SIRT, do inglês *Simultaneous Iterative Reconstruction Technique*) e a Técnica de Reconstrução Simultânea Algébrica (SART, do inglês *Simultaneous Algebraic Reconstruction Technique*).

4.3.1 Técnica de Reconstrução Algébrica (ART)

A reconstrução algébrica é uma técnica iterativa na qual considera a reconstrução de imagens como um problema de resolução de sistemas lineares. Em um sistema discreto, a integral de linha dos raios emitidos através do corpo imageado é chamado de raios-soma. Cada componente do sistema é definido como a atenuação do raio causado pelos voxels da imagem.

Figura 27 – Representação do Método ART.



Fonte – Adaptado de Rangayyan (2004).

Para facilitar a discussão desse método, uma grade foi adicionada ao objeto imageado, como pode ser visto na Figura 27. É considerado que essa grade tenha N células, cuja n -ésima célula seja representada por f_n . Cada célula possui dimensão horizontal, dada por Δx , e vertical, dada por Δy . Na obtenção de uma projeção, é considerado que o feixe de raios X tenha M raios. Ao contrário das abordagens analíticas, aqui um raio possui largura expressa por τ . Portanto, os raios podem não abranger toda a célula, sendo necessário calcular a contribuição parcial que uma determinada célula exerce sobre o raio. Para isso, basta calcular a área no qual o raio sobrepõe o objeto em uma determinada célula. Na Figura 27, o trapézio preto, dado pelos pontos $ABCD$, representa a contribuição exercida pela célula f_n no raio p_m . Dessa maneira, w_{mn} fornece o fator de contribuição do n -ésimo elemento da imagem para o m -ésimo raio. Assim, é possível obter o valor dos raios-soma somando todas as contribuições das células ao longo dos raios, como

expressa a Equação 4.22 (RANGAYYAN, 2004):

$$p_m = \sum_{n=1}^N w_{mn} f_n, \quad m = 1, 2, 3, \dots, M. \quad (4.22)$$

Observe que um determinado raio-soma passa através de poucas células da imagem, portanto, a contribuição de grande parte das células são nulas.

Ao calcular p_m para todos os raios-soma tem-se um sistema com M equações e N incógnitas (fornecidas por f_n) que deve ser resolvido para obter a imagem reconstruída. Caso M e N sejam pequenos, o sistema pode ser resolvido normalmente por métodos tradicionais. Caso contrário, o maior problema se encontra no custo computacional necessário para encontrar a solução. É possível ainda que $M < N$, tornando o sistema subdeterminado, sendo necessário a utilização de outros métodos.

Um entendimento geométrico de sistemas lineares pode ser realizado considerando cada equação como um hiperplano no espaço, sendo a solução desse sistema um ponto, que é determinado pela intersecção de todos os hiperplanos (em caso de sistemas com uma única solução). Portanto, a imagem reconstruída pode ser considerada como o ponto $\vec{f} = (f_1, f_2, \dots, f_n)$. Como o ART é um método iterativo, o ponto é alterado durante várias etapas para convergir na solução final em algum intervalo de tempo. Assim, o método de Kaczmarz (KAK; SLANEY, 1988) propõem projetar a solução atual (representada por \vec{f}) em um hiperplano do sistema, resultando em um outro plano mais próximo da solução. A realização desse procedimento continuamente leva a encontrar o ponto ótimo, que é a solução do problema.

Inicialmente, uma estimativa preliminar da imagem $\vec{f}^0 = (f_1^0, f_2^0, \dots, f_N^0)$ é escolhida, geralmente com valores constantes. No método de Kaczmarz, esse ponto é projetado no hiperplano definido pela primeira equação, resultando em \vec{f}^1 . Em seguida, \vec{f}^1 é projetado no segundo hiperplano, resultando em \vec{f}^2 , que por sua vez é projetado no terceiro hiperplano, gerando \vec{f}^3 . Esse procedimento, ilustrado na Figura 28, é repetido até que a solução do sistema seja encontrada. Caso o sistema não tenha solução, os pontos irão oscilar entre os hiperplanos até encontrar a mínima variância. A componente n , do ponto resultante da i -ésima iteração, pode ser definida como na Equação 4.23 (RANGAYYAN, 2004),

$$f_n^i = f_n^{i-1} + \Delta f_n^{i-1}, \quad (4.23)$$

sendo Δf_n^{i-1} expresso pela Equação 4.24 (RANGAYYAN, 2004):

$$\Delta f_n^{i-1} = \frac{p_m - f^{(i-1)} \cdot \vec{w}_m}{\vec{w}_m \cdot \vec{w}_m} w_m. \quad (4.24)$$

Na Equação acima, p_m e $\vec{w}_m = (w_{m1}, w_{m2}, \dots, w_{mN})$ são o raio-soma e o vetor de pesos do hiperplano m no qual se deseja projetar o ponto, respectivamente. A operação “ \cdot ” é o produto escalar/produto interno entre vetores, sendo $\vec{w}_m \cdot \vec{w}_m$ o produto escalar entre \vec{w}_m e ele mesmo. Por sua vez, Δf_n^{i-1} pode ser entendido como a correção do tom de cinza da n -ésima componente da nova solução.

Substituindo a Equação 4.24 na Equação 4.23, obtem-se a Equação 4.25 (RANGAYYAN, 2004):

$$f_n^{(i)} = f_n^{(i-1)} + \frac{p_m - f^{(i-1)} \cdot \vec{w}_m}{\vec{w}_m \cdot \vec{w}_m} w_m. \quad (4.25)$$

Dessa maneira, a Equação acima representa o valor da n -ésima componente do novo ponto, projetado sobre o m -ésimo hiperplano. Resolvendo a Equação 4.25 para $n = 1, 2, \dots, N$, tem-se a nova solução. A escolha do ponto inicial é um importante aspecto para agilizar a convergência, sendo que em alguns casos é utilizada uma simples reconstrução pelo FBP.

4.3.2 Técnica de Reconstrução Iterativa Simultânea (SIRT) e Técnica de Reconstrução Simultânea Algébrica (SART)

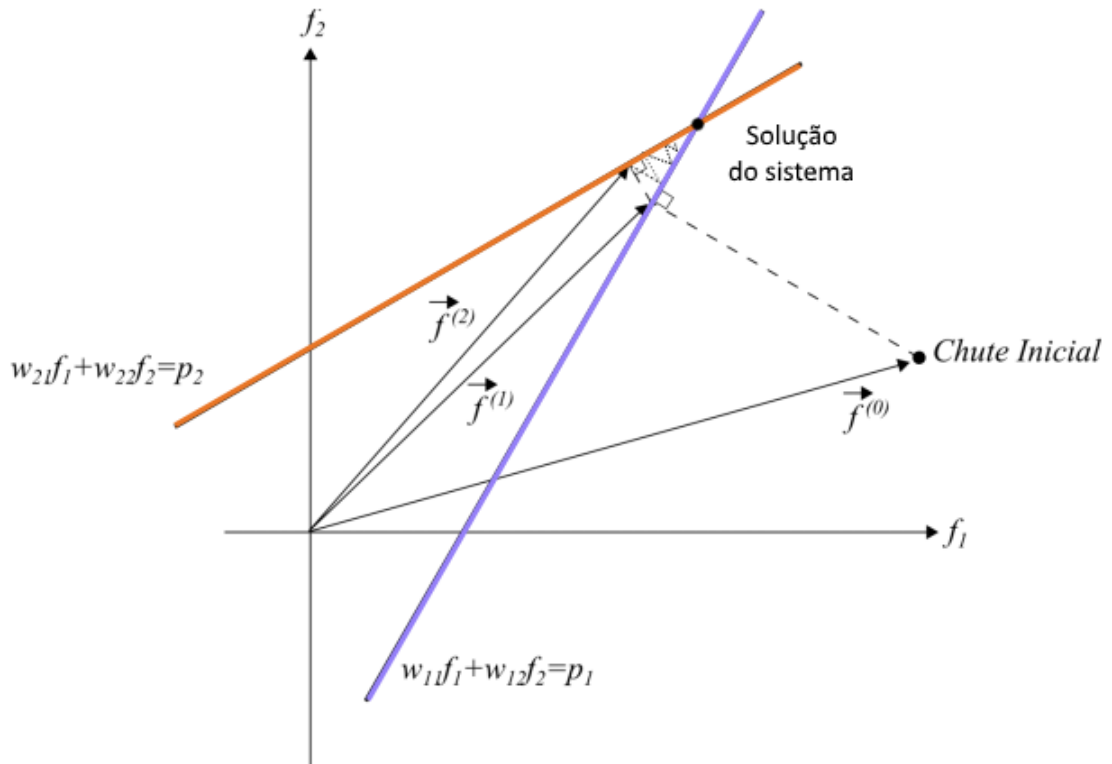
Para iniciar, as Equações 4.23 e 4.25 podem ser generalizadas como

$$f_n^i = f_n^{i-1} + \lambda \Delta f_n^{i-1} = f_n^{(i-1)} + \lambda \frac{p_m - f^{(i-1)} \cdot \vec{w}_m}{\vec{w}_m \cdot \vec{w}_m} w_m, \quad (4.26)$$

sendo λ um fator de relaxamento para limitar amplas variações nos resultados.

No método SIRT, em vez de calcular individualmente o novo valor de cada célula na imagem, esta técnica percorre todas as equações presentes no sistema e calcula a média das contribuições para obter o novo valor da célula na imagem. Esta técnica possui um custo computacional superior à técnica ART, porém consegue melhores resultados em algumas

Figura 28 – Exemplificação do método de Kaczmarz.



Fonte – Adaptado de Kak e Slaney (1988).

aplicações. A Equação 4.27 descreve o método SIRT (YANG, 2012):

$$\begin{aligned}
 f_n^i &= f_n^{i-1} + \lambda \frac{1}{M_n} \sum_{m=1}^M \Delta f_n^{i-1} \\
 &= f_n^{i-1} + \lambda \frac{1}{M_n} \sum_{m=1}^M \frac{p_m - f^{(i-1)} \cdot \vec{w}_m}{\vec{w}_m \cdot \vec{w}_m} w_m.
 \end{aligned}
 \tag{4.27}$$

Conforme Yang (2012), o método SART é uma combinação do ART e SIRT, e pode ser expresso pela Equação 4.28:

$$f_n^i = f_n^{i-1} + \lambda \frac{\sum_{m=1}^M \Delta f_n^{i-1}}{\sum_{m=1}^M w_{mn}}.
 \tag{4.28}$$

Finalmente, substituindo a Equação 4.24 na Equação 4.28, tem-se a Equação 4.29:

$$f_i^n = f_n^{i-1} + \lambda \frac{\sum_{m=1}^M \frac{p_m - f^{(i-1)} \cdot \vec{w}_m}{\vec{w}_m \cdot \vec{w}_m} w_m}{\sum_{m=1}^M w_{mn}}.
 \tag{4.29}$$

4.4 Reconstrução Estatística Iterativa

Este tipo de reconstrução tomográfica considera o problema de reconstrução como um problema de otimização, mas leva em considerações informações estatísticas da imagem previamente conhecidas para melhorar a qualidade dos resultados, tal com o método da Máxima Verossimilhança..

Como nos outros métodos iterativos, o método da Máxima Verossimilhança considera a imagem como sendo um sistema linear, sendo a imagem final reconstruída uma possível solução desse sistema. Cada elemento da imagem exerce uma atenuação no raio que passa através do corpo imageado, semelhante ao que acontece no método ART (Seção 4.3.1).

Considere que Y_m seja uma variável aleatória que armazena a contagem dos fótons incidentes ao longo de um raio X , y_m é uma observação de Y_m , μ_m é o vetor de atenuação do raio m , d_m é a intensidade do raio X incidente e r_m representa a variância do ruído de leitura. Além disso, cada voxel possui um coeficiente de atenuação, sendo l_m o vetor de coeficientes do m -ésimo raio (XU, 2014). Assim, a partir da distribuição de Poisson, é possível chegar à Equação 4.30:

$$L(\mu) = \sum_m^M \{d_m e^{-\langle \mu, l_m \rangle} - y_m \log(d_m e^{-\langle \mu, l_m \rangle} + r_m)\} + c. \quad (4.30)$$

Para minimizar a Equação 4.30, uma saída seria utilizar a função *surrogate* de $L(\mu)$ para relaxar o problema. Assim, é possível minimizar a função *surrogate* ao invés da minimização da função $L(\mu)$. Usando uma iteração de Newton, é possível chegar na Equação 4.31:

$$\mu_n^{i+1} = \mu_n^i - \frac{\sum_{m=1}^M l_{mn} (-d_m e^{\langle \mu^i, l_m \rangle} + y_m)}{\sum_{m=1}^M (l_{mn} \sum_{n=1}^N l_{mn} d_m e^{\langle \mu^i, l_m \rangle})}. \quad (4.31)$$

Um estudo comparativo entre os algoritmos BP, MLEM e SART foi efetuado por Zhang et al. (2006), no qual foi avaliado o contraste, nitidez e a presença de artefatos na imagem reconstruída final, utilizando phantoms, na modalidade de DBT com feixes cônicos. Os autores utilizaram 21 projeções, com uma distância angular de $\pm 30^\circ$ do emissor de radiação, e utilizaram as métricas de avaliação *Contrast-to-Noise Ratio* (CNR), *Artefact Spread Function* (ASF), *Noise Power Spectrum* (NPS), e a *Line Object Spread Function* (LOSF). Os autores mostraram que o algoritmo BP resultou em imagens reconstruídas com baixo ruído no plano de fundo e, para objetos com baixo contraste, o valor de CNR foi maior do que nos métodos SART e MLEM.

Entretanto, os métodos iterativos apresentaram boa qualidade nas bordas de objetos, aumentaram o contraste geral da imagem, diminuíram o borramento entre as seções transversais e a quantidade de artefatos, obtendo melhores valores de ASF.

Utilizando um corpo de prova de alto contraste e com um fundo da imagem que imita o tecido humano, o método BP ocasionou artefatos de borramento na direção do movimento do emissor de radiação, o que prejudicou a análise das características do objeto imageado, enquanto que os outros métodos facilitaram a análise removendo objetos sobrepostos Zhang et al. (2006). Um ponto importante a destacar é que, com um fator de relaxamento escolhido adequadamente, o método SART teve a convergência mais rápida do que o MLEM. Em apenas 1 iteração, o SART obteve resultados comparáveis com MLEM executado com 7 iterações, quando a solução inicial foi uma reconstrução a partir do algoritmo FBP.

4.5 Reconstrução Baseada em Sensoriamento Comprimido

O funcionamento dos métodos baseados em Sensoriamento Comprimido está relacionado com a resolução de sistemas lineares, assim como os métodos iterativos discutidos anteriormente. Entretanto, eles utilizam o gradiente descendente para guiar a reconstrução da imagem a uma solução aceitável. O principal método nesta abordagem é o Total Variation (TV).

O algoritmo TV se sobrepõem aos métodos analíticos devido à sua capacidade de gerar imagens reconstruídas com maior qualidade a partir de poucas projeções (PARK et al., 2014). Considerando que em uma projeção tenha M raios e a imagem a ser reconstruída tenha N voxels, é possível calcular os raios-soma (soma das contribuições que cada voxel efetua sobre um determinado raio), como feito na Seção 4.3.1, para cada projeção, e ir ajustando os valores das contribuições de cada voxel, por meio do Gradiente Descendente, até que uma solução ideal seja encontrada. Sendo assim, o problema de reconstrução de imagens utilizando TV pode ser modelado como (PARK et al., 2014)

$$\mathbf{b} = A\mathbf{x}, \quad (4.32)$$

onde \mathbf{b} é um vetor de tamanho M contendo os raios-soma de uma projeção, \mathbf{x} é o vetor contendo os voxels da imagem a ser reconstruída, e A é uma matriz de dimensões $M \times N$ contendo a atenuação que o n -ésimo voxel exerce sobre o m -ésimo raio.

Dessa maneira, a reconstrução da imagem se transforma em um problema de otimização, sendo o objetivo encontrar o ponto de mínimo que satisfaça a Equação 4.33 (PARK et al., 2014):

$$\mathbf{x}' = \underset{\mathbf{x} \in \mathbf{Q}}{\operatorname{argmin}} \phi(\mathbf{x}), \quad (4.33)$$

onde $\phi(\mathbf{x})$ é definido pela Equação 4.34:

$$\phi(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \alpha \sum_{i=1}^N \|\mathbf{D}_i \mathbf{x}\|_2, \quad (4.34)$$

e o termo $\mathbf{Ax} - \mathbf{b}$ reflete que as projeções da nova estimativa de imagem devem ser iguais às projeções adquiridas no início do algoritmo. Ou seja, as poucas projeções disponíveis servem como parâmetros para avaliar a qualidade da solução que está sendo gerada, sendo então possível calcular o Gradiente Descendente e encontrar uma imagem boa o suficiente. O valor de α representa o parâmetro de regularização e D_i a aproximação de diferença do gradiente ao i -ésimo voxel. É importante observar que a função de custo utilizada é a ℓ_2 (PARK et al., 2014).

4.6 Reconstrução Baseada em Aprendizado de Máquina

O desenvolvimento de métodos de reconstrução de imagens médicas utilizando Aprendizado de Máquina se tornou uma abordagem emergente muito ativa. Isto pode ser notado pela publicação em 2018 de uma Seção especial (WANG et al., 2018) na revista *IEEE Transactions on Medical Imaging*. Assim, muitos trabalhos recentes surgiram com o intuito de aplicar essa abordagem nas mais diversas modalidades de imageamento (WANG et al., 2018), como PET, MRI e CT.

Assim como descrito em Adler e Öktem (2018), no contexto de reconstrução tomográfica baseada em Aprendizado de Máquina, três abordagens tem se destacado na literatura. A primeira se caracteriza pela utilização de um modelo de Aprendizado de Máquina na etapa de pós-processamento, podendo ser chamada de *filtragem de ruído aprendido*. Essa abordagem tem por objetivo utilizar uma técnica de reconstrução tradicional para obter uma reconstrução τ^\dagger e depois utilizar uma técnica de Aprendizado de Máquina para aprender um operador Λ_θ para melhorar a qualidade e diminuir o ruído da reconstrução (YANG et al., 2018; JIN et al., 2017). Dessa

maneira, essa abordagem pode ser expressa por

$$\tau_{\theta}^{\dagger} = \Lambda_{\theta} \circ \tau^{\dagger} \quad (4.35)$$

sendo τ_{θ}^{\dagger} a reconstrução final e θ um conjunto de parâmetros estimados na fase de treinamento que melhor leva a imagem do domínio das projeções ao domínio da reconstrução. Ainda nesta seção, uma abordagem de reconstrução também será considerada como pertencente a categoria *filtragem de ruído aprendido* se ela aplicar o operador Λ_{θ} antes da reconstrução, diretamente nas projeções.

A segunda abordagem se caracteriza por aprender o termo regularizador da Equação 4.1, utilizado em abordagens iterativas. No decorrer deste texto, esta abordagem será referida como *regularizador aprendido*.

Por último, há os trabalhos que buscam modelar todo o processo de reconstrução por meio de uma abordagem de Aprendizado de Máquina, não apenas a componente de regularização de abordagens iterativas ou etapas de pós-processamento individuais. Isso leva ao aprendizado de *esquemas iterativos* (como será referida no restante deste texto), mas utilizam Aprendizado de Máquina para encontrar a melhor correção da reconstrução a cada iteração, dado a iteração anterior (SUN et al., 2016; PUTZKY; WELLING, 2017; ADLER; ÖKTEM, 2017). Porém, é importante notar que o processo iterativo, que é o de maior custo computacional, vai para a etapa de treinamento, sendo que uma vez que a rede esteja treinada, utilizá-la é rápido. Ou seja, em comparação com métodos iterativos de abordagens clássicas, tal como o MLEM ou TV, onde o processo iterativo ocorre a cada reconstrução, nas abordagens de reconstrução por aprendizado de máquina o processo iterativo ocorre uma única vez para aquele conjunto de dados de treinamento, independente do número de reconstruções posteriores. Finalmente, alguns métodos de reconstrução usando aprendizado de máquina encontrados na literatura são apresentados nas próximas Seções.

4.6.1 FPB Simulado via RNA

Devido ao grande sucesso alcançado pelas DNNs em diversas tarefas, é natural a avaliação das mesmas também no processamento de imagens médicas. Com isso, a tarefa de reconstrução de imagens tomográficas é uma das diversas aplicações que podem fazer uso das DNNs, sendo que alguns trabalhos já vem sendo desenvolvidos nesta área.

Entre esses trabalhos, Würfl et al. (2016) propôs a simulação do algoritmo FBP (ver Seção 4.2.3) por meio de redes neurais artificiais, utilizando uma arquitetura com 3 camadas para efetuar a reconstrução de imagens de CT com raios paralelos. A primeira é uma camada convolucional para simular a etapa de filtragem das projeções. Na segunda, os autores implementaram uma camada completamente conectada que simula a operação de retroprojeção. Por fim, na terceira, foi anexada uma camada de não-negatividade ReLU (Equação 3.6). Os autores também propuseram uma camada específica para reconstruir projeções adquiridas com raios do tipo leque (*fan beam*). O número de neurônios em cada camada é igual ao número de pixels no sinograma.

Um dos principais problemas com o algoritmo FBP na reconstrução de imagens é que a discretização das projeções resulta na necessidade de efetuar interpolações, originando problemas como a presença de artefatos. Dessa maneira, Würfl et al. (2016) propuseram o uso das redes neurais para encontrar a melhor maneira de interpolar os dados ausentes, resultando em uma reconstrução melhor. A seguir, será feita a derivação da camada completamente conectada que implementa a operação de retroprojeção.

Considere a Equação 4.21, que representa o FBP com valores discretos. Sendo que $t = x \cos \theta + y \sin \theta$ (Equação 4.3), tem-se que $Q_{\theta_i(t)}$ fornece apenas valores discretos de t , os quais podem ser representados por $q_{i,j}$, sendo necessária a utilização de uma interpolação unidimensional, como expresso pela Equação 4.36 (WÜRFL et al., 2016),

$$\hat{f}(x, y) = \frac{\pi}{K} \sum_{i=1}^K \sum_{j=1}^L w_j(x, y, \theta_i) q_{i,j}, \quad (4.36)$$

onde $w_j(x, y, \theta_i)$ são os pesos de interpolação, L é um inteiro par que representa o número de coeficientes de interpolação utilizado, K é o número de projeções e $q_{i,j}$ é expresso pela Equação 4.37 (WÜRFL et al., 2016):

$$q_{i,j} = q \left[x \cos \theta_i + y \sin \theta_i - \frac{L+2}{2} + j \right], i. \quad (4.37)$$

Considere agora a função de ativação utilizada por neurônios artificiais representada pela Equação 3.1. É possível adaptar essa função para o contexto de reconstrução de imagens, tomando o bias $\theta = 0$ (neste contexto de RNAs), trocando $W_i X_i$ pela notação das projeções e trocando Y por $f(x_a, y_b)$, que representa o pixel (a, b) da imagem reconstruída de dimensões

$A \times B$. Assim, a Equação 4.36 é equivalente à Equação 4.38:

$$f(x_a, y_b) = \sum_{i=1}^K \sum_{j=1}^L w_{a+(b-1) \cdot A, j+(i-1) \cdot L} \cdot q_{i,j}, \quad (4.38)$$

se a relação expressa pela Equação 4.39 for verdadeira (WÜRFL et al., 2016):

$$w_{i+(j-1) \cdot L, j+(i-1) \cdot L} = \frac{\pi}{K} w_j(x_a, y_b, \theta_n). \quad (4.39)$$

Conectando a camada convolucional na camada completamente conectada, cuja função de ativação é expressa pela Equação 4.36, tem-se a RNA resultante expressa pela Equação 4.40:

$$f(x_a, y_b) = \max \left[0, \sum_{i=1}^K \sum_{j=1}^L \frac{\pi}{K} w_j(x_a, x_b, \theta_i) \cdot \left(\sum_{k=-\frac{L}{2}}^{\frac{L}{2}} w_k p_{j-k,i} \right) \right]. \quad (4.40)$$

Dessa maneira, pode-se dizer que a RNA resultante implementa o algoritmo FBP para feixes de raios paralelos. Assim, esse trabalho se encaixa na categoria de esquemas aprendidos. É importante mencionar que, na Equação 4.40, não tem-se valores negativos devido ao operador \max , pois o resultado representa um valor de atenuação linear.

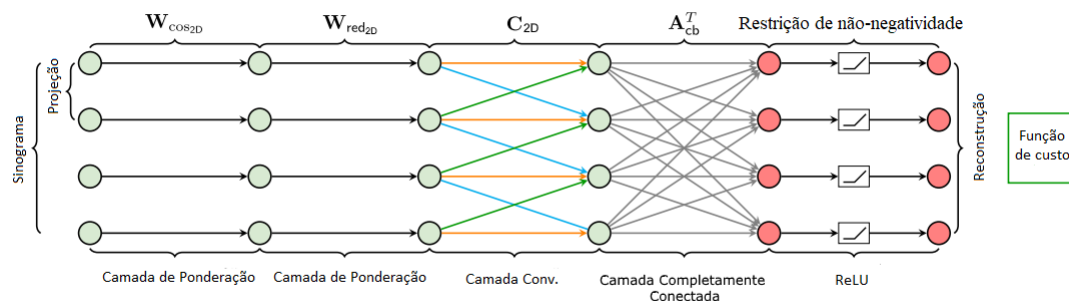
Como muitos equipamentos modernos de CT utilizam os raios do tipo cone, Würfl et al. (2018) expandiram a ideia para o algoritmo Feldkamp-Davis-Kress (FDK) (FELDKAMP; DAVIS; KRESS, 1984). A reconstrução via FDK pode ser expressa, portanto, pela notação de operadores (WÜRFL et al., 2018)

$$f = A_{CB}^T C_{2D} W_{red_{2D}} W_{cos_{2D}} P, \quad (4.41)$$

sendo que p representa o conjunto das projeções de entrada (sinograma), $W_{cos_{2D}}$ representa um ponderamento bidimensional por cossenos, necessário quando a geometria dos raios é do tipo Cone, $W_{red_{2D}}$ representa uma camada de redundância, que pode ser iniciada com a clássica ponderação de Parker (PARKER, 1982), para tratar os raios redundantes devido a arquitetura dos raios tipo cone, C_{2D} é uma camada convolucional para filtrar as projeções antes da convolução, que pode ser iniciada com o filtro de Ram-Lak (KAK; SLANEY, 1988) e A_{CB}^T é o operador proposto por Würfl et al. (2018) que executa a retroprojeção tridimensional levando em consideração os raios do tipo cone.

A Equação 4.41 pode ser expressa na RNA mostrada na Figura 29. Nesta figura, os nós preenchidos com a cor verde representam os pixels do sinograma, portanto, em cada camada, a quantidade de neurônios é igual ao número de pixels do sinograma. Como a camada A_{CB}^T realiza a retroprojeção, os neurônios das camadas posteriores, preenchidos com a cor vermelha, representam os pixels da imagem reconstruída. Portanto, a quantidade de neurônios em cada camada é igual à quantidade total de voxels em todas as fatias do volume reconstruído. É importante mencionar que, como a abordagem de raios paralelos expressa por Würfl et al. (2016), existe apenas 1 camada treinável, que é a camada W_{red2D} .

Figura 29 – Rede neural que simula o algoritmo FDK.



Fonte – Adaptado de Würfl et al. (2018).

A arquitetura dos equipamentos de DBT se assemelha aos equipamentos de CT do tipo cone, pois o primeiro pode ser considerado como um cone truncado, com algumas modificações adicionais. Portanto, a próxima subseção irá mostrar com mais detalhes como essa abordagem foi treinada, quais dados foram utilizados e os resultados obtidos.

4.6.1.1 Experimentos e Resultados

O conjunto de dados utilizado foi provido pela *Mayo Clinic* para o *AAPM Low Dose CT Grand Challenge* (MCCOLLOUGH, 2016), e possui imagens simuladas de CT de cabeças, abdômens e tórax humanos adquiridos de 10 pacientes a partir de equipamentos da Siemens, GE e Philips¹. Nesse conjunto de dados, o movimento do emissor de radiação utilizado foi do tipo hélice, portanto, os autores fizeram a transformada de Radon dos volumes reconstruídos (a partir da dose total de radiação) utilizando a trajetória do tipo cone. Assim, projeções e reconstruções de CT utilizando o emissor com trajetória do tipo cone se tornaram disponíveis, sendo possível utilizar esses dados nos experimentos. Nesta transformação, os autores utilizaram 360 projeções

¹ Mais detalhes sobre esse conjunto em <https://www.aapm.org/GrandChallenge/LowDoseCT/>

com espaçamento angular de 1° para simular o conjunto de treinamento. Para os testes, com o objetivo de simular uma reconstrução obtida a partir de poucas projeções, eles amostraram obtendo apenas 180 projeções para cada volume. O detector de radiação utilizado tem tamanho 720×880 pixels, com cada pixel tendo 1×1 mm. Os voxels do volume reconstruído possuem tamanho de $0,664 \times 0,664 \times 1$ mm. Os autores realizaram validação cruzada dividindo o conjunto de dados em 10 partições. Portanto, devido ao custo computacional elevado, eles optaram por reconstruir um volume com tamanho reduzido de $256 \times 256 \times 256$ voxels. O treinamento demorou cerca de 50 horas no total. A função de custo e o método de otimização utilizados foram o MSE e o Gradiente Descendente, respectivamente.

Os resultados obtidos foram comparados com outras abordagens do estado da arte, como (i) Weighted Total Variation (wTV), usando os mesmos parâmetros descritos em Huang et al. (2018) (utilizando 200 iterações), (ii) uma abordagem de pós-processamento descrita por Gu e Ye (2017), na qual utiliza uma rede neural baseada na *U-Net* (RONNEBERGER; FISCHER; BROX, 2015), que tem por objetivo melhorar o volume reconstruído retirando ruído e artefatos e (iii) a abordagem analítica clássica que utiliza ponderações de Parker (PARKER, 1982) para tratar os raios redundantes devido à trajetória do tipo cone. As métricas de avaliação empregadas foram SSIM e PSNR. O método wTV obteve os melhores resultados em 100% dos testes de acordo com a métrica PSNR, enquanto que a abordagem proposta obteve os melhores resultados em 100% dos casos levando em consideração a métrica SSIM. Segundo os autores, essa divergência se deve ao fato de que a métrica SSIM penaliza muito forte a suavização inserida pelo wTV, enquanto que é menos penalizada pelo PSNR.

4.6.2 Reconstrução de Primal-Dual

O trabalho de Adler e Öktem (2018) propôs um *framework* para reconstrução de imagens de CT, sendo classificada como um esquema iterativo aprendido, pois utiliza uma RNA para aprender todo o operador de reconstrução. Para isso, o trabalho usa um método Proximal, utilizado para minimizar uma função de custo não diferenciável, que pode ser resultado da utilização de uma função regularizadora também não diferenciável. Os métodos Proximais substituem a utilização do gradiente descendente para encontrar o mínimo da função de custo, pois não é possível aplicá-lo em funções não-diferenciáveis (ver Seção 3.2.1). Portanto, não é comum a utilização de métodos proximais para a minimização da Equação 4.1, pois o operador proximal não possui uma solução fechada.

Entretanto, Adler e Öktem (2018) argumentam que uma possível estratégia é a utilização do proximal *Primal-Dual*, implementada no algoritmo Gradiente Híbrido Primal Dual (PDHG, do inglês, Primal Dual Hybrid Gradient) (CHAMBOLLE; POCK, 2011). Foi observado ainda que os trabalhos de Romano, Elad e Milanfar (2017), Heide et al. (2014) substituíram o operador proximal por outras abordagens de operadores com sucesso. Por isso, a ideia de Adler e Öktem (2018) é substituir o operador PDHG por um operador aprendido por meio de um conjunto de treinamento.

Algoritmo 1: PRIMAL-DUAL APRENDIDOS. ADAPTADO DE ADLER E ÖKTEM (2018)

Data: $f_0 \in X^{N_{\text{primal}}}$, $h_0 \in U^{N_{\text{dual}}}$

```

1 begin
2   for  $i = 1, \dots, I$  do
3      $h_i \leftarrow \Gamma_{\theta^d}(h_{i-1}, K(f_{i-1}^{(2)}), g)$ 
4      $f_i \leftarrow \Lambda_{\theta^p}(f_{i-1}, [\partial K(f_{i-1}^{(1)})]^*(h_i^{(1)}))$ 
5   end
6 end
7 return  $f_I$ 

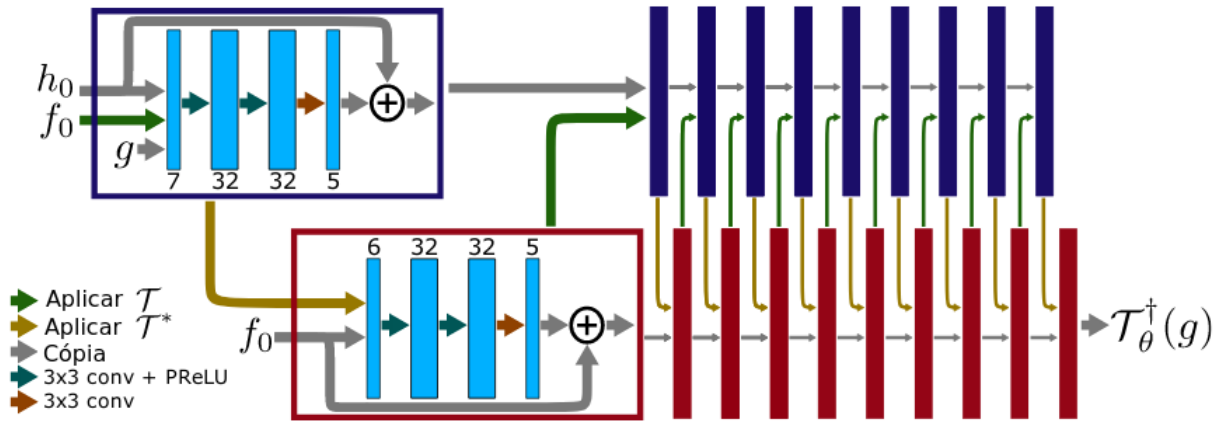
```

No Algoritmo 1, $K : X \rightarrow U$ é um operador possivelmente não linear, U é um espaço de Hilbert, $h \in U$ é a variável dual, $[\partial K(f_{i-1}^{(1)})]^* : U \rightarrow X$ é o adjunto da derivada (de Fréchet) de K no ponto f_i e g é dado pela Equação 4.1. A quantidade de iterações I foi estabelecida em um número fixo. A imagem reconstruída final é apresentada por f_I , e é possível diminuir o tempo de processamento fornecendo uma reconstrução inicial como ponto de partida (f_0), que pode ser uma reconstrução feita pelo FBP, por exemplo.

Os operadores primal e dual foram simulados a partir de redes neurais convolucionais (Γ_{θ^d} e Λ_{θ^p} , respectivamente), sendo θ^d e θ^p os parâmetros estimados para os respectivos operadores na fase de treinamento. Ambas CNNs possuem filtros de tamanho 3x3 contendo 4 camadas cada uma, sendo que o número de filtros para o operador primal aprendido, em cada camada, é de 6, 32, 32 e 5, respectivamente. O número de filtros para o operador dual aprendido é de 7, 32, 32 e 5, para cada camada, respectivamente. A Figura 30 mostra a arquitetura empregada para cada operador e a sequência de ligação das unidades, assim como exemplificado pelo Algoritmo 1. Os retângulos azuis escuros representam o operador primal aprendido, e os vermelhos representam o operador dual aprendido. Todos os retângulos representam a mesma arquitetura de RNA, que está demonstrada nas caixas do lado esquerdo da figura. Ainda nessa Figura, a variável τ representa a

projeção do objeto (ou a transformada de Radon), τ^* representa a retroprojeção da respectiva projeção e τ_θ^\dagger representa a imagem reconstruída.

Figura 30 – Rede Neural que simula os operadores Primal e Dual.



Fonte – Traduzido de Adler e Öktem (2018).

4.6.2.1 Experimentos e Resultados

Os autores executaram o treinamento e as comparações da rede neural utilizando: (i) um conjunto de dados sintético formado por elipses geradas aleatoriamente e (ii) um conjunto de dados realísticos de CT do abdômen humano. Em (i), as imagens geradas possuem 128×128 pixels, de onde foram obtidas 30 projeções unidimensionais utilizando feixes de raios paralelos e um detector de 182 pixels. O intervalo angular, o largura e comprimento exato do detector não foram informados. Em (ii), o conjunto de treinamento utilizado foi provido pela *Mayo Clinic*, para o *AAPM Low Dose CT Grand Challenge* (MCCOLLOUGH, 2016), utilizado também pelo método da Seção 4.6.1, e possui imagens realísticas de CT de abdômens humanos. Os autores utilizaram fatias reconstruídas com espessura de 3mm, totalizando 2168 imagens de treinamento. Para a reconstrução das fatias, foi utilizada uma arquitetura com raios do tipo leque bidimensional, adquirindo 1000 ângulos. O intervalo angular não foi mencionado. Em todos os casos, a rede foi treinada utilizando a função de custo MSE. A rede demorou 4 horas para ser treinada no conjunto de dados sintético e 11 horas no conjunto de dados reais.

O método proposto, chamado de *Primal Dual Reconstruction*, foi comparado com diversas outras abordagens de reconstrução, como TV, FBP, FBP junto com uma abordagem de pós-processamento baseada na U-Net (RONNEBERGER; FISCHER; BROX, 2015), Gradiente Aprendido (do inglês, *Learned Gradient*) (ADLER; ÖKTEM, 2017) e algumas simplificações do

método proposto por Gupta et al. (2018), que são *Learned PDHG*, *Learned Primal* e *FBP com residual noising*, que é uma versão simplificada do *Primal Dual Reconstruction* com uma solução inicial dada pelo FBP. Os métodos foram comparados em relação às métricas PSNR, SSIM, tempo de execução e complexidade da abordagem, expressa em quantidade de parâmetros².

Para o conjunto de dados sintético, o método proposto obteve uma melhora muito relevante em relação às métricas PSNR e SSIM em comparação com o FBP e o método de estado da arte TV. Para o conjunto de dados real, o método proposto também apresentou melhores valores de PSNR e SSIM, embora com uma relevância menor que no conjunto anterior. Como o FBP é uma abordagem analítica, ele obteve o melhor tempo de execução em todos os testes. Entretanto, sendo uma abordagem iterativa, o TV apresentou um tempo de execução muito maior que todas as abordagens avaliadas. O *Primal Dual Reconstruction* é mais rápido que o TV, pois a primeira é baseada em RNA e, embora a etapa de treinamento seja custosa, após os parâmetros serem aprendidos, a etapa de teste é muito mais rápida.

4.6.3 Método Estatístico Iterativo para CBCT baseado em CNNs

Segundo Chen et al. (2018a), algoritmos estatísticos iterativos que fazem uso de termos de regularização apresentam bons resultados na reconstrução de imagens utilizando raios do tipo cone. Em especial, o algoritmo TV apresenta bons resultados na remoção de ruído e preservação de bordas, mas uma limitação é a presença de artefatos do tipo escada (*staircase effect*) (JIA et al., 2010; SIDKY; KAO; PAN, 2006). Assim, Chen et al. (2018a) apresenta uma nova abordagem de reconstrução estatística iterativa para imagens tomográficas do tipo cone (*Cone Beam CT*), utilizando uma RNA para simular o termo de regularização (Equação 4.1), evitando sua escolha manual. Portanto, essa abordagem se encaixa na classificação de regularizador aprendido.

O autor utiliza o método de reconstrução iterativo Mínimo Quadrado Ponderado Penalizado (PWLS, do inglês, *Penalized Weighted Least-Square*), que leva em consideração o borramento encontrado em reconstrução de feixe cônico, e não somente o ruído encontrado nas imagens reconstruídas, introduzindo um termo de borramento na função de custo a ser otimizada pelo processo iterativo de reconstrução. Dessa maneira, os autores derivam uma Equação para

² Para abordagens que não são baseadas em Rede Neural, como o FBP e TV, a quantidade de parâmetros foi definida como 1.

alcançar esse objetivo, tal como

$$\mathbf{x}_{sol} = \arg \min_{\mathbf{x}} (\mathbf{p} - \mathbf{Ax})^T \Sigma^{-1} (\mathbf{p} - \mathbf{Ax}) + \lambda \|\mathbf{x} - \mathbf{Gu}\|_2^2, \quad (4.42)$$

$$\mathbf{u}_{sol} = \arg \min_{\mathbf{u}} \lambda \|\mathbf{x} - \mathbf{Gu}\|_2^2 + \beta R(u), \quad (4.43)$$

cujo primeiro termo da Equação 4.42 é o Mínimo Quadrado Ponderado (WLS, do inglês *Weighted Least-Square*), \mathbf{x}_{sol} é a imagem reconstruída, \mathbf{p} são as projeções ruidosas iniciais, \mathbf{A} é a matriz de projeção, \mathbf{u} é uma função que fornece a atenuação sofrida pelos raios em cada ponto do corpo imageado. O termo \mathbf{G} corresponde a um kernel de borrimento introduzido pelo autor para que a Equação resultante leve em consideração o borrimento e não somente o ruído. O termo Σ é a matriz diagonal com seu i -ésimo elemento sendo a variância do ruído $\sigma_i^2 = \exp(\bar{p}_i)/N_{i0}$, com N_{i0} sendo o número de fótons incidentes no i -ésimo raio.

Na Equação 4.43, o termo λ pondera a relação entre o borrimento e a reconstrução, $R(u)$ é o termo de regularização que, neste trabalho, foi substituído por uma CNN, e β é um termo que pondera a relação do termo de regularização com a imagem reconstruída.

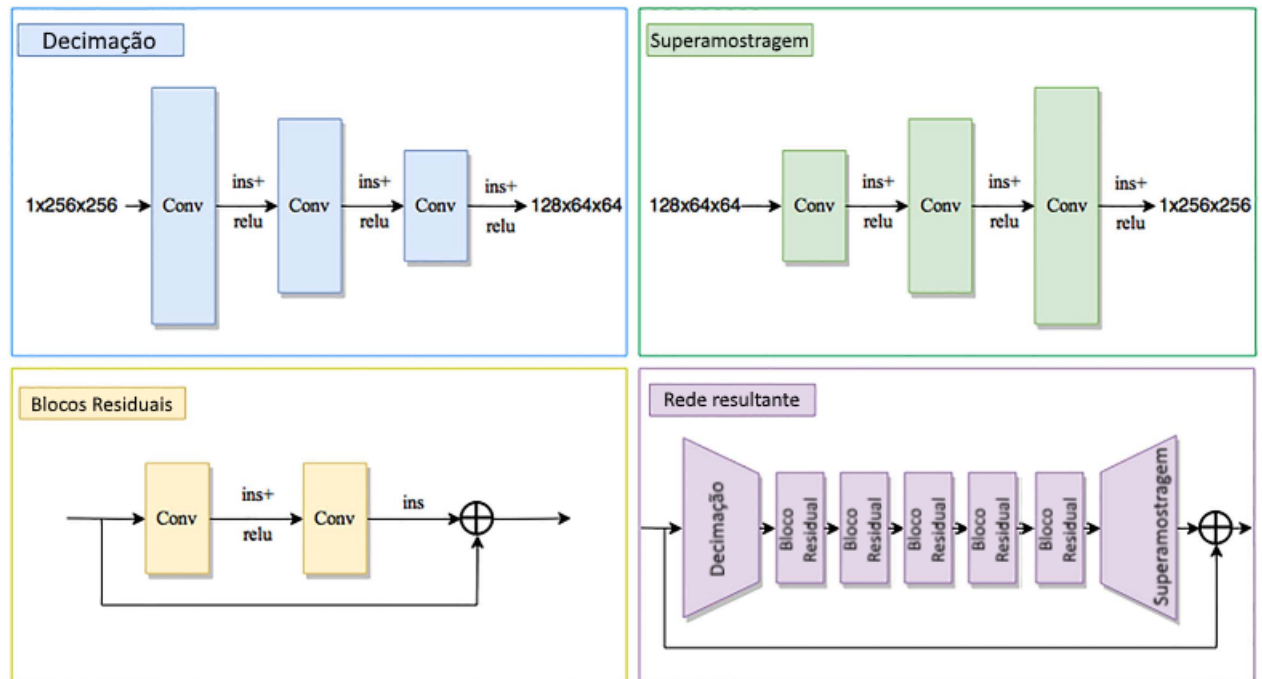
A RNA representada pelo termo $R(u)$ na Equação 4.43 é composta por um bloco de *Decimação* (ou *Subamostragem*) e *Superamostragem*. O primeiro é responsável por diminuir a dimensionalidade das imagens de 256 para 64, enquanto que o último é responsável por passar de volta as imagens de tamanho 64 para 256. No meio desses blocos há 5 camadas convolucionais com aprendizado residual. Essa arquitetura é mostrada na Figura 31.

Dessa maneira, a RNA tem por objetivo aprender a mapear as imagens ruidosas para uma versão sem ruído, levando em consideração o nível de ruído presente na imagem a ser reconstruída em cada iteração.

4.6.3.1 Experimentos e Resultados

Ao contrário da maioria das abordagens descritas neste capítulo, os autores não treinaram essa abordagem em um conjunto de imagens médicas, mas no conjunto de imagens COCO (do inglês *Common Objects in Context*) (LIN et al., 2014), contendo aproximadamente 80 mil imagens de objetos cotidianos. Entretanto, como essa rede neural tem por objetivo mapear e retirar o ruído de imagens, não tendo compromisso com as estruturas específicas que compõem

Figura 31 – Rede Neural aplicada no processo de reconstrução proposto por Chen et al. (2018a).



Fonte – Adaptado de Chen et al. (2018a).

imagens de CT, essa escolha se justifica. No processo de reconstrução, a medida que as iterações vão sendo executadas, a rede neural treinada no conjunto COCO irá retirar progressivamente o ruído das imagens de CT. Contudo, os autores se preocuparam em converter as imagens de treinamento para tons de cinza. O processo de treinamento durou apenas 4 horas, um tempo curto, segundo os autores, por causa da presença de camadas de subamostragem, que diminuem o número de parâmetros a serem processados. A função de custo utilizada foi a MSE, assim como o algoritmo de otimização ADAM (KINGMA; BA, 2014).

As métricas de avaliação PNSR (*Peak Signal-To-Noise ratio*), ISNR (*Improvement in Signal-to-Noise ratio*), CNR (*Contrast-to-Noise Ratio*) e SSIM foram utilizadas para avaliar o desempenho dessa abordagem utilizando 3 categorias de corpos de prova: (i) sintéticos, (ii) físicos e (iii) dados reais de pacientes. Para os phantoms sintéticos, simulou-se em computador 2 deles: o Compressing Sensing (CS) e o Shepp-Logan, este último com linhas horizontais modificadas para apresentarem diferença na intensidade, ambos contendo $0,766 \times 0,766 \times 0,766mm^3$ e $350 \times 350 \times 16$ pixels. Para os corpos de prova físicos, foram utilizados o *CatPhan 600* e o *Anthropomorphic Head*. Foram adquiridas 634 projeções utilizando o aparelho *Varian Acuity*, com o intervalo angular total de 360. Cada projeção possui o tamanho de 1024×768 pixels. Os voxels possuem tamanhos de $0,776 \times 0,776 \times 0,776mm$ para o *CatPhantom 600* e

$0,388 \times 0,388 \times 0,388\text{mm}$ para o *Anthropomorphic Head*. Por último imagens de pescoço e cabeça também foram adquiridas de pacientes reais, utilizando novamente o aparelho *Varian Acuity*. Para que uma comparação pudesse ser feita, essas imagens foram adquiridas utilizando um protocolo de alta dose de radiação, gerando menos ruído, e um protocolo de baixa dose de radiação, gerando mais ruído. Foram adquiridas 365 projeções em um intervalo angular de 200. O volume reconstruído possui $500 \times 500 \times 16$ voxels, sendo que cada voxel possui tamanho de $0,5 \times 0,5 \times 0,5\text{mm}$.

Os autores avaliaram os resultados individualmente para os 3 tipos de corpos de prova descritos anteriormente, comparando a abordagem proposta com os algoritmos de reconstrução TV e FDK. Os resultados levando em consideração os corpos de prova sintéticos e físicos revelaram que o método proposto possui maior capacidade em suprimir ruído, borramento e preservar estruturas. Os resultados obtidos a partir dos dados reais de pacientes revelaram que o método proposto possui resolução espacial maior que o TV.

4.6.4 Aperfeiçoamento de parâmetros via Deep Learning

Em problemas de otimização que são resolvidos por métodos iterativos, a qualidade da solução está diretamente relacionada com a escolha de parâmetros presentes na abordagem utilizada. Entretanto, essa escolha é manual na maioria das vezes (SHEN et al., 2017). Em reconstrução de imagens tomográficas utilizando técnicas estatísticas iterativas, como o TV, não é diferente. Assim, Shen et al. (2017) propuseram uma abordagem utilizando RNAs e aprendizado por reforço que tem por objetivo descobrir os melhores parâmetros para algoritmos de reconstrução iterativos, assim, otimizando o processo de reconstrução. Por exemplo, levando em consideração um problema que busca minimizar a Equação 4.1, o usuário deve selecionar manualmente o melhor parâmetro λ de acordo com a qualidade da imagem resultante.

No trabalho de Shen et al. (2017) foi considerada a reconstrução iterativa de CT dada por

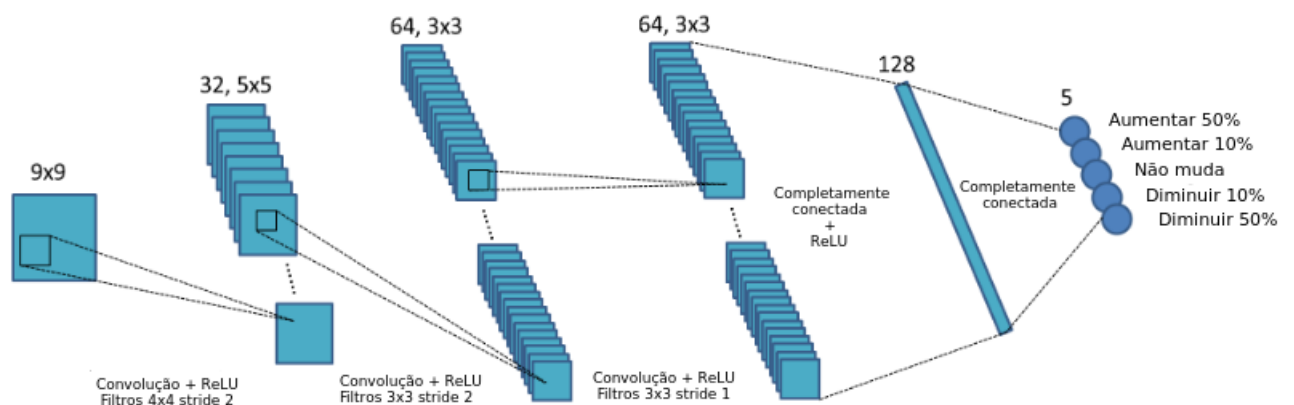
$$f^* = \arg \min_f \frac{1}{2} |Pf - g|^2 + \lambda \nabla, \quad (4.44)$$

sendo esta uma reconstrução baseada no método TV, com o segundo termo uma função de regularização que é aplicado a cada pixel da imagem. A variável λ é um vetor que controla a relação entre a função de regularização e a reconstrução. Neste caso, o usuário deveria estabelecer um valor para λ para cada pixel, ao contrário da abordagem simples de TV regularizado, no

qual um usuário deveria escolher apenas um valor que governa a relação inteira da função de regularização com a reconstrução. Dessa maneira, é possível visualizar a importância de uma abordagem que automatize a escolha dos parâmetros de maneira eficiente. Assim, os autores propuseram uma rede neural com aprendizado por reforço para escolher, em tempo de execução, os parâmetros para λ que resultassem na melhor reconstrução. Assim, a principal diferença entre esse método e o método proposto no trabalho de trabalho Chen et al. (2018a) é que o termo de regularização não é substituído por uma RNA, mas somente o fator de relação entre a regularização e a imagem reconstruída. Dessa maneira, esse trabalho também mostra que a qualidade da imagem reconstruída depende de uma boa escolha desse parâmetro. O método ADAM foi utilizado para a minimização da Equação 4.44, resultando na imagem reconstruída.

Os autores utilizaram uma CNN com 6 camadas, sendo 4 delas ocultas. A CNN tem segmentos da imagem reconstruída de tamanho 9x9 pixels como entrada e cinco opções de ajuste do valor do parâmetro como saída, que são: aumentar em 50%, diminuir em 50%, aumentar em 10%, diminuir em 10% e manter como está. Conforme mais amostras vão sendo colocadas na entrada, a arquitetura vai aprendendo o melhor valor para o parâmetro. A Figura 32 ilustra a RNA implementada pelos autores.

Figura 32 – Rede neural proposta para escolher os melhores parâmetros em abordagens de reconstrução iterativas.



Fonte – Adaptado de (SHEN et al., 2017).

Os experimentos foram realizados a partir de reconstruções de CT do tipo leque, com 180 projeções espaçadas igualmente no intervalo angular de 2π , com a resolução das imagens sendo de 128×128 pixels. Os autores avaliaram a qualidade geral da imagem com a métrica PSNR. Os autores escolheram 6 casos onde os parâmetros foram escolhidos manualmente, e foram comparados com os parâmetros ajustados automaticamente pela abordagem proposta e

com o parâmetro inicial. Os resultados foram agrupados no conjunto de teste e no conjunto de treinamento. Em somente um caso (do conjunto de testes) a abordagem proposta obteve resultado inferior que os valores escolhidos manualmente, porém, a diferença foi pequena.

4.6.5 PGD baseado em Rede Neural

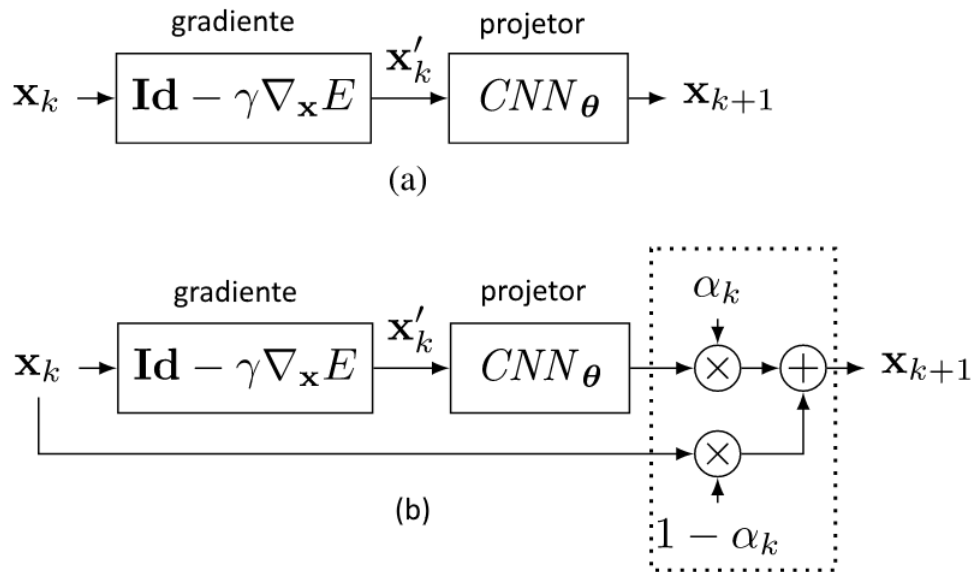
O trabalho de Gupta et al. (2018) teve por objetivo utilizar o método Gradiente Descendente Projetado (PGD, do inglês *projected gradient descent*) e trocar o projetor por uma CNN, apresentando um novo método de reconstrução. Os autores argumentam que os métodos atuais que utilizam RNAs como parte do processo de reconstrução não possuem garantia que a imagem reconstruída resultante é consistente. Mesmo no caso ideal, em que as projeções não possuem ruído, tais métodos não possuem um mecanismo de *feedback*. Portanto, não pode haver total segurança que a imagem reconstruída é consistente com as projeções. Assim, outra contribuição do trabalho de Gupta et al. (2018) é a proposta de um fator de relaxamento que garante a convergência sob certas condições, garantindo o alcance de um mínimo local para problemas de reconstrução não-convexos. O problema de reconstrução é considerado por Gupta et al. (2018) como

$$x^* = \arg \min_{x \in \mathbb{R}^N} (E(Hx, y) + \lambda R(x)), \quad (4.45)$$

sendo H o operador que efetua as projeções do objeto, $y = Hx + n$ são as projeções com ruído n , e $y, H : \mathbb{R}^N \rightarrow \mathbb{R}^M$. Ainda, a variável $x \in \mathbb{R}^N$ é o objeto que se deseja reconstruir e $n \in \mathbb{R}^M$ é o ruído inerente ao processo de aquisição. Por sua vez, $R(x)$ é uma função de regularização, λ é um fator que controla a contribuição da regularização com a imagem reconstruída e $E : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}^+$ é um termo de fidelidade dos dados que beneficia soluções consistentes.

A Figura 33 representa o método de reconstrução proposto pelos autores, onde a Figura 33 (a) mostra que o resultado final é dado pelo cálculo do gradiente descendente, sendo que a CNN projeta o resultado para o conjunto de soluções viáveis. O termo CNN_θ representa uma CNN treinada sobre um conjunto de parâmetros θ . Portanto, os autores argumentam que a CNN é um projetor estimado, podendo divergir algumas vezes. Então foi proposto um relaxador que utiliza um termo α_k (Figura 33 (b)) que força a solução sempre convergir, sendo este método chamado de Gradiente Descendente Projetado Relaxado. O símbolo Id representa o operador identidade.

Figura 33 – Diagrama do Gradiente Descendente Projetado. Uma CNN é utilizada para fazer o papel do projetor.



Fonte – Adaptado de Gupta et al. (2018).

4.6.5.1 Experimentos e Resultados

Foram utilizados dois conjuntos para a avaliação dos resultados de Gupta et al. (2018). O primeiro foi disponibilizado pela *Mayo Clinic* para o *AAPM Low Dose CT Grand Challenge* (MCCOLLOUGH, 2016), já descrito na Seção 4.6.1.1. Os autores utilizaram 500 imagens realísticas do pulmão, cabeça e abdômen dos 10 pacientes presentes no conjunto, com tamanho de 512×512 pixels. Os autores dividiram 475 imagens de 9 pacientes para a etapa de treinamento e 25 imagens do paciente restante para testes. O conjunto de dados contém exames com 720 projeções. Para simular exames com poucas projeções, os autores amostraram 144 e 45 projeções para simular exames com doses 5 vezes e 16 vezes menores. O tamanho de cada projeção se manteve o mesmo. O segundo conjunto utilizado contém o sinograma real de CT do cérebro de um rato, contendo 720 projeções unidimensionais com 1493 pixels cada, totalizando 377 fatias. Os autores redimensionaram as projeções para 729 pixels, criando um novo conjunto para o treinamento e teste da rede, havendo um intervalo de 25 fatias que não são usadas entre o conjunto de treinamento e o de teste. Os autores compararam o método proposto com os algoritmos FBP, TV e FBP com pós-processamento via CNN. As métricas utilizadas para avaliação foram SNR e SSIM.

Para o conjunto de dados *AAPM Low Dose CT Grand Challenge*, os autores inseriram

de maneira artificial ruído Gaussiano aos sinogramas para que eles tivessem o SNR igual a 35, 40, 45, 70 e infinito dB, separando os 3 primeiros como conjunto de nível baixo de ruído e os dois últimos como conjunto de nível alto de ruído. Para o conjunto com baixo ruído, o método proposto obteve melhor resultado em todas as métricas e sobre todos os algoritmos avaliados. Para o conjunto com nível alto de ruído, o método proposto obteve melhor desempenho em 83% dos casos.

Para o conjunto de dados do cérebro do rato, os autores novamente inseriram manualmente ruído de Poisson nos sinogramas, correspondendo a 25, 30 e 35 dB de SNR. O método proposto obteve melhores resultados em todas as métricas de avaliação sobre todos os métodos comparados.

4.6.6 Método Learn

O Sensoriamento Comprimido (do inglês, *Compressed Sensing*), na aquisição de projeções tomográficas, é uma abordagem que obtém projeções esparsas e sub-amostradas. Em métodos de reconstrução iterativos, é comum a utilização de regularizadores para melhorar a qualidade da reconstrução, sendo que esses regularizadores devem ser escolhidos manualmente, o que pode se apresentar como uma tarefa difícil. Assim, Chen et al. (2018b) propuseram utilizar uma RNA para estimar um termo de regularização baseado em campos de especialistas (do inglês, *fields of experts*) (ROTH; BLACK, 2009), que é expressa por

$$R(x) = \sum_{k=1}^K \phi_k(G_k, x), \quad (4.46)$$

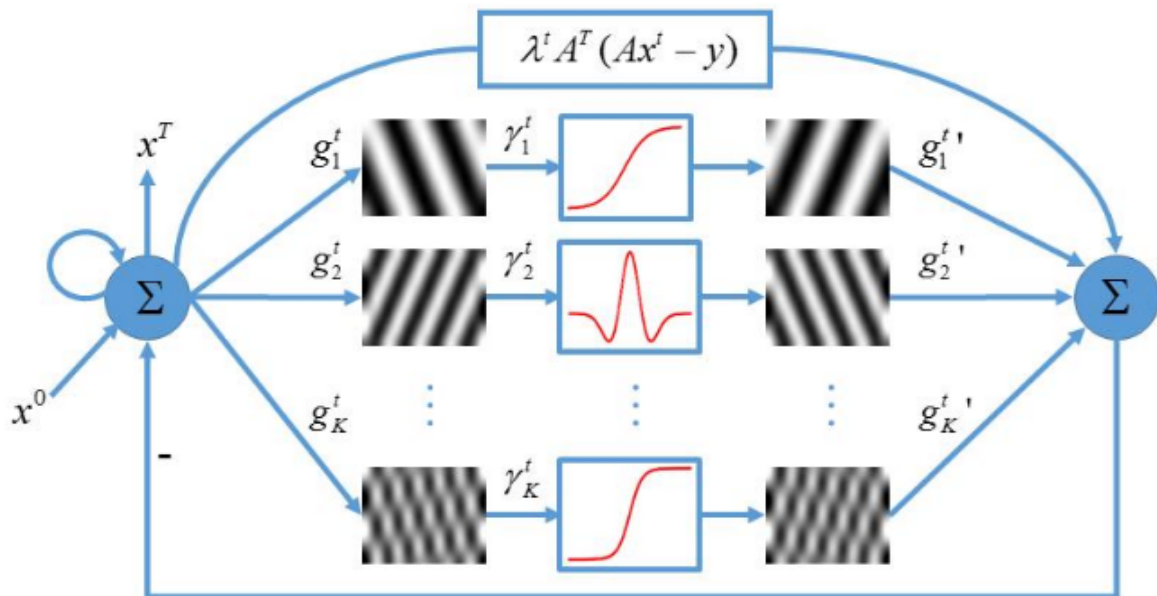
onde K é o número de regularizadores, G_k é uma matriz de transformação, que pode ser entendida como um operador convolucional para a imagem x , e $\phi_k(\cdot)$ é uma função potencial. Os termos G_k e ϕ_k podem ser aprendidos a partir dos dados. Dessa maneira, aplicando a função de regularização FoE na definição de reconstrução iterativa, tem-se que a imagem reconstruída x na iteração $t + 1$ é dada por

$$x^{t+1} = x^t - (\lambda^t A^T (Ax^t - y) + \sum_{k=1}^K (G_k^t)^T \gamma_k^t (G_k^t x^t)), \quad (4.47)$$

onde $\gamma(\cdot) = \phi'(\cdot)$, a letra sobrescrita T representa a operação de transposição de matriz, A^T representa o operador de retroprojeção, y representa a transformada de Radon do objeto. Os símbolos com a letra t sobrescrita representam que são dependentes da iteração anterior.

O termo $\sum_{k=1}^k (G_k^t)^T \gamma_k^t (G_k^t x^t)$ da Equação 4.47 pode ser representado por uma CNN. A cada iteração, a imagem é submetida à CNN e é somada com a iteração anterior, podendo ser considerada como uma CNN recorrente. Há duas camadas convolucionais, a primeira representa o termo $(G_k^t)^T$ e a segunda representa o termo G_k^t , que podem ser substituídas pelos *kernels* g_k^t e $g_k^{t'}$, respectivamente. A função de ativação é representada pelo termo γ_k^t . O termo $\lambda^t A^T (Ax^t - y)$ é somado com a saída da CNN, que pode ser considerado como a componente residual da RNA. A Figura 34 representa a rede neural resultante. Uma reconstrução inicial utilizando o algoritmo FBP pode ser disponibilizado como entrada.

Figura 34 – Rede neural recorrente residual que representa a Equação 4.47.



Fonte – Extraído de Chen et al. (2018b).

4.6.6.1 Experimentos e Resultados

A rede neural de Chen et al. (2018b) foi avaliada no contexto de CT utilizando poucas projeções, o que resultaria na presença de artefatos e redução da visibilidade de características importantes da imagem. O conjunto de dados utilizado também foi o disponibilizado pela *Mayo Clinics* para o *2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge*. No total, existem 5936 fatias de 1mm cada³, pertencentes a volumes reconstruídos a partir de 2304 projeções. Portanto, para simular uma reconstrução com dose reduzida, as projeções utilizadas para reconstruir as fatias utilizadas nos testes foram amostradas em 64 e 128, respectivamente.

³ Devido ao tamanho pequeno da fatia reconstruída, nota-se a presença de ruído até mesmo quando a reconstrução é feita a partir do conjunto completo de 2304 projeções.

Para os seus experimentos, os autores optaram por amostrar 250 do total de 5936 fatias reconstruídas, selecionando aleatoriamente 25 fatias de cada paciente. A abordagem proposta por Chen et al. (2018b) tem como entrada as projeções reduzidas e como saída as fatias reconstruídas. Assim, essa abordagem utiliza uma rede neural para reconstruir diretamente das poucas projeções dadas como entrada em uma imagem que representa uma fatia do interior do corpo imageado, diminuindo a quantidade de artefatos. Para que isso seja possível, é necessário treinar essa abordagem utilizando amostras contendo, de um lado, as projeções reduzidas e, do outro lado, as fatias reconstruídas com as projeções completas, pois a rede neural precisa estimar a melhor função para mapear as projeções sub-amostradas em uma fatia de qualidade. Os autores utilizaram 200 amostras para o treinamento e 50 amostras para o teste, e executaram validação cruzada para aumentar a confiança nos resultados. A função de custo MSE foi utilizada durante o treinamento e o algoritmo Gradiente Descendente foi utilizado para a otimização.

Os resultados obtidos foram comparados com outras cinco abordagens de reconstrução de imagens tomográficas, que são (i) ASD-POCS (SIDKY; PAN, 2008), (ii) aprendizado duplo de dicionário (Dual-DL) (LU; ZHAO; WANG, 2011), (iii) mínimos quadrados ponderados penalizados baseado em variação total generalizada (PWLS-TGV) (NIU et al., 2014), (iv) Gamma-Reg (ZHANG et al., 2017a) e (v) FBP com pós-processamento baseado em Aprendizado Profundo (FBPConvNet) (JIN et al., 2017). As métricas RMSE, PNSR e SSIM foram utilizadas para a avaliação dos métodos. Em todos os casos, o método de (CHEN et al., 2018b) obteve melhor resultado. Contudo, em relação ao tempo de processamento, ele teve resultado pior apenas em relação aos métodos FBP e FBPConvNet, mas teve um ganho considerável de velocidade em relação aos outros métodos.

4.6.7 PWLS-ULTRA e PWLS-ST

Para reconstrução de imagens utilizando Tomografia com Dose Baixa (LDCT), o trabalho de Zheng et al. (2018) apresentou dois métodos baseados nos Mínimos Quadrados Ponderados Penalizados (PWLS, do inglês *Penalized Weighted Least Squares*) para a reconstrução tomográfica de imagens, que utilizam duas funções de regularização diferentes, que são: (i) Transformação Quadrática (PWLS-ST, do inglês *Penalized Weighted Least Squares - Square Transform*) e (ii) União de Transformações Aprendidas (PWLS-ULTRA, do inglês *Penalized Weighted Least Squares - Union of Learned TRANSforms*). Para isso, a reconstrução de imagens tomográficas foi

considerada como um problema de otimização, como mostra a seguir,

$$\min_{x < 0} \frac{1}{2} \|y - Ax\|_W^2 + \beta R(x), \quad (4.48)$$

sendo $x \in \mathbb{R}^{N_p}$ a imagem a ser reconstruída a partir dos sinogramas ruidosos $y \in \mathbb{R}^{N_d}$, $W = \text{diag}\{w_i\} \in \mathbb{R}^{R_d \times R_d}$ sendo uma matriz diagonal de pesos, sendo cada peso a variância inversa estimada de y_i , $A \in \mathbb{R}^{N_d \times N_p}$ é a matriz de projeção e β é o parâmetro que controla a relação entre o regularizador $R(x)$ e o restante da equação.

Para o método PWLS-ST, é necessário aprender a transformação quadrática $\Omega \in \mathbb{R}^{l \times l}$ para ser aplicada na função de regularização, que é dada pela equação

$$\min_{\Omega, Z} \|\Omega X - Z\|_F^2 + \lambda Q(\Omega) + \sum_{i=1}^{N'} \eta^2 \|Z_i\|_0, \quad (4.49)$$

sendo l a quantidade de pixels da imagem, $\lambda = \lambda_0 \|X\|_F^2$, $\lambda_0 > 0$ uma constante, $\eta > 0$ são parâmetros escalares, $\{Z_i\}_{i=1}^{N'}$ são códigos esparsos das imagens de treinamento, $\{X_i\}_{i=1}^{N'}$ definidas em forma de vetor, sendo que a matriz $X \in \mathbb{R}^{l \times N'}$ contém as imagens de treinamento e a matriz $Z \in \mathbb{R}^{l \times N'}$ contém os códigos esparsos. O símbolo $\|\cdot\|_0$ efetua a contagem do número de valores não-zeros em um vetor (norma L_0). O termo $\|\Omega X - Z\|_F^2$ calcula o erro dos sinais transformados em relação às suas transformações esparsas. O regularizador $Q(\Omega) \triangleq \|\Omega\|_F^2 - \log|\det\Omega|$ evita soluções triviais.

Após isso, é possível aplicar Ω na função de regularização $R(x)$, com o objetivo de minimizar a Equação 4.48, sendo $R(x)$ então definido por

$$R(x) \triangleq \min_{z_j} \sum_{j=1} \tau_j \{ \|\Omega P_j x - z_j\|_2^2 + \gamma^2 \|z_j\|_0 \}, \quad (4.50)$$

sendo segmentos da imagem, o operador $P_j \in \mathbb{R}^{l \times N_p}$ responsável por extrair o j -ésimo segmento com o tamanho de l voxels de x , e o vetor $z_j \in \mathbb{R}^l$ denota a transformação esparsa de $P_j x$. A componente γ^2 ($\gamma > 0$) é um peso para a penalidade de dispersão baseada em ℓ_0 . Segundo os autores, o termo $\{\tau_j\}$ induz ruído e resolução espacial uniformes, e é definido como

$$\tau_j \triangleq \frac{\|P_j k\|_1}{l}, \quad (4.51)$$

sendo k definido a partir das entradas da matriz A (a_{ij}) e da matriz W , como

$$k \triangleq \sqrt{\frac{\sum_{i=1}^{N_d} a_{ij} w_i}{\sum_{i=1}^{N_d} a_{ij}}}. \quad (4.52)$$

Da mesma maneira, os autores proporam o arcabouço PWLS-ULTRA, que utiliza a união das transformações esparsificantes $\{\Omega_k\}_{k=1}^K$, junto com os coeficientes esparsos das imagens $\{Z_i\}$, sendo definido como

$$\min_{\Omega_k, Z_i, C_k} \sum_{k=1}^K \sum_{i \in C_k} \{ \|\Omega_k X_i - Z_i\|_2^2 + \eta^2 \|Z_i\|_0 \} + \sum_{k=1}^K \lambda_k Q(\Omega_k) \text{ sendo } \{C_k\} \in G, \quad (4.53)$$

onde K representa o número de classes, C_k é o conjunto de índices das imagens que pertencem a k -ésima classe, G o conjunto de todas as possíveis partições de $\{1, 2, \dots, N'\}$ e N' o número de imagens.

Assim, é possível encontrar uma solução para a Equação 4.48 utilizando um regularizador $R(x)$, que tem por objetivo aprender um conjunto de transformações esparsificantes $\{\Omega_k\}_{k=1}^K$, junto com os coeficientes esparsos das imagens $\{Z_i\}$, sendo definido como:

$$R(x) \triangleq \min_{z_j, C_k} \sum_{k=1}^K \left\{ \sum_{j \in C_k} \tau_j \{ \|\Omega_k P_j x - z_j\|_2^2 + \gamma^2 \|z_j\|_0 \} \right\} \text{ sendo } \{C_k\} \in G. \quad (4.54)$$

Dessa maneira, ao utilizar a função de regularização expressa pela Equação 4.54 no problema de otimização definido pela Equação 4.48, a imagem reconstruída x , os coeficientes esparsos das imagens $\{z_j\}$ e um conjunto de grupos $\{C_k\}$ são estimados a partir do sinograma com dados ruidosos y . A contribuição dos autores é, em uma etapa anterior à reconstrução, aprender as transformações $\{\Omega_k\}$ por meio de um conjunto de dados e utilizá-las na reconstrução.

4.6.7.1 Experimentos e Resultados

Os experimentos compararam os métodos PWLS-ULTRA e PWLS-ST com o corpo de prova XCAT (SEGARS et al., 2008) para CT 2D com feixe em leque e CT 3D com feixe cônico e com trajetória axial. Além disso, outro teste foi realizado com um conjunto de dados reais de CT helicoidal, contendo imagens do peito e do abdômen.

A utilização de CT 2D significa que o treinamento e os testes foram realizados levando em consideração apenas fatias 2D do corpo de prova XCAT, e não o volume inteiro. Portanto, nesse caso, a etapa de treinamento ocorre ao aprender as transformações quadráticas e união de transformações quadráticas a partir de segmentos de imagens sobrepostas com dimensões 8×8 pixels extraídas de 5 fatias de dimensões 512×512 pixels do corpo de prova XCAT, utilizando um deslocamento de 1×1 pixel para cada segmento extraído. Para os testes, outra fatia, agora com dimensões 840×840 pixels, foi extraída do corpo de prova XCAT, com pixels de tamanho $0,4883 \times 0,4883$ mm. As projeções foram realizadas com um equipamento *GE LightSpeed* com feixe em leque. O detector possui tamanho de 888×984 mm, com as projeções adquiridas tendo o mesmo tamanho. A quantidade de projeções e o intervalo angular percorrido não foram informados.

A utilização de CT 3D significa que o treinamento e os testes foram realizados considerando as 3 dimensões do corpo de prova XCAT. Portanto, a etapa de treinamento se resume em aprender as transformações quadráticas e a união das transformações quadráticas levando em consideração segmentos de $8 \times 8 \times 8$ pixels do corpo de prova XCAT (com sobreposição), com deslocamento de $2 \times 2 \times 2$ pixels em cada segmento, possuindo $420 \times 420 \times 54$ voxels. Para o conjunto de testes foram geradas projeções de CT axial usando o XCAT com dimensões $840 \times 840 \times 96$ voxels, com cada voxel possuindo dimensões de $0,4883 \times 0,4883 \times 0,625$ mm. Foram geradas 64 projeções de dimensões 888×984 pixels. Os autores não deixaram claro o intervalo angular utilizado e o tamanho de cada pixel do detector.

Os dados de pacientes reais foram utilizados para avaliar o método proposto com CT com trajetória helicoidal. As transformações quadráticas e a união das transformações quadráticas das imagens clínicas do abdômen, na etapa de treinamento, foram aprendidas em segmentos de imagens de $8 \times 8 \times 8$ pixels, com um deslocamentos de $3 \times 3 \times 3$ pixels. O volume reconstruído possui dimensões de $420 \times 420 \times 222$ voxels, tendo cada voxel dimensões de $1,1667 \times 1,1667 \times 0,625$ mm. Foram utilizadas 64 projeções de tamanho 888×2952 mm. Para melhor avaliar os resultados, um subconjunto de imagens do abdômen humano foram reconstruídas, utilizando o mesmo método de CT helicoidal, porém as transformações quadráticas e a união das transformações quadráticas das imagens clínicas do tórax, na etapa de treinamento, foram aprendidas em segmentos de imagens de $8 \times 8 \times 8$ pixels, com um deslocamentos de $3 \times 3 \times 3$ pixels. O volume reconstruído possui dimensões de $512 \times 512 \times 200$ voxels, tendo cada voxel dimensões de $1 \times 1 \times 0,625$ mm. Foram utilizadas 64 projeções de tamanho 888×2952 mm.

As imagens reconstruídas utilizando PWLS-ST e PWLS-ULTRA foram comparadas com outros algoritmos de reconstrução, entre eles (i) FBP, (ii) PWLS com um regularizador de preservação de bordas (PWLS-EP, do inglês, *Penalized Weighted Least Squares - Edge Preserving*) (CHO; FESSLER, 2014) e (iii) PWLS com regularização baseada em síntese de dicionário supercompleta aprendida (PWLS-DL, do inglês *Penalized Weighted Least Squares - Dictionary Learning*) (XU et al., 2012).

Os resultados foram avaliados em termos das medidas RMSE (do inglês *Root Mean Square Deviation*) e SSIM. Os resultados levando em consideração os dados provenientes de CT 2D com feixe em leque, utilizando o XCAT, demonstram que os métodos adaptativos PWLS-ULTRA e PWLS-ST apresentam excelentes resultados em comparação ao FBP e aos métodos PWLS-EP e PWLS-DL, que não são adaptativos. Em relação ao método PWLS-DL, que também é adaptativo, os métodos PWLS-ULTRA e PWLS-ST apresentaram resultados semelhantes ou melhores. Para as imagens provenientes de CT 3D com feixe cônico, os métodos PWLS-ULTRA e PWLS-ST aumentaram significativamente os valores de RMSE e SSIM em relação a todos os métodos comparados. Para as imagens provenientes de imagens realísticas do tórax, observou-se que as reconstruções provenientes do método PWLS-ULTRA possuem menor nível de ruído e artefatos, além de ter as bordas e outras características melhor preservadas. Nas imagens do abdômen humano, observou-se também que o método PWLS-ULTRA proveu uma reconstrução com menor nível de ruído e maior resolução, assim como melhor preservação de detalhes estruturais.

4.6.8 Discussões e avaliações

Nas seções anteriores foi observado como técnicas de Aprendizado de Máquina e Aprendizado Profundo podem ser incorporadas no processo de reconstrução de imagens tomográficas, resultando em imagens com menos ruído e maior qualidade. Entretanto, os vários métodos apresentados foram testados ou treinados em diferentes conjuntos de dados e confrontados com diferentes abordagens. Dessa maneira, as Tabelas 3 e 4 tem por objetivo padronizar algumas informações e possibilitar uma conclusão sobre o benefício desses métodos.

Tabela 3 – Classificação e detalhes experimentais dos métodos discutidos nas seções anteriores.

Método	Categoria	Tipo de Imageamento	Conjunto testado	Medidas de avaliação
(WÜRFL et al., 2018)	Esquema Iterativo Aprendido	CT Cone Beam	Dados reais	PSNR e SSIM
(ADLER; ÖKTEM, 2018)	Esquema Iterativo Aprendido	(i) CT Raios Paralelos (ii) CT Cone Beam	(i) Dados sintéticos (ii) Dados reais	PSNR, SSIM e tempo de execução
(CHEN et al., 2018a)	Regularizador Aprendido	CT Cone Beam	(i) Dados sintéticos (ii) Dados físicos	PNSR, CNR, ISNR e SSIM
(GUPTA et al., 2018)	Esquema Iterativo Aprendido	CT Cone Beam	Dados reais	SNR e SSIM
(CHEN et al., 2018b)	Regularizador Aprendido	CT Cone Beam	Dados reais	ISNR, PSNR, SSIM e tempo de execução
(ZHENG et al., 2018)	Regularizador Aprendido	CT Cone Beam	(i) Dados sintéticos	RMSE e SSIM

Tabela 4 – Detalhes sobre a performance dos métodos discutidos nas seções anteriores.

Método	Métodos confrontados	Resultados
(WÜRFL et al., 2018)	wTV Huang et al. (2018), Parker (PARKER, 1982) e uNet Gu e Ye (2017), Ronneberger, Fischer e Brox (2015)	(a) Superior em relação ao SSIM (b) Inferior ao wTV em relação ao PSNR
(ADLER; ÖKTEM, 2018)	FBP, FBP + uNet (RONNEBERGER; FISCHER; BROX, 2015) e TV, Learned Gradient (ADLER; ÖKTEM, 2017)	(a) Superior em relação ao SSIM e PSNR (b) Inferior ao FBP e superior ao TV em termos de execução
(CHEN et al., 2018a)	TV e FDK	Superior em relação ao PNSR, CNR, ISNR e SSIM.
(GUPTA et al., 2018)	FBP, FBP + convNet (JIN et al., 2017) e TV	Superior em relação ao SNR e SSIM
(CHEN et al., 2018b)	FBP + ConvNet (JIN et al., 2017), ASD-POCS (SIDKY; PAN, 2008), Dual-DL (LU; ZHAO; WANG, 2011), PWLS-TGV (NIU et al., 2014) e Gamma-Reg (ZHANG et al., 2017a)	(a) Superior em relação ao ISNR, PSNR, SSIM (b) inferior ao FBP + ConvNet e superior aos demais em relação ao tempo de execução
(ZHENG et al., 2018)	FBP, PWLS-EP (CHO; FESSLER, 2014) e PWLS-DL (XU et al., 2012)	Resultados superiores ou semelhantes ao PWLS-DL, e superior aos demais.

Na Tabela 3, a coluna “Método” se refere às abordagens explicadas nas Seções 4.6.1, 4.6.2, 4.6.3, 4.6.5, 4.6.6 e 4.6.7, respectivamente. Observe que o método explicado na Seção 4.6.4 não foi inserido, pois ele não apresenta uma abordagem nova de reconstrução, mas, como explicado, um método para otimizar os parâmetros de uma função de regularização pré-definida. A coluna “Categoria” representa o tipo de abordagem, como explicado no Capítulo 4.6. A coluna “Tipo de Imageamento” mostra o método em questão no qual os experimentos foram executados. A coluna “Conjunto testado” mostra o tipo do conjunto de dados no qual os experimentos foram realizados, cujos valores podem ser: (i) “dados realísticos”, que se refere a um conjunto que possui imagens realísticas de pacientes; (ii) “dados sintéticos”, que se refere a um conjunto de imagens que foram simuladas computacionalmente, normalmente contendo elementos específicos para o teste da qualidade da reconstrução; (iii) “dados físicos”, que se refere a corpos de prova físicos que foram imageados em um aparelho de CT real. A coluna “Medidas de avaliação” mostra as medidas que os autores utilizaram para avaliar a qualidade da reconstrução e comparar com outras abordagens.

Na Tabela 4, a coluna “Método” mostra, novamente, as abordagens explicadas na seção anterior. A coluna “Métodos confrontados” mostra as outras abordagens utilizadas para avaliar o desempenho da reconstrução. Por último, a coluna “Resultados” mostra alguns resultados que podem ser afirmados, de acordo com os valores expressos pelas medidas de avaliação (presentes na coluna “Medidas de avaliação” da Tabela 3), informados pelos autores. É importante mencionar que essas afirmações se originaram de uma avaliação quantitativa, e podem ser conferidas com mais detalhes na Seção anterior. Observe que a Tabela 4 é uma extensão da Tabela 3.

Dessa maneira, observando a coluna “Resultados” da Tabela 4, nota-se que os resultados adquiridos por abordagens baseadas em Aprendizado de Máquina, em sua maioria, supera os resultados de abordagens tradicionais, como o FBP e TV.

4.7 Considerações Finais do Capítulo

Neste Capítulo foram apresentadas as principais técnicas de reconstrução de imagens tomográficas. A vantagem da abordagem analítica é o tempo de reconstrução, que é menor do que da abordagem iterativa, o que fez com que uma ampla quantidade de equipamentos de mamografia implementassem tais abordagens.

Além disso, a Seção 4.6.1 mostrou uma técnica de reconstrução de imagens médicas

utilizando redes neurais artificiais, algo promissor considerando o grande avanço das RNAs nos últimos anos. Em relação à velocidade de reconstrução, as RNAs demandam muito tempo e capacidade de processamento na etapa de treinamento, mas na etapa de inferência o modelo pode ser processado rapidamente, com o tempo se aproximando dos obtidos por abordagens analíticas. Além disso, as RNAs possuem grande capacidade de paralelização, o que pode ajudar na velocidade da reconstrução tomográfica.

Além das abordagens que utilizam RNA diretamente na reconstrução, pode-se notar também aquelas que utilizam RNAs na etapa de Pós-Processamento. Por exemplo, na modalidade PET (*Positron Emission Tomography*), Yang, Ying e Tang (2018) utilizaram uma RNA para melhorar imagens reconstruídas por meio do algoritmo MAP EM. Eles utilizaram imagens cerebrais simuladas para compor o conjunto de treinamento, sendo que cada amostra possui a imagem reconstruída pelo MAP, contendo ruído, e a respectiva imagem sem o ruído. Dessa maneira, o autor interpreta o problema de melhorar as imagens como um problema de regressão, onde uma função deve ser aproximada pela RNA para levar as imagens ruidosas a um estado no qual elas não tenham ruído. Assim, a rede neural seria capaz de aprender as relações implícitas e melhorar novas imagens que não estejam no conjunto de treinamento. A RNA proposta teria como entrada uma imagem ruidosa e na saída seria fornecida a mesma imagem melhorada. A arquitetura de RNA utilizada possui uma camada de entrada, uma intermediária e uma de saída, as duas últimas sendo completamente conectadas. A função de custo utilizada pelo autor foi o Erro Quadrático Médio. Em todos os experimentos realizados, a imagem melhorada com a RNA atingiu um nível de ruído semelhante ou inferior do que as imagens reconstruídas sem a utilização da RNA.

Assim, com base na fundamentação teórica apresentada até o momento, uma proposta para a reconstrução tomográfica de DBT utilizando Redes Neurais Artificiais com Aprendizado Profundo foi desenvolvida e encontra-se discutida com maiores detalhes no próximo Capítulo.

5 Desenvolvimento do Projeto

Como discorrido no Capítulo 1, o câncer de mama é uma doença que afeta uma grande quantidade de pessoas ao redor do mundo. Por isso, um grande esforço foi e tem sido empregado para desenvolver exames de imageamento capazes de diagnosticar essa doença em estágios iniciais, como explicado no Capítulo 2. Em especial, a Tomossíntese Mamária é um método que teve sua utilização difundida recentemente, que diminui a sobreposição de tecidos em uma imagem tridimensional da mama, permitindo ao médico elaborar diagnósticos mais precisos.

Por outro lado, nos últimos anos, observou-se um grande progresso na área de Inteligência Artificial, mais especificamente na utilização de Redes Neurais com Aprendizado Profundo, como discorrido no Capítulo 3. Naturalmente, os especialistas em Processamento de Imagens Médicas tem avaliado DNNs para melhorar as imagens geradas por reconstrução tomográfica, como discorrido no Capítulo 4. Dessa maneira, é justificável os esforços empregados neste trabalho para investigar outros métodos de reconstrução baseados em DNNs para melhorar ainda mais o estado da arte em reconstrução de DBT, visto a importância do assunto e a relevância científica da área de Reconstrução Tomográfica e Redes Neurais com Aprendizado Profundo, como a metodologia específica neste Capítulo.

No decorrer deste Capítulo, a Seção 5.1 discute a metodologia empregada para reconstruir imagens de DBT. A Seção 5.2 discute a metodologia empregada nos experimentos. A Seção 5.3 descreve os materiais e métodos empregados nos experimentos. Por fim, a Seção 5.4 descreve como os resultados experimentais são avaliados.

5.1 Metodologia realizada para reconstrução de DBT

Würfl et al. (2018) já simularam o algoritmo FBP para reconstrução de imagens de CT com feixe cônico em uma Rede Neural Artificial, sendo que sua metodologia pode ser encontrada com mais detalhes na Seção 4.6.1.

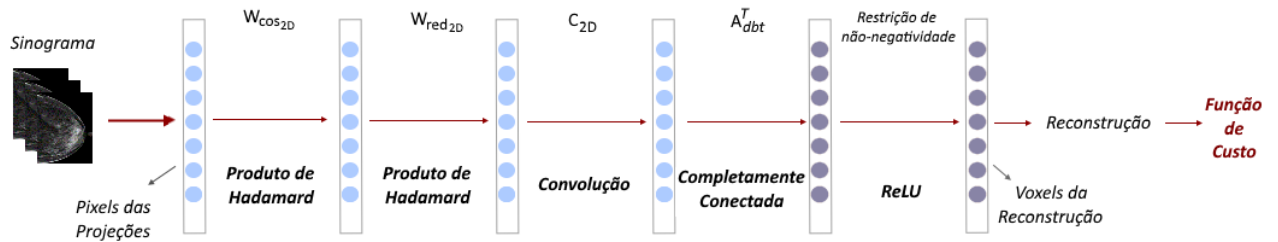
O software original foi desenvolvido para reconstruir projeções de CT, utilizando o operador A_{cb}^T , como descrito na Seção 4.6.1, não sendo possível reconstruir projeções de DBT. Desta maneira, coube a este trabalho desenvolver o operador A_{dbt}^T , tornando possível a reconstrução de projeções de DBT, a partir da adaptação e extensão da implementação do

operador A_{cb}^T para a geometria de DBT. A principal diferença entre A_{cb}^T e A_{dbt}^T é que, no primeiro, os raios advindos do emissor de radiação são do tipo cone, enquanto que, no segundo, são do tipo cone truncado. Assim, foi necessário ajustar as equações de projeção para que os voxels reconstruídos correspondessem corretamente aos pixels das projeções. Dessa maneira, a Equação para reconstrução de imagens de DBT passa a ser descrita como

$$f = A_{dbt}^T C_{2D} W_{red_{2D}} W_{cos_{2D}} P, \quad (5.1)$$

sendo A_{dbt}^T o operador de retroprojeção para DBT, implementado eficientemente em GPU, utilizando a biblioteca de computação paralela CUDA (NICKOLLS et al., 2008). A arquitetura de RNA que representa esta Equação está presente na Figura 35.

Figura 35 – Arquitetura de reconstrução de DBT utilizando os pesos de Parker.



Fonte – Autoria própria.

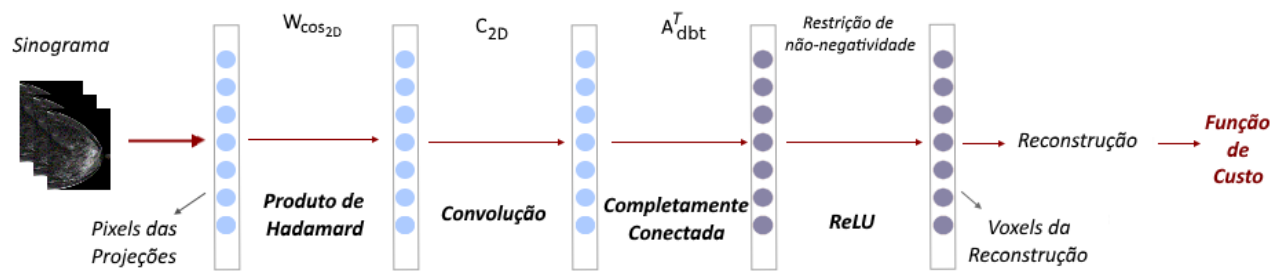
Implementações do FBP para DBT normalmente não utilizam as ponderações de Parker (VIMIEIRO; BORGES; VIEIRA, 2019; BARUFALDI et al., 2018b), representadas pelo operador $W_{red_{2D}}$. Entretanto, optou-se por deixar este operador como uma camada treinável, inicializada com as ponderações de Parker, para avaliar se a RNA conseguiria modificá-los para torná-los úteis na reconstrução de DBT, em termos de diminuição de ruído.

Para a reconstrução de imagens de DBT sem a utilização das ponderações de Parker, a Equação resultante fica

$$f = A_{dbt}^T C_{2D} W_{cos_{2D}} P, \quad (5.2)$$

cujas arquitetura de RNA correspondente está representada na Figura 36.

Figura 36 – Arquitetura de reconstrução de DBT sem a utilização dos pesos de Parker.



Fonte – Autoria própria.

É importante mencionar que, na Equação 5.1, a única camada treinável é a $W_{red_{2D}}$, enquanto que na Equação 5.2, a única camada treinável é a C_{2D} , inicializada com o filtro da rampa (Ram-Lak) (KAK; SLANEY, 1988). No restante deste texto, a arquitetura de reconstrução representada pela primeira Equação será chamada de Arquitetura Parker, enquanto que a arquitetura representada pela segunda equação será chamada de Arquitetura Ram-Lak.

Com a implementação do operador A_{dbt}^T , houve também a necessidade de implementar o operador A_{dbt} , que representa o gradiente da camada de retroprojeção na etapa de treinamento. Assim, por meio do operador A_{dbt} , os gradientes advindos das camadas após a retroprojeção podem ser transferidos para as camadas anteriores, possibilitando o ajuste dos parâmetros das camadas treináveis de toda a rede. Este operador também foi desenvolvido eficientemente utilizando a biblioteca CUDA.

Ressalta-se que, na referência base deste trabalho (WÜRFL et al., 2018), os autores desenvolveram o software para utilizar como entrada matrizes de projeção para descrever a geometria do equipamento de CT utilizado. Essas matrizes fornecem as informações sobre as posições do emissor de radiação, distância entre o detector e emissor de radiação, incremento angular entre cada projeção, etc. Entretanto, os autores não disponibilizaram nenhum modelo de arquivo de entrada. Portanto, coube a este trabalho deduzir as configurações de entrada, adaptar e padronizar para DBT, facilitando o uso para futuros usuários.

As projeções de DBT consideradas neste trabalho possuem pixels de dimensões $0,14 \times 0,14$ mm, seguindo o padrão dos equipamentos de DBT da *Hologic*. Entretanto, o software original detalhado na Seção 4.6.1 não possui implementação funcional para pixels menores que 1×1 mm. Dessa maneira, para que o trabalho pudesse ser desenvolvido, foi necessário efetuar

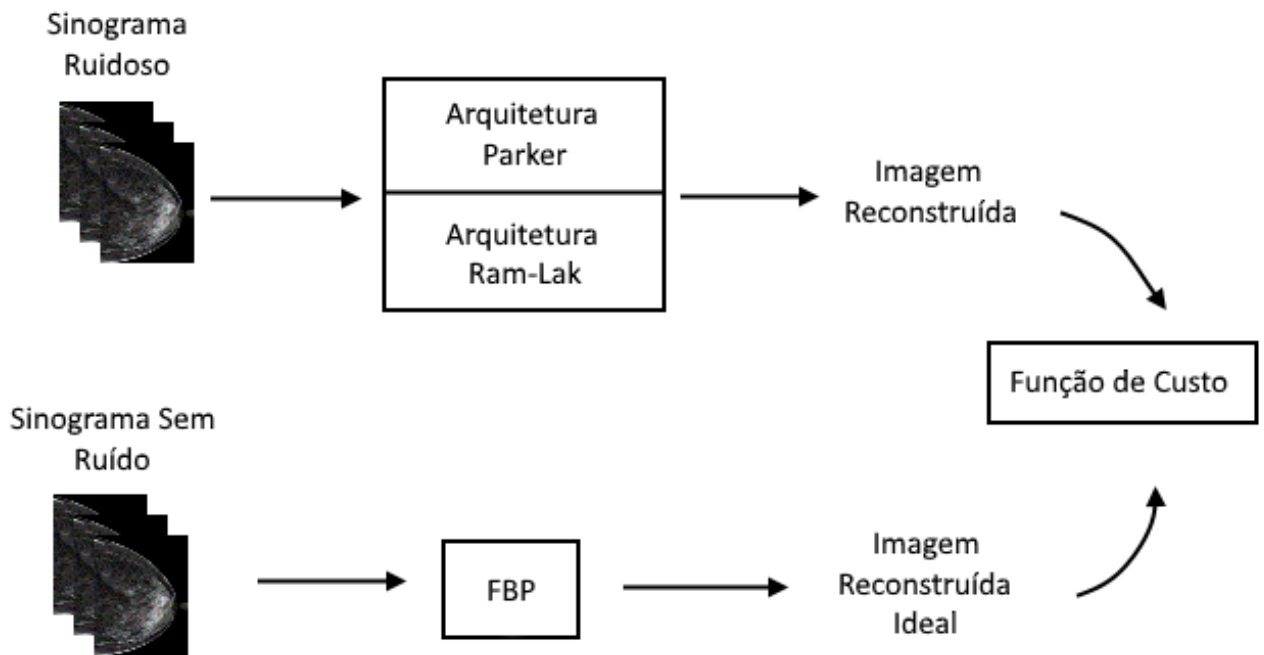
uma nova implementação das ponderações de Parker levando em consideração aplicações de DBT.

Além disso, no software original, não havia a implementação da lei de *Beer-Lambert* para o ajuste das projeções. Portanto, coube a este trabalho desenvolver esta funcionalidade e disponibilizá-la junto com o restante do software.

5.2 Metodologia de Avaliação Experimental

Como meta e objetivo principal, este trabalho buscou avançar o estado da arte de reconstrução de DBT, uma vez que as recentes propostas de reconstrução de imagens tomográficas utilizando Aprendizado Profundo apresentam bons resultados (Seção 4.6), mas ainda não haviam sido avaliadas para DBT, pelo nosso conhecimento. Com isso, seguindo o descrito na Seção anterior, é necessário avaliar a capacidade da Arquitetura Parker (Equação 5.1) e da Arquitetura Ram-Lak (Equação 5.2) quanto às suas capacidades na diminuição de ruído da reconstrução. Essas arquiteturas recebem o sinograma degradado com ruído como entrada, e, como rótulo, as imagens reconstruídas a partir das versões livre de ruído dos sinogramas. Dessa maneira, as RNAs devem mapear adequadamente o sinograma ruidoso para uma imagem reconstruída de melhor qualidade. Essas arquiteturas foram comparadas com as imagens reconstruídas pelo FBP, que recebe como entrada os sinogramas livres de ruído. Este processo de comparação pode ser visualizado na Figura 37. O otimizador utilizado para treinar as Arquiteturas Parker e Ram-Lak foi o Gradiente Descendente (KIEFER; WOLFOWITZ et al., 1952) (função `tf.train.GradientDescentOptimizer()` do Tensorflow).

Figura 37 – Metodologia de avaliação das arquiteturas de reconstrução propostas.

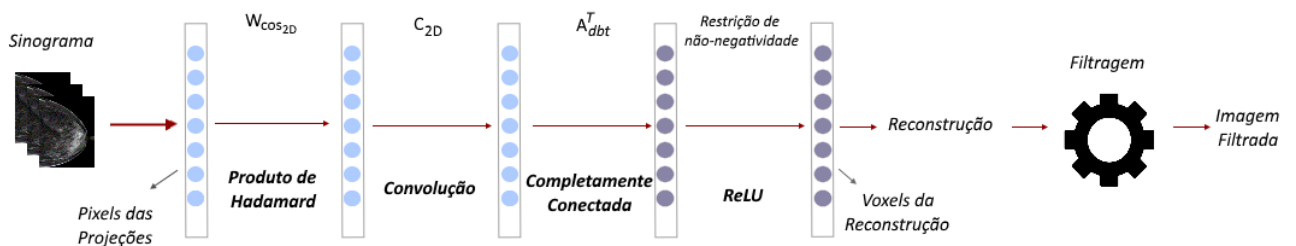


Fonte – Autoria própria.

Além disso, uma análise adicional foi efetuada utilizando a Arquitetura Ram-Lak. Essa análise avalia se a utilização de um método de filtragem junto com a Arquitetura Ram-Lak apresenta resultados melhores em comparação com a mesma imagem reconstruída pelo FBP e filtrada pelo mesmo método. Optou-se por utilizar o método de filtragem Rede Neural Convolutacional de Filtragem (do inglês, DnCNN) (ZHANG et al., 2017b) contendo 20 camadas, 64 filtros em cada camada, treinada utilizando o otimizador Adam (KINGMA; BA, 2014) e a função de custo L_2 . Esta abordagem utiliza *Deep Learning* para mapear imagens ruidosas em suas versões sem ruído. A DnCNN foi treinada separadamente dos métodos de reconstrução, e a filtragem foi efetuada após a reconstrução, como um método de pós-filtragem, assim como mostra a Figura 38. Devido aos custos computacionais, foi utilizada uma Região de Interesse (RoI, do inglês *Region of Interest*) de 180×180 px dos interiores das mamas, para as etapas de treinamento e testes da DnCNN. Para comparar a eficiência da arquitetura Ram-Lak na eliminação do ruído em relação a outras abordagens de filtragem, os métodos *Block-Matching and 3D Collaborative Filtering* (BM3D) (DABOV et al., 2007) e o Nonlocal Means (NLM) (VIEIRA et al., 2015) foram também utilizados para filtrar as 15 projeções antes da reconstrução, sendo então classificados como métodos de pré-filtragem. A transformada de Anscombe (VIEIRA et al., 2013) foi efetuada nas projeções antes da filtragem para aproximar a distribuição de probabilidade do ruído, que

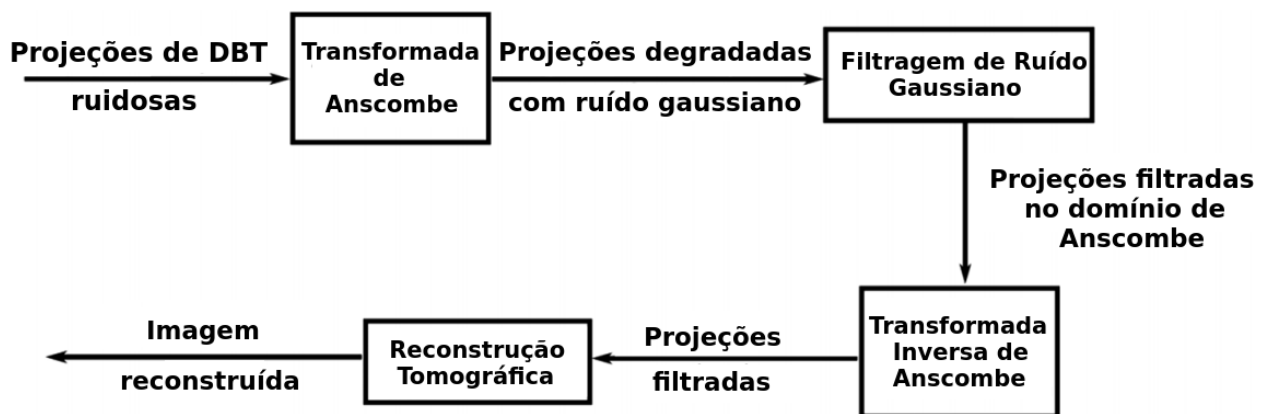
possui inicialmente uma distribuição de Poisson, em uma distribuição Gaussiana com média igual a zero e desvio padrão igual a 1. Após a aplicação do método de filtragem, a transformada inversa de Anscombe foi executada para voltar as projeções ao domínio original. Estas projeções filtradas por BM3D ou NLM foram reconstruídas por FBP, sendo uma abordagem tradicional de reconstrução tomográfica. A Figura 39 ilustra o procedimento explicado.

Figura 38 – RNA para reconstrução de DBT seguida de um método de filtragem.



Fonte – Autoria própria.

Figura 39 – Procedimento de filtragem das projeções pré-reconstrução usado nos métodos de comparação (filtros NLM ou BM3D para filtrar projeções), com a reconstrução por FBP.



Fonte – Adaptado de Scarparo et al. (2019).

5.3 Materiais e Métodos

Os conjuntos de dados disponíveis desempenham papel importante na qualidade dos resultados gerados por uma RNA. Conjunto de imagens clínicas em várias doses de radiação, como necessárias para a realização deste projeto, são inviáveis de serem adquiridas, devido aos

empecilhos impostos pela radiação em exames de DBT. Entretanto, uma solução é a utilização de imagens que simulam computacionalmente projeções de DBT, possibilitando um amplo conjunto de imagens com diversas variações estruturais da mama, algo necessário para o treinamento e teste de Redes Neurais Profundas. Essa abordagem permite a geração de imagens com ruído e sem ruído. Para este trabalho foram utilizados um conjunto de 100 corpos de prova gerados pelo software *OpenVCT* (BARUFALDI et al., 2018b), desenvolvido pela Universidade da Pensilvânia, sendo que as imagens geradas são muito próximas das imagens de pacientes reais (BARUFALDI et al., 2018a). Portanto, devido à facilidade de gerar novos modelos virtuais de mamas por meio desse software, as 100 amostras do conjunto de phantoms simulados contém as devidas variações de tamanhos, densidade e lesões, sem prejudicar a saúde de nenhum paciente e mantendo-se muito próximo da realidade.

O trabalho de Würfl et al. (2018) foi implementado em Python e em C++, utilizando a biblioteca de Deep Learning Tensorflow¹. Por isso, as implementações desse trabalho utilizaram o código de Würfl et al. (2018) como base, mantendo as mesmas linguagens de programação e bibliotecas.

Todos os dados utilizados neste trabalho para os testes da RNA consideram projeções adquiridas a partir do equipamento de DBT da *Hologic*, com o detector de radiação fixo e com o emissor de radiação tendo uma trajetória angular a 700mm acima da mama da paciente, indo de -7.5 até $7,5$, considerando o eixo vertical perpendicular ao detector de radiação como sendo 0. Em cada exame, foram geradas 15 projeções (em intervalos igualmente espaçados) de tamanho 2048×1792 pixels, que é o mesmo tamanho do detector, sendo que cada pixel possui $0,14 \times 0,14$ mm. O volume reconstruído tem dimensões de $2048 \times 2048 \times 64$ voxels, cada voxel tendo dimensões de $0,2 \times 0,2 \times 1$ mm.

5.4 Avaliação dos Resultados

O rótulo das amostras pertencentes aos conjuntos de dados simulados foram reconstruídos a partir das projeções sem ruído, pois, como explicado, as RNAs utilizaram essas reconstruções como base para o treinamento.

Para garantir a eficácia da avaliação da RNA, o conjunto de dados foi dividido em

¹ O código fonte se encontra disponível para a comunidade na URL <https://github.com/ma0ho/Deep-Learning-Cone-Beam-CT>.

conjunto de treinamento, conjunto de testes e conjunto de validação. O conjunto de treinamento é utilizado para treinar a RNA e o conjunto de validação é utilizado para definir o ponto de parada do treinamento. Após o treinamento, a avaliação da acurácia da rede neural é efetuada utilizando o conjunto de testes. Para a avaliação da Arquitetura Parker (Seção 6.1) e da Arquitetura Ram-Lak (Seção 6.2), o conjunto de dados foi dividido em 79 amostras para o conjunto de treinamento, 1 amostra para o conjunto de validação e 20 amostras para o conjunto de testes. O treinamento das arquiteturas avaliadas foi realizado até que o erro no conjunto de validação apresentasse pouca variação. Para a avaliação das abordagens de filtragem (Seção 6.2.1), o conjunto de dados foi dividido em 80 amostras para treinamento e 20 amostras para testes.

As imagens reconstruídas obtidas pela arquitetura de RNA proposta foram comparadas com as imagens reconstruídas pelo método FBP. Esta comparação foi feita utilizando as medidas quantitativas *Peak Signal-to-Noise Ratio* (PSNR), expressa pela Equação 5.3:

$$\begin{aligned} PSNR(x, y) &= 10 \cdot \log_{10}\left(\frac{MAX_x^2}{MSE(x, y)}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_x}{MSE(x, y)}\right) \\ &= 20 \cdot \log_{10}(MAX_x) - 10 \cdot \log_{10}(MSE(x, y)), \end{aligned} \quad (5.3)$$

e *Structural Similarity Index* (SSIM) (WANG; BOVIK, 2009), expressa pela Equação 5.4:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (5.4)$$

A medida de avaliação SSIM foi escolhida pois ela utiliza características conhecidas do sistema visual humano (WANG et al., 2004), enquanto que a medida de avaliação PSNR foi escolhida pois ela é uma medida tradicional na área de Processamento de Imagens para avaliar a quantidade de ruído em uma imagem.

Por fim, no próximo Capítulo a metodologia foi seguida, apresentando os resultados experimentais detalhadamente.

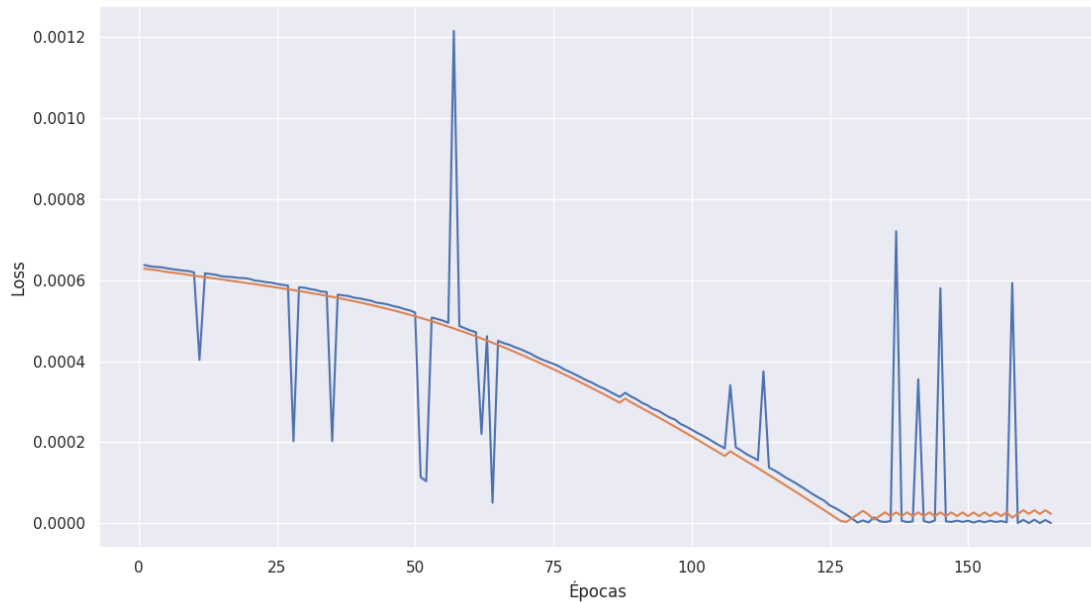
6 Experimentos e Resultados

Neste Capítulo foram avaliados os desempenhos da Arquitetura Parker (Equação 5.1) e da Arquitetura Ram-Lak (Equação 5.2) na diminuição de ruído das reconstruções de DBT, utilizando as métricas de comparação quantitativas SSIM e PSNR, além de análises qualitativas. Os resultados da Arquitetura Parker estão detalhados na Seção 6.1, enquanto que os resultados da Arquitetura Ram-Lak estão detalhada na Seção 6.2, incluindo a combinação de filtragem com reconstrução.

6.1 Arquitetura Parker

O treinamento durou cerca de 12 horas utilizando uma GPU K80 com 12GB de memória, em uma máquina com 6 CPUs e 56 GB de memória. O erro utilizado foi L_1 , sendo que os valores para o conjunto de treinamento e validação no decorrer das épocas de treinamento podem ser encontrados na Figura 40. Uma busca foi feita para encontrar a melhor Taxa de Aprendizado (TA), indo de $TA = 1 * 10^{-9}$ até $TA = 1 * 10^3$, com intervalos de 10, sendo que $TA = 1$ apresentou os melhores resultados e com a convergência mais rápida.

Figura 40 – Valores de perda no conjunto de treinamento e validação. Linha azul é o erro no conjunto de treinamento, enquanto a linha laranja é o erro no conjunto de validação.



Fonte – Autoria própria.

Repare na Figura 40 que a convergência da arquitetura acontece a partir da época 128, sendo a Loss no conjunto de validação igual a $5,5 * 10^{-6}$ e a Loss no conjunto de treinamento igual a $2,92 * 10^{-6}$, sendo esta época escolhida para a análise do PSNR e SSIM.

A média dos valores de PSNR e SSIM foram obtidas para as 20 amostras presentes no conjunto de testes, as quais podem ser encontradas na Tabela 5, onde a linha FBP se refere ao método de reconstrução analítico utilizando o método FBP e as linhas Arquitetura Parker e Arquitetura Ram-Lak se referem as reconstruções utilizando às referidas arquiteturas, treinadas no conjunto de corpos de prova simulados, sendo os resultados desta última arquitetura discutidos na Seção 6.2.

Tabela 5 – Desempenho da Arquitetura Parker para a diminuição do ruído.

Métodos	SSIM	PSNR
FBP	0,396 ± 0,021	26,931 ± 0,18
Arquitetura Parker	0,202 ± 0,009	21,162 ± 0,273
Arquitetura Ram-Lak + L1	0,456 ± 0,0155	28,054 ± 0,131
Arquitetura Ram-Lak + L2	0,446 ± 0,0184	27,827 ± 0,186

Observando a Tabela 5, é possível notar a perda de qualidade que a Arquitetura Parker ocasionou nas imagens reconstruídas, obtendo um PSNR 5,769 dB e um SSIM 0,194 menor que o FBP. Isso pode ser confirmado inspecionando visualmente a fatia central de uma amostra do conjunto de teste reconstruída, mostrada na Figura 41a. Nesta Figura, é possível observar diversos artefatos (linhas verticais) predominante na metade direita da amostra. Entretanto, a Figura 41b mostra a respectiva fatia reconstruída utilizando os pesos originais de Parker (PARKER, 1982), e nota-se que ela está incorreta, pois a distribuição da luminosidade está diferente da fatia central da amostra reconstruída sem a utilização das ponderações de Parker, mostrada na Figura 41c. Esse efeito negativo nas reconstruções pode estar relacionado com o fato de que as Ponderações de Parker (PARKER, 1982) foram desenvolvidas para serem utilizadas com as projeções adquiridas em um amplo intervalo angular.

Desta maneira, a Arquitetura Parker não foi capaz de corrigir os pesos iniciais de Parker, na camada $W_{red_{2D}}$, para que a ela se tornasse útil na reconstrução de DBT. Isso pode ser resultante da pequena complexabilidade da rede e do fato que a camada $W_{red_{2D}}$ não é completamente conectada, mas sim, um Produto de Hadamard.

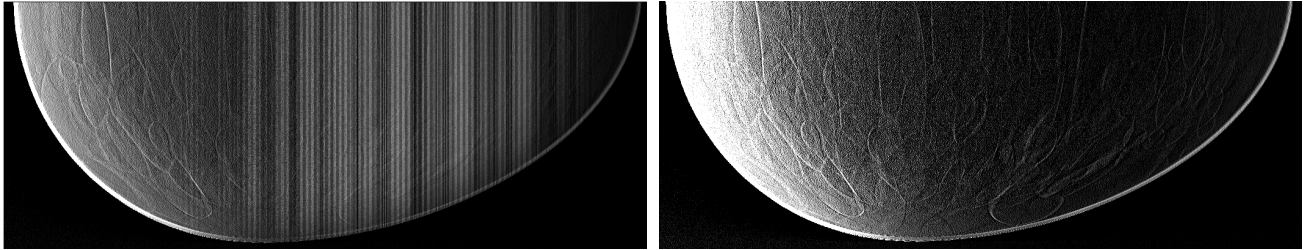
6.2 Arquitetura Ram-Lak

A arquitetura Parker, descrita na Seção anterior, se mostrou ineficiente para a reconstrução de imagens de DBT. Portanto, uma alternativa é a utilização da Arquitetura Ramlak, expressa pela Equação 5.2.

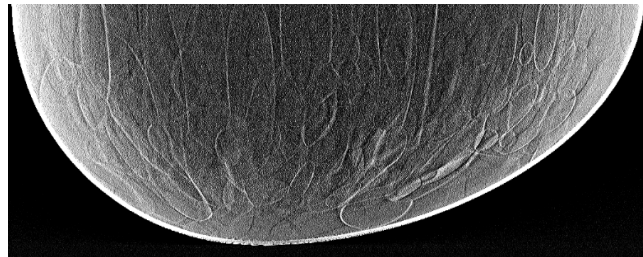
A máquina cujos experimentos foram executados foi a mesma descrita na Seção anterior.

Figura 41 – Reconstruções de uma amostra do conjunto de teste.

- (a) Amostra do conjunto de teste reconstruída usando a Arquitetura Parker. (b) Amostra do conjunto de teste reconstruída usando FBP e as ponderações originais de Parker (PARKER, 1982).



- (c) Amostra do conjunto de teste reconstruída usando o FBP.



Fonte – Autoria própria.

O treinamento durou cerca de 1 hora. Uma busca foi feita para encontrar a melhor Taxa de Aprendizado, indo de $TA = 1 * 10^{-9}$ até $TA = 1 * 10^{-6}$, com intervalos de 10, sendo que a que apresentou melhores resultados com a convergência mais rápida foi $TA = 1 * 10^{-9}$.

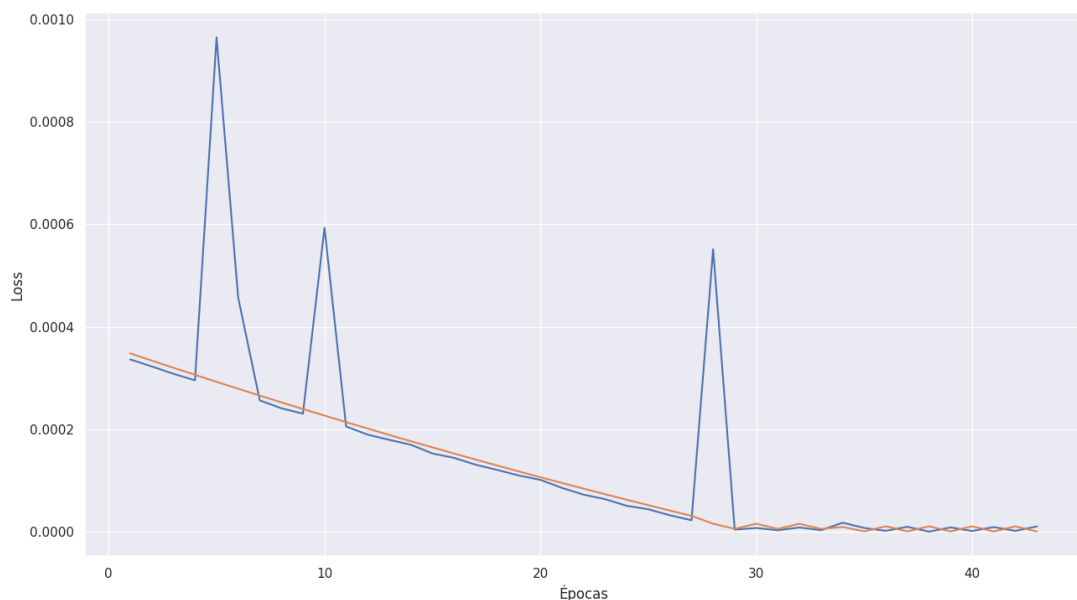
Repare na Figura 42 que a convergência da arquitetura acontece a partir da época 29. Portanto, para a avaliação dos resultados, foi escolhido os parâmetros referente a época 40, sendo a Loss no conjunto de validação igual a $4,28 * 10^{-7}$ e a Loss no conjunto de treinamento igual a $8,17 * 10^{-6}$.

A média dos valores de PSNR e SSIM foram obtidas para as 20 amostras presentes no conjunto de testes, as quais podem ser encontradas na Tabela 5, onde a linha FBP se refere ao método de reconstrução analítico FBP, a linha Arquitetura Ram-Lak + L_1 se refere à reconstrução utilizando a arquitetura treinada com a norma L_1 como função de custo, e a linha Arquitetura Ram-Lak + L_2 se refere à reconstrução utilizando a arquitetura treinada com a norma L_2 como função de custo. Repare que a Arquitetura Ram-Lak + L_1 obteve melhores resultados em termos de SSIM e PSNR, alcançando um SSIM de 0,06 e um PSNR 1,123 dB maior que o FBP, um SSIM de 0,01 e PSNR 0,227 dB maior que a Arquitetura Ram-Lak + L_2 e um SSIM 0,254 e um PSNR 6,892 dB maior que a Arquitetura Parker.

Desta maneira, a Figura 43 permite a realização de uma inspeção visual. A Figura 43a

mostra a fatia central de um corpo de prova ruidoso do conjunto de teste, reconstruída pelo FBP, a Figura 43b mostra a fatia do corpo de prova reconstruída pela Arquitetura Ram-Lak + L_1 , a Figura 43c mostra a fatia do corpo de prova reconstruída pela Arquitetura Ram-Lak + L_2 , e a Figura 43d mostra a fatia ideal (sem ruído). É possível observar que a imagem reconstruída pela Arquitetura Ram-Lak obteve uma melhora em termos de ruído, mas ainda está distante da qualidade apresentada pela amostra ideal.

Figura 42 – Valores de perda no conjunto de treinamento e validação. Gráfico azul é o erro no conjunto de treinamento, enquanto o gráfico laranja é o erro no conjunto de validação.



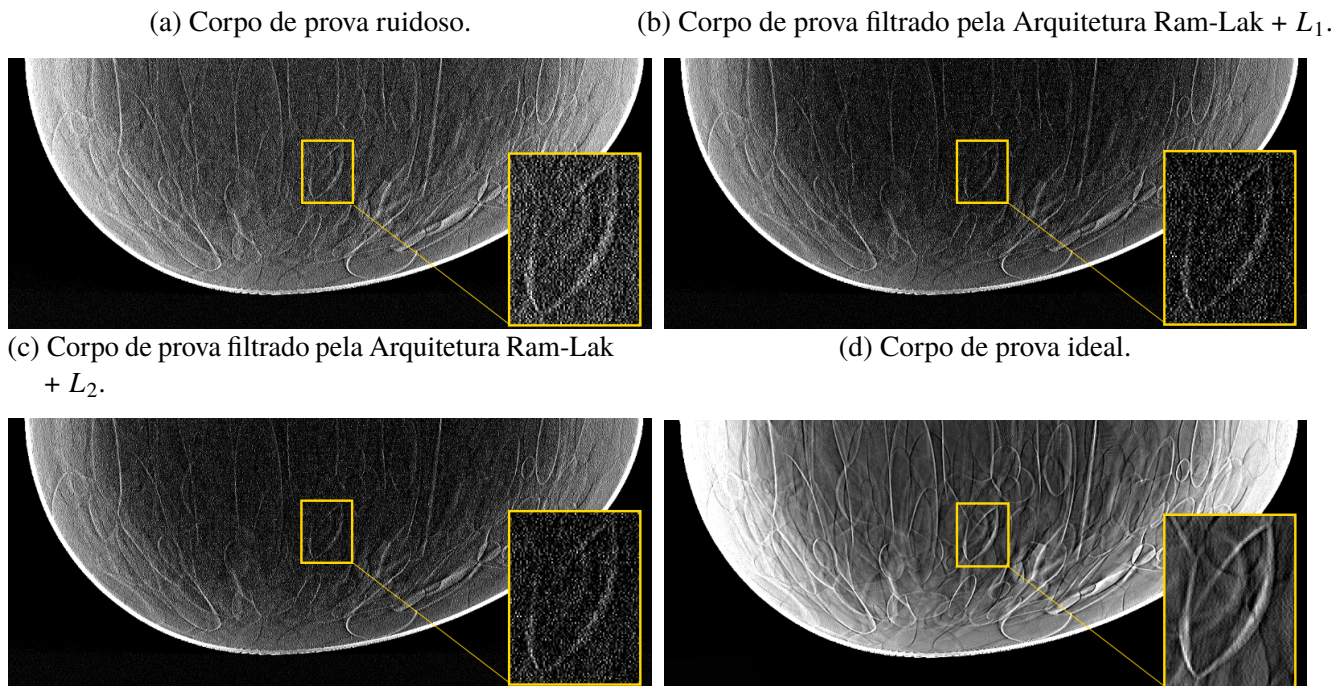
Fonte – Autoria própria.

6.2.1 Comparação com Abordagens de filtragem

Na Seção anterior, foi possível observar que a Arquitetura Ram-Lak + L_1 foi capaz de reconstruir as imagens com uma leve melhora em relação ao FBP, mas ainda muito degradada em relação às imagens ideais. Esta Seção compara a Arquitetura Ram-Lak + L_1 + DnCNN com outras abordagens de filtragem, assim como descrito na Seção 5.2.

Na Tabela 6 encontram-se a média e o desvio padrão dos valores de SSIM e PSNR para os segmentos de 180×180 px das 20 amostras do conjunto de testes. A linha BM3D+FBP apresenta os resultados da pré-filtragem feita pelo BM3D com a reconstrução feita pelo FBP, a

Figura 43 – Comparação de um corpo de prova presente no conjunto de teste.



Fonte – Autoria própria.

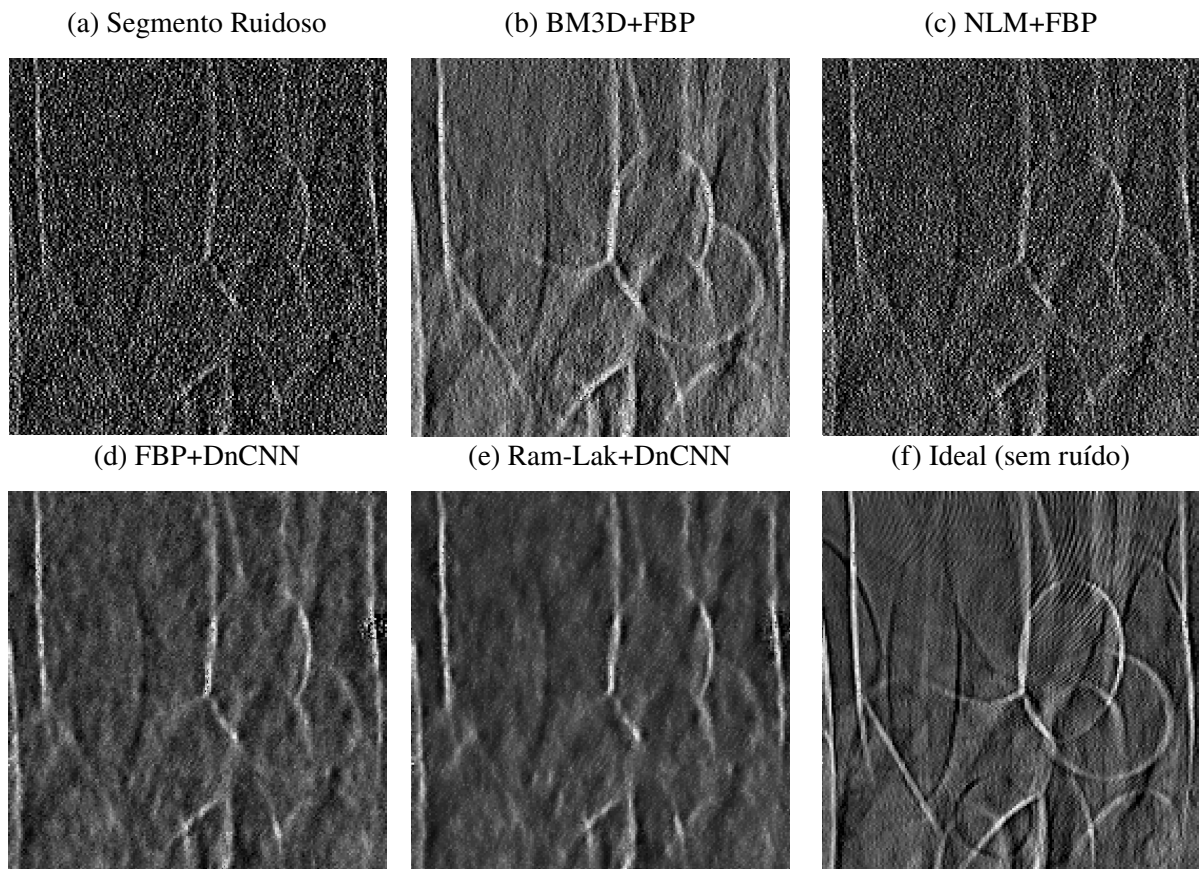
Tabela 6 – Resultados experimentais para os métodos de filtragem.

Métodos	SSIM	PSNR
BM3D+FBP	0,854 ± 0,006	37,102 ± 0,334
NLM+FBP	0,636 ± 0,009	31,91 ± 0,160
FBP+DnCNN	0,864 ± 0,010	37,412 ± 0,623
Arquitetura Ram-Lak + L1 + DnCNN	0,869 ± 0,010	37,644 ± 0,521

linha NLM+FBP apresenta os resultados da pré-filtragem feita pelo NLM e com a reconstrução feita pelo FBP, a linha FBP+DnCNN apresenta os resultados da pós-filtragem feita pelo DnCNN e com a reconstrução feita pelo FBP, e a linha Arquitetura Ram-Lak + L_1 + DnCNN apresenta a pós-filtragem feita pela DnCNN e a reconstrução feita pela Arquitetura Ram-Lak + L_1 .

Por outro lado, a Figura 44 mostra os segmentos de 180x180px de uma amostra do conjunto de teste que foram submetidos aos métodos de filtragem dispostos na Tabela 6, possibilitando uma análise visual. A Figura 44a é o segmento com ruído, a Figura 44b é o segmento do método BM3D+FBP, a Figura 44c é o segmento do método NLM+FBP, a Figura 44d é o segmento do método FBP+DnCNN, a Figura 44e é o segmento reconstruído pela Arquitetura Ram-Lak + L_1 e pós-filtrado pelo método DnCNN, e a Figura 44f é o segmento ideal (rótulo).

Figura 44 – Comparação dos resultados em relação a uma etapa adicional para filtragem pre- ou pós-reconstrução.



Fonte – Autoria própria.

Observa-se que o método BM3D+FBP obteve bons resultados, diminuindo o ruído e preservando detalhes sutis. O método NLM+FBP não foi eficiente em eliminar o ruído, apresentando um SSIM 0,23 e PSNR 5,734 menor que a Arquitetura Ram-Lak + L_1 + DnCNN. Os métodos FBP+DnCNN e Arquitetura Ram-Lak + L_1 + DnCNN apresentaram bons resultados na eliminação do ruído, embora seja possível observar um borramento maior em comparação com o método BM3D+FBP, mesmo que na Tabela 6 eles apresentaram um aumento de 0,01 de SSIM e 0,31 dB de PSNR e 0,015 de SSIM e 0,542 dB de PSNR, respectivamente, em comparação com o BM3D+FBP. Por fim, o método Arquitetura Ram-Lak + L_1 + DnCNN apresentou um aumento de 0,005 de SSIM e 0,232 de PSNR em comparação com o método FBP+DnCNN.

7 Conclusões e Trabalhos Futuros

Os índices de mortalidade do câncer vem aumentando no Brasil¹, o que tem relação com o aumento da expectativa de vida da população e pode ter influência dos novos hábitos resultantes da urbanização (PORTER, 2008). De igual maneira, o câncer de mama também apresenta um aumento nos índices de incidência e mortalidade, tornando necessário explorar novos tratamentos e métodos de diagnósticos que diminuam o avanço dessa doença. Por isso, nos Capítulos anteriores, esse trabalho explorou como a tomossíntese mamária (Capítulo 2) pode trazer benefícios para o diagnóstico precoce do câncer de mama. No Capítulo 3, foi detalhado como as técnicas de Inteligência Artificial ganharam espaço por seus bons resultados em diversas áreas. Assim, é natural aplicar essas técnicas na área médica, na tentativa de levar para a população ferramentas diagnósticas contra o câncer mais eficazes, especialmente visando melhoria na qualidade das imagens, para permitir redução da necessidade de biópsia e redução da dose de radiação. Muitos trabalhos envolvendo a utilização de Inteligência Artificial em imagens médicas já foram propostos (LITJENS et al., 2017), em especial o trabalho de Würfl et al. (2018), discutido na Seção 4.6.1, que descreve como o algoritmo de reconstrução tomográfica FBP pode ser simulado em uma RNA, que é base para a metodologia proposta no Capítulo 5.

Dessa maneira, este trabalho apresentou duas arquiteturas de RNAs para reconstruir imagens de DBT, as quais foram avaliadas em termos da capacidade de diminuição de ruído. A primeira, chamada de Arquitetura Parker, se mostrou não adequada para a reconstrução de imagens de DBT. A segunda, chamada de Arquitetura Ram-Lak, mostrou resultados relevantes em comparação aos métodos de estado-da-arte em reconstrução de imagens de DBT, especialmente com a adição de uma etapa adicional de filtragem realizada por Rede Neural Artificial com Aprendizado Profundo. Em uma análise quantitativa, esta arquitetura apresentou os melhores resultados. Entretanto, a análise qualitativa demonstrou uma tendência de borramento maior que abordagem utilizando a pré-filtragem BM3D com reconstrução por FBP, que é uma das abordagens de estado-da-arte para reconstrução de DBT. Assim, a Arquitetura Ram-Lak ficou aquém do esperado, pois esperava-se que o filtro Ram-Lak se adaptasse para um filtro mais eficiente e adequado para gerar imagens reconstruídas de DBT com ruído reduzido.

¹ O painel iterativo da Organização Mundial da Saúde contendo dados sobre o câncer pode ser encontrado em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt10uf.def>

Apesar disso, foi efetuada a implementação de uma biblioteca na linguagem de programação Python² que, baseando-se no trabalho de Würfl et al. (2018), utiliza uma DNN para simular analiticamente o FBP para reconstruir imagens de DBT.

Ainda, como parte do plano inicial, estava previsto para este trabalho considerar como volume alvo (rótulos das amostras) os dados originais reconstruídos por métodos iterativos, tal como o SART. Entretanto, como a rede não foi capaz de aprender o dado original reconstruído por FBP, entendeu-se que o mesmo deveria ocorrer com estes métodos. Unindo a isto, a ideia de que a DnCNN, contendo múltiplas camadas, foi adicionada após a arquitetura Ram-Lak e foi capaz de melhorar consideravelmente a qualidade da imagem, concluiu-se que há a necessidade de adicionar camadas adicionais (possivelmente mais de 15 camadas, sejam elas antes e/ou após a reconstrução) à arquitetura Ram-Lak básica, a fim de gerar uma DNN mais poderosa para o problema de reconstrução de DBT.

Dessa maneira, o desenvolvimento de uma DNN capaz de aprender satisfatoriamente o volume livre de ruído pode ser reavaliado no futuro. Assim, como trabalhos futuros, espera-se: (i) avaliar a geração dos rótulos utilizando projeções adquiridas em todo o espaço angular, oferecendo uma referência mais precisa para o treinamento da DNN, possibilitando também avaliar a capacidade da DNN em aprender dados incompletos; (ii) avaliar se a utilização de dados clínicos junto com o conjunto de dados simulados melhora o desempenho da DNN; (iii) avaliar se o treinamento fim-a-fim da DnCNN junto com a arquitetura Ram-Lak melhora a qualidade das imagens reconstruídas; (iv) usar um conjunto de dados maior, (v) efetuar uma filtragem pré-reconstrução e pós-reconstrução na arquitetura Ram-Lak (dupla-filtragem) e (vi) comparar o tempo de execução da arquitetura Ram-Lak com abordagens de reconstrução de DBT analíticas e iterativas. Dito isso, a força desta dissertação está justamente no desenvolvimento e codificação da arquitetura básica para a realização de reconstrução de DBT, além da indicação de caminhos futuros para a solução deste problema.

Por fim, como limitação do método proposto, vale ressaltar que no uso de equipamentos com outras configurações/geometrias, tal como equipamentos de DBT da GE, a rede precisaria ser treinada totalmente com dados nesta arquitetura.

² Link do repositório no GitHub: <https://github.com/daviduarte/DeepLearningDBT>

Referências

- ABADI, M. et al. Tensorflow: A system for large-scale machine learning. *CoRR*, abs/1605.08695, 2016. Disponível em: <<http://arxiv.org/abs/1605.08695>>.
- ADLER, J.; ÖKTEM, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, IOP Publishing, v. 33, n. 12, p. 124007, 2017.
- ADLER, J.; ÖKTEM, O. Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1322–1332, 2018.
- ANDOLINA, V.; LILLÉ, S. *Mammographic imaging: a practical guide*. [S.l.]: Lippincott Williams & Wilkins, 2011.
- ANIL, R. et al. Large scale distributed neural network training through online distillation. *CoRR*, abs/1804.03235, 2018. Disponível em: <<http://arxiv.org/abs/1804.03235>>.
- AZZARI, L.; BORGES, L. R.; FOI, A. Modeling and estimation of signal-dependent and correlated noise. In: *Denoising of Photographic Images and Video*. [S.l.]: Springer, 2018. p. 1–36.
- BARUFALDI, B. et al. Developing populations of software breast phantoms for virtual clinical trials. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *14th International Workshop on Breast Imaging (IWBI 2018)*. [S.l.], 2018. v. 10718, p. 107181U.
- BARUFALDI, B. et al. Openvct: a gpu-accelerated virtual clinical trial pipeline for mammography and digital breast tomosynthesis. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2018: Physics of Medical Imaging*. [S.l.], 2018. v. 10573, p. 1057358.
- BENGIO, Y. et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, Now Publishers, Inc., v. 2, n. 1, p. 1–127, 2009.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, v. 5, n. 2, p. 157–166, 1994.
- BERNS, E. A. et al. Digital and screen-film mammography: comparison of image acquisition and interpretation times. *American Journal of Roentgenology*, Am Roentgen Ray Soc, v. 187, n. 1, p. 38–41, 2006.
- BERRY, D. A. et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine*, Mass Medical Soc, v. 353, n. 17, p. 1784–1792, 2005.
- BORGES, L. R. *Dose savings in digital breast tomosynthesis through image processing*. Tese (Doutorado) — Universidade de São Paulo, 2017.
- BORGES, L. R. et al. Pipeline for effective denoising of digital mammography and digital breast tomosynthesis. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2017: Physics of Medical Imaging*. [S.l.], 2017. v. 10132, p. 1013206.
- BORGES, L. R. et al. Simulation of dose reduction in digital breast tomosynthesis. In: SPRINGER. *International Workshop on Digital Mammography*. [S.l.], 2016. p. 343–350.

- BORGES, L. R. et al. Method for simulating dose reduction in digital mammography using the anscombe transformation. *Medical physics*, Wiley Online Library, v. 43, n. 6Part1, p. 2704–2714, 2016.
- BRASIL; SAÚDE, M. da. *Diretrizes para a detecção precoce do câncer de mama no Brasil*. [S.l.]: INCA Rio de Janeiro, 2015.
- BRAY, F. et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, Wiley Online Library, v. 68, n. 6, p. 394–424, 2018.
- BROEDERS, M. et al. The impact of mammographic screening on breast cancer mortality in europe: a review of observational studies. *Journal of medical screening*, SAGE Publications Sage UK: London, England, v. 19, n. 1_suppl, p. 14–25, 2012.
- BURR, I. W. Cumulative frequency functions. *The Annals of mathematical statistics*, JSTOR, v. 13, n. 2, p. 215–232, 1942.
- CAFFE2. <<https://caffe2.ai/>>. Accessed: 2018-12-09.
- CAO, Z. et al. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*, 2016.
- CHAMBOLLE, A.; POCK, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, Springer, v. 40, n. 1, p. 120–145, 2011.
- CHEN, B. et al. Statistical iterative cbct reconstruction based on neural network. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1511–1521, 2018.
- CHEN, H. et al. Learn: Learned experts' assessment-based reconstruction network for sparse-data ct. *IEEE transactions on medical imaging*, IEEE, 2018.
- CHEUNG, L. et al. Image performance of a new amorphous selenium flat panel x-ray detector designed for digital breast tomosynthesis. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2005: Physics of Medical Imaging*. [S.l.], 2005. v. 5745, p. 1282–1291.
- CHO, J. H.; FESSLER, J. A. Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-d x-ray ct. *IEEE transactions on medical imaging*, IEEE, v. 34, n. 2, p. 678–689, 2014.
- COHEN, G.; DIBIANCA, F. A. The use of contrast–detail–dose evaluation of image quality in a computed tomographic scanner. *Journal of computer assisted tomography*, v. 3, n. 2, p. 189–195, 1979.
- DABOV, K. et al. Image denoising by sparse 3-d transform-domain collaborative filtering. *image processing, iee transactions on* 16 (8), pp. 2080-2095. 2007.
- DATASUS. *Sistema de Informações sobre Mortalidade - SIM*. Disponível em: <<http://tabnet.datasus.gov.br/cgi/defthtm.exe?sim/cnv/obt10uf.def>>.
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. [S.l.], 2009. p. 248–255.

- DENG, L.; YU, D. et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014.
- DROMAIN, C.; BALLEYGUIER, C. Contrast-enhanced digital mammography. In: *Digital mammography*. [S.l.]: Springer, 2010. p. 187–198.
- EPSTEIN, C. L. *Introduction to the mathematics of medical imaging*. [S.l.]: Siam, 2008. v. 102.
- FELDKAMP, L. A.; DAVIS, L.; KRESS, J. W. Practical cone-beam algorithm. *Josa a*, Optical Society of America, v. 1, n. 6, p. 612–619, 1984.
- FRIEDEWALD, S. M. et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *Jama*, American Medical Association, v. 311, n. 24, p. 2499–2507, 2014.
- FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, Springer, v. 36, n. 4, p. 193–202, 1980.
- Gatys, L. A.; Ecker, A. S.; Bethge, M. Image style transfer using convolutional neural networks. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 2414–2423.
- GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, v. 9, p. 249–256, 01 2010.
- GLOROT, X.; BORDES, A.; BENGIO, Y. Deep sparse rectifier neural networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011*, v. 15, p. 315–323, 01 2011.
- GU, J.; YE, J. C. Multi-scale wavelet domain residual learning for limited-angle ct reconstruction. *arXiv preprint arXiv:1703.01382*, 2017.
- GUPTA, H. et al. Cnn-based projected gradient descent for consistent ct image reconstruction. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1440–1453, 2018.
- HAHNLOSER, R. H. et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, Nature Publishing Group, v. 405, n. 6789, p. 947, 2000.
- HANSON, K. M. Noise and contrast discrimination in computed tomography. *Radiology of the Skull and Brain*, v. 5, n. 1, p. 3941–3955, 1981.
- HE, K.; SUN, J. Convolutional neural networks at constrained time cost. *CoRR*, abs/1412.1710, 2014. Disponível em: <<http://arxiv.org/abs/1412.1710>>.
- HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>.
- HEIDE, F. et al. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (TOG)*, ACM, v. 33, n. 6, p. 231, 2014.
- HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- HORNIK, K. Approximation capabilities of multilayer feedforward networks. *Neural networks*, Elsevier, v. 4, n. 2, p. 251–257, 1991.

- HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- HUANG, Y. et al. Scale-space anisotropic total variation for limited angle tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*, IEEE, v. 2, n. 4, p. 307–314, 2018.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, Wiley Online Library, v. 160, n. 1, p. 106–154, 1962.
- INCA. *Detecção precoce do câncer de mama*. 2018. Disponível em: <<https://www.inca.gov.br/controle-do-cancer-de-mama/acoes-de-controle/deteccao-precoce#1>>.
- INCA - Instituto Nacional de Câncer José Alencar Gomes da Silva. *Incidência de Câncer no Brasil - Estimativa de 2018*. 2018. Disponível em: <<http://www.inca.gov.br/estimativa/2018/casos-taxas-brasil.asp>>.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- JIA, X. et al. Gpu-based fast cone beam ct reconstruction from undersampled and noisy projection data via total variation. *Medical physics*, Wiley Online Library, v. 37, n. 4, p. 1757–1760, 2010.
- JIA, Y. et al. Caffe: Convolutional architecture for fast feature embedding. In: *ACM. Proceedings of the 22nd ACM international conference on Multimedia*. [S.l.], 2014. p. 675–678.
- JIN, K. H. et al. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, IEEE, v. 26, n. 9, p. 4509–4522, 2017.
- JOHELSON, M. Contrast-enhanced digital mammography. *Radiologic Clinics*, Elsevier, v. 52, n. 3, p. 609–616, 2014.
- JOUPPI, N. P. et al. In-datacenter performance analysis of a tensor processing unit. In: *IEEE. Computer Architecture (ISCA), 2017 ACM/IEEE 44th Annual International Symposium on*. [S.l.], 2017. p. 1–12.
- KAK, A. C.; SLANEY, M. *Principles of computerized tomographic imaging*. [S.l.]: IEEE press New York, 1988.
- KALENDER, W. A. Computed tomography: fundamentals, system technology, image quality, applications. *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications*, by Willi A. Kalender, pp. 220. ISBN 3-89578-081-2. Wiley-VCH, November 2000., p. 220, 2000.
- KC, O. et al. Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society. *JAMA*, v. 314, n. 15, p. 1599–1614, 2015. Disponível em: <<http://dx.doi.org/10.1001/jama.2015.12783>>.
- KIEFER, J.; WOLFOWITZ, J. et al. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 23, n. 3, p. 462–466, 1952.

- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.
- LEWIN, J. M. et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *American Journal of Roentgenology*, Am Roentgen Ray Soc, v. 179, n. 3, p. 671–677, 2002.
- LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: SPRINGER. *European conference on computer vision*. [S.l.], 2014. p. 740–755.
- LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017.
- LU, Y.; ZHAO, J.; WANG, G. Few-view image reconstruction with dual dictionaries. *Physics in Medicine & Biology*, IOP Publishing, v. 57, n. 1, p. 173, 2011.
- MAINPRIZE, J. G. et al. Dependence of image quality on geometric factors in breast tomosynthesis. *Medical physics*, Wiley Online Library, v. 38, n. 6Part1, p. 3090–3103, 2011.
- MAINPRIZE, J. G. et al. Resolution at oblique incidence angles of a flat panel imager for breast tomosynthesis. *Medical physics*, Wiley Online Library, v. 33, n. 9, p. 3159–3164, 2006.
- MAINPRIZE, J. G.; WANG, X.; YAFFE, M. J. The effect of lag on image quality for a digital breast tomosynthesis system. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2009: Physics of Medical Imaging*. [S.l.], 2009. v. 7258, p. 72580R.
- MCCOLLOUGH, C. Tu-fg-207a-04: Overview of the low dose ct grand challenge. *Medical physics*, Wiley Online Library, v. 43, n. 6Part35, p. 3759–3760, 2016.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.
- MIGOWSKI, A. et al. Diretrizes para detecção precoce do câncer de mama no brasil. ii-novas recomendações nacionais, principais evidências e controvérsias. *Cadernos de Saúde Pública*, SciELO Public Health, v. 34, p. e00074817, 2018.
- MUNOZ, D. et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. *JNCI: Journal of the National Cancer Institute*, Oxford University Press, v. 106, n. 11, 2014.
- NICKOLLS, J. et al. Scalable parallel programming with cuda. *Queue*, ACM New York, NY, USA, v. 6, n. 2, p. 40–53, 2008.

- NIKLASON, L. T. et al. Digital tomosynthesis in breast imaging. *Radiology*, v. 205, n. 2, p. 399–406, 1997.
- NIU, S. et al. Sparse-view x-ray ct reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, IOP Publishing, v. 59, n. 12, p. 2997, 2014.
- ORGANIZATION, W. H. *Cancer control: knowledge into action: WHO guide for effective programmes*. [S.l.]: World Health Organization, 2007. v. 2.
- PARK, Y. et al. Compressed-sensing (cs)-based digital breast tomosynthesis (dbt) reconstruction for low-dose, accurate 3d breast x-ray imaging. *Journal of the Korean Physical Society*, Springer, v. 65, n. 4, p. 565–571, 2014.
- PARKER, D. L. Optimal short scan convolution reconstruction for fan beam ct. *Medical physics*, Wiley Online Library, v. 9, n. 2, p. 254–257, 1982.
- PASZKE, A. et al. Automatic differentiation in pytorch. 2017.
- PISANO, E. D. et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, Mass Medical Soc, v. 353, n. 17, p. 1773–1783, 2005.
- PLANTES, B. Z. D. Eine neue methode zur differenzierung in der roentgenographie (planigraphie). *Acta Radiologica*, SAGE Publications Sage UK: London, England, n. 2, p. 182–192, 1932.
- PORTER, P. “westernizing” women’s risks? breast cancer in lower-income countries. *New England Journal of Medicine*, Mass Medical Soc, v. 358, n. 3, p. 213–216, 2008.
- PRASAD, K.; COLE, W.; HAASE, G. Radiation protection in humans: extending the concept of as low as reasonably achievable (alara) from dose to biological damage. *The British journal of radiology*, British Institute of Radiology, v. 77, n. 914, p. 97–99, 2004.
- PUTZKY, P.; WELLING, M. Recurrent inference machines for solving inverse problems. *arXiv preprint arXiv:1706.04008*, 2017.
- RADON, J. Uber die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten. Berichte uber die Verhandlungen Gesellschaft der Wissenschaften zu Leipzig. *Journal of Mathematical Physics*, v. 69, p. 262–277, 1917.
- RAFFERTY, E. A. et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology*, Radiological Society of North America, Inc., v. 266, n. 1, p. 104–113, 2013.
- RAMACHANDRAN, G.; LAKSHMINARAYANAN, A. Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of fourier transforms. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 68, n. 9, p. 2236–2240, 1971.
- RAMACHANDRAN, H. et al. Estimation of mass moment of inertia of human body, when bending forward, for the design of a self-transfer robotic facility. v. 11, p. 166–176, 02 2016.
- RANGAYAN, R. M. *Biomedical image analysis*. [S.l.]: CRC press, 2004.

- REN, B. et al. A new generation ffdm/tomosynthesis fusion system with selenium detector. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2010: Physics of Medical Imaging*. [S.l.], 2010. v. 7622, p. 76220B.
- ROMANO, Y.; ELAD, M.; MILANFAR, P. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, SIAM, v. 10, n. 4, p. 1804–1844, 2017.
- 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.
- 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.
- ROSS, J. et al. *Neural network processor*. [S.l.]: Google Patents, 2017. US Patent 9,747,546.
- ROTH, S.; BLACK, M. J. Fields of experts. *International Journal of Computer Vision*, Springer, v. 82, n. 2, p. 205, 2009.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, Springer, v. 115, n. 3, p. 211–252, 2015.
- SALVADEO, D. H. P. et al. Filtragem de ruído em imagens tomográficas com baixa taxa de contagem utilizando uma abordagem bayesiana contextual. Universidade Federal de São Carlos, 2013.
- SAMARASINGHE, S. *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. [S.l.]: Auerbach publications, 2016.
- SCARPARO, D. C. et al. Evaluation of denoising digital breast tomosynthesis data in both projection and image domains and a study of noise model on digital breast tomosynthesis image domain. *Journal of Medical Imaging*, International Society for Optics and Photonics, v. 6, n. 3, p. 031410, 2019.
- SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural networks*, Elsevier, v. 61, p. 85–117, 2015.
- SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015. Disponível em: <<http://arxiv.org/abs/1503.03832>>.
- SECHOPOULOS, I. A review of breast tomosynthesis. part i. the image acquisition process. *Medical physics*, Wiley Online Library, v. 40, n. 1, 2013.
- SEGARS, W. P. et al. Realistic ct simulation using the 4d xcat phantom. *Medical physics*, Wiley Online Library, v. 35, n. 8, p. 3800–3808, 2008.
- SHEN, C. et al. Intelligent parameter tuning in optimization-based iterative ct reconstruction via deep reinforcement learning. *arXiv preprint arXiv:1711.00414*, 2017.
- SIBIO, A. D. et al. Female breast cancer in central and south america. *Cancer epidemiology*, Elsevier, v. 44, p. S110–S120, 2016.
- SICKLES, E. A. et al. Performance benchmarks for diagnostic mammography. *Radiology*, Radiological Society of North America, v. 235, n. 3, p. 775–790, 2005.

- SIDKY, E. Y.; KAO, C.-M.; PAN, X. Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct. *Journal of X-ray Science and Technology*, IOS Press, v. 14, n. 2, p. 119–139, 2006.
- SIDKY, E. Y.; PAN, X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, IOP Publishing, v. 53, n. 17, p. 4777, 2008.
- SILVA, G. A. et al. Tendência da mortalidade por câncer nas capitais e interior do brasil entre 1980 e 2006. *Revista de Saúde Pública*, SciELO Public Health, v. 45, p. 1009–1018, 2011.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- SMITH, R. A.; COKKINIDES, V.; EYRE, H. J. American cancer society guidelines for the early detection of cancer, 2003. *CA: a cancer journal for clinicians*, Wiley Online Library, v. 53, n. 1, p. 27–43, 2003.
- SMITH, S. W. et al. The scientist and engineer's guide to digital signal processing. California Technical Pub. San Diego, 1997.
- SPRAGUE, B. L. et al. National performance benchmarks for modern diagnostic digital mammography: update from the breast cancer surveillance consortium. *Radiology*, Radiological Society of North America, v. 283, n. 1, p. 59–69, 2017.
- SRIVASTAVA, R. K.; GREFF, K.; SCHMIDHUBER, J. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015.
- SUN, J. et al. Deep admm-net for compressive sensing mri. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2016. p. 10–18.
- 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.
- TABÁR, L. et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*, Radiological Society of North America, Inc., v. 260, n. 3, p. 658–663, 2011.
- ULYANOV, D.; VEDALDI, A.; LEMPITSKY, V. Deep image prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 9446–9454.
- VEDANTHAM, S. et al. Digital breast tomosynthesis: state of the art. *Radiology*, Radiological Society of North America, v. 277, n. 3, p. 663–684, 2015.
- VIEIRA, M. A. et al. Feasibility study of dose reduction in digital breast tomosynthesis using non-local denoising algorithms. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2015: Physics of Medical Imaging*. [S.l.], 2015. v. 9412, p. 94122C.
- VIEIRA, M. A. d. C. et al. Investigating poisson noise filtering in digital breast tomosynthesis. In: UNIVERSIDADE FEDERAL FLUMINENSE (UFF). *IX Workshop de Visão Computacional-WVC, IX*. [S.l.], 2013.

- VIMIEIRO, R. B.; BORGES, L. R.; VIEIRA, M. A. Open-source reconstruction toolbox for digital breast tomosynthesis. In: SPRINGER. *XXVI Brazilian Congress on Biomedical Engineering*. [S.l.], 2019. p. 349–354.
- WANG, G. et al. Image reconstruction is a new frontier of machine learning. *IEEE transactions on medical imaging*, v. 37, n. 6, p. 1289–1296, 2018.
- WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, IEEE, v. 26, n. 1, p. 98–117, 2009.
- WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.
- WERBOS, P. Beyond regression: "new tools for prediction and analysis in the behavioral sciences. *Ph. D. dissertation, Harvard University*, 1974.
- WÜRFL, T. et al. Deep learning computed tomography. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2016. p. 432–440.
- WÜRFL, T. et al. Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1454–1463, 2018.
- XU, Q. et al. Low-dose x-ray ct reconstruction via dictionary learning. *IEEE transactions on medical imaging*, IEEE, v. 31, n. 9, p. 1682–1697, 2012.
- XU, S. *Tomographic image reconstruction: implementation, optimization and comparison in digital breast tomosynthesis*. [S.l.]: Southern Illinois University at Carbondale, 2014.
- YANG, B.; YING, L.; TANG, J. Artificial neural network enhanced bayesian pet image reconstruction. *IEEE transactions on medical imaging*, IEEE, 2018.
- YANG, G. *Numerical Approaches for Solving the Combined Reconstruction and Registration of Digital Breast Tomosynthesis*. Tese (Doutorado) — UCL (University College London), 2012.
- YANG, Q. et al. Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, IEEE, 2018.
- ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: SPRINGER. *European conference on computer vision*. [S.l.], 2014. p. 818–833.
- ZEILER, M. D.; TAYLOR, G. W.; FERGUS, R. Adaptive deconvolutional networks for mid and high level feature learning. In: IEEE. *Computer Vision (ICCV), 2011 IEEE International Conference on*. [S.l.], 2011. p. 2018–2025.
- ZENG, G. L. *Medical image reconstruction: a conceptual tutorial*. [S.l.]: Springer, 2010.
- ZHANG, J. et al. Sparse-view x-ray ct reconstruction with gamma regularization. *Neurocomputing*, Elsevier, v. 230, p. 251–269, 2017.
- ZHANG, K. et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, IEEE, v. 26, n. 7, p. 3142–3155, 2017.

ZHANG, Y. et al. A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis. *Medical physics*, Wiley Online Library, v. 33, n. 10, p. 3781–3795, 2006.

ZHAO, B.; ZHAO, W. Imaging performance of an amorphous selenium digital mammography detector in a breast tomosynthesis system. *Medical physics*, Wiley Online Library, v. 35, n. 5, p. 1978–1987, 2008.

ZHENG, X. et al. Pwls-ultra: An efficient clustering and learning-based approach for low-dose 3d ct image reconstruction. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1498–1510, 2018.