



**RECONSTRUÇÃO DE IMAGENS UTILIZANDO
APRENDIZADO PROFUNDO E
TÉCNICAS DE COMPRESSIVE SENSING**

ALINE ALVES SOARES THOMAZ

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**RECONSTRUÇÃO DE IMAGENS UTILIZANDO
APRENDIZADO PROFUNDO E
TÉCNICAS DE COMPRESSIVE SENSING**

ALINE ALVES SOARES THOMAZ

Orientador: PROF. DR. MYLENE CHRISTINE QUEIROZ DE FARIAS, ENE/UNB

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

**PUBLICAÇÃO PPGENE.DM - 793/2022
BRASÍLIA-DF, 23 DE SETEMBRO DE 2022.**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**RECONSTRUÇÃO DE IMAGENS UTILIZANDO
APRENDIZADO PROFUNDO E
TÉCNICAS DE COMPRESSIVE SENSING**

ALINE ALVES SOARES THOMAZ

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA.

APROVADA POR:

Prof. Dr. Mylene Christine Queiroz de Farias, ENE/UnB
Orientador

Dr. Jonathan Alis Salgado Lima, GAMA/UnB
Co-Orientador

Dr. Pedro Garcia Freitas, Samsung-Brasil
Examinador externo

Prof. Dr. Vinícius de Carvalho Rispoli, GAMA/UnB
Examinador interno

Prof. Dr. Adson Ferreira da Rocha, ENE/UNB
Examinador interno

BRASÍLIA, 23 DE SETEMBRO DE 2022.

FICHA CATALOGRÁFICA

ALINE ALVES SOARES THOMAZ

Reconstrução de imagens utilizando Aprendizado Profundo e técnicas de Compressive Sensing

2022xv, 72p., 201x297 mm

(ENE/FT/UnB, Mestre, Engenharia Elétrica, 2022)

Dissertação de Mestrado - Universidade de Brasília

Faculdade de Tecnologia - Departamento de Engenharia Elétrica

REFERÊNCIA BIBLIOGRÁFICA

ALINE ALVES SOARES THOMAZ (2022) Reconstrução de imagens utilizando Aprendizado Profundo e técnicas de Compressive Sensing. Dissertação de Mestrado em Engenharia Elétrica, Publicação 793/2022, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 72p.

CESSÃO DE DIREITOS

AUTOR: Aline Alves Soares Thomaz

TÍTULO: Reconstrução de imagens utilizando Aprendizado Profundo e técnicas de Compressive Sensing.

GRAU: Mestre ANO: 2022

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor se reserva a outros direitos de publicação e nenhuma parte desta dissertação de Mestrado pode ser reproduzida sem a autorização por escrito do autor.

Aline Alves Soares Thomaz

Brasília, Distrito Federal, Brasil

Agradecimentos

Na realização da presente dissertação, contei com o apoio direto ou indireto de múltiplas pessoas às quais estou profundamente grata. Correndo o risco de injustamente não mencionar algum dos contributos quero deixar expresso os meus agradecimentos:

À orientadora desta dissertação a Doutora Professora Mylene Christine Queiroz de Farias, pela orientação prestada, pelo seu incentivo, disponibilidade e apoio que sempre demonstrou. Aqui lhe exprimo a minha gratidão.

Aos co-orientadores Doutor Jonathan Alis Salgado Lima e Doutor Professor Cristiano Jacques Miosso, pelas suas disponibilidades nos trabalhos, pelo seu incentivo, e igualmente pelo seu apoio na elaboração deste trabalho.

A todos os amigos e colegas que de uma forma direta ou indireta, contribuíram, ou auxiliaram na elaboração do presente estudo, pela paciência, atenção e força que prestaram em momentos menos fáceis. Para não correr o risco de não enumerar algum não vou identificar ninguém, aqueles a quem este agradecimento se dirige sabê-lo-ão, desde já os meus agradecimentos.

Não poderia deixar de agradecer à minha família por todo o apoio, pela força e pelo carinho que sempre me prestaram ao longo de toda a minha vida académica, bem como, à elaboração da presente dissertação a qual sem o seu apoio teria sido impossível.

Por fim, a Deus não venho aqui pedir nada, mas agradecer por tudo. Agradeço por essa vitória, pois reconheço que toda conquista será sempre nossa, nunca apenas minha! Dando graças constantemente a Deus Pai por todas as coisas, em nome de nosso Senhor Jesus Cristo.

A todos o meu sincero e profundo Muito Obrigada!

RESUMO

O uso de imagens de ressonância magnética é amplamente difundido no diagnóstico de tumores e auxilia no tratamento de câncer, uma doença que causa a produção de células anormais que agrupam-se formando tumores e gerando metástases. Um exame de prevenção ou diagnóstico precoce de uma doença como o câncer aumentam consideravelmente suas chances de cura. É sempre importante considerar que a qualidade das imagens obtidas em exames médicos como a ressonância magnética (RMN) é de suma importância para o diagnóstico de doenças. A Engenharia, através de linhas de pesquisa como o processamento de sinais, fornece meios de aumentar a robustez frente a interferências no processamento de imagens com ferramentas como aprendizado profundo. Dentro desta vertente, o uso de redes neurais convolucionais, como a U-Net e os autoencoders, alinhados ao *Compressive Sensing*, tem extrema valia no desenvolvimento de sistemas de diagnósticos mais acurados a doenças tão oprobriosas. Através de testes realizados com códigos baseados em linguagem Python e utilizando imagens de exames médicos provou-se por inspeção visual e por métricas de qualidade como a razão sinal-ruído de pico e o índice de semelhança de estrutura que o uso destas técnicas é capaz de produzir resultados satisfatórios frente a tentativa de eliminar artefatos de imagem e possibilitar um diagnóstico mais contundente, incrementando o potencial de recuperação de um paciente. Para tanto, o modelo proposto conta ainda com técnicas de regularização baseadas nas normas de *Compressive Sensing* e em padrões de subamostragem.

ABSTRACT

The use of magnetic resonance imaging is widespread in the diagnosis of tumors and aids in the treatment of cancer, a disease that causes the production of abnormal cells that group together to form tumors and generate metastases. An examination for prevention or early diagnosis of a disease such as cancer considerably increases your chances of cure. It is always important to consider that the quality of the images obtained in medical examinations such as magnetic resonance imaging (MRI) is of paramount importance for the diagnosis of diseases. Engineering, through research lines such as signal processing, provides ways to increase robustness against interferences in image processing with tools such as deep learning. Within this strand, the use of convolutional neural networks, such as U-Net and autoencoders, aligned to the *Compressive Sensing*, has extreme value in the development of more accurate diagnosis systems to such opprobrious diseases. Through tests performed with codes based on Python language and using medical exam images, it was proven by visual

inspection and quality metrics such as peak signal-to-noise ratio and structure similarity index that the use of these techniques is capable of producing satisfactory results when trying to eliminate image artifacts and enable a more accurate diagnosis, increasing the recovery potential of a patient. To this end, the proposed model also relies on regularization techniques based on the norms of *Compressive Sensing* and on subsampling patterns.

RESUMEN

El uso de la resonancia magnética está muy extendido en el diagnóstico de tumores y ayuda en el tratamiento del cáncer, una enfermedad que provoca la producción de células anormales que se agrupan para formar tumores y generar metástasis. Un examen para la prevención o el diagnóstico precoz de una enfermedad como el cáncer aumenta considerablemente sus posibilidades de curación. Siempre es importante tener en cuenta que la calidad de las imágenes obtenidas en exámenes médicos como la resonancia magnética (RMN) es de suma importancia para el diagnóstico de enfermedades. La ingeniería, a través de líneas de investigación como el procesamiento de señales, proporciona medios para aumentar la robustez frente a las interferencias en el procesamiento de imágenes con herramientas como el aprendizaje profundo. En este contexto, el uso de redes neuronales convolucionales, como la U-Net y los autocodificadores, alineados con *Compressive Sensing*, tiene un valor extremo en el desarrollo de sistemas de diagnóstico más precisos para tales enfermedades oprobiosas. A través de pruebas realizadas con códigos basados en el lenguaje Python y utilizando imágenes de exámenes médicos se demostró mediante inspección visual y mediante métricas de calidad como la relación señal/ruido de pico y el índice de similitud de estructura que el uso de estas técnicas es capaz de producir resultados satisfactorios en un intento de eliminar los artefactos de imagen y permitir un diagnóstico más preciso, aumentando el potencial de recuperación de un paciente. Por lo tanto, el modelo propuesto cuenta aún con técnicas de regularización basadas en las normas de *Compressive Sensing* y en los padrones de subamostragem.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	MOTIVAÇÃO	1
1.2	ANÁLISE DO ESTADO DA ARTE	3
1.3	PROPOSTA DESTA DISSERTAÇÃO	6
2	FUNDAMENTAÇÃO TEÓRICA	7
2.1	RESSONÂNCIA MAGNÉTICA	7
2.2	APRENDIZADO PROFUNDO	10
2.3	COMPRESSIVE SENSING	20
2.3.1	NOÇÕES PRELIMINARES E NOTAÇÃO	21
2.3.2	PROPRIEDADE DE ISOMERIA RESTRITA	23
2.3.3	COERÊNCIA	24
2.4	PARÂMETROS DE AVALIAÇÃO DE DESEMPENHO	26
2.5	ARTEFATOS DE IMAGEM	27
3	METODOLOGIA E IMPLEMENTAÇÃO	32
3.1	ESCOLHA ENTRE A U-NET E O AUTOENCODER	32
3.2	RECONSTRUÇÃO DE IMAGENS UTILIZANDO AUTOENCODER	33
3.3	RECONSTRUÇÃO DE IMAGENS UTILIZANDO A U-NET	35
3.3.1	REGULARIZAÇÃO L1	37
3.3.2	REGULARIZAÇÃO L2	37
3.3.3	REGULARIZAÇÃO TV	38
3.3.4	SUBAMOSTRAGEM	39
3.3.5	CONFIGURAÇÃO BÁSICA DO MODELO PROPOSTO	41
4	RESULTADOS	43
4.1	RESULTADOS DE MODELO SIMPLIFICADO BASEADO EM AUTOENCODER	43
4.2	RESULTADOS DO MODELO BASEADO NA ARQUITETURA U-NET	44
4.2.1	CASO 1 - COMPARAÇÃO DIRETA	45
4.2.2	CASO 2 - COMPARAÇÃO CRUZADA	48
4.3	CONSIDERAÇÕES	54

5	ANÁLISE FINAL	58
5.1	ESTUDOS FUTUROS.....	59
	REFERÊNCIAS BIBLIOGRÁFICAS.....	60

LISTA DE FIGURAS

2.1	Estágio 1 - Alinhamento dos elétrons após incidência do campo magnético. Retirado de [1].	8
2.2	Estágio 2 - Excitação de um elétron após tornar-se instável por conta da absorção de energia liberada pelo campo magnético. Retirado de [1].	8
2.3	Estágio 3 - Detecção de radiofrequência de um tecido humano mostrando as diferenças de intensidade de energia dos átomos. Retirado de [1].	9
2.4	Exemplo de IRM cerebral adquirida durante uma exame de ressonância magnética corriqueira. Retirado do banco de dados BIRN [2].	10
2.5	Anatomia de um Neurônio [3].	12
2.6	Modelo de neurônio em uma rede neural artificial [4].	13
2.7	Exemplos de tumores cancerígenos na pele humana. Retirado de: [5].	15
2.8	Exemplo de arquitetura profunda com duas camadas ocultas, gerando representações intermediárias função de ativação softmax [6].	17
2.9	Modelo básico de um autoencoder. Retirado de: [7].	18
2.10	Modelo básico de uma U-net. Adaptada de [8].	19
2.11	Representação geométrica das bolas l_p para alguns valores de p . Retirado de:[9].	22
2.12	Exemplo de artefato causado pelo movimento do fluxo sanguíneo. Retirado de[10].	28
2.13	Exemplo de artefato causado pelo movimento do paciente. Retirado de [10].	28
2.14	Artefatos causados por material metálico. Retirado de[11].	29
2.15	Exemplo de artefato causados por Aliasing. Retirado de[10].	29
2.16	Artefatos causados por desvio químico. Retirado de[1].	30
2.17	Artefatos causados por interferência de radiofrequência. Retirado de [12].	30
2.18	Exemplo de artefato causados por oscilação (Fenômeno de Gibbs). Retirado de[12].	30
2.19	Exemplos de artefato causado por spikes. Retirado de[12].	31
3.1	Diagrama de blocos do modelo proposto com o autoencoder.	34
3.2	Diagrama de blocos do modelo proposto com a U-Net.[13]	35
3.3	Gráfico de distribuição de dados no plano cartesiano.	36
3.4	Vários padrões de subamostragem: (b) subamostragem cartesiana, (c) subamostragem radial e (d) subamostragem em espiral. Retirada de:[14].	40

3.5	Círculo trigonométrico com alguns valores de N da equação do ângulo dobrado indicados. Adaptada de:[15]	41
3.6	Processo de medição no espaço k e resultados da regularização de imagem. Retirado de: [13].....	42
4.1	Imagem de ressonância magnética cerebral original retirado do banco de dados BIRN.	44
4.2	Imagens corrompidas pelo ruído gaussiano branco com média zero e $\sigma = 0.01, 0.05, \text{ e } 0.1$ (a, d, g), imagens reconstruídas com 10.000 épocas (b, e, h), e as imagens reconstruídas com 20.000 épocas (c, f, i).	45
4.3	Resultado de reconstruções em comparação com a imagem original e com a imagem diferença para o caso direto. (a), (c) e (i) mostram a imagem original. (b), (f) e (j) representam as imagens reconstruídas com os casos diretos de L1, L2 e TV. As imagens (c), (g) e (k) são as imagens diferença para os três casos.	46
4.4	Resultado de reconstruções em comparação com a imagem original e com a imagem diferença para o caso direto. (a), (c) e (i) mostram a imagem original. (b), (f) e (j) representam as imagens reconstruídas com os casos diretos de L1, L2 e TV. As imagens (c), (g) e (k) são as imagens diferença para os três casos.	47
4.5	Resultado de reconstruções para os modelos de reconstrução do caso cruzado em comparação com as imagens de entrada quando o caso cruzado envolve a reconstrução L1-Unet. As imagens (a), (d) e (g) mostram a imagem original. (b), (e) e (h) representam as imagens reconstruídas com os casos cruzados de L1-Unet e (c), (f) e (i) são as imagens diferenças para os três casos.	50
4.6	Resultado de reconstruções para os modelos de reconstrução do caso cruzado em comparação com as imagens de entrada quando o caso cruzado envolve a reconstrução L2-Unet. As imagens (a), (d) e (g) mostram a imagem original. (b), (e) e (h) representam as imagens reconstruídas com os casos cruzados de L2-Unet e (c), (f) e (i) são as imagens diferenças para os três casos.	51
4.7	Resultado de reconstruções para os modelos de reconstrução do caso cruzado em comparação com as imagens de entrada quando o caso cruzado envolve a reconstrução TV-Unet. As imagens (a), (d) e (g) mostram a imagem original. (b), (e) e (h) representam as imagens reconstruídas com os casos cruzados de TV-Unet e (c), (f) e (i) são as imagens diferenças para os três casos.....	52
4.8	Gráficos boxplot para os valores de PSNR (dB) e SSIM para o caso cruzado... ..	54
4.9	Gráfico representando a média da função de perda nos processo de treinamento e validação pelo número de épocas.....	56
4.10	Gráfico da métrica T-SNE com 10 classes de agrupamento por semelhança para os pixels das imagens de tamanho 512x512.	57

LISTA DE TABELAS

4.1	Média PSNR e SSIM das saídas U-Net frente a imagem original para as reconstruções da Figura 4.3.	48
4.2	Média PSNR e SSIM das saídas U-Net frente a imagem original para as reconstruções da Figura 4.4.	48
4.3	Média PSNR e SSIM das saídas U-Net para o caso reto frente todas as imagens originais presentes no banco de dados BIRN.	48
4.4	Média PSNR e SSIM das saídas U-Net para o caso cruzado frente todas as imagens originais do exemplos das figuras 4.5, 4.6 e 4.7.	49
4.5	Média PSNR e SSIM das saídas U-Net para o caso cruzado frente todas as imagens originais presentes no banco de dados BIRN.	52

LISTA DE TERMOS E SIGLAS

ACR American College of Radiology

AE Autoencoder

ANN Redes Neurais Artificiais

CCC Centro de Combate ao Câncer

CNN Redes Neurais Convolucionais

CNR Relação Contraste-Ruído

CS Compressive Sensing

DANT Gerência Doenças e Agravos não Transmissíveis

DFT Transformada Discreta de Fourier

DL Deep Learning

ECG Eletrocardiograma

ECT Tomografia de Capacitância Elétrica

EM Esclerose Múltipla

EUSR European Society Urogenital Radiology

FCN Redes Totalmente Convolucionais

GR Proporção Áurea

IA Inteligência Artificial

INCA Instituto Nacional de Câncer

IRM Imagens de Ressonância Magnética

LASSO Least Absolute Shrinkage and Selection Operator

MAE Erro Quadrático Absoluto

ML Machine Learning

MSE Erro Quadrático Médio

NN Redes Neurais

NSP Propriedade do Espaço Nulo

PI – RADS Prostate Imaging Reporting & Data System

PR Projeção Subamostrada

PSNR Relação Sinal- Ruído de Pico

RIP Propriedade de Isomeria Restrita

RMF Ressonância Magnética Funcional

RMN Ressonância Magnética Nuclear

RMSE Raiz do Erro Quadrático Médio

RNN Redes Neurais Recorrentes

SB Substância Branca

SNR Razão Sinal-Ruído

SSIM Índice de Semelhança de Estrutura

t – SNE Incorporação de Vizinhos Estocásticos t-Distribuídos

TV Variação Total

U – Net Redes Convolucionais para Segmentação de Imagens

Capítulo 1

Introdução

1.1 Motivação

O diagnóstico de doenças geralmente é um processo lento e dependente, muitas vezes, de diversos médicos especialistas (como, por exemplo, um pediatra e um endocrinologista para diagnóstico de diabetes em uma criança), o que pode atrasar o tratamento específico de determinadas enfermidades. No Brasil, o Ministério da Saúde e o Instituto nacional de combate ao câncer (INCA), corroboram que as Doenças e agravos não transmissíveis (DANT) já são as principais responsáveis pelo adoecimento e óbito da população no mundo [16].

Uma das doenças com diagnóstico mais delicado é o câncer. O Brasil estaria registrando 625 mil novos casos de câncer para cada ano do triênio 2020/2022, segundo o INCA. Estima-se que, em 2008, 36 milhões dos óbitos (63%) ocorreram em consequência das DANT, com destaque para as doenças cardiovasculares (48%) e o câncer (21%). A estimativa mundial mostra que, em 2012, ocorreram 14,1 milhões de casos novos de câncer e 8,2 milhões de óbitos. Houve um discreto predomínio do sexo masculino tanto na incidência (53%) quanto na mortalidade (57%). A distribuição da incidência por região geográfica mostra que as regiões Sul e Sudeste concentram 70% da ocorrência de casos novos. Nestas regiões, o padrão da incidência mostra que predominam os cânceres de próstata e de mama feminina, bem como os cânceres de pulmão e de intestino. A região Centro-Oeste, apesar de semelhante, incorpora em seu perfil os cânceres do colo do útero e de estômago entre os mais incidentes. Nas regiões Norte e Nordeste, apesar de também apresentarem os cânceres de próstata e mama feminina entre os principais, a incidência dos cânceres do colo do útero e estômago tem impacto importante nessa população [16].

Atualmente, muitas pessoas tentam realizar o “auto-diagnóstico,” pesquisando os sintomas na Internet e tomando medicamentos por conta própria. Assim, a procura por um médico é feita após as tentativas iniciais de cura não se concretizarem. Ao procurar um médico, o paciente, muitas vezes, não entende ainda a natureza da sua doença e, assim, não procura diretamente um especialista. Para se ter uma ideia, segundo o Centro de combate ao

câncer (CCC), 70% dos diagnósticos de câncer são feitos por médicos não-cancerologistas como, por exemplo, dermatologistas, ginecologistas e ortopedistas. Muitas vezes, este cenário acaba propiciando um diagnóstico tardio, pois não é comum um paciente procurar um oncologista por imaginar que esteja com câncer. O médico chega a uma suposição diagnóstica por meio de várias etapas, durante as quais deve proceder a uma análise cuidadosa, com base principalmente em seu conhecimento do caso e da patologia [17], olhando sempre o paciente como um todo, não se restringindo à sua especialidade.

Na busca de um diagnóstico são utilizados recursos tecnológicos como tomografias computadorizadas, ultrassonografias, ressonâncias magnéticas e exames laboratoriais gerais e específicos, além dos exames físicos (sinais e sintomas do paciente)[18]. Entretanto, em exames de imagens, podem ocorrer a presença de artefatos: tudo e qualquer coisa que ou altera a imagem original, ou prejudica a visualização da imagem, ou acrescenta informações a imagem. Os artefatos podem ser provenientes de movimentos, corpos estranhos, interferências e etc. Os artefatos não são bem vindos nos exames de imagem, em especial nos obtidos através de Ressonância Magnética. Porém alguns artefatos são provenientes de movimentos fisiológicos ou de próteses dos pacientes, sendo dessa forma impossíveis de serem eliminados. Contudo, estes artefatos podem ser compensados diminuindo as suas interferências no resultado final da imagem. Faz-se portanto necessário, para a efetividade do tratamento de câncer, um diagnóstico preciso que aumente as chances de cura. Essa objetividade no diagnóstico passa pela realização de imagens de ressonância magnética com a menor taxa de erro possível.

Dentro do contexto de diagnósticos de doenças afetadas negativamente por artefatos, a técnica de aprendizado profundo (do inglês *Deep Learning* - DL), que é um ramo do aprendizado de máquina (do inglês *Machine Learning* - ML), tem sido amplamente utilizada para o desenvolvimento de modelos de reconstrução de imagens médicas. A grande quantidade de estudos bem-sucedidos que usam Aprendizado Profundo para desenvolver ferramentas diagnósticas sugere sua viabilidade para detecção de artefatos de imagens obtidas por ressonância magnética em exames médicos, reduzindo as dificuldades em alguns diagnósticos [12]. O presente trabalho busca apresentar essa técnica e algumas aplicações para melhoria de exames de imagens através do uso do Aprendizado profundo a fim de se propor e implementar um algoritmo de reconstrução de imagens capaz de gerar em sua saída um resultado aprimorado, auxiliando a detecção de doenças[19], isto é, o objetivo desta dissertação é apresentar um método de reconstrução de imagens de ressonância magnética baseado em aprendizado profundo e detecção de compressão para auxiliar o diagnóstico médico de doenças severas como o câncer cerebral.

1.2 Análise do Estado da Arte

Um dos maiores objetivos da tecnologia é melhorar a qualidade e prolongar a vida humana. Assim, a Engenharia tem buscado aprimorar as técnicas utilizadas pela Medicina para incrementar o diagnóstico de doenças. Uma das vertentes mais utilizadas vem sendo a reconstrução de imagens. A reconstrução de imagens adveio das pesquisas na área de processamento de imagens, que tiveram início em 1917 com o estudo de construção de imagens em projeção de problemas gravitacionais. Já a técnica de reconstrução de imagens propriamente dita foi primeiro desenvolvida por Ronald N. Bracewell (1921-2007), físico e matemático australiano, em 1956, para uso na radioastronomia, com a finalidade de identificar regiões do sol, emissoras de radiação de micro-onda [20]. Somente em 1961 surgiram as primeiras aplicações na Medicina com a reconstrução de imagens de tomografia, quando Godfrey Hounsfield (1919 -2004), engenheiro eletricitista britânico e ganhador do prêmio Nobel de Medicina juntamente com Allan McLeod Cormack (1924 -1998), um físico sul-africano, desenvolveu o primeiro equipamento de tomografia computadorizada, que era capaz de processar imagens cerebrais [21]. Com o passar dos anos, as pesquisas e estudos realizados foram capazes de aprimorar as técnicas de processamento e reconstrução de imagens. Em termos práticos, a reconstrução de imagens pode ser definida como a geração de uma figura bidimensional ou tridimensional a partir de elementos dispersos ou incompletos como, por exemplo, as interpretações de radiação adquiridas no decorrer da análise de imagens médicas.

Um artigo escrito por Yuen *et al.*[22], intitulado *Reconstruction artifacts in digital video compression*, analisou as distorções visuais em vídeos que foram introduzidas por compressão por movimento (MC), modulação de código de pulso diferencial (DPCM) e transformada discreta de cosseno (DCT). A ideia era criar um processo de reconstrução de vídeo para objetos afetados por artefatos de imagem estabelecendo sua correlação com as características espaciais/temporais locais dentro da sequência de imagens que formam o vídeo. Através do uso de técnicas de compressão, as sequências de imagens CCITT H.261 e MPEG foram reconstruídas na saída observando que o nível de gravidade dos artefatos que os tornavam visualmente perceptíveis foi neutralizado [22].

Outro estudo realizado por Peters *et al.* [23], intitulado *Undersampled projection reconstruction applied to MR angiography*, verifica a reconstrução de imagens de angiografia por ressonância magnética (ARM). Os autores citam que na angiografia convencional com Transformada de Fourier (FT) tridimensional (3D), a resolução na direção de codificação de fase é proporcional ao tempo de aquisição. Propondo uma reconstrução de imagens por projeção subamostrada (PR), o estudo revelou ser possível proporcionar alta resolução para as imagens geradas e os artefatos que antes causavam problemas visuais nos exames de angiografia, onde predominam vasos claros e cheios de contraste, foram amortecidos, favorecendo uma maior resolução por unidade de tempo fornecida pela reconstrução de imagens por projeção subamostrada elevando os resultados dos exames em pacientes [23].

Já o estudo realizado por Hyun *et al.*[24], denominado *Deep learning for undersampled MRI reconstruction*, apresenta um método de reconstrução de imagens baseado em técnicas de aprendizado profundo para imagens de ressonância magnética, reduzindo os dados do espaço k com estratégias de subamostragem. A subamostragem é usada para a codificação de fase na captura de características de imagem de alta resolução, enquanto lida com o problema de dobra de imagem ditado pela fórmula Poisson. Para lidar com a incerteza de localização devido ao dobramento da imagem, um pequeno número de dados de espaço k [25] de baixa frequência é adicionado. O treinamento da rede de aprendizado profundo envolve imagens de entrada e saída que são pares das transformadas de Fourier dos dados de espaço k subamostrados. Os experimentos realizados mostraram o desempenho notável: apenas 29 dos dados do espaço k podem gerar imagens de alta qualidade de forma tão eficaz quanto a reconstrução padrão de IRM com os dados totalmente amostrados [24].

Ehrhardt *et al.* desenvolveu um método parecido com Hyun *et al.*[26] no artigo *Multicontrast MRI Reconstruction with Structure-Guided Total Variation*. Muitos estudos de imagens clínicas adquirem dados de ressonância magnética para muitos níveis de contraste, o que aumenta o tempo médio de varredura do exame. Como todas essas imagens mostram a mesma anatomia subjacente, os autores omitiram medidas desnecessárias levando em consideração a semelhança durante a reconstrução. Para tanto, os pesquisadores utilizam a abordagem de normalização por variação total (TV). Foram realizados testes em seis conjuntos de dados baseados em fantasmas e imagens reais de ressonância magnética. Em todos os casos, foi mais proveitoso explorar as informações estruturais das imagens para melhores resultados do que a utilizar a reconstrução separada com variação total em termos de métricas padrão, como relação sinal-ruído de pico e índice de similaridade estrutural. Outro detalhe da abordagem foi que a exploração das informações bidimensionais resultou em imagens com bordas bem definidas, resultando em uma qualidade superior a das imagens reconstruídas apenas com informações a priori sobre a localização das bordas [26].

O artigo *Undersampled radial MRI with multiple coils: Iterative image reconstruction using a total variation constraint*, de autoria de Block *et al.* [27], propõe como uma alternativa para a reconstrução de imagens de ressonância magnética codificadas radialmente. Sabendo que os conjuntos de dados subamostrados têm impacto sobre as imagens, os autores desenvolveram um método para reconstrução de imagens de ressonância magnética baseado em uma otimização não-linear que permite a incorporação de conhecimento prévio e o uso de funções de penalidade. O procedimento usa um mecanismo em duas etapas, que consiste em primeiramente estimar os perfis das bobinas eletromagnéticas e, em seguida, renderizar a imagem final que, de acordo com métricas de qualidade, são semelhantes às observações completas. O conhecimento prévio é introduzido pela penalização das arestas nos perfis das bobinas e por uma restrição de variação total da imagem final. A última condição leva a uma supressão efetiva de artefatos de subamostragem e um certo grau de redução de ruído. Além das simulações, os resultados experimentais para uma sequência de ressonância magnética radial spin-echo foram realizados utilizando dados de fantasmas e imagens reais do cérebro

humano. Em comparação com as reconstruções convencionais como, por exemplo o *regridding*, o método de Block produziu uma qualidade de imagem visualmente melhor em todos os casos [27].

O trabalho *Computed tomography image reconstruction using stacked U-Net* de Misuzawa *et al.* [28] trabalha com o desenvolvimento de métodos de aprendizado profundo, em especial na melhoria da qualidade da imagem usando redes neurais convolucionais, devido a eficácia destas redes na redução de ruído, super-resolução de imagem única e segmentação. Neste estudo, os autores utilizaram uma rede neural conhecida como U-Net na reconstrução de imagens de tomografia computadorizada, gerando imagens de alta qualidade em um curto espaço de tempo e com um pequeno número de projeções. Utilizando várias imagens de tomografia computadorizada, os autores obtiveram resultados mostrando que uma imagem de corte transversal pode ser recuperada com sucesso. Diante dos resultados obtidos, o método de Misuzawa apresenta baixo tempo de reconstrução e boa qualidade de imagem, quando comparado aos métodos existentes [28].

Uma outra técnica muito utilizada no processamento de imagens de ressonância magnética é o *Compressive Sensing - CS*, pois acelera o processo de aquisição de imagens. Assim, as redes neurais profundas e algoritmos de CS estão sendo integrados para redefinir a forma de obtenção de imagens de ressonância com alta qualidade para diagnóstico. O artigo *AI-Based Reconstruction for Fast MRI—A Systematic Review and Meta-Analysis* produzido por Chen *et al.* [29], revisa sistematicamente as técnicas de Compressive Sensing baseadas em aprendizado profundo para ressonância magnética, descrevendo os principais modelos projetados, destacando avanços e discutindo direções promissoras. Os autores também realizaram uma análise abrangente de um sistema de classificação para avaliar o papel fundamental do aprendizado profundo na aceleração baseada em detecção de compressão para imagens de ressonância magnética [29].

Já a proposta de Hammernik *et al.* [30] no artigo *Learning a variational network for reconstruction of accelerated MRI data* permite a reconstrução rápida e de alta qualidade de dados de ressonância magnética através de uma rede variacional que combina a estrutura matemática de modelos variacionais com aprendizado profundo. Para tanto, os autores incorporaram a reconstrução generalizada de imagens médicas como um modelo variacional a um esquema de descida de gradiente. Todos os parâmetros desta formulação, incluindo o modelo anterior definido por filtros e funções de ativação, bem como os pesos dos termos de dados, são aprendidos no treinamento da rede. A abordagem de rede variacional é avaliada em um protocolo de análise clínica da imagem do joelho usando dados subamostrados para diferentes fatores de aceleração e padrões de amostragem. As reconstruções de rede variacional tiveram desempenho superior aos algoritmos de reconstrução padrão, como concluíram métricas de desempenho como a relação sinal-ruído de pico (do inglês *Peak Signal to Noise Ratio - PSNR*) e o índice de semelhança de estrutura (do inglês *Structural Similarity Index Measure - SSIM*) [31]. Os autores mostraram que as reconstruções de rede variacional preservam a aparência natural das imagens de ressonância magnética, ajudando a identificação

de patologias [30].

Um último trabalho realizado é o estudo feito por Zheng *et al.* [32] chamado de *An Autoencoder-Based Image Reconstruction for Electrical Capacitance Tomography*. A pesquisa utiliza uma rede neural supervisionada, conhecida como Autoencoder, para resolver o problema de reconstrução de imagem por Tomografia de Capacitância Elétrica (ECT). Um conjunto de dados baseado em simulação consistindo em 40.000 pares de instâncias, dos quais cada par de amostra tem um vetor de capacitância e um vetor de distribuição de permissividade correspondente, é usado para treinar e testar o desempenho do autocodificador por validação cruzada. Os autores também utilizaram ruído artificial, dados sobre o padrão de fluxo que não estão no conjunto de dados de treinamento e dados experimentais de um sistema ECT prático para testar a capacidade de generalização e prática da rede. Os resultados preliminares mostram que o algoritmo de reconstrução de imagem baseado em autocodificador proposto para ECT está fornecendo melhores resultados de reconstrução destas imagens, segundo os autores [32].

1.3 Proposta desta Dissertação

Nesta dissertação, levando em consideração o estado da arte na área de reconstrução de imagens de ressonância magnética, utilizamos técnicas de *Compressive Sensing* e aprendizado profundo para propor um processo de reconstrução de imagens de ressonância magnética, aumentando a sua robustez imagens a artefatos. Buscamos responder as seguintes perguntas nesta pesquisa:

- É possível reconstruir uma imagem de ressonância magnética sem perder detalhes importantes da fisiologia do corpo humano?
- Utilizando métodos de aprendizado profundo, é possível construir um algoritmo de reconstrução que possa ser implementado em exames médicos corriqueiros?

O principal interesse desta dissertação é responder a estes questionamentos de forma técnica apresentando testes e resultados com imagens reconstruídas. A dissertação é dividida em 6 capítulos, com o primeiro capítulo sendo esta introdução. No capítulo 2 são apresentados os fundamentos teóricos dessa dissertação para um maior entendimento das técnicas utilizadas no trabalho. No capítulo 3 é apresentada a proposta do algoritmo de reconstrução de imagens de ressonância magnética baseado em técnicas de aprendizado profundo e, no capítulo 4, os resultados obtidos. No Capítulo 5 são discutidas as conclusões e propostas de futuros trabalhos.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta os conceitos teóricos fundamentais para o desenvolvimento dessa dissertação e para a compreensão do método proposto.

2.1 Ressonância Magnética

A ressonância magnética nuclear (RMN), ou simplesmente ressonância magnética, é uma técnica de formação de Imagens de Ressonância Magnética (IRM) que usa as características dos núcleos atômicos de alguns elementos químicos para, na presença de um campo magnético, formar padrões de imagem que se assemelham à forma humana [33, 34].

As características fundamentais de ressonância magnética têm início na relação entre um campo magnético exterior e um átomo; Os átomos mais substanciais no organismo humano são: hidrogênio, oxigênio, carbono, fósforo, cálcio, flúor, sódio, potássio e nitrogênio. Apesar de todos estes núcleos possuírem propriedades que permitam as suas utilizações em IRM, o hidrogênio é o mais empregado por dois motivos básicos: é o mais abundoso no corpo humano e o seu próton possui o maior momento magnético proporcionando uma suscetibilidade superior à RMN. Os prótons são partículas carregadas positivamente, que possuem uma propriedade chamada de *spin* ou momento angular, que nada mais é que uma rotação similar à rotação da Terra sob o seu próprio eixo. Outra propriedade do próton é o chamado momento magnético, que permite que o próton se comporte como um ímã [35, 34, 36].

Os equipamentos de RMN possuem componentes que proporcionam o fenômeno da ressonância. O campo magnético é gerado por ímãs e distribuído pelas bobinas de gradiente. Já as bobinas de radiofrequência recebem e transmitem os sinais dos tecidos. Assim que um paciente está posicionado para a execução de um exame de IRM através de um aparelho de ressonância, ele fica sob a ação do campo magnético. Neste instante, os prótons dos átomos de hidrogênio orientam-se em direção à incidência do campo, apontando tanto paralelamente quanto antiparalelamente ao campo após a aplicação de um pulso de radiofrequência, em um estágio conhecido como alinhamento. Caso o pulso de radiofrequência seja enviado com

valor de frequência exatamente igual à frequência de agitação dos prótons de hidrogênio, o tecido exposto ao campo será excitado em um processo denominado excitação. Durante o processo de excitação, os prótons de hidrogênio presentes receberão energia do pulso, perdendo sua estabilidade. Com essa perda, os prótons retornarão um sinal para a bobina. Este retorno de sinal caracteriza o processo de detecção de radiofrequência. Estes três processos – alinhamento, excitação e detecção de radiofrequência – constituem o procedimento para composição de uma IRM. As Figuras 2.1, 2.2 e 2.3, retiradas do texto escrito pelo médico radiologista Dr. Flávio Pereira das Posses intitulado *Ressonância Magnética (RMN): o que é, tipos e como funciona* ilustram as três fases do processo de ressonância magnética [1]. A Figura 2.1 apresenta os elétrons desalinhados em seu estágio inicial e, após a incidência do campo magnético, alinhados de acordo com a orientação do campo. Já a Figura 2.2 ilustra a passagem de um elétron estável para uma condição instável, pulando para uma camada mais distante do núcleo do átomo. Essa excitação é explicada pela absorção de energia, que possibilita que o elétron se afaste do núcleo. Na Figura 2.3 é apresentada um exemplo de imagem de ressonância magnética. A imagem está dividida em quatro partes, representando as diferenças de intensidade da energia dos átomos nos tecidos humanos na fase de detecção de radiofrequência.

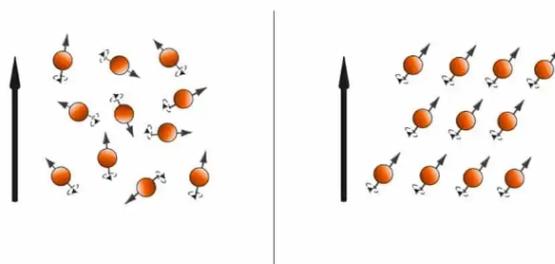


Figura 2.1: Estágio 1 - Alinhamento dos elétrons após incidência do campo magnético. Retirado de [1].

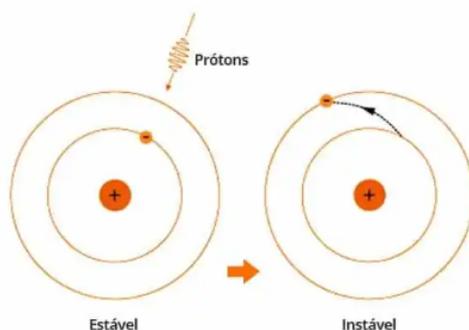


Figura 2.2: Estágio 2 - Excitação de um elétron após tornar-se instável por conta da absorção de energia liberada pelo campo magnético. Retirado de [1].

Entretanto, há um questionamento que poder ser feito: Como é possível distinguir as diferenças dos tecidos humanos durante a realização de um exame? A explicação consiste em variar a direção da intensidade do campo magnético, pois esta variação permite diferen-

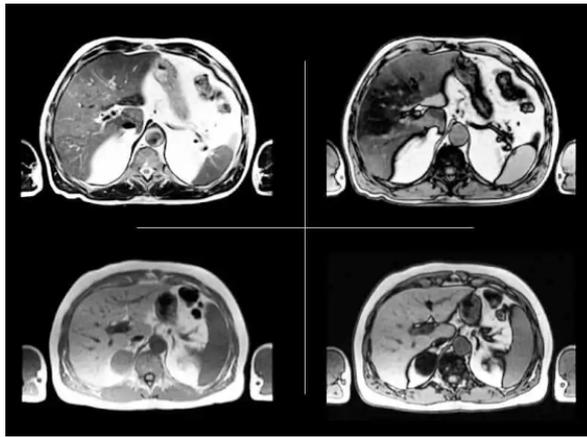


Figura 2.3: Estágio 3 - Detecção de radiofrequência de um tecido humano mostrando as diferenças de intensidade de energia dos átomos. Retirado de [1].

ciar a anatomia. Com a introdução dos chamados gradientes de campo magnético, podemos variar linearmente, em uma dada direção, a intensidade do campo e, conseqüentemente, a excitação dos átomos. Para tanto, é necessário codificar o sinal. São exigidas três etapas para a codificação do sinal de forma a adquirir uma IRM: seleção de corte, codificação de fase e codificação de frequência. Cada parte representa o acionamento de gradientes em uma dada direção. Se o gradiente de seleção de corte for acionado na direção z , cada arranjo ao longo do eixo da mesa irá processar com um valor distinto de frequência dividindo o paciente em cortes axiais. Os outros dois gradientes serão acionados nos eixos que restaram (x e y). Quando o gradiente de codificação de fase é acionado, um dos eixos do corte fica mapeado com a fase. É indispensável demandar n vezes o gradiente de codificação de fase. Cada vez que é acionado, altera-se a amplitude do gradiente. Por fim, no instante da leitura do sinal, o gradiente de codificação de frequência é acionado na direção excedente. Desta forma, o segundo eixo do corte ficará mapeado em frequência [36]. Dependendo da forma e do tempo de excitação dos átomos, as imagens poderão tornar-se mais suscetíveis a diferentes propriedades dos tecidos [33, 34]. Um exemplo citado por Madureira *et al.* [34] e por Arcuri *et al.* [37] diz a respeito à região do cérebro onde, por exemplo, torna-se possível diferenciar os sinais emitidos pela substância cinzenta ou branca e também pelo líquido céfalo-raquidiano, gerando uma imagem de grande precisão espacial [37, 34, 38].

Existem diversos equipamentos de ressonância magnética, que diferem de acordo com a intensidade ou tipo de campo magnético que produzem. Quanto ao campo de magnetismo, os equipamentos podem ser classificados como equipamento de campo aberto, onde o paciente não entra no aparelho, sendo utilizados para pacientes com obesidade, claustrofobia ou que necessitem de exames nas extremidades do corpo, e equipamentos de campo fechado, o tipo mais comum, com abertura de 60 a 70 centímetros para a entrada do paciente. De acordo com a potência do campo magnético, o aparelho pode possuir intensidade de sinal magnético de valores variados como por exemplo:

- inferior a 1 Tesla: sendo estes aparelhos usados para realizar exames de extremidades

corporais, com baixa qualidade de imagens

- 1 Tesla: pouco eficientes e utilizados regularmente em exames veterinários
- 1,5 Tesla: os mais comuns nos locais que oferecem exames de ressonância magnética
- 3 Tesla: muito utilizados em pesquisas embora ofereçam maiores riscos a paciente que possuem componentes como marca-passos.

Há alguns anos, as IRMs restringiam-se a exames de regiões específicas do corpo. Tal restrição devia-se aos longos períodos necessários à realização do exame. No entanto, as IRMs de todo o corpo têm evoluído para testes diagnósticos que envolvem uma variedade de indicações nos últimos anos e, atualmente, com a utilização de bobinas de superfícies especiais, são geradas imagens de corpo inteiro para pesquisar metástases, por exemplo. Nesta dissertação, daremos ênfase às IRMs do cérebro humano, como a Figura 2.4. A geração destas imagens dura aproximadamente 15 minutos. Como aplicações para o uso das IRMs, têm-se [34]: estudos cognitivos e emocionais que utilizam a IRM da pupila para medição de índice de emoção ou atenção; diagnóstico de enfermidades relacionadas a problemas auditivos; exames de ressonância magnética funcional (RMF); eletrocardiografia (ECG) ou oxímetro de pulso; diagnóstico da doença músculo-esqueléticas; diagnóstico de pacientes com esclerose múltipla (EM) com envolvimento da medula espinhal; estudos anatômicos do cérebro *in vivo*; evidenciamento de hiperintensidade de substância branca (SB), reportada em indivíduos com transtornos afetivos, esquizofrenia e doença de Alzheimer; diagnóstico de enfermidades relacionadas ao sistema vascular e diagnóstico de tumores cancerígenos.

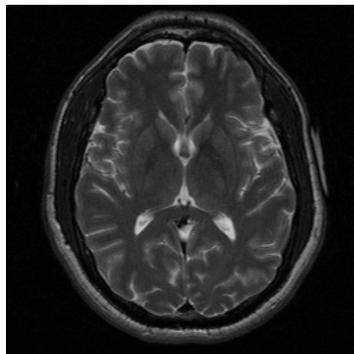


Figura 2.4: Exemplo de IRM cerebral adquirida durante uma exame de ressonância magnética corriqueiro. Retirado do banco de dados BIRN [2].

2.2 Aprendizado Profundo

Uma das maiores capacidades dos seres humanos é o aprendizado. O aprendizado permite que os seres humanos adquiram informações valiosas acerca das situações que vivem e consigam alcançar novos níveis de entendimento do ambiente. Os autores Shai Shalev-Shwartz e Shai Ben-David, no livro *Understanding Machine Learning: From theory*

to algorithms [39], fizeram uma revisão interessante sobre o aprendizado humano em comparação com aprendizado de máquina. Segundo eles, as máquinas teriam um aparato diferente dos seres humanos. Enquanto os seres humanos podem confiar no senso comum para chegar a conclusões, para realizar a mesma tarefa com o aprendizado de máquina devemos fornecer princípios nítidos e bem definidos que evitarão que o programa chegue a conclusões sem sentido ou inúteis [39]. O desenvolvimento de tais princípios é um dos objetivos da teoria do aprendizado de máquina (do inglês *Machine Learning* - ML) que pode ser descrito como uma tecnologia que permite o desenvolvimento de algoritmos e modelos estatísticos para realização de tarefas automaticamente, ou seja sem interferência humana, através da inferência de padrões [19, 40, 41, 42, 43]. Ainda segundo Shalev & Ben-David [39], o aprendizado de máquina alcançou importantes avanços devido a dois aspectos: capacidade de resolver problemas complexos e a capacidade de adaptabilidade a diversos cenários. Já que é uma área de estudo interdisciplinar, possuindo muitas afinidades com a estatística e a teoria da informação, por exemplo. Consequentemente, o aprendizado de máquina se ramificou em vários subcampos que lidam com diferentes tipos de aprendizagem por reconhecimento de padrões [39].

Indo mais além nos conceitos que envolvem o aprendizado de máquina, surge a concepção das redes neurais (do inglês *Neural Networks* - NNs). As redes neurais são estruturas de processamento baseadas nas redes neurais humanas. Existem alguns tipos de redes neurais como as redes neurais artificiais (do inglês *Artificial Neural Networks* - ANNs), que consistem em um grande número de dispositivos básicos de computação (imitando os neurônios) conectados entre si em uma complexa rede de comunicação, tal qual o cérebro é constituído [44, 39]. Por sua vez, as redes neurais recorrentes (do inglês *Recurrent Neural Networks* - RNNs) criam laços durante o processamento fazendo com que não apenas novas entradas sejam utilizadas mais também as saídas das camadas anteriores da rede. Finalmente, as redes neurais convolucionais (do inglês *Convolutional Neural Networks* - CNNs), objeto de estudo desta dissertação, são utilizadas mais frequentemente e compostas basicamente de camadas convolucionais, que processam as entradas considerando campos receptivos locais, conforme definido pelos autores Ian Goodfellow, Yoshua Bengio e Aaron Courville, no livro "*Deep Learning*" [45].

Para entender melhor como as redes neurais artificiais funcionam é importante compreender a estrutura do cérebro. O cérebro humano é constituído por estruturas celulares menores chamadas de neurônios, encontradas aos bilhões. Os neurônios são células capazes de transmitir informações através de impulsos elétricos e reações químicas, podendo, cada neurônio individualmente, formar até 10 mil conexões com outros neurônios [46, 47]. O neurônio é formado por três componentes, conforme a Figura 2.5 [3]:

- Soma ou corpo celular: do qual emanam algumas ramificações denominadas de dendritos;
- Axônio: responsável pela transmissão das mensagens nervosas;

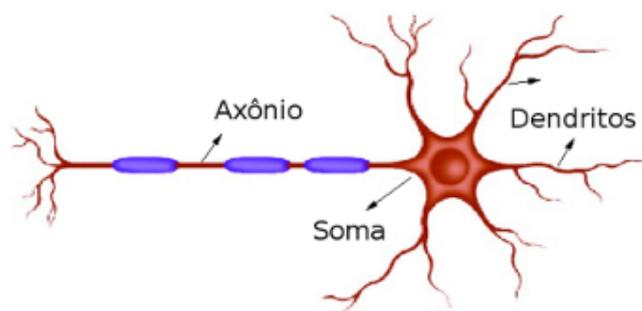


Figura 2.5: Anatomia de um Neurônio [3].

- Nervos terminais: terminações dos neurônios.

A região entre os neurônios por onde ocorre a transmissão de mensagens nervosas é conhecida como sinapse [48]. É através das sinapses que os neurônios se unem em redes neurais [49]. O efeito das sinapses é variável e é esta variação que dá ao neurônio capacidade de adaptação. Todas as experiências, práticas e acontecimentos aos quais o ser humano se submete são processadas e guardadas pelo cérebro, formando a memória. O cérebro está em constante aprendizado.

O modelo de transmissão que ocorre nos neurônios biológicos motivou os estudiosos das áreas de Engenharia e Computação a elaborar um modelo de neurônio matemático, que influenciou o desenvolvimento da inteligência artificial. Se as redes neurais são a base da inteligência humana, por conseguinte, as redes neurais artificiais foram utilizadas como bases da inteligência artificial [50, 51, 52], partindo da premissa de que os dados seriam processados por diversas camadas de processamento não-lineares, tal como ocorre na rede de neurônios[53].

Mais especificamente, o neurônio matemático foi modelado a partir dos processos biológicos formadores e propagadores de impulsos elétricos pela membrana celular dos neurônios. Esse modelo de neurônio funciona da seguinte maneira: ele recebe os sinais de entrada e devolve um único sinal de saída, podendo este sinal ser propagado como saída do sistema ou como entrada para demais neurônios das camadas subsequentes da rede neural artificial. Os dendritos e axônios não possuem uma representação matemática no modelo de neurônio artificial e são representados matematicamente pelas operações das sinapses. A intensidade da ligação é representada por uma grandeza denominada peso sináptico. A Figura 2.6 mostra o modelo de neurônio artificial extraído da dissertação de mestrado de Saulo Favoretto intitulada *Máquinas de aprendizado extremo aplicadas à identificação de pessoas através de eletrocardiograma (ECG)* [4]. Segundo Favoretto, um conjunto de entradas x_1, \dots, x_n representam os dendritos e uma única saída y representa o terminal do axônio. As sinapses nesse contexto são os pesos sinápticos w_1, \dots, w_n , que ponderam cada uma das entradas. A junção somadora soma todos os sinais ponderados e o *viés*, produzindo um potencial de ativação u . Finalmente, uma função de ativação $g(u)$ restringe a amplitude do sinal de saída e adiciona a não-linearidade ao sistema.

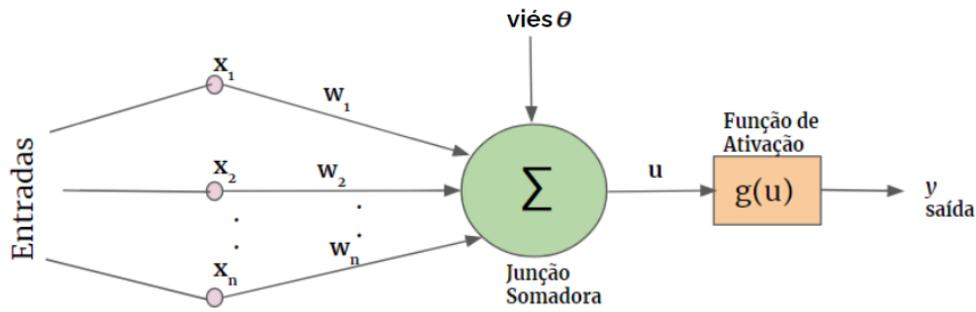


Figura 2.6: Modelo de neurônio em uma rede neural artificial [4].

Em suma, como exposto por Favoretto e ilustrado na Figura 2.6, temos as seguintes variáveis e funções no modelo rede neural artificial:

- Sinais de entrada $\{x_1, x_2, \dots, x_n\}$: São os sinais externos, geralmente normalizados, que são as entradas do modelo e usados, alimentando o modelo preditivo.
- Pesos sinápticos $\{w_1, w_2, \dots, w_n\}$: São valores usados para ponderar os sinais de cada entrada da rede. Esses valores são aprendidos durante o treinamento.
- Combinador linear Σ : Agrega todos sinais de entrada, ponderados pelos respectivos pesos sinápticos, para produzir um potencial de ativação.
- Limiar de ativação Θ : Especifica qual será o patamar apropriado para que o resultado produzido pelo combinador linear gere um valor de disparo de ativação.
- Potencial de ativação u : É o resultado obtido pela diferença do valor produzido pelo combinador linear e o limiar de ativação. Se o valor for positivo ($u \geq 0$), então o neurônio produz um potencial excitatório. Caso contrário, o potencial será inibitório.
- Função de ativação g : Seu objetivo é limitar a saída de um neurônio em um intervalo de valores.
- Sinal de saída y : É o valor de saída podendo ser usado como entrada de outros neurônios que estão sequencialmente interligados [4].

Os neurônios artificiais, como recém explicado constituem todas as redes neurais, incluindo, logicamente, as CNNs, a quem daremos mais ênfase nesta dissertação. As CNNs possuem um objetivo principal: filtrar linhas, curvas e bordas das imagens de entrada e, em cada camada, transformar essa filtragem em uma imagem mais complexa garantindo o reconhecimento visual. Quando falamos em reconstrução de imagens, as entradas são usualmente imagens com altura, largura e profundidade, determinadas pela quantidade de canais de cores. Em geral as imagens utilizam três canais, RGB, com os valores de cada pixel. Para a reconstrução é necessário processar as imagens e, para isso, as CNNs contam com

operações específicas. A primeira delas é a convolução. A convolução funciona como um filtro que enxerga pequenos quadrados e percorre por toda a imagem captando os traços mais marcantes. Exemplificando para facilitar o entendimento, considere uma imagem $32 \times 32 \times 3$ e um filtro que cobre uma área de 5×5 da imagem com movimento de 2 saltos (chamado de *stride*). O filtro passará pela imagem inteira, por cada um dos canais, formando no final um mapa de ativação de $28 \times 28 \times 1$. A profundidade da saída de uma convolução é igual a quantidade de filtros aplicados, ou seja, quanto mais profundas são as camadas das convoluções, mais detalhados são os traços identificados. O filtro, que também é conhecido por *kernel*, é formado por pesos inicializados aleatoriamente, atualizando-os a cada nova entrada durante o processo de *backpropagation*. A pequena região da entrada onde o filtro é aplicado é chamada de *receptive field*. Além do tamanho do filtro e o *stride* da convolução como hiperparâmetros, o engenheiro que está modelando uma CNN deve escolher como será o *padding*, que tem por objetivo evitar que as camadas não diminuam antes que a rede complete o aprendizado. O *padding* pode ser nulo, no qual a saída da convolução ficará no seu tamanho original, ou zero *pad*, onde uma borda é adicionada e preenchida com zeros.

Além dos parâmetros citados, existem ainda as funções de ativação. Elas servem para trazer a não-linearidade ao sistema, para que a rede consiga aprender diversas funcionalidades. Há muitas funções, como sigmoid, tanh e softmax, entretanto a mais indicada para uso com as CNNs é a ReLu. A ReLu zera todos os valores negativos da saída da camada anterior, sendo eficiente computacionalmente sem grandes diferenças de acurácia quando comparada a outras funções. A função *pooling* serve para simplificar a informação da camada anterior. Por exemplo, se a saída da camada anterior for 24×24 , a saída do *pooling* será 12×12 . Além disso, é preciso escolher como será feita a sumarização. O método mais utilizado é o *maxpooling*, no qual apenas o maior número da unidade é passado para a saída. Essa sumarização de dados serve para diminuir a quantidade de pesos a serem aprendidos e também para evitar *overfitting*. Ao final da rede é colocada uma camada *Fully connected*, onde sua entrada é a saída da camada anterior e sua saída são N neurônios, com N sendo a quantidade de classes do seu modelo para finalizar a reconstrução.

Um exemplo da aplicação das técnicas apresentadas até aqui é a detecção de câncer. O câncer, também chamado de tumor maligno, consiste na rápida criação de células anormais formadas a partir do resultado da interação entre os fatores genéticos de uma pessoa e uma ou mais categorias de agentes externos, incluindo:

- Cancerígenos físicos, tais como radiação ultravioleta e ionizante;
- Substâncias químicas cancerígenas, como o amianto, componentes do fumo do tabaco, aflatoxina (um contaminante alimentar) e arsênio (um contaminante da água potável);
- Cancerígenos biológicos, tais como infecções por certos vírus, bactérias ou parasitas;
- O envelhecimento é outro fator fundamental para o desenvolvimento do câncer devido a uma acumulação de riscos para cânceres específicos.

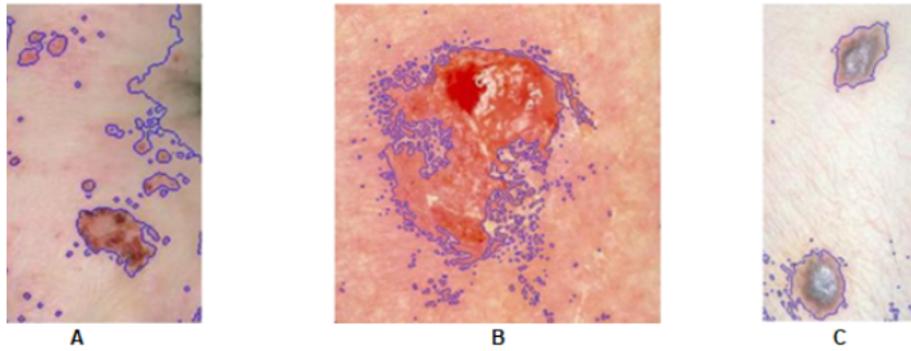


Figura 2.7: Exemplos de tumores cancerígenos na pele humana. Retirado de: [5].

Estas células anormais crescem e invadem diversas áreas do corpo podendo ser encontradas em vários locais da fisiologia, em um processo chamado de metástase [54, 55]. O diagnóstico de câncer é realizado através de exames de imagens como as IRMs. Considere um exemplo: o câncer de pele [40]. Em seu diagnóstico, notamos o que são denominados melanomas e não-melanomas. Embora muito menos comuns, os melanomas são responsáveis pelo maior índice de mortalidade pelo câncer de pele. Na Figura 2.7 são apresentados exemplos de imagens geradas por algoritmos que utilizaram CNNs para categorização destes tipos de tumores tegumentares. Em A é ilustrado o carcinoma espinocelular, em B o carcinoma basocelular (A e B são os dois tipos mais comuns de câncer de pele não-melanoma) e em C o melanoma maligno (o tipo mais letal de câncer de pele). As mudanças na pigmentação encontradas na pele do paciente na Figura 2.7 são as características mais relevantes para o diagnóstico do câncer de pele uma vez que permitem distinguir os tipos de tumor e proporcionar ao paciente o início de tratamento mais adequado à criticidade do câncer. A descoberta de doenças em estágio inicial pode aumentar consideravelmente as chances de cura de um paciente e está diretamente ligada à realização de exames de rotina e acompanhamento da saúde. Em casos de câncer, os exames podem ser clínicos (onde o médico palpa e sente uma lesão), laboratoriais (exame de sangue) ou por imagem, como os exames de RMN. Porém, muitas vezes devido ao modo como a imagem é registrada, o resultado pode induzir ao erro. Um algoritmo de detecção otimizado por aprendizado profundo (em inglês *Deep Learning* - DL) [56] pode melhorar o aspecto da imagem para reduzir este erro[57].

O aprendizado profundo busca descobrir um padrão em um determinado conjunto de dados de forma a guiar o aprendizado do modelo a partir destes dados. Ao final do processamento tem-se uma função capaz de receber dados brutos e fornecer uma representação adequada para um problema em questão [6]. O aprendizado profundo pode ser aplicado a diferentes paradigmas de aprendizado, segundo LeCun *et al.* [58], incluindo aprendizado supervisionado e aprendizado auto supervisionado. O aprendizado profundo é considerado supervisionado em situações como classificação de imagens ou detecção de objetos, pois a rede é usada para prever um rótulo ou um número (a entrada e a saída são conhecidas). Como os rótulos das imagens são conhecidos, a rede é utilizada para reduzir a taxa de erro, por isso é “supervisionada”. Porém, quando a técnica de aprendizado profundo foca nas redes neu-

rais para compactação e reconstrução de imagens por meio de uma representação de espaço latente de dados compactados, o aprendizado profundo é considerado auto-supervisionado pois o sistema aprende a prever parte da sua entrada a partir de outras partes da mesma entrada[59].

O aprendizado profundo dá às máquinas a capacidade de aprender conforme são introduzidas a novas informações sem a necessidade de intervenção humana. Esta técnica vem sendo usada em diversas aplicações, tais como reconhecimento facial, detecção de fraudes, sistemas de recomendação e diagnóstico por imagens, que é o assunto desta dissertação [60, 61]. Todavia, o grande segredo do aprendizado profundo está, justamente, na sua “profundidade”. Por exemplo, considere que se queira projetar uma função que seja capaz de diferenciar imagens que contém cães, gatos, tartarugas e corujas. Pode-se concluir que, matematicamente, o objetivo é aprender uma função $f(x) = y$, em que x representa a imagem e y a classificação mais plausível de x .

Segundo Ponti *et al.* [6], o que diferencia o aprendizado profundo do aprendizado de máquina é, propriamente, a função $f(x)$. Em métodos de aprendizado “superficiais” ou “rasos” a busca é por uma única função que possa, a partir de um conjunto de parâmetros, gerar o resultado desejado. Já no aprendizado profundo, os parâmetros de aprendizado são adquiridos por meio de funções compostas $f(x) = f_L(\dots f_2(f_1(x_1))\dots)$, onde cada função f_L toma como entrada um vetor de dados x_L gerando como saída o próximo vetor x_{l+1} . Sendo assim, no aprendizado profundo, o intuito é aprender sucessivas representações dos dados intermediários. Assim, a profundidade permite aprender uma sequência de funções que transformam vetores mapeando-os de um espaço a outro, até atingir o resultado desejado, pois cada função opera sobre uma entrada gerando uma representação que é então passada para a próxima função[6]. A Figura 2.8 mostra uma exemplo de arquitetura profunda com duas camadas ocultas, gerando representações intermediárias extraído do artigo *Como funciona o Deep Learning*, dos autores Moacir A. Ponti e Gabriel B. Paranhos da Costa.

Um exemplo de arquitetura profunda de redes neurais são os chamados *Autoencoders* ou Autoencodificadores (AE) [49]. Os autoencoders fornecem um dos paradigmas fundamentais para a aprendizagem não-supervisionada [62], nos quais não existem resultados pré-definidos para o modelo utilizar como referência para aprender. Nos autoencoders, as modificações sinápticas são induzidas por ocorrências locais que podem ser coordenadas de forma auto-organizada para que haja uma aprendizagem global [63]. Os autoencoders são redes neurais que visam copiar suas entradas para suas saídas, como pode ser visto no seu modelo básico apresentado na Figura 2.9 [7]. Eles trabalham compactando a entrada em uma representação de espaço latente e, em seguida, reconstruindo a saída dessa representação [64, 65]. Dessa forma, este tipo de rede é composto de duas partes:

- Codificador (*Encoder*): É o segmento da rede que condensa a entrada em uma exposição de espaço latente (codificando a entrada). Pode ser representado por uma função de codificação $H = f(x)$.

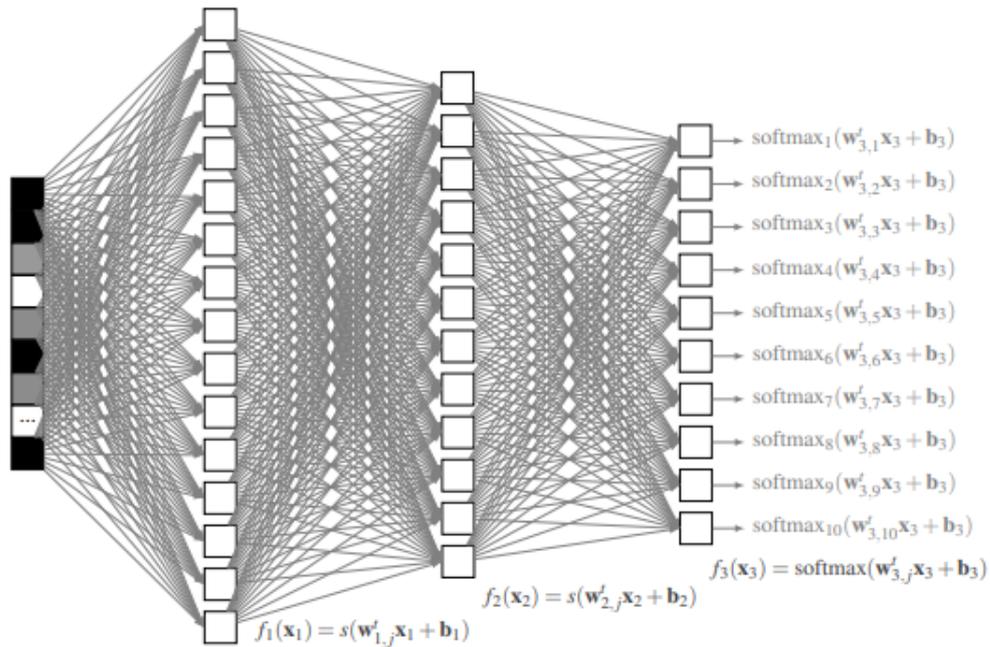


Figura 2.8: Exemplo de arquitetura profunda com duas camadas ocultas, gerando representações intermediárias função de ativação softmax [6]

- Decodificador(*Decoder*): Este segmento tem como objetivo reconstruir a entrada da exposição do espaço latente. Pode ser representado por uma função de decodificação $R = g(h)$ [32].

No autoencoder, os dados de entrada são representados por $\{x_1, x_2, \dots, x_m\}$, os dados de saída são representados por $\{y_1, y_2, \dots, y_m\}$, e ambos os dados podem possuir as mesmas dimensões n e m . Sendo que, n representa o número de observações da base de dados e m representa o número da variável. Assim, tem-se que:

$$H = f(x) = S(WX + b) \quad (2.1)$$

e

$$R = g(H) = S(W'Z + b') \quad (2.2)$$

nas quais W e W' representam a matriz de pesos, b o vetor de viés e S representa uma função de ativação.

Os autoencoders possuem as seguintes propriedades:

- São específicos para um conjunto de dados, o que significa que eles só poderão compactar dados semelhantes aos que foram treinados;
- Geram perdas, o que significa que as saídas descomprimidas serão degradadas em comparação com as entradas originais;

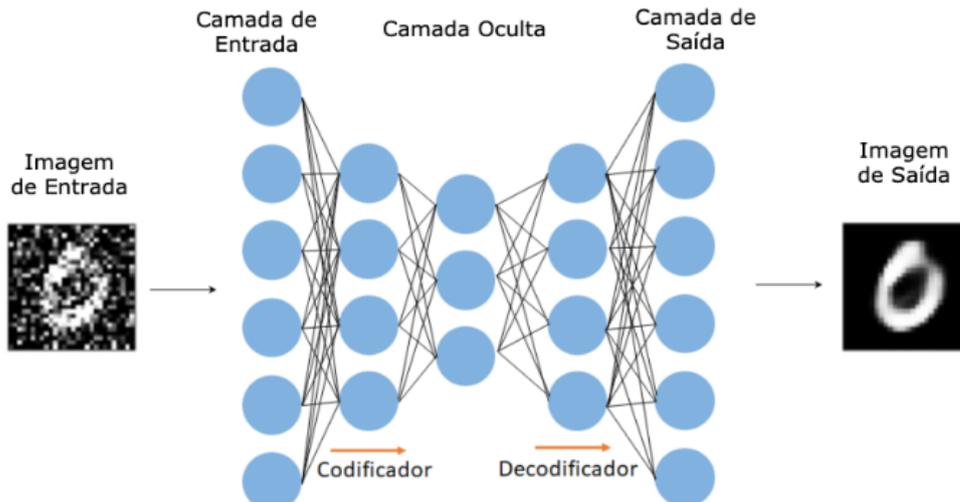


Figura 2.9: Modelo básico de um autoencoder. Retirado de: [7].

- Aprendem automaticamente a partir de exemplos de dados.

Outra arquitetura de redes neurais é a U-Net. Segundo Ronneberger *et al.* [66], existe um amplo consenso de que o treinamento bem-sucedido de redes neurais requer muitas amostras para treinamento. No artigo *U-Net: Convolutional Networks for Biomedical Image Segmentation*, os autores apresentam uma estratégia de rede e treinamento que se baseia no forte uso de maiores quantidades de dados para usar as amostras com mais eficiência, a nomeada U-Net. A arquitetura consiste em duas estruturas: a via de contração – para capturar o contexto em que a rede está inserida, seguindo a arquitetura típica de uma CNN – e um caminho de expansão simétrico – que permite a localização precisa dos dados. Um diagrama na Figura 2.10 ilustra como a U-Net atua [8].

Seu funcionamento ocorre da seguinte maneira [67]. No início da via de contração, a imagem de entrada passa por:

- 2 convoluções com kernel de tamanho 3x3 com stride 1;
- Camada ReLU como função de ativação, gerando 8 mapas de características (usa-se 8 kernels diferentes para se produzir os mapas de características citados); e
- Uma operação de max pooling 2x2 com stride 2, que é um processo de discretização baseado em amostras com objetivo de reduzir a amostragem de uma representação de entrada e reduzir sua dimensionalidade.

Após cada operação de max pooling, a quantidade de mapa de características é aumentada por um fator de dois e o tamanho da entrada é reduzido pelo mesmo fator devido aos efeitos do max pooling [68]. Na via de contração, uma etapa é definida por duas convoluções e uma operação de max pooling. Sendo assim, após 4 etapas, a saída resultante do processo na via

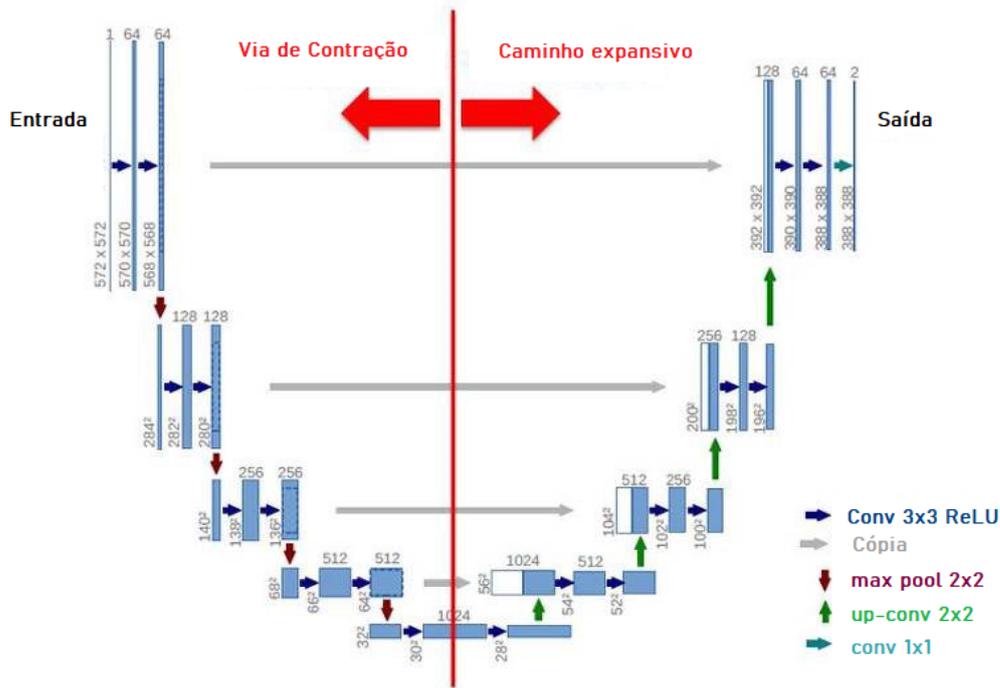


Figura 2.10: Modelo básico de uma U-net. Adaptada de [8].

de contração é passada como entrada para uma transposição convolucional (up-convolution, ou “convolução ascendente” [69]) com kernel 2x2, stride 2 e 64 detectores de características no primeiro passo. Essa transposição convolucional dá início a etapa do caminho de expansão. Cada etapa desta fase consiste na transposição deconvolucional com os parâmetros descritos acima, a concatenação com sua parte correspondente da parte de contração e convoluções assim como as aplicadas na via de contração. Após cada etapa, o número de canais de recursos é reduzido por um fator de dois. Na camada final, uma convolução 1x1 é usada para mapear cada vetor de 64 componentes para o número desejado de classes. No total, a rede possui 23 camadas convolucionais[67].

Embora possa aparentar que a U-Net é um autoencoder especializado, alguns autores como Spolti *et al.* [70] consideram que a U-Net se difere do autoencoder devido a chamadas conexões de salto, que conectam cada estágio da via de contração ao estágio correspondente do caminho de expansão. Essas conexões de salto tendem a simplificar ou dividir o treinamento. Por exemplo, durante as iterações iniciais de treinamento, se a rede não tiver conexões de salto, ela deverá ajustar todos os pesos da rede para gerar qualquer resultado significativo. Esse ajuste é feito pelo autoencoder mas não pela U-Net. Tanto os autoencoders quanto a U-net possuem diversas aplicações. No caso dos autoencoders, é promissora aplicação para remoção de ruídos e reconhecimento de imagens. Já a U-net tem grande potencial na segmentação e reconstrução de imagens biomédicas.

2.3 Compressive Sensing

O aumento da tecnologia permitiu a humanidade obter cada vez mais dados em um curto espaço de tempo. Deduzir a estrutura de um sistema a partir de seus dados é uma tarefa fundamental na Engenharia. Uma das técnicas mais relevantes para o processamento digital de sinais é a amostragem, que permite a conversão de dados analógicos (uma função contínua no tempo ou espaço) em sequências de bits (uma uma função discreta no tempo ou espaço) no domínio digital. Em 1948, Claude Elwood Shannon, baseado nas ideias de Harry Nyquist, formulou o chamado teorema de Shannon-Nyquist ou teorema da amostragem. Segundo o teorema, para se evitar a perda da informação na conversão digital um sinal analógico, deve-se ter uma taxa de amostragem de pelo menos duas vezes o valor da largura de banda do sinal original. Em termos matemáticos:

$$f_s = \frac{1}{T_s} > 2B \quad (2.3)$$

onde B é a largura de banda do sinal original. Cabe ressaltar que o valor mínimo dado pela relação $f_s = 2B$ é chamado de taxa de Nyquist, que é a mínima taxa de amostragem necessária para reconstruir um sinal. Outro passo frequentemente necessário para trabalhar dados digitais é compressão. A compressão pode ser definida como uma codificação de informações usando menos bits do que o representação original. Ambas as etapas são necessárias para armazenar, manipular, processar, transmitir e interpretar dados. Embora as técnicas de amostragem e compressão sejam notoriamente qualificadas para processar dados, a possível esparsidade dos dados não é muitas vezes levada em consideração [71, 72].

Na década de 2010 foi proposta a técnica de *Compressive Sensing* - CS. Segundo Cláudio Mayrink Verdun, CS expressa que a aquisição e a compactação de dados podem ser realizadas simultaneamente de forma que os algoritmos de reconstrução explorem a estrutura dos dados [71, 73]. A CS baseia-se na observação empírica de que muitos tipos de sinais e imagens do mundo real têm uma expansão esparsa em termos de uma base adequada. Isto significa que a expansão tem apenas um pequeno número de termos significativos, ou seja, que o vetor do coeficiente pode ser bem aproximado com um pequeno número de entradas [74, 75]. A CS apoia-se na reconstrução de sinais esparsos, utilizando um pequeno número de medidas (menor do que as necessárias segundo o teorema da amostragem), produzindo um problema de reconstrução de imagens. Uma das primeiras aplicações foi na área de ressonância magnética [76].

A CS, embora recente no meio tecnológico, já possui muitas aplicações graças a uma das suas principais características: a utilização de poucas medidas não adaptativas para reconstruir um sinal arbitrário. Como exemplos de aplicações de CS podemos citar a reconstrução a partir da compressão e aquisição de imagens médicas, a análise de dados geofísicos, o projeto de códigos de correções de erro em comunicações, entre outros ¹. Em especial, a

¹Uma lista com mais aplicações pode ser encontrada no trabalho *Compressive sensing in medical imaging*,

aplicação de CS em imagens médicas geradas por ressonância magnética contribui muito para o seu processo de reconstrução pois estas imagens são esparsas. No cenário de CS, uma representação dos sinais pode beneficiar a pesquisa por algoritmos de otimização a depender da maneira como um sinal é caracterizado. Nesta seção explicamos os conceitos técnicos e matemáticos do CS, introduzindo o conceito de esparsidade e a recuperação de vetores de medições lineares incompletas e não adaptativas.

2.3.1 Noções Preliminares e Notação

Uma matriz $M \in \mathbb{K}^{m \times n}$ pode ser usada para explicar a relação de linearidade entre dois vetores m e n . Aqui, a letra \mathbb{K} é usada para representar os números que podem ser complexos ou reais. Se tivermos um vetor $x \in \mathbb{K}^n$ e o submetermos a matriz M teremos um vetor de saída $b \in \mathbb{K}^m$, sendo b uma combinação linear das n colunas de M . Assim, a matriz M pode fornecer um mapeamento dos conjuntos \mathbb{K}^m e \mathbb{K}^n . Outro detalhe é que o produto $Mx = b$ é explicado como um sistema de equações lineares que estabelece o mapeamento dos conjuntos. As soluções possíveis para o sistema são dadas ao se estudar a dependência linear das linhas e colunas de M . Considere um sistema linear modelado pela matriz M que relaciona o sinal x em acordo com o teorema de amostragem com um conjunto de medidas $b = Mx$. Sabendo que os sistemas mapeiam os sinais originais buscando a menor dimensão, é possível supor que existam redundâncias no sinal e tais redundâncias podem ser exploradas na compressão. Quando um sistema é subdeterminado, no entanto, as soluções são infinitas. Para isso, a CS surge para restringir o domínio do mapeamento linear procurando a solução que mais esparsa possível [78].

Para reconstruir um sinal usando CS é necessário certo conhecimento de álgebra linear. Segundo esta área de estudo, a reconstrução de um sinal é possível através do cálculo do vetor mais esparsa a partir de um vetor de menor dimensão dado que o mapa linear entre eles seja conhecido. Um vetor $x \in \mathbb{K}^n$ é chamado de k -esparso se, ao menos, k valores de entrada são não nulos. O conjunto $\Omega_x = \{i \mid x[i] \neq 0\}$ é chamado de conjunto suporte de x . Considere, a partir deste ponto da explicação, que a métrica para um espaço vetorial E , conhecida como $\|\cdot\|$, é uma função que associa um valor real não negativo a um vetor $x \in E$. Para calcular a esparsidade de um sinal, recorreremos a métricas l_p pois elas satisfazem algumas condições tais:

1. Se $\|x\| = 0$, então x é um vetor nulo de E ;
2. Homogeneidade: Sabendo que $\alpha \in \mathbb{R}$, temos que $\|\alpha x\| = |\alpha| \times \|x\|$;
3. Desigualdade triangular: $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$.

Definição 2.1 - Métrica l_p : Seja $x = x[k]$ um vetor no domínio \mathbb{R}^N e $p \in \mathbb{R}$. A métrica l_p calcula do comprimento de x de acordo com:

Applied Optics dos autores C. G. Graff e E. Y. Sidky [77].

$$\|x\|_p = \left(\sum_{k=0}^{N-1} |x[k]|^p \right)^{\frac{1}{p}} \quad (2.4)$$

Supondo $p = 0$, a métrica l_0 é dada por:

$$\|x\|_0 = \lim_{p \rightarrow 0} \left(\sum_{k=0}^{N-1} |x[k]|^p \right) \quad (2.5)$$

Se $p \leq 1$, então l_p também será uma norma. A norma l_0 se fundamenta na contagem de entradas diferentes de 0 no vetor x . Assim, quanto mais esparsos for o sinal, menor será sua norma l_0 .

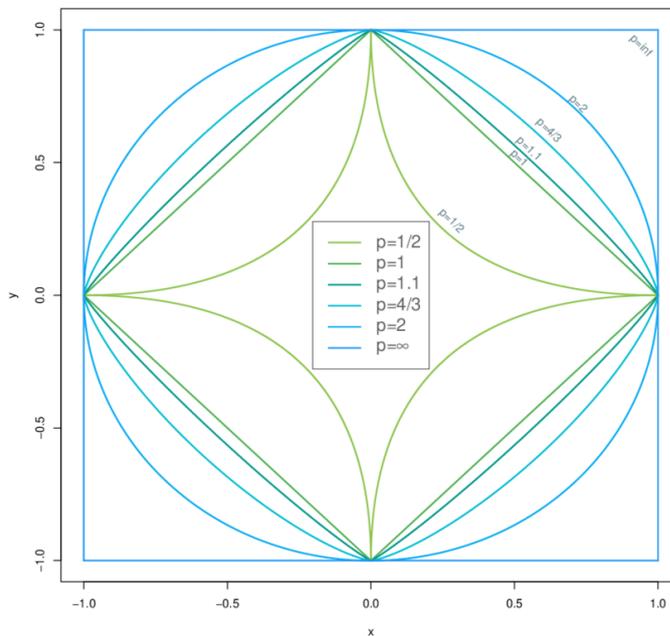


Figura 2.11: Representação geométrica das bolas l_p para alguns valores de p . Retirado de:[9].

Podemos ver na Figura 2.11 uma representação geométrica das bolas l_p no domínio \mathbb{R}^N para alguns valores de p . Ao observar essas bolas podemos entender de forma ilustrada como os algoritmos de CS fazem seus cálculos. Isso é possível pois a matriz M define as soluções do sistema linear $Mx = b$. A reconstrução buscará entre todas as soluções aquela com a menor métrica l_p . Note que à medida que p se aproxima de zero, a bola, conseqüentemente, se aproxima dos eixos. Isso possibilita que as normas l_p , com $0 \leq l_p \leq 1$ encontrem a solução mais esparsa para os conjuntos de soluções de um sistema linear indeterminado.

Como já explicado, a CS trata sistemas onde a relação entre o sinal $x \in \mathbb{K}^n$ e suas medidas $b \in \mathbb{K}^m$, que pode ser modelado como um sistema linear subamostrado $Mx = b$, onde M é a matriz de medidas. Contudo, uma restrição é necessária para garantir que o algoritmo de reconstrução possa calcular uma única solução apropriada das medições b . Quando fa-

lamos de minimização, definir uma restrição para calcular a solução de um sistema linear subdeterminado significa escolher uma função de custo de forma que a solução que possui o menor valor de custo seja a escolhida entre as infinitas soluções. Na CS, essa minimização é forçada para esparsidade. A solução comprimida para um sistema subdeterminado é o sinal x que satisfaz a relação linear $Mx = b$, cujas entradas são nulas. Sendo assim, quanto mais esparsa for o sinal, menos coeficientes são necessários. Entretanto, há situações em que a classe dos sinais de interesse não atende aos critérios de esparsidade. Para que os algoritmos ainda possam ser usados, recorreremos a uma base transformadora. Em outras palavras, supomos que o sinal x tenha uma representação esparsa $\hat{x} = Tx$ sob a transformação linear $T \in \mathbb{K}^{m \times n}$. Definindo A como uma matriz de aquisição dada por $A = MT^{-1}$, relação $Mx = b$ pode ser reescrita como $A\hat{x} = b$.

Contudo, uma dúvida pode surgir: Quais são as possíveis funções de custo que se adequariam bem a esparsidade? Suponha que o conjunto de suporte Ω seja conhecido. Nesse caso, a solução pode ser calculada por Mínimos Quadrados, porque o problema pode ser reduzido a um sistema determinado. Outra alternativa também utilizada são os algoritmos gulosos. A ideia é aproximar a solução mais esparsa melhorando a estimativa do suporte iterativamente. Todavia, como já discutido, a norma l_0 se fundamenta na contagem de entradas diferentes de 0 no vetor x . Assim, esta norma pode expressar a soluções de CS como:

$$\min \|\hat{x}\|_0 \text{ sujeito a } A\hat{x} = b. \quad (2.6)$$

Nesse sentido, uma métrica l_p , com $(0 \leq p \leq 1)$, pode ser usada como função de custo para o problema de minimização tal que:

$$\min \|\hat{x}\|_p^p \text{ sujeito a } A\hat{x} = b. \quad (2.7)$$

Mesmo sabendo que temos uma definição para o problema de encontrar a solução mais esparsa de um sistema linear subdeterminado $A\hat{x} = b$, ainda precisamos satisfazer certas condições para garantir que a solução mais esparsa exista e seja única. Nesse sentido, existem critérios que avaliam a matriz A para fornecer parâmetros sobre a relação entre a dispersão e o número de medições necessárias para a reconstrução.

2.3.2 Propriedade de Isomeria Restrita

Um critério relativo à matriz A que pode garantir a reconstrução sinais a partir de um número reduzido de medições resolvendo uma otimização l_0 ou l_1 é a propriedade de isometria restrita (do inglês *Restricted Isomorphism Property* - RIP). Candès *et al.* [76] introduziu este conceito.

Definição 2.2 - constante de isomeria restrita: Seja a matriz $A \in \mathbb{K}^{m \times N}$. A constante de isomeria restrita de A de ordem k é o menor número δ_k positivo tal que, para todos os

valores k -esparsos de \hat{x} ocorrer,

$$(1 - \delta_k) \|\hat{x}\|_2^2 \leq \|A\hat{x}\|_2^2 \leq (1 + \delta_k) \|\hat{x}\|_2^2. \quad (2.8)$$

A constante de isometria restrita reflete a dependência linear dos vetores coluna da matriz de aquisição. Além disso, caracteriza o espaço nulo de A analisando seus valores singulares para garantir a unicidade da solução mais esparsa. A ideia de usar a constante de isometria restrita como critério para unicidade tem base em fornecer um valor máximo para δ_{2k} para o qual o comportamento aproximadamente ortonormal da matriz de aquisição garante que existe um único vetor k esparsos recuperável. Assim, é possível definir a propriedade de isometria restrita.

Definição 2.3 - propriedade de isometria restrita: Seja δ_{2k} a constante de isometria restrita para a matriz de aquisição A . A minimização l_1 terá uma solução única se $\delta_{2k} \sqrt{2} - 1$. Indo mais além, as soluções para as minimizações l_1 e l_0 são equivalentes.

Se quisermos avaliar a propriedade de isometria restrita para outras constantes de isometria restrita, usando os parâmetros $p \in (0, 1]$, referindo-se à métrica adotada no minimizador, e $t \in (1, 2]$, referindo-se a ordem da constante de isometria restrita, uma solução única para

$$\min \|\hat{x}\|_p^p \text{ sujeito a } A\hat{x} = b \quad (2.9)$$

será dada pela combinação entre p e t de forma que,

$$\delta_{tk} \leq \frac{\eta}{\frac{2-p}{t-1} - \eta}. \quad (2.10)$$

onde η é a única solução positiva para:

$$\frac{p}{2} \eta^{\frac{2}{p}} + \eta - \frac{2-p}{2(t-1)} = 0. \quad (2.11)$$

2.3.3 Coerência

A coerência é um índice que estuda a dependência linear das colunas de uma dada matriz. Esta propriedade é usada para estimar a esparsidade garantindo que os sinais sejam recuperados unicamente pela minimização l_0 . Portanto, a coerência relaciona a quantidade máxima de entradas não nulas na saída à estrutura da matriz que modela o sistema linear em questão. Dadas as características da coerência, é possível defini-la.

Definição 2.4 - coerência: Seja a matriz $A \in \mathbb{K}^{m \times N}$ e defina a matriz \bar{A} como uma

matriz com colunas $(\overline{a}_1, \overline{a}_2, \dots, \overline{a}_m)$ normalizadas com a minimização l_2 a partir das colunas da matriz A . A Coerência (μ) é o valor absoluto máximo do produto escalar entre duas colunas distintas \overline{a}_i e \overline{a}_j tal que,

$$\mu(A) = \max_{1 \leq i, j \leq m; i \neq j} |\overline{a}_i^H \overline{a}_j|. \quad (2.12)$$

É importante lembrar que o produto escalar mede a direção relativa de dois vetores. Isso significa que a coerência associa a distância angular máxima entre dois vetores-coluna distintos da matriz A . Podemos questionar: como a coerência é utilizada como critério para avaliar a unicidade da solução mais esparsa de A ? Ao definirmos *spark* como o menor número de colunas de A que pode ser selecionada para formar um conjunto linearmente dependente e afirmamos que não há um vetor nulo $y \in \mathbb{K}^n$ em A , além de definir os valores das colunas de A como (a_1, a_2, \dots, a_n) e o conjunto suporte de y como Ω_y , é válida a relação:

$$A_y = \sum_{i=1}^n y[i] a_i = \sum_{i \in \Omega_y} y[i] a_i = 0. \quad (2.13)$$

Isso nos mostra que o subconjunto $\{a_i | i \in \Omega_y\}$ das colunas de A é linearmente dependente. Aplicando a definição de *spark*, se $|\Omega_y| < \text{spark}(A)$, então não há um vetor y diferente de zero que satisfaça a equação 2.13. Portanto, $\|y_o\| \geq \text{spark}(A)$.

Podemos usar essa desigualdade para relacionar as colunas de A a esparsidade das soluções do sistema em questão. Para fazer isso, vamos tomar x^1 e x^2 como duas prováveis soluções para o sistema subdeterminado $A\hat{x} = b$. Lembrando ainda da desigualdade triangular, temos que $\|\hat{x}_1 + \hat{x}_2\|_0 \geq \|\hat{x}_1\|_0 + \|\hat{x}_2\|_0 \geq \text{spark}(A)$. É importante frisar que, se $\|\hat{x}_1\|_0 < \frac{1}{2}\text{spark}(A)$, então não há uma solução com menos valores diferentes de zero, isto é, a solução mais esparsa será única. Alguns autores chamam este critério de propriedade do espaço nulo (do inglês *Null Space Property* - NSP) [79]. Aplicando o teorema do disco de Gershgorin [80], podemos estender a propriedade do espaço nulo até a coerência. Para estabelecer uma condição para a unicidade, precisamos relacionar a coerência com a norma l_0 .

Teorema 2.1: coerência - Seja $A\hat{x} = b$ um sistema subdeterminado que possui uma solução \hat{x}_s . Se a relação abaixo for satisfeita, então \hat{x}_s é a solução mais esparsa possível.

$$\|\hat{x}_s\|_0 \leq \frac{1}{2} \left(1 + \frac{1}{\mu(A)}\right). \quad (2.14)$$

O que se pode concluir é que a propriedade de isomeria restrita e a coerência são capazes de deduzir da estrutura da matriz de aquisição dado que o grau de subamostragem ainda permite a reconstrução de acordo com uma esparsidade máxima esperada para os sinais que se esteja trabalhando [76].

2.4 Parâmetros de Avaliação de Desempenho

Para a avaliação das imagens reconstruídas, diversos parâmetros de avaliação têm sido utilizados pelos pesquisadores [27]. As métricas de avaliação mais comuns são: a relação sinal-ruído (do inglês *Signal-to-Noise Ratio* -SNR), a relação contraste-ruído (do inglês *Contrast-to-Noise Ratio* - CNR), o erro quadrático médio (do inglês *Mean Absolute Error* - MSE), a raiz do erro quadrático médio (do inglês *Root Mean Square Error* - RMSE), a relação sinal-ruído de pico (do inglês *Peak Signal-to-noise Ratio* - PSNR), o índice de similaridade estrutural (do inglês *Structure Similarity Index* - SSIM) e o erro absoluto médio (do inglês *Mean Absolute Error* - MAE). A seguir, apresenta-se a definição dos mesmos.

O MAE calcula a média da diferença absoluta entre a previsão do modelo e o valor alvo, através da equação a seguir:

$$MAE = \frac{1}{N} \left(\sum_{i=1}^N |I_{oi} - I_{ri}| \right), \quad (2.15)$$

onde N representa o número total de coeficientes de dados de imagem, I_{oi} é a intensidade ou amplitude do i-ésimo pixel da imagem original e I_{ri} é a intensidade ou amplitude do i-ésimo pixel da imagem processada.

A SNR [81] é usada para avaliar a qualidade da imagem com a definição como:

$$SNR = 10 \log_{10} \left(\frac{MAE}{\delta} \right). \quad (2.16)$$

onde δ representa o desvio padrão do ruído.

Uma das métricas de desempenho mais utilizadas é o MSE, que consiste na média do erro das diferenças ao quadrado. Em outras palavras, pega-se a diferença entre o valor previsto pelo modelo e o valor real, eleva-se o resultado ao quadrado. Faz-se a mesma coisa com todos os outros pontos, soma-os, e divide-se pelo número de elementos preditos. Quanto maior esse número, pior o modelo. Essa métrica apresenta valor mínimo 0, sem valor máximo e é dada pela equação abaixo:

$$MSE = \frac{1}{N} \sum_{i=1}^N (I_{oi} - I_{ri})^2. \quad (2.17)$$

A RMSE é uma regra de pontuação quadrática que fornece a média do erro. A RMSE é usado para medir a precisão de variáveis contínuas. Na equação abaixo, I_s é a intensidade do pixel da imagem original e I_{ri} é a intensidade dos pixels da imagem processada. Matematicamente, a RMSE é escrito da seguinte forma:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_{oi} - I_{ri})^2} \quad (2.18)$$

A PSNR é a relação entre a potência máxima possível de uma imagem e a potência do ruído de corrupção que afeta a qualidade de sua representação [82]. Para estimar a PSNR de uma imagem é necessário comparar essa imagem com uma imagem limpa ideal, isto é, sem ruído, com a máxima potência possível.

$$\text{PSNR} = 10 \log_{10} \left(\frac{(L-1)^2}{MSE} \right) = 20 \log_{10} \left(\frac{(L-1)}{RMSE} \right), \quad (2.19)$$

onde L é o número de níveis de intensidade máximos possíveis em uma imagem.

O SSIM [81] é uma medida da degradação originada devido à compressão de dados ou devido a perdas de dados durante a transmissão. Duas imagens são necessárias para o cálculo do SSIM: a imagem original e a reconstruída. O SSIM é expresso matematicamente como:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (2.20)$$

onde μ_x e μ_y caracterizam a média de x e y , respectivamente, e μ_1 expressa o valor médio da imagem original I . Os termos σ_x e μ_y representam a variância e σ_{xy} representa a covariância de x e y , respectivamente. Aqui na equação, c_1 e c_2 estabilizam a divisão com um denominador fraco. Se o SSIM resultar no valor 0, as imagens serão totalmente diferentes e duas imagens serão totalmente correspondidas se o valor for 1.

A CNR determina a qualidade da imagem reconstruída em comparação com a imagem original. A CNR é semelhante a SNR com a diferença de que ele subtrai algum termo de intensidade da imagem e então calcula a razão. A CNR é dado pela seguinte equação:

$$\text{CNR} = \frac{|S_O - S_P|}{\sigma_0} \quad (2.21)$$

onde S_O representa a intensidade da imagem da estrutura de produção O , S_P indica a intensidade da imagem para estruturas produtoras de imagem P na região de interesse e σ_0 representa o desvio padrão do ruído da imagem original [83]

2.5 Artefatos de Imagem

Os artefatos de imagem são alterações na imagens que causam uma piora na qualidade das mesmas. É importante que estes artefatos sejam conhecidos para que possam ser compensados. Alguns artefatos podem ser eliminados, porém outros são irreversíveis e podem apenas ser reduzidos [84]. Os artefatos podem, ainda, ser relacionados ao paciente (movimento, metais ou fluxo), processamento de imagens (químico, dobradura, aliasing, etc.) ou aparelho (radiofrequência, gradiente) [5]. Abaixo, têm-se alguns exemplos de artefatos com suas respectivas fontes citadas:

Os artefatos de movimento são adquiridos através da movimentação de um paciente no

momento exato em que a imagem está sendo formada, causando perda de sinal e de parâmetros de localização. A Figura 2.12 apresenta um artefato de movimento causado pelo fluxo sanguíneo nos vasos humanos e a Figura 2.13 mostra um artefato causado pelo próprio paciente que se moveu durante a realização de um exame de ressonância magnética.

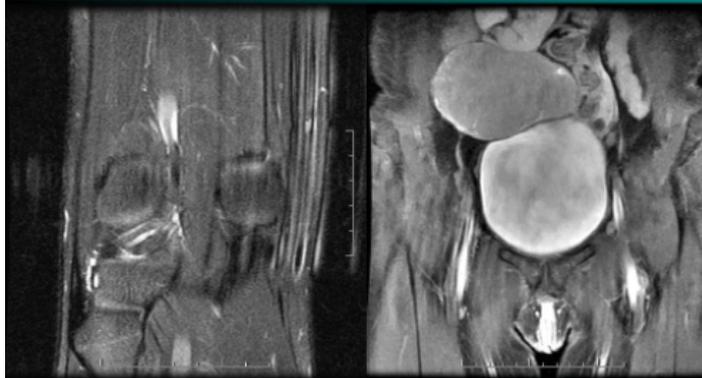


Figura 2.12: Exemplo de artefato causado pelo movimento do fluxo sanguíneo. Retirado de[10].

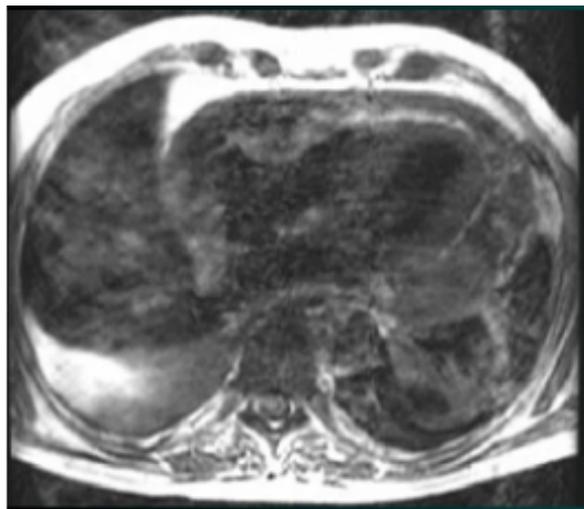


Figura 2.13: Exemplo de artefato causado pelo movimento do paciente. Retirado de [10].

Os artefatos metálicos são adquiridos através da presença de objetos metálicos no corpo humano (como próteses dentárias ou *piercings*, por exemplo), causando a possibilidade de criação de um campo eletromagnético que cause interferência. A Figura 2.14 mostra um exemplo de artefato metálico causado por um aparelho dentário do paciente.

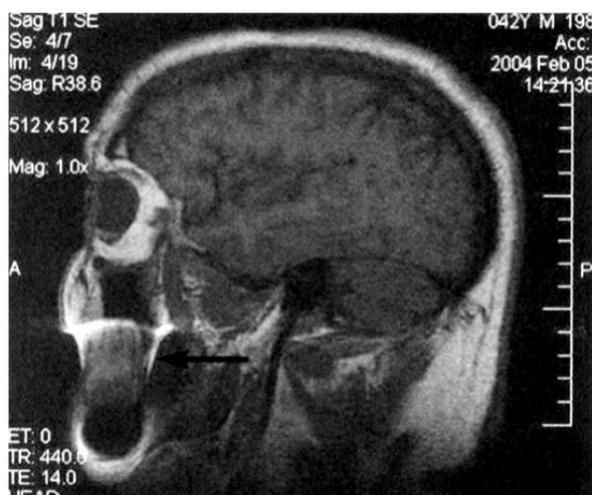


Figura 2.14: Artefatos causados por material metálico. Retirado de[11]

O artefato de *aliasing* também conhecido como artefato por dobramentos, é produzido quando a imagem é adquirida com informações adicionais a área externa. A imagem aparece com "dobramentos", dando, assim, a nomenclatura. Um exemplo deste artefato está na Figura 2.15, onde tem-se a impressão de haver a sobreposição de duas imagens do crânio em uma só.



Figura 2.15: Exemplo de artefato causados por Aliasing. Retirado de[10]

O artefato de Desvio ou Deslocamento Químico é causado por diferentes ambientes químicos de gordura e água no corpo humano. Este artefato causa uma margem escura entre os lipídios e a água. Na Figura 2.16 encontra-se um exemplo de artefato causado por fluidos no tecido humano.

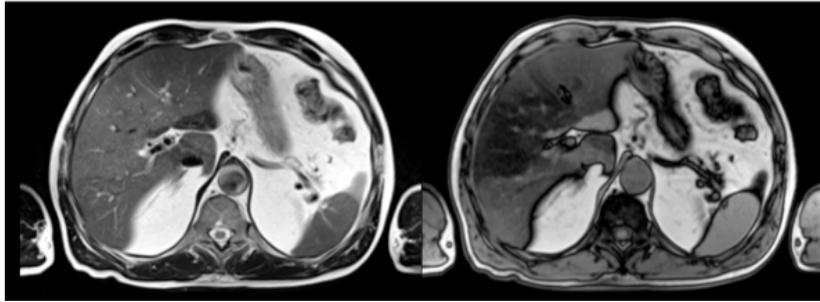


Figura 2.16: Artefatos causados por desvio químico. Retirado de[1].

O artefato por fuga de Radiofrequência é causado pela saída ou interferência de rádio frequência (RF) externa. Conforme a Figura ref2.19, onde a interferência causou a presença de linhas anormais na imagem.

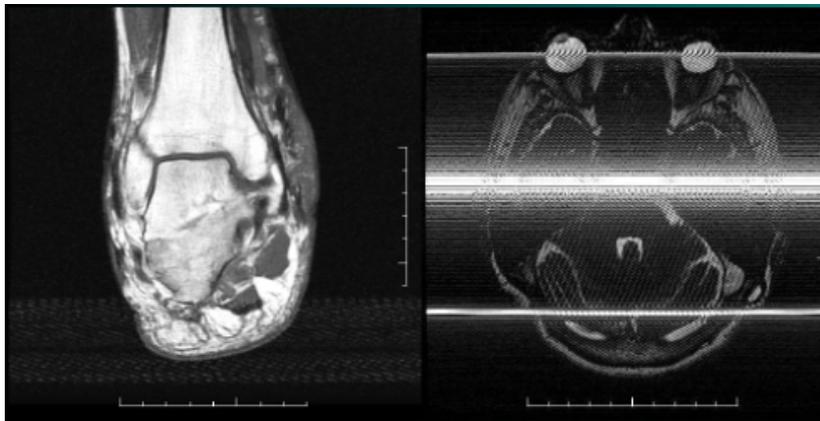


Figura 2.17: Artefatos causados por interferência de radiofrequência. Retirado de [12].

O artefato oscilação (Fenômeno de Gibbs) é causado pela sub-amostragem de altas frequências espaciais. A falta de apropriados componentes de alta frequência leva a uma oscilação brusca. Ele ocorre próximo às fronteiras nítidas, onde existe alto contraste nas transições do objeto. Um exemplo deste artefato é mostrado na Figura 2.18, onde as setas brancas ilustram oscilações na imagem.

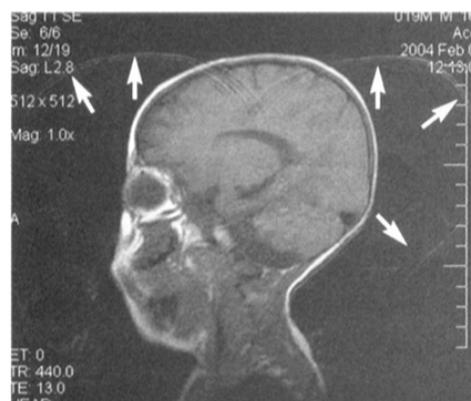


Figura 2.18: Exemplo de artefato causados por oscilação (Fenômeno de Gibbs). Retirado de[12].

O artefato *spikes* tem este nome por causa da forma como este artefato aparece nos dados não processados, como pontos brancos. O artefato *spike* é caracterizado por um aumento repentino no sinal recebido, que não foi causado pelo objeto de estudo. A Figura 2.19 ilustra um exemplo onde a imagem possui linhas destacadas que não condizem com o esperado.

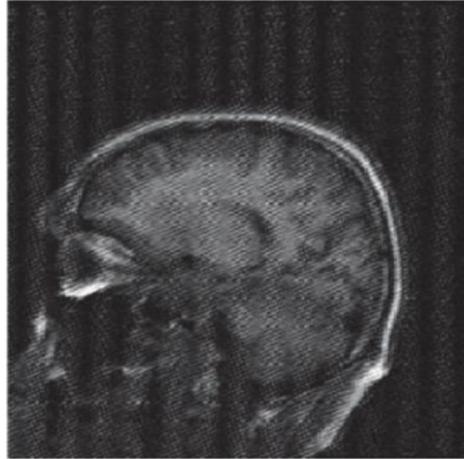


Figura 2.19: Exemplos de artefato causado por spikes. Retirado de[12].

Capítulo 3

Metodologia e Implementação

Embora os métodos de aprendizagem profunda possam aprender de forma automática e sejam capazes de obter implicitamente as características espaço-temporais de um sinal, o custo computacional para serem treinados é elevado e aumentam o nível de complexidade das diversas arquiteturas propostas devido ao elevado número de parâmetros a serem aprendidos. Além disso, as técnicas de CS desenvolvidas por engenheiros são muito úteis na detecção, segmentação e, principalmente, na reconstrução de imagens. O presente trabalho tem como proposta apresentar um algoritmo de reconstrução de imagens de ressonância magnética, de complexidade computacional mediana, utilizando técnicas de aprendizado profundo para aumentar a qualidade das imagens geradas por exames de imagem clínicos, auxiliando o diagnóstico de doenças como o câncer.

3.1 Escolha entre a U-net e o Autoencoder

O primeiro passo necessário neste trabalho era escolher a técnica de aprendizado profundo mais eficaz para o problema proposto. Sabendo que o uso das CNNs teria alta aplicabilidade para o processo de reconstrução de imagens de ressonância magnética, foi necessário responder ao questionamento: autoencoder ou U-Net? Apesar das propostas serem semelhantes, há uma importante diferença entre as elas. Nos autoencoders, há um espaço latente L com um mapeamento não linear da entrada X para o espaço de entrada $E : X \rightarrow L$ e um mapeamento correspondente deste espaço latente para o espaço de saída $D : L \rightarrow Y$ [65, 85]. Neste caso, há uma distinção clara entre o codificador e o decodificador. O codificador altera a representação de cada amostra em algum parâmetro no espaço latente e o decodificador é capaz de construir saídas com apenas estes parâmetros. Assim, pode-se desmontar essa rede e usar o codificador e o decodificador separadamente [86].

Na U-Net, no entanto, o mapeamento de saída depende diretamente do espaço de entrada, mas em vez de L corresponder diretamente a Y , $[X + L] \rightarrow Y$. Isso significa que não há blocos de codificador e decodificador, que mapeiam as amostras em algum espaço latente

bem definido e depois computam a saída correspondente [66]. Ou seja, não se pode dividir uma U-Net em partes e usá-las separadamente, pois para calcular a saída, a entrada também é necessária. Assim, como em todas as suas representações intermediárias, existem vários espaços latentes na U-Net: $X \rightarrow L_1 \rightarrow L_2 \rightarrow \dots \rightarrow L_n$.

Os Autoencoders são treinados para preservar o máximo de informação possível quando uma entrada é passada pelo codificador e depois pelo decodificador, mas também são treinados para fazer com que a nova representação tenha várias propriedades interessantes. Essa característica apresenta um risco para a reconstrução de imagens médicas, pois o Autoencoder pode acabar sacrificando detalhes da fisiologia de um órgão ou tecido humano para que a imagem de saída obtenha uma representação interessante. Já a U-Net foi pensada inicialmente como uma ferramenta capaz de processar imagens biomédicas onde o objetivo não é apenas identificar se há ou não a presença de uma doença mas também identificar a área que viria a ser afetada pela doença. Para isso, a U-Net leva em conta todos os aspectos da imagem de entrada e entrega, na saída, a imagem mais completa possível, já que a segmentação ajuda a identificar onde objetos de diferentes classes estão presentes em uma imagem. Assim, o uso da U-net foi o mais adequado para implementar o método proposto.

3.2 Reconstrução de Imagens Utilizando Autoencoder

Através de testes computacionais feitos na elaboração deste método e apresentador em seguida no Capítulo 4, seção 4.1, chegou-se a conclusão de que o uso de autoencoder como rede neural artificial não apresentava resultados satisfatórios para o algoritmo de reconstrução de imagens que se desejou implementar. Ainda assim, apresenta-se a metodologia que foi usada para chegar a esta conclusão. Primeiramente, utilizou-se um conjunto de cerca de 4.000 imagens extraídas do banco de dados BIRN [87]. O código utilizado foi baseado no código *open source* desenvolvido pela cientista de dados Eugenia Anello, em linguagem de programação Python¹ [88]. O código desenvolvido usa um autoencoder convolucional composto por:

- Codificador: etapa pela qual o modelo aprende a criar uma versão comprimida dos dados. O codificador do modelo utilizado possuía 8 camadas convolucionais onde cada camada realizava uma convolução 2x2 seguida de uma Leaky ReLu (é um tipo de função de ativação baseada em uma ReLU);
- Representação Comprimida: armazena a representação comprimida calculada pelo codificador. Durante o processo de codificação as imagens eram comprimidas em fator de 2^n ;
- Decodificador: etapa responsável por reconstruir os dados previamente codificados. O decodificador do modelo utilizado também possuía 8 camadas convolucionais onde

¹Acessível em: <https://www.python.org/>

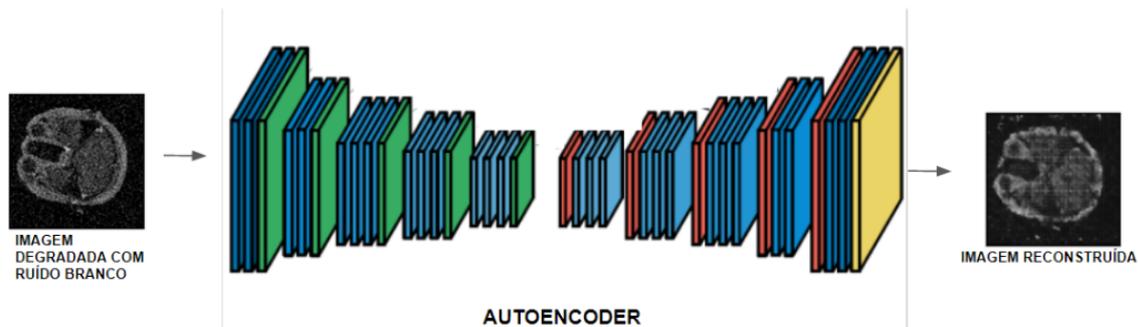


Figura 3.1: Diagrama de blocos do modelo proposto com o autoencoder.

cada camada realizava uma convolução 2x2 seguida de uma ReLu;

- Cálculos de perda: métrica que mede a qualidade da reconstrução feita pelo decodificador. O objetivo era minimizar este número a medida que os testes eram realizados com diversos números de épocas.

Antes de treinar o autoencoder convolucional foi necessário definir blocos de construção: as redes codificadoras e decodificadoras, a função de perda baseada em MSE, a função de otimização Adam e o uso da GPU através das ferramentas do Pytorch. Definimos então uma função para treinar o modelo com o autoencoder. Primeiro, passamos as imagens de entrada para o codificador. Mais tarde, os dados codificados são passados para o decodificador e então calculamos a perda de reconstrução. Uma vez criada a função de treinamento, definimos uma função para avaliar o desempenho do modelo. Como antes, passamos a imagem para o codificador. A imagem codificada é passada para o decodificador. Em seguida, armazenamos todos os lotes de imagens e a reconstrução em duas listas diferentes, que serão usadas para calcular a perda do teste. Para tentar minimizar a função de perda, utilizamos cada vez mais épocas nos testes, indo de 30 a 20000 épocas.

A Figura 3.1 ilustra um diagrama de blocos da arquitetura proposta para testes com o autoencoder. Treinou-se a rede com 80% dos dados para treino, 10% para teste e 10% para validação. Antes mesmo de serem submetidas a primeira camada do autoencoder, as imagens originais foram acrescidas de ruído branco com desvios padrões de gaussiana iguais a 0.01, 0.05 e 1, respectivamente.

No processo de treinamento, foi utilizado o MSE como o função de perda e um tamanho de *lote* igual a 32. Além disso, foi utilizado o otimizador Adam com uma taxa de aprendizado constante de 0,0001 e um decaimento de peso de 0,00001. Foram testados ainda diversos números de épocas pois era viável ver as imagens reconstruídas durante cada época do treinamento. Os testes computacionais foram feitos primeiramente com imagens de tamanho 32x32. Já sabendo do funcionamento da rede e quais adaptações deveriam ser feitas para alcançar maior eficácia para o método de reconstrução de imagens de ressonância magnética, realizaram-se testes com imagens maiores, de tamanhos 128x128 e, por fim, de

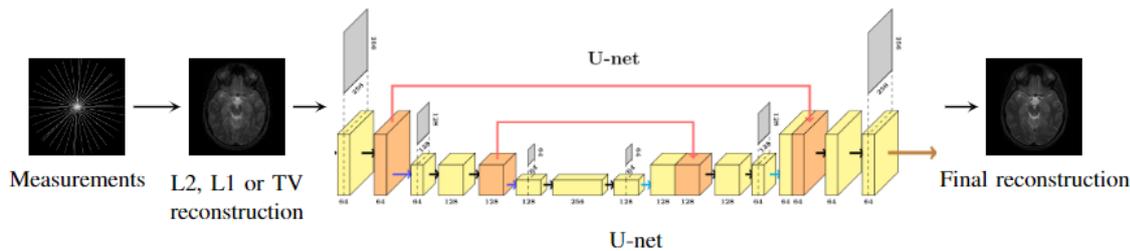


Figura 3.2: Diagrama de blocos do modelo proposto com a U-Net.[13]

tamanho 512x512, que é o tamanho das imagens que foram utilizadas para implementar o método proposto.

3.3 Reconstrução de Imagens Utilizando a U-Net

Melhorar o desempenho dos algoritmos de reconstrução de imagens é fundamental dentro da área de processamento de imagens. O desempenho do tempo de execução da maioria das métricas utilizadas para avaliar um algoritmo depende da qualidade da imagem original bem como parâmetros específicos. Propõe-se aqui uma estrutura de Aprendizado Profundo, através do uso de uma rede neural artificial treinada e testada, a U-Net. A U-Net pode ser usada para aprender a não só identificar erros nas imagens, bem como possibilitar um processamento com resultados satisfatórios. O objetivo deste processamento em imagens médicas é auxiliar o diagnóstico de um médico, já que as imagens de ressonância magnética estão suscetíveis a presença de artefatos que causam sua degradação.

Para a construção deste método de reconstrução de imagens é necessário utilizar ferramentas de otimização nas imagens a serem reconstruídas, para que as imagens de saída sejam mais similares às imagens de entrada. Um fenômeno negativo comum nos modelos de aprendizado de máquina é o *overfitting*. O *overfitting* ocorre quando um modelo de aprendizado de máquina é extremamente adaptado a um conjunto de dados específico e não apresenta resultados consistentes em outros conjuntos de dados diferentes, ou seja, os pontos de dados que não refletem as propriedades dos dados são considerados irrelevantes. Isso geralmente acontece em modelos complexos, como redes neurais profundas. Para evitar o *overfitting* recorre-se a técnicas como amostragem de validação cruzada, redução do número de recursos, poda e, uma das mais utilizadas, a *regularização*. A regularização é um processo de introdução de informações adicionais no algoritmo de otimização para minimizar a função de perda. É um tipo de regressão que minimiza as estimativas dos coeficientes a zero para reduzir a capacidade de um modelo. Neste contexto, a redução da capacidade de um modelo envolve a retirada de pesos extras de modo que o modelo generalize os dados. A generalização acaba por evitar o *overfitting*.

Para entender como as ferramentas de otimização funcionam, deve-se primeiramente levar em consideração alguns conceitos importantes. Para uma abordagem mais clara, os

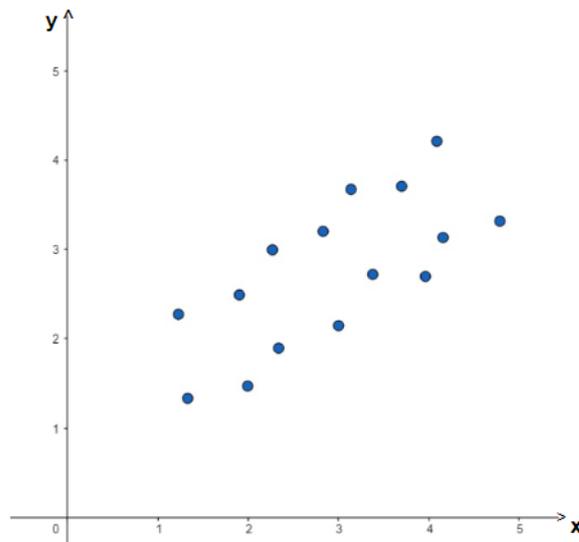


Figura 3.3: Gráfico de distribuição de dados no plano cartesiano.

conceitos serão apresentados matematicamente. O primeiro é o de regressão linear simples, que busca indicar qual o comportamento de uma variável dependente como uma função que contenha uma ou mais variáveis independentes para um modelo \hat{y} . A relação matemática da regressão linear é dada pela seguinte equação:

$$\hat{y} = zx + b. \quad (3.1)$$

onde z é o peso e b é o viés. O valor de \hat{y} representa um valor dependente, aquilo que o modelo tenta prever. A variável x é dita explicativa ou independente, são os valores de entrada do modelo.

Outro conceito importante é o da função de perda ou função de custo, que é calculada tirando-se como a diferença entre o valor final y e o valor previsto \hat{y} . Esta função calcula o nível de erro de predição dos resultados adquiridos com os valores iniciais. Sua representação matemática é dada pela seguinte equação:

$$\text{Loss} = (\hat{y} - y)^2 = (zx + b - y)^2. \quad (3.2)$$

A regularização utiliza ainda a otimização por gradiente descendente z . Quando utilizamos a regressão linear estamos tentando descrever uma função que se adeque a um conjunto de dados. Essa função pode ser descrita por uma reta que explique o comportamento dos dados buscando melhor a relação de x com o \hat{y} . Observe a Figura 3.3. Na figura temos vários pontos que representam os dados (x e respectivos valores de y). Com a regressão linear, o que queremos é uma função que descreva o comportamento destes dados. Na verdade, estamos procurando uma reta que explique bem o comportamento dos nossos dados sendo a mais ajustada possível, de forma a ser a que melhor mostre a relação do x com o \hat{y} .

Sabendo como a regressão linear foi descrita na Equação 3.1, o gradiente pode ser explicado como um algoritmo inteligente que começa em um valor e, a partir dele, tenta calibrar

as variáveis da regressão linear z e b de tal forma que basta seguir a direção oposta do gradiente para encontrarmos o mínimo da função, isto é, o gradiente descendente. Este conceito é representado pela seguinte equação, onde η é a taxa de aprendizado:

$$z_{\text{final}} = z - \eta \frac{\partial \text{Loss}}{\partial z}. \quad (3.3)$$

Dados os conceitos fundamentais apresentados nesta seção, escolhemos três métodos de otimização por regularização, que são descritos a seguir. A ideia de utilizar três tipos de métodos de otimização diferentes é permitir o estudo de efeitos de uma certa rede neural nos diferentes tipos de algoritmos de reconstrução.

3.3.1 Regularização L1

A ideia da regularização L1 é aplicar uma norma L1 no processo de regularização para encontrar o melhor vetor z para minimizar uma função de perda, ou seja, incorporar a norma L1 de z encontrando o menor valor para \hat{y} . Um detalhe crucial é que um vetor z com uma norma L1 pequena tende a ser uma solução esparsa, pois a maioria de seus componentes são zeros. A regularização L1 leva este nome devido a norma L1 do vetor z a seguir:

$$\|z\|_1 = |z_1| + |z_2| + \dots + |z_n|. \quad (3.4)$$

O modelo de regressão linear que implementa a norma L1 para regularização é chamado de regressão de Lasso e é dado pela seguinte equação:

$$\hat{y} = z_1 x_1 + z_2 x_2 + \dots + z_N x_N + b. \quad (3.5)$$

O cálculo da função de perda de L1 é dado seguinte equação:

$$\text{Loss} = \text{Error}(y, \hat{y}) + \lambda \sum_{i=1}^N |z_i|, \quad (3.6)$$

e o gradiente descendente pela seguinte equação

$$z_{\text{final}} = z - \eta \frac{\partial L1}{\partial w} = z - \eta \cdot (2x(zx + b - y) \pm \lambda), \quad (3.7)$$

onde o ideal para evitar o *overfitting* é que $\eta = 1$.

3.3.2 Regularização L2

A aplicação de uma norma L2 no processo de regularização tem o objetivo de encontrar o melhor vetor z para minimizar uma função de perda, ou seja, incorporar a norma L2 de z , encontrando o menor valor para \hat{y} e forçando suas componentes a terem valores pequenos

mas não necessariamente iguais a zero. A regularização L2 leva este nome devido a norma L2 do vetor z dada pela seguinte equação:

$$\|z\|_2 = (|z_1|^2 + |z_2|^2 + \dots + |z_n|^2)^{\frac{1}{2}}. \quad (3.8)$$

O modelo de regressão linear que implementa a norma L2 para regularização é chamado de regressão de cume e é dado pela seguinte equação:

$$\hat{y} = z_1x_1 + z_2x_2 + \dots + z_Nx_N + b. \quad (3.9)$$

O cálculo da função de perda de L2 é dado por:

$$\text{Loss} = \text{Error}(y, \hat{y}) + \lambda \sum_{i=1}^N (z_i)^2, \quad (3.10)$$

e o gradiente descendente pela seguinte equação:

$$z_{final} = z - \eta \frac{\partial L2}{\partial z} = z - \eta \cdot (2x(zx + b - y) + 2z\lambda), \quad (3.11)$$

onde o ideal para evitar o *overfitting* é que $\eta < 1$.

Comparando as regularizações L1 e L2, observa-se que a regularização L1 penaliza a soma dos valores absolutos dos pesos, enquanto a regularização L2 penaliza a soma dos quadrados dos pesos. Além disso, a solução de regularização L1 é esparsa, já a solução de regularização L2 não é esparsa. Outra diferença se dá na seleção de recursos, já que a regularização L2 não realiza a seleção de recursos, pois os pesos são reduzidos apenas para valores próximos a 0 (em vez de 0), enquanto que a regularização L1 possui uma seleção de recursos incorporada.

3.3.3 Regularização TV

A regularização de Variação Total (TV) é uma técnica determinística que protege descontinuidades nas tarefas de processamento de imagens [89]. A regularização TV auxilia na redução de ruído da imagem sob ruído gaussiano. O primeiro passo da regularização TV é estimar o grau de ruído gaussiano que impactou a imagem original, com a seguinte equação:

$$y = Hx + n, \quad (3.12)$$

onde H é o operador linear e n é o ruído gaussiano. A conhecida norma de Variação Total é dada pela seguinte equação:

$$\sum_{i,j} \sqrt{(X_{i+i,j} - X_{i,j})^2 + (X_{i,j+1} - X_{i,j})^2} = \sum_{i,j} |(\nabla X)_{i,j}|, \quad (3.13)$$

onde ∇X denota a norma no espaço gradiente. O cálculo do gradiente descendente é descrito pelas seguintes equações:

$$\begin{aligned}(X_x)_{i,j} &= X_{i,j+1} - X_{i,j} \\ (X_y)_{i,j} &= X_{i+1,j} - X_{i,j}\end{aligned}\tag{3.14}$$

3.3.4 Subamostragem

Segundo o trabalho de Hyun *et al.* [24], uma imagem pode ser reconstruída com amostragem de espaços K cartesianos dada pela relação $y \in C^{N \times N}$, onde N^2 é o número de pixels da imagem e C , representa o conjunto dos números complexos. A imagem de ressonância magnética y pode ser reconstruída a partir dos dados de espaço k correspondentes. O espaço K pode trazer uma melhor compreensão da codificação espacial multidimensional [24].

Evocando o que foi explicado na Seção 2.3, o teorema da amostragem fornece uma frequência mínima na qual um sinal limitado em banda deve ser amostrado a fim de que possa ser reconstruído sem perdas. De acordo com o teorema, $f_s \geq 2f_{max}$, em que f_s é a taxa de amostragem do sinal e f_{max} é a frequência da componente de maior frequência do sinal. Deste modo, a fim de que o sinal seja perfeitamente reconstruído, ele deve ser amostrado com pelo menos o dobro da sua maior componente em frequência. Quando utiliza-se uma imagem de ressonância magnética subamostrada, viola-se o critério de Nyquist introduzindo artefatos na imagem, em especial, *aliasing*. É comum que os modelos de processamento de imagem atendam ao critério de amostragem Nyquist com relação ao seu padrão de amostragem. O engenheiro que estiver atuando na reconstrução de imagens tem liberdade para escolher o padrão de amostragem do espaço K . Popularmente, o padrão mais utilizado é o cartesiano pois é mais robusto evitando a introdução de imperfeições na imagens. Embora os padrões cartesianos sejam de longe os mais usados, alguns outros padrões estão em uso, incluindo amostragem ao longo de linhas radiais e amostragem ao longo de padrões espirais. Padrões radiais são menos suscetíveis a artefatos de movimento do que padrões cartesianos e podem ser significativamente subamostradas, especialmente para objetos de alto contraste, sendo mais utilizadas em reconstruções de imagens porque são menos suscetíveis a artefatos. Padrões Espirais fazem uso eficiente do hardware do sistema de gradiente e são usadas em aplicações de imagem rápida e em tempo real [90]. A Figura 3.4, retirada do artigo *Compressed sensing MRI: a review from signal processing perspective*, de Jong Chul Ye, mostra os padrões de subamostragem citados.

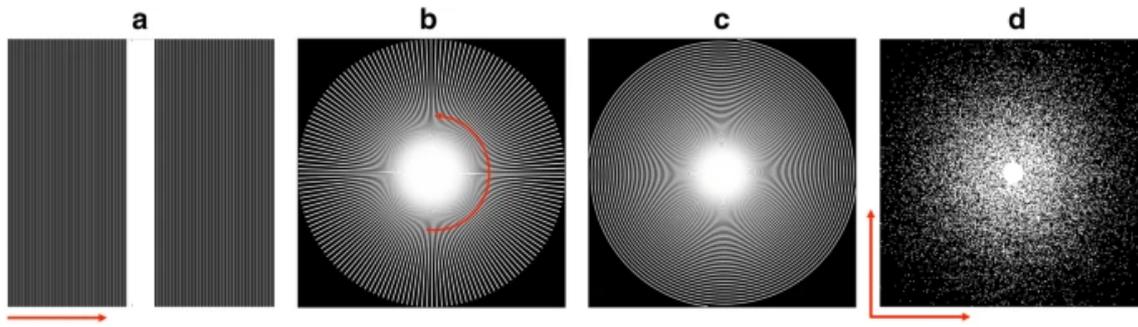


Figura 3.4: Vários padrões de subamostragem: (b) subamostragem cartesiana, (c) subamostragem radial e (d) subamostragem em espiral. Retirada de:[14].

No caso particular do padrão radial, ao invés de coletar linhas de espaço k em um esquema paralelo, a coleta das linhas no espaço k é feita em forma de estrela, com cada linha passando pelo centro. Semelhante ao padrão cartesiano, todos os raios radiais rotativos também são adquiridos linha por linha sequencialmente, mas com sobreposição ao redor do centro do espaço podem ser processadas para reconstruir uma imagem de ressonância magnética. Este padrão de amostragem oferece flexibilidade para projetar como estes raios radiais são girados de um para o outro. Um dos esquemas de rotação mais conhecido e mais usado é a chamada amostragem radial de ângulo dourado, na qual os raios radiais são girados por um chamado “ângulo dourado”, por exemplo, $111,25^\circ$. Para amostragem radial 2D, o ângulo de cada linha é dado pela seguinte equação [15]:

$$X = \frac{180(N-1)}{GR} \cdot \text{mod } 360 \quad (3.15)$$

onde GR corresponde a proporção áurea ², N é um ponto sobre o círculo trigonométrico, *mod* representa que a equação é modular e 360 equivale ao número de graus do círculo trigonométrico. A Figura 3.5, adaptada do artigo *Golden-Angle Radial MRI: Basics, Advances, and Applications* de Li Feng [15] ilustra a equação do ângulo dourado onde cada ponto indicado sobre o círculo representa o valor de N na equação 3.15.

Este padrão foi o escolhido para ser implementado no método de reconstrução de imagens proposto por suas vantagens. Uma delas é que a geometria exclusiva da amostragem radial de ângulo dourado incluiu uma maior robustez ao movimento em comparação com a amostragem cartesiana. O comportamento de subamostragem incoerente deste padrão pode, ainda, ser combinado sinergicamente com uma reconstrução esparsa de ressonância magnética e aquisição contínua de dados com classificação flexível de dados. Além disso, a amostragem radial também oferece auto-navegação que pode ser explorada para melhorar o gerenciamento de movimento [15].

²constante real algébrica irracional obtida no momento em que segmentamos uma reta em duas partes de forma que a divisão do segmento maior pelo menor seja igual à reta completa dividida pelo segmento mais longo. Seu valor é constituído pela dízima não periódica 1,6180339887...

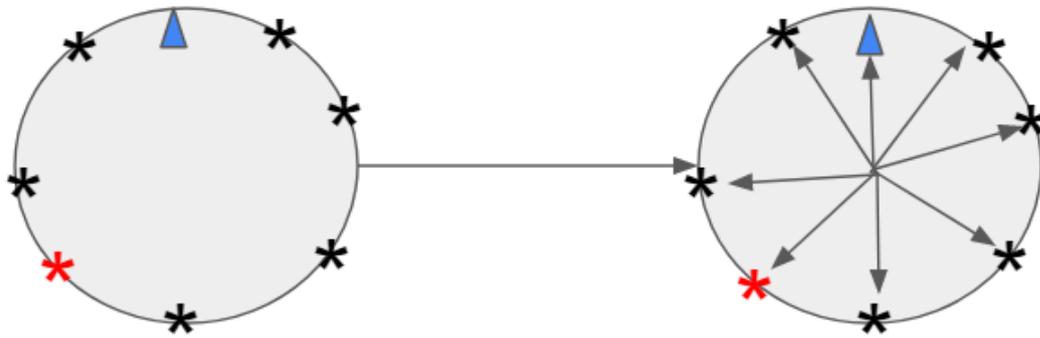


Figura 3.5: Círculo trigonométrico com alguns valores de N da equação do ângulo dourado indicados. Adaptada de:[15]

3.3.5 Configuração Básica do Modelo Proposto

Para a construção do método de reconstrução de imagens de ressonância magnética proposto implementou-se o modelo U-net usando o framework Pytorch³. A arquitetura da U-net utilizada possuía 23 camadas, divididas em via de contração (4 etapas contendo: 2 convoluções 3x3, uma ReLU, uma operação de max pooling 2x2) e caminho expansivo (4 etapas contendo: 2 convoluções 3x3, uma ReLU, uma operação de max pooling 2x2, além de uma camada final com uma convolução 1x1). Entre a via de contração e o caminho expansivo, foi realizada uma convolução ascendente 2x2. Após cada etapa, o número de canais de recursos foi reduzido por um fator de 2^n . Para executar as simulações, foram utilizadas 4.020 imagens de ressonâncias magnéticas cerebrais tiradas do conjunto de dados BIRN [87]. O conjunto de dados foi dividido da seguinte forma: 80% foi utilizado para treinamento, 10% para validação e 10% para teste. No processo de treinamento, foi utilizado o Erro Quadrático Médio (MSE) como o função de perda e um *lote* de tamanho igual a 8. Além disso, foi utilizado o otimizador Adam com uma taxa de aprendizado constante de 0,0001 e um decaimento de peso de 0,00001. Ao final, treinou-se o modelo por 300 épocas, selecionando os parâmetros da época com o menor valor de MSE.

Como mencionado anteriormente, através do processo de regularização, as imagens originais foram otimizadas com as normas L2, L1 e TV, gerando 3 modelos:

- L2-UNet: modelo treinado com imagens obtidas com o processo de regularização L2;
- L1-UNet: modelo treinado com imagens obtidas com o processo de regularização L1;
- TV-UNet: modelo treinado com imagens obtidas com o processo de regularização TV.

Após os modelos de otimização serem gerados, as imagens, agora regularizadas, foram introduzidas na rede U-Net e passaram por todo o seu processo matemático gerando uma imagem

³Acessível em <https://pytorch.org/>

de saída reconstruída. Por fim, como métricas de desempenho foram utilizadas a PSNR e o SSIM. Essas métricas são calculadas entre as imagens de saída finais e a imagens originais, com valores mais altos indicando melhores qualidades de imagem. O modelo completo proposto é ilustrado pela Figura 3.2.

Para realizar os testes neste trabalho, utilizamos apenas 20 pseudo-linhas radiais e uma grade retangular de 256×256 . Isso corresponde a 5.436 amostras retiradas de uma grade de 256×256 , correspondendo para uma taxa de amostragem de 8,365%. Vale ressaltar que este é um número muito pequeno de linhas radiais, o que gera medições sub-amostradas que são um desafio para o processo de reconstrução de sensoriamento pressionado. A Figura 3.6, retirada de Thomaz *et al.* [13], ilustra as informações citadas, onde (a) mostra o padrão de amostragem radial do espaço k , (b) a imagem original da qual as amostras recolhidas, (c) mostra a imagem amostrada correspondente e (d), (e), (f) mostram, respectivamente, as reconstruções da imagem original (b) utilizando as regularizações L2, L1 e TV.

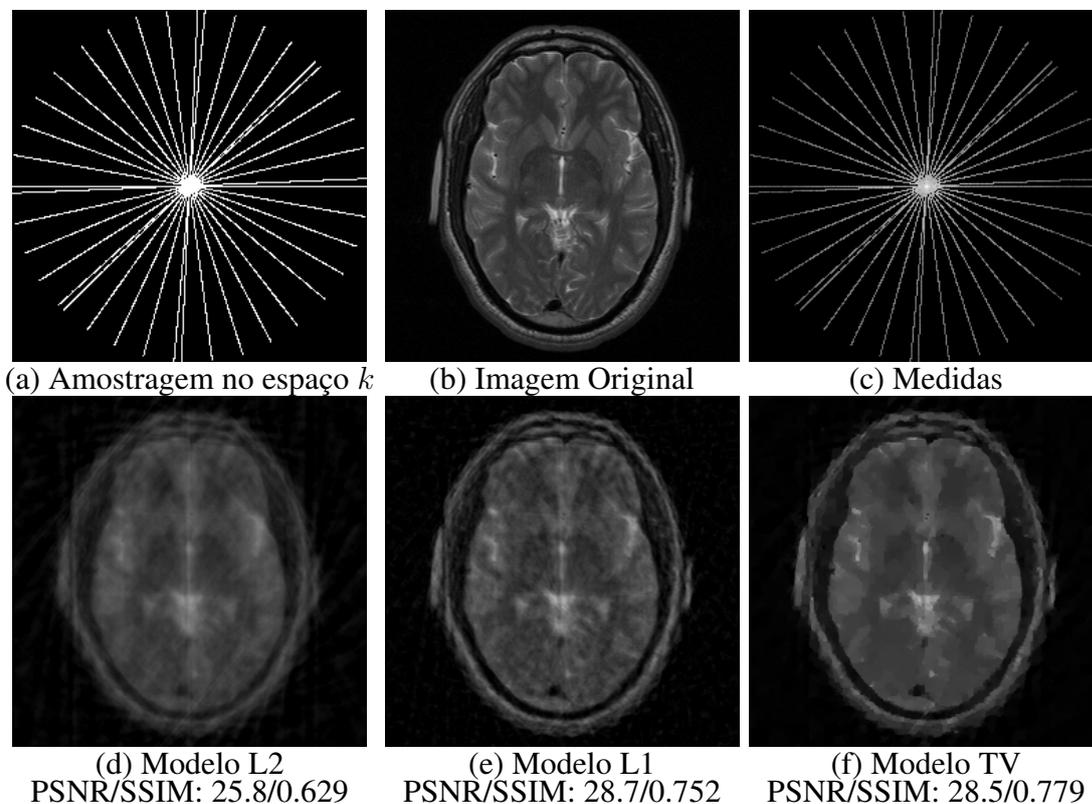


Figura 3.6: Processo de medição no espaço k e resultados da regularização de imagem. Retirado de: [13]

Capítulo 4

Resultados

Como mencionado nos capítulos anteriores, um dos objetivos para os métodos de reconstrução de imagens é a qualidade da imagem reconstruída em relação a imagem original. Estes métodos estão se tornando mais complexos ao longo das décadas, com alguns dos algoritmos implementados buscando aumentar a acurácia e a precisão na geração de imagens reconstruídas. Com base na ideia de aumentar as chances de um diagnóstico mais preciso de doenças severas, como câncer, propomos um método que melhora o desempenho da reconstrução de imagens digitais, sem afetar seu desempenho de precisão. Neste capítulo, serão apresentados os resultados das reconstruções de imagens de ressonância magnética utilizando o método proposto. Os resultados foram avaliados de forma qualitativa, através da comparação visual, assim como quantitativamente, com o cálculo do SSIM e da PSNR em decibéis.

Este capítulo está dividido em três seções. Na primeira seção, apresentamos os testes realizados com autoencoders justificando a escolha de a rede convolucional U-Net em seu detrimento. A seção seguinte apresenta os resultados para os dois casos propostos, divididos em duas subseções: comparação direta e comparação cruzada, elencando análise escrita e exemplificada com imagens. A última seção apresenta considerações relevantes.

4.1 Resultados de Modelo Simplificado Baseado em Auto-encoder

Como exposto no capítulo anterior, uma análise a respeito das variações de resultados obtidos entre autoencoders e U-Net foi necessária para uma maior eficácia na implementação do método proposto. A ideia inicial deste trabalho era a utilização dos autoencoders. Para isso, adaptamos um código de um autoencoder convolucional produzido por Eugenia Anello, pesquisadora na Universidade de Pádua, Itália [88] em linguagem Python ¹, realizando testes com imagens de banco de imagens de ressonância magnética. Para isso, foram utilizadas imagens de tamanho 512×512 . Os parâmetros são os mesmos utilizados por Anello *et al.*

¹Acessível em: <https://www.python.org/>

[88]. Realizamos testes para diferentes números de épocas, como por exemplo 30, 90, 120, 1000, 20000, 10000 e 20000. Considere a Figura 4.1 como a imagem original para ambos os resultados com 10.000 e 20.000 épocas. A título de comparação, apresentamos, na Figura 4.2, exemplos de resultados obtidos para 10.000 e 20.000 épocas, que obtiveram os melhores resultados visíveis. Neste testes, a cada imagem de entrada foi adicionado ruído branco Gaussiano, com média zero e desvios padrões de gaussiana iguais a 0.01, 0.05 e 1, respectivamente.

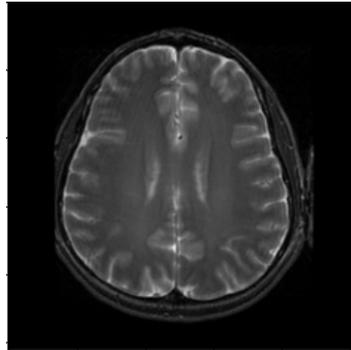


Figura 4.1: Imagem de ressonância magnética cerebral original retirado do banco de dados BIRN.

Como resultados, percebemos de forma clara que, embora o autoencoder tenha conseguido relativo sucesso na reconstrução de imagens, a reconstrução perdeu praticamente todos os detalhes da fisiologia do cérebro além de perdas no contraste e brilho. A medida que o fator de ruído aumenta, a reconstrução apresenta resultados piores. Sendo assim, após análise do material obtido, comprovamos o que foi exposto na subseção 3.1 do capítulo anterior: autoencoders perdem os detalhes das imagens durante o seu processo de reconstrução e não são muito viáveis para uso com imagens de ressonância magnética. Assim, optamos por usar a U-Net como Rede Neural Artificial.

4.2 Resultados do Modelo Baseado na Arquitetura U-Net

A análise dos resultados obtidos deve ser dividida em dois casos. O primeiro um caso de comparação direta, onde os três métodos de otimização L1, L2 e TV são comparados a eles mesmos. Em seguida, em uma segunda análise, estudamos os efeitos da reconstrução de imagens utilizando, novamente, os três métodos de otimização L1, L2 e TV em comparação cruzada, gerando um total de 9 combinações. Por fim, apresentamos os testes adicionais sobre o desempenho de métodos de avaliação de qualidade de imagem ao utilizar a reconstrução de imagens pela U-Net.

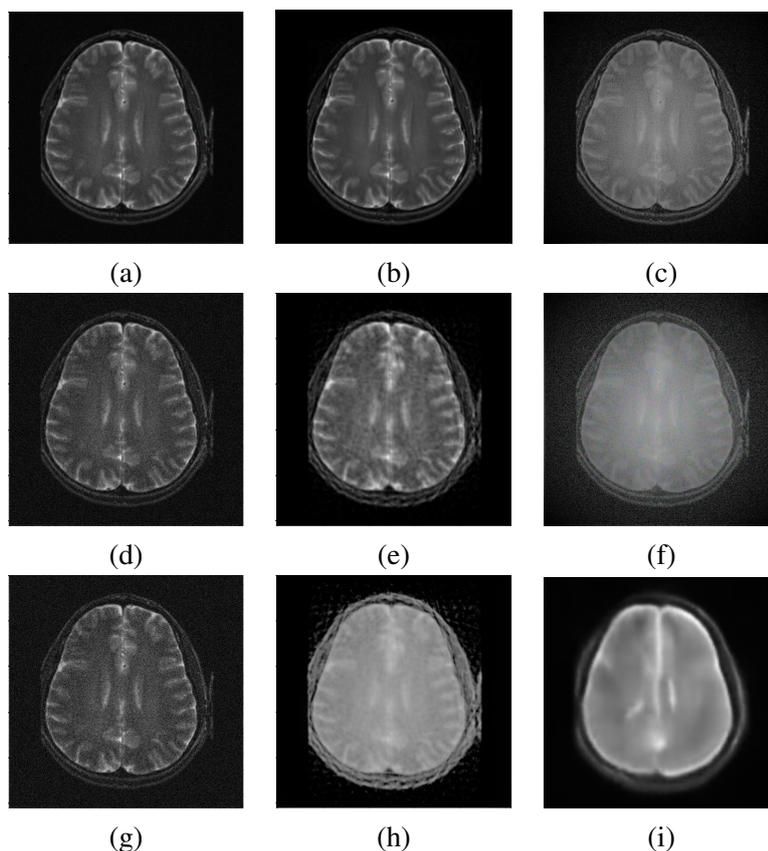


Figura 4.2: Imagens corrompidas pelo ruído gaussiano branco com média zero e $\sigma = 0.01, 0.05, \text{ e } 0.1$ (a, d, g), imagens reconstruídas com 10.000 épocas (b, e, h), e as imagens reconstruídas com 20.000 épocas (c, f, i).

4.2.1 Caso 1 - Comparação Direta

As Figuras 4.3 e 4.4 mostram os resultados de reconstrução de amostras obtidas do conjunto de dados completo do teste, correspondendo cada linha aos resultados de uma amostra de conteúdo da ressonância magnética. A primeira coluna mostra as imagens originais das amostras retiradas do conjunto de dados. A segunda coluna mostra os resultados da reconstrução correspondente obtidos com a arquitetura U-net quando as entradas são, respectivamente, as imagens reconstruídas de L2, L1, e TV. Já na terceira coluna é representado a imagem diferença de cada caso. A imagem diferença é dada pela subtração da imagem original das imagens reconstruídas. Os valores PSNR e SSIM para cada imagem reconstruída são mostrados nas tabelas 4.1 e 4.2. Notamos que os três modelos foram capazes de obter imagens com uma qualidade visual considerada satisfatória. Além disso, as imagens obtidas com os 3 modelos são semelhantes à imagem original e uma à outra. Isto é apoiado pelos valores comunicados da PSNR e SSIM.

Através das imagens obtidas, observamos que reconstrução L1 gerou imagens com uma riqueza de detalhes da fisiologia cerebral inferior as imagens geradas pela reconstrução L2 e TV. A diferença de densidade em áreas adjacentes de as imagens reconstruídas em relação as imagens originais foi considerável, ou seja, essa variação na densidade, resultou em um

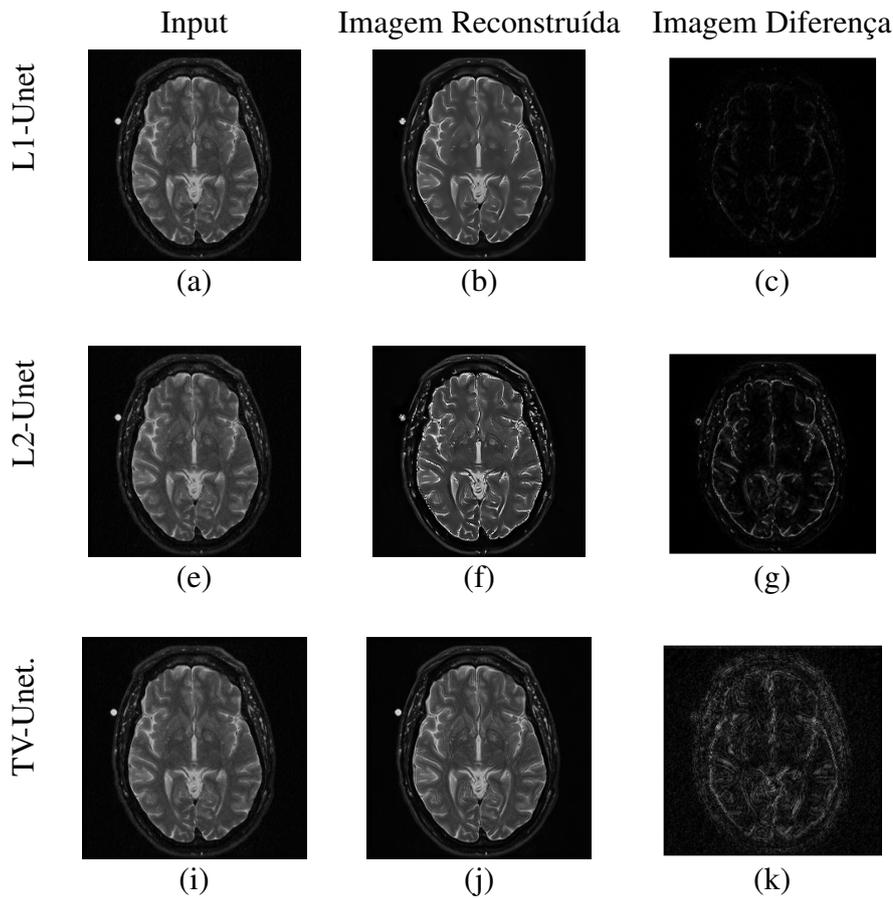


Figura 4.3: Resultado de reconstruções em comparação com a imagem original e com a imagem diferença para o caso direto. (a), (c) e (i) mostram a imagem original. (b), (f) e (j) representam as imagens reconstruídas com os casos diretos de L1, L2 e TV. As imagens (c), (g) e (k) são as imagens diferença para os três casos.

nível de contraste mais baixo em relação às outras reconstruções. A reconstrução L1 mostrou que quanto menor esta variação ou menor a diferença de densidade de áreas adjacentes, menor o contraste. Em alguns exemplos, as imagens reconstruídas eram até mesmo opacas. A reconstrução L2, por sua vez, apresentou um nível de acuracidade frente às características cerebrais abaixo do esperado, dificultando a determinação de grande parte das ranhuras a olhos vistos, isto é, o fator de qualidade que falhou na reconstrução com L2 foi a nitidez. Neste caso, o incongruente é que a reconstrução com L2 poderia esconder lesões, microfraturas, pequenos tumores e outros sintomas importantes para o diagnóstico, indo contra a proposta do método. Com a reconstrução baseada no modelo TV, a imagem perdeu precisão e sensibilidade. As falhas características na utilização do TV são evidentes. O contraste entre as intensidades estava presente, mas mesmo assim, era quase impossível perceber as ranhuras devido à baixa resolução da imagem gerada. Contudo, a maioria das imagens geradas através da reconstrução com TV tiveram um índice de brilho considerável.

Entretanto, é possível notar em algumas imagens a presença de uma degradação que não agregou positivamente: o ruído. O ruído na imagem de ressonância magnética é uma variação aleatória da densidade de fundo da imagem. Esta variação pode dar às imagens uma apa-

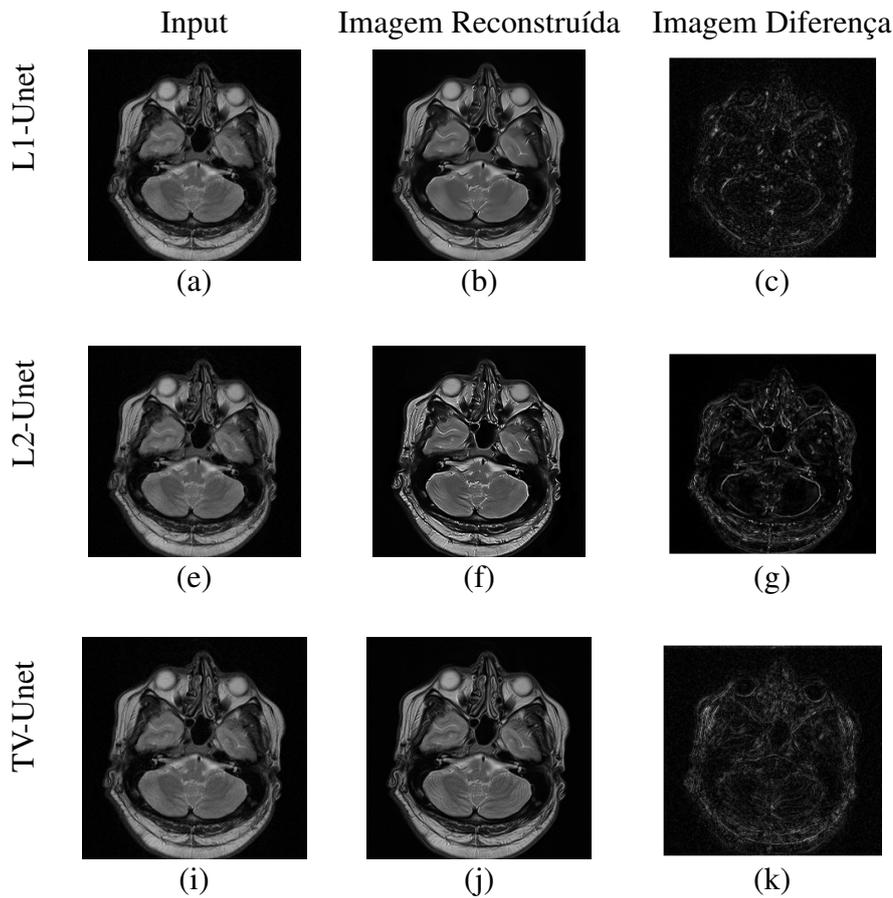


Figura 4.4: Resultado de reconstruções em comparação com a imagem original e com a imagem diferença para o caso direto. (a), (c) e (i) mostram a imagem original. (b), (f) e (j) representam as imagens reconstruídas com os casos diretos de L1, L2 e TV. As imagens (c), (g) e (k) são as imagens diferença para os três casos.

rência granulada ou texturizada, mais ou menos como aqueles aparelhos televisivos antigos sem sinal. Essa variação representa informação indesejável, pois não agrega nenhum valor para o diagnóstico. A reconstrução L1 foi a mais suscetível a este ruído, seguido pela reconstrução TV e, enfim, a reconstrução L2. Ainda sim, os resultados obtidos no caso direto foram considerados com grau alto de excelência, pois as falhas encontradas não afetaram vigorosamente a reconstrução das imagens originais. Todas as imagens reconstruídas conseguiram ser geradas sem distorção, ou seja, mantiveram o formato e estruturas das imagens originais com alta performance. Mais especificamente, em termos de métrica de qualidade, as U-Nets melhoraram significativamente a qualidade visual das imagens reconstruídas, especialmente para os métodos de reconstrução que inicialmente apresentavam mau desempenho.

Para visualização melhorada, as tabelas 4.1 e 4.2 apresentam os valores de PSNR e SSIM das saídas U-net em comparação com as imagens originais para todas as imagens dos Exemplo 1 e 2 (Figuras 4.3 e 4.4). A Tabela 4.3 mostra a média dos valores de PSNR e SSIM das saídas U-net em comparação com as imagens originais para todas as imagens do banco de dados utilizado. Em termos quantitativos, as reconstruções apresentaram um desempenho significativo. A reconstrução L2 apresentou uma média PSNR/SSIM de 25,8/0,655, mas após a

aplicação da L2-Unet treinada a média PSNR/SSIM aumentou para 30,9/0,867. A reconstrução L1 apresentou inicialmente uma média PSNR/SSIM de 29,2/0,799, que aumentou para 33,5/0,891 após a utilização da L1-unet. A reconstrução da TV teve uma PSNR/SSIM média de 31,5/0,832, que foi aumentada para 33,3/0,877 após a utilização da TV-Unet.

Em todos os 3 casos, os valores médios de SSIM se aproximam de 1, o que implica uma maior correspondência entre as imagens originais e as respectivas imagens reconstruídas. Com referência ainda aos valores SSIM, observamos que o desempenho da reconstrução L2 é ligeiramente inferior ao desempenho das reconstruções L1 e TV, que têm valores muito próximos. Ao analisarmos os valores de PSNR, observamos que a relação entre a máxima energia da imagem e o artefato que afeta sua representação estavam em intervalos menores que 34dB, o que é considerado um resultado consistente. Note que, o salto de maior qualidade, diante das métricas, ocorre para a reconstrução do L2. No entanto, como esperado, a TV-Unet e a L1-Unet apresentaram pontuações de qualidade final mais elevadas.

Tabela 4.1: Média PSNR e SSIM das saídas U-Net frente a imagem original para as reconstruções da Figura 4.3.

Métrica	Original	L1-unet	L2-unet	TV-unet
PSNR	27.9	29.2	28.7	29.7
SSIM	0.672	0.773	0.745	0.821

Tabela 4.2: Média PSNR e SSIM das saídas U-Net frente a imagem original para as reconstruções da Figura 4.4

Métrica	Original	L2-unet	L1-unet	TV-unet
PSNR	24.9	27.9	26.1	27.8
SSIM	0.579	0.716	0.681	0.732

Tabela 4.3: Média PSNR e SSIM das saídas U-Net para o caso reto frente todas as imagens originais presentes no banco de dados BIRN.

Métrica	Original	L2-unet	L1-unet	TV-unet
PSNR	26.4	30.9	33,5	33,3
SSIM	0.699	0.867	0,891	0,877

4.2.2 Caso 2 - Comparação Cruzada

Analisamos também como diferentes modelos treinados de U-net transformam as imagens geradas por diferentes métodos de otimização realizando um teste de comparação entre as imagens de reconstrução cruzada. Neste teste, a comparação cruzada é formada por todas as combinações de métodos de otimização L2, L1 e TV com entradas L2-Unet, L1-Unet e

TV-Unet, inclusive os casos de comparação direta já apresentados. Essa comparação cruzada gera um total de 9 combinações que são:

- L2 + L2-UNet;
- L2 + L1-UNet;
- L2 + TV-UNet;
- L1 + L1-UNet;
- L1 + L2-UNet;
- L1 + TV-UNet;
- TV + TV-UNet;
- TV + L2-UNet;
- TV + L1-UNet;

Notamos que os melhores resultados ocorrem para tipos de correspondência direta como, por exemplo, reconstrução de otimização L2 + L2-Unet. Os métodos baseados em L1 e TV apresentaram melhores resultados do que os baseados em L2.

Nas Figuras 4.5, 4.6 e 4.7 podemos ver um exemplo de imagem reconstruída com os casos cruzados de otimização e suas respectivas imagens diferença. Cada imagem tem suas reconstruções apresentadas na figura. As U-nets foram treinadas para remover o tipo de degradação apresentada nas imagens geradas pela reconstrução da otimização correspondente. A Tabela 4.4 apresenta os valores médios de PSNR e SSIM das saídas U-Net para o caso cruzado.

Tabela 4.4: Média PSNR e SSIM das saídas U-Net para o caso cruzado frente todas as imagens originais do exemplos das figuras 4.5, 4.6 e 4.7.

	PSNR				SSIM			
	Orig.	L2-unet	L1-unet	TV-unet	Orig.	L2-unet	L1-unet	TV-unet
Original	27.4	26.6	29.0	30.7	0.603	0.636	0.705	0.793
L2 rec.	26.8	31.4	28.1	27.1	0.698	0.834	0.743	0.717
L1 rec.	32.0	26.4	32.3	29.5	0.785	0.781	0.848	0.786
TV rec.	26.8	23.7	29.9	32.7	0.813	0.741	0.815	0.851

A utilização da U-net para diferentes tipos de otimização gerou resultados intrigantes. Por exemplo, a L2-Unet é capaz de gerar imagens com um nível de nitidez abaixo do esperado. Devemos isto ao fato de a combinação L2+L2-Unet ter a capacidade de distinção entre elementos inferior. Observamos também que L1-Unet e TV-Unet geram imagens com resolução inferior, isto é, uma quantidade de detalhes menor, quando comparadas com a resolução das imagens originais quando a entrada é a reconstrução de L2 (Figura 4.6 (b)),

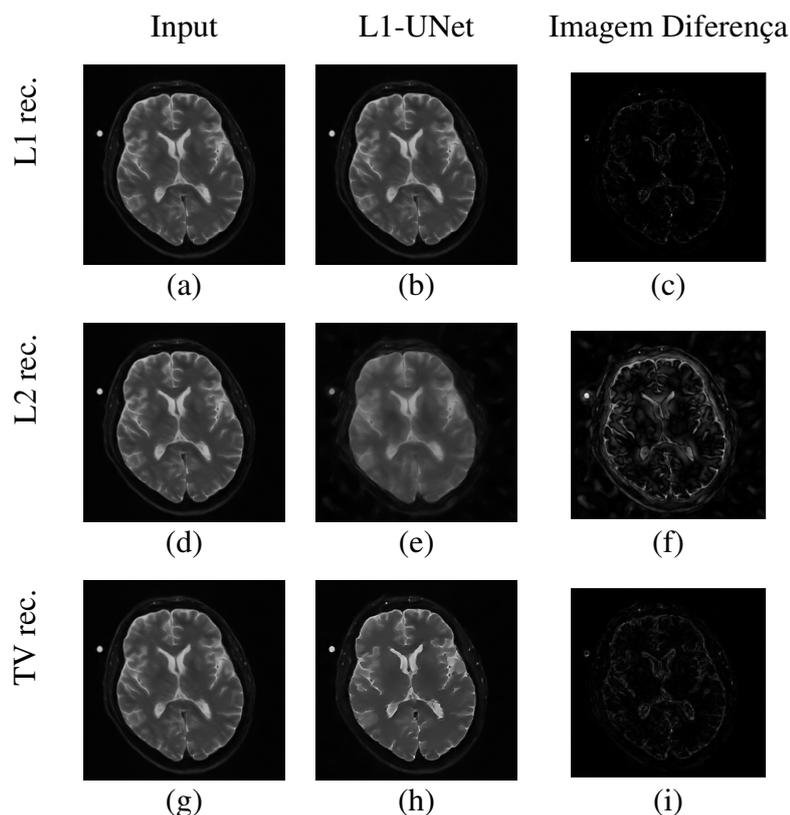


Figura 4.5: Resultado de reconstruções para os modelos de reconstrução do caso cruzado em comparação com as imagens de entrada quando o caso cruzado envolve a reconstrução L1-Unet. As imagens (a), (d) e (g) mostram a imagem original. (b), (e) e (h) representam as imagens reconstruídas com os casos cruzados de L1-Unet e (c), (f) e (i) são as imagens diferenças para os três casos.

(h)), ao contrário da L2-Unet. Contudo, os principais detalhes da fisiologia cerebrais são visualmente perceptíveis. Outro exemplo é que a L2-Unet gerou imagens com estruturas cerebrais que não eram apresentadas na imagem original, como as ranhuras encontradas nas imagens finais da combinação TV + L2-Unet na Figura 4.7 (e). Isso se deve pois as imagens de entrada regularizadas com TV possui uma característica visual conhecida como ‘estética de pintura’, onde as imagens parecem ter sido pintadas ao invés de processadas digitalmente.

As imagens de entrada para os 3 modelos U-net são mostradas na primeira coluna das figuras 4.5, 4.6 e 4.7. Notamos que os melhores resultados ocorrem quando o tipo de modelo aplicado corresponde ao tipo de reconstrução. As imagens resultantes tiveram um resultado satisfatório em termos de reconstrução. As semelhanças com as imagens originais podem ser observadas na diagonal principal das imagens.

Realizando uma análise visual das imagens obtidas no processo de reconstrução no caso cruzado e considerando a percepção cromática da imagem em preto e branco, podemos avaliar a sensação cromática, ou seja, as impressões cromáticas físicas da imagem. Observamos que, em termos do fator brilho (descriptor subjetivo que incorpora a noção acromática de intensidade), as imagens reconstruídas com L1-Unet na figura 4.5 (b), (e) e (h) tiveram o pior desempenho, ao passo que as imagens reconstruídas com TV-Unet na figura 4.7 (b), (e) e

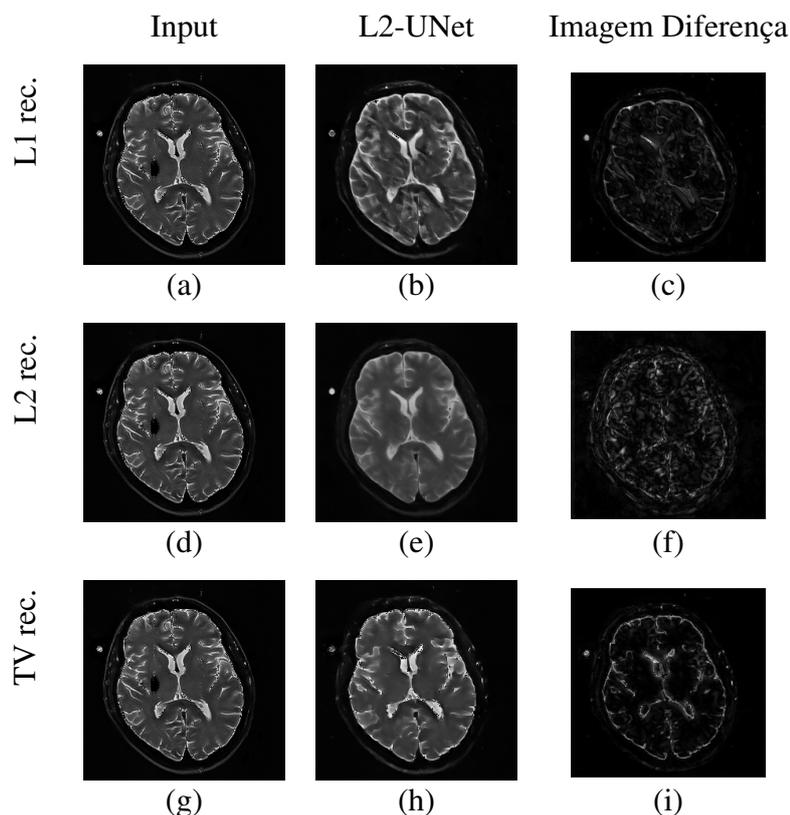


Figura 4.6: Resultado de reconstruções para os modelos de reconstrução do caso cruzado em comparação com as imagens de entrada quando o caso cruzado envolve a reconstrução L2-UNet. As imagens (a), (d) e (g) mostram a imagem original. (b), (e) e (h) representam as imagens reconstruídas com os casos cruzados de L2-UNet e (c), (f) e (i) são as imagens diferenças para os três casos.

(h) o melhor, passando uma percepção visual de que são mais intensas. O primeiro evento de cor (percepção de elementos componentes da imagem em preto e branco através do grau de contraste entre o branco, o preto e os cinzas, gerando a percepção da textura), também pode ser analisado nas imagens. A percepção da contraste nas reconstruções cruzadas com L2-U-Net na figura 4.6 - (b), (e) e (h) - e TV-U-Net na figura 4.7 - (d), (e), (h) - tiveram resultados muito próximos, levando a um reconhecimento de objetos diversos e de texturas como ranhuras, formas e estruturas do cérebro com maior facilidade, tornando evidente a estrutura da forma e gerando significado rapidamente, apenas com os recursos cromáticos do branco, do preto e dos cinzas e através dos contrastes entre eles. Mais uma vez, as imagens (b), (e) e (h) da figura 4.5 tiveram resultados abaixo das outras imagens reconstruídas.

Ainda acerca dos fatores de qualidade das imagens reconstruídas, outras análises podem ser feitas. É visualmente notável que as reconstruções do caso cruzado tiveram, como já mencionado, sucesso em manter os detalhes da imagem original de modo geral, ainda que não totalmente. Contudo, o método gerou melhores resultados com as reconstruções cruzadas com L2-U-Net, apresentadas na Figura 4.6 (b), (e) e (h). Nessas imagens reconstruídas, foi possível manter a nitidez da linhas finas e das bordas de tecidos ou estruturas visíveis na imagem. As imagens reconstruídas a partir da combinação L1+U-Net, apresentadas na

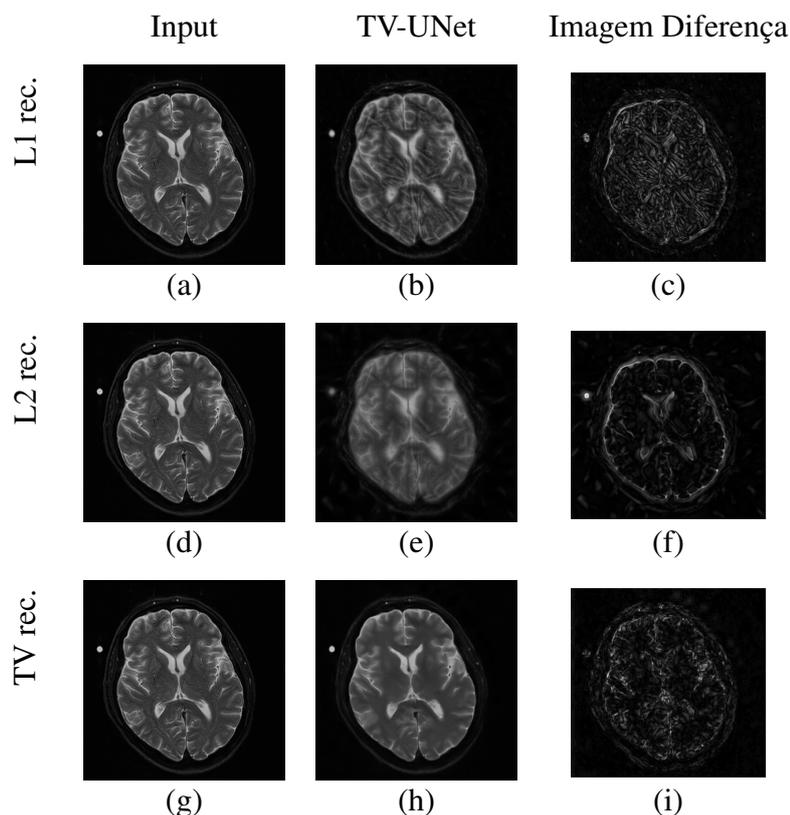


Figura 4.7: Resultado de reconstruções para os modelos de reconstrução do caso cruzado em comparação com as imagens de entrada quando o caso cruzado envolve a reconstrução TV-Unet. As imagens (a), (d) e (g) mostram a imagem original. (b), (e) e (h) representam as imagens reconstruídas com os casos cruzados de TV-Unet e (c), (f) e (i) são as imagens diferenças para os três casos.

Figura 4.5 (b), (e), (h), mostram um alto grau de borramento, ou seja, ausência de nitidez, passando a sensação de serem opacas. As imagens reconstruídas pela combinação TV-U-Net apresentaram, em termos de nitidez, um efeito de pintura, como se as imagens tivessem sido reproduzidas em uma tela. Em suma, em termos de distorção, as imagens reconstruídas obtiveram resultados de qualidade satisfatórios. A Tabela 4.5 apresenta os valores médios de PSNR e SSIM das saídas U-Net para o caso cruzado frente todas as imagens do banco de dados BIRN.

Tabela 4.5: Média PSNR e SSIM das saídas U-Net para o caso cruzado frente todas as imagens originais presentes no banco de dados BIRN.

	PSNR				SSIM			
	Orig.	L2-unet	L1-unet	TV-unet	Orig.	L2-unet	L1-unet	TV-unet
Original	26.5	24.3	27.9	30.2	0.599	0.614	0.718	0.768
L2 rec.	28.0	32.1	28.7	28.8	0.675	0.865	0.755	0.764
L1 rec.	31.2	26.7	33.5	31.4	0.799	0.793	0.891	0.837
TV rec.	31.5	24.6	29.8	33.3	0.832	0.786	0.854	0.877

Os gráficos boxplot possuem componentes que auxiliam a disposição das medidas medidas de estatísticas tais como: valor mínimo, valor máximo, primeiro quartil, segundo quartil

(também chamado de mediana) e o terceiro quartil. A Figura 4.8 mostra os gráficos boxplot ou diagramas de caixa das métricas PSNR e SSIM, respectivamente, para todas as amostras do conjunto de teste. Um boxplot permite visualizar a distribuição dos valores dos dados, fornecendo assim um meio complementar para desenvolver uma perspectiva sobre o caráter dos dados.

Analisando, primeiramente, o boxplot da Figura 4.8(a), para os valores de PSNR, é possível observar que os casos de comparação cruzada L2+L2-Unet, TV+TV-Unet e L2+L1-Unet obtiveram os valores simétricos de mediana distribuídos, pois a linha do segundo quartil (ou linha da mediana) destes casos está na metade do bloco. Isso demonstra que o valor da mediana destes casos é aproximadamente igual ao valor médio deles. Essa simetria é considerada um bom resultado visto que implicam que o modelo foi coeso nas decisões tomadas por ele, isto é, a análise dos dados de entrada foi devidamente considerada. O modelo aparentou ser capaz de reconstruir as imagens satisfatoriamente com qualquer forma de otimização. Isso pode ser útil caso as imagens de entrada do modelo sejam provenientes de outro banco de dados, por exemplo, pois o modelo mostrou, através da simetria e igualdade entre as reconstruções, que o processo de regularização surtiu o efeito esperado, isto é, o modelo aparenta ser resiliente a mudança dos dados que podem lhe são apresentados. Este fato também é observado com a entrada TV.

Os demais casos apresentaram assimetria positiva, isto é, existe uma maior concentração de valores na zona de valores mais reduzidos da amostra, já que a linha do segundo quartil é próxima da linha do primeiro quartil. Isso demonstra que o valor da mediana destes casos é menor que o valor médio deles. Em suma, embora não tão simétricos como os casos já citados, os demais casos apresentam valores de PSNR plausíveis. Com relação a dispersão dos dados, dada pela diferença entre o terceiro quartil e o primeiro quartil, é perceptível que a maioria dos casos apresentou uma dispersão razoável a exceção dos casos L1 e L1+L1-Unet pois os seus blocos são mais longos, atribuindo uma maior variabilidade para os valores PSNR. Isso se dá ao fato de que a dispersão dos dados pode ser representada pelo intervalo interquartil: uma estatística mais robusta para medir variabilidade uma vez que não sofre influência de valores extremos. Contudo uma observação pertinente é que todos os casos apresentaram os chamados outliers: valores discrepantes que extrapolaram o limite superior do blocos, podendo ser produto de um erro de observação ou de arredondamento. As maiores incidências de outliers aconteceram em casos que envolviam o L2 (entrada L2, L2+L2-Unet, L2+L1-Unet e L2+TV-Unet).

Analisando o gráfico boxplot da métrica SSIM apresentado na Figura 4.8(b) é possível perceber que existe uma variação menor nos valores de SSIM entre as reconstruções quando comparados com valores do gráfico (a). A maioria das comparações apresentou a linha do segundo quartil no meio do bloco. Em suma, a maioria dos modelos (mais precisamente, 6 dos 9 modelos) obtiveram os valores de mediana distribuídos de forma mais uniforme, o que indica que os valores de mediana das reconstruções são muito próximos dos seus valores médios. Para os casos onde a linha da mediana não se localiza ao centro, é observável a

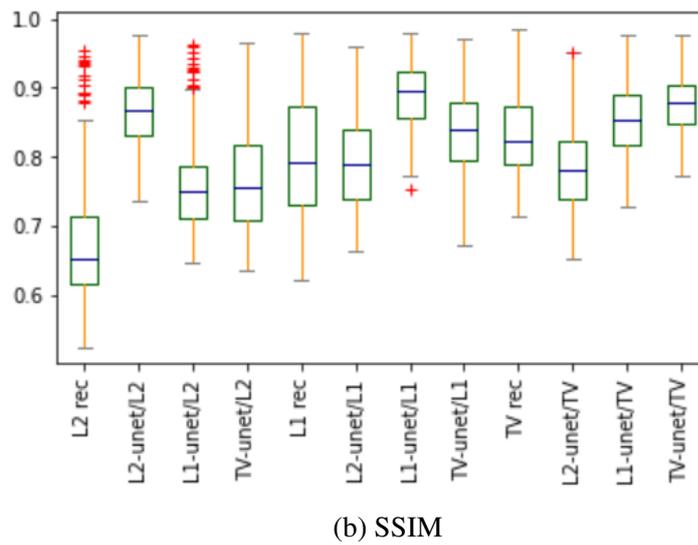
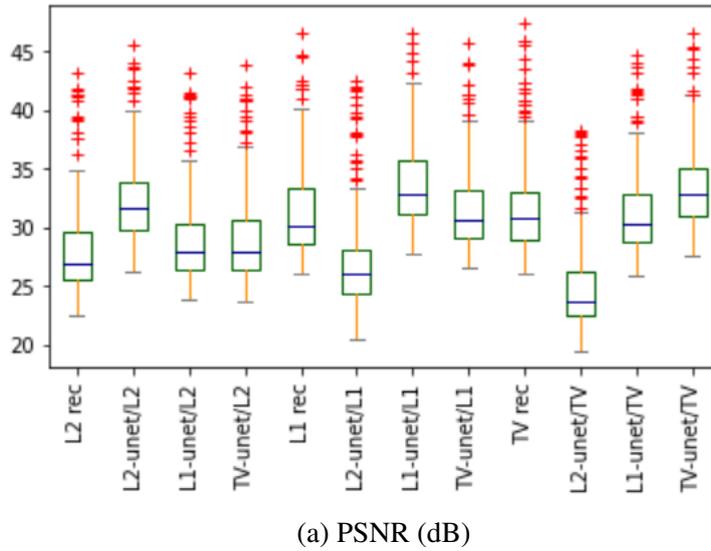


Figura 4.8: Gráficos boxplot para os valores de PSNR (dB) e SSIM para o caso cruzado.

assimetria positiva, apresentando valores de SSIM com intervalos pequenos. Com relação a dispersão dos valores de SSIM, os casos L2, TV+L2-Unet, L1 e L2+L1-Unet são os que apresentaram maior índice, com destaque para o caso L1, que possui um bloco duas vezes maior que o bloco da reconstrução TV+TV-Unet, por exemplo, acarretando uma maior amplitude, ou seja, a diferença entre o seu valor máximo e seu valor mínimo é alta. Isso implica que os valores de SSIM das imagens reconstruídas usando o modelo L1 apresentaram maior variabilidade. Os demais casos salientaram valores de dispersão modestos.

4.3 Considerações

Este capítulo apresentou os resultados obtidos com o algoritmo proposto baseado em aprendizado profundo e redes neurais artificiais. A análise foi feita com o banco de dados BIRN de 4.020 imagens de ressonância magnética do cérebro humano, todas em escalas de

cinza. As reconstruções foram apresentadas tanto para o conjunto de treino, quanto para o conjunto de imagens de testes. Foram realizados testes de comparação diretos e cruzados. Os valores de PSNR e o SSIM foram usados como métricas de qualidade para comparação das imagens. As reconstruções tiveram grande semelhança com as reconstruções padrão-ouro, com valores de SSIM próximos a entre 0.5 e 0.9 e valores PSNR entre 30dB e 40dB. Isto indica que o método proposto produz imagens reconstruídas com boa qualidade, mantendo as características da estrutura do tecido cerebral tais como ranhuras ou ambientes químicos de gordura e água. As imagens reconstruídas apresentaram também características semelhantes as imagens originais referentes a seu posicionamento na imagem, com os tecidos humanos na mesma posição e com tamanho similar as imagens de entrada. Isto é, a estrutura do tecido cerebral das imagens de saída estava na mesma direção, sentido e dimensão da estrutura da imagem de entrada.

Outra análise realizada é da função de perda ao longo das épocas, pois embora a U-Net seja capaz de reconstruir bem as imagens após 300 épocas, ainda existem alguns detalhes divergentes entre a imagem de saída e a imagem de entrada. Durante o processo de treino do modelo, o valor médio da função de perda ao longo de cada época encontrado foi de 0.005234 e durante o processo de validação, 0.004397. O gráfico da Figura 4.9 apresenta o comportamento destes valores da função perda médios ao longo das épocas. A linha na cor vermelha corresponde ao processo de treino do modelo, enquanto que a linha verde corresponde ao processo de validação. Observamos um decrescimento exponencialmente característico, como as perdas de reconstrução diminuindo ao longo das épocas para ambos os processos. Isto corrobora a eficiência do modelo proposto em reconstruir as imagens de ressonância magnética mantendo detalhes cruciais da fisiologia, pois a perda de informações durante o processo de reconstrução das imagens foi branda. Por outro lado, a curva de erros correspondente ao processo de treino possui um decaimento mais abrupto. Isso diz que, durante o processo de treino, a perda foi mais frequente e com maior intensidade e o processo de validação teve valores de perda mais próximos uns dos outros sem grandes variações.

Uma outra técnica de redução dimensional que favorece a visualização dos dados e que é utilizada para avaliar o desempenho de um modelo de reconstrução é a t-SNE. A t-SNE permite visualizar dados de alta dimensão dando a cada ponto uma localização em um mapa bidimensional ou tridimensional. Segundo o artigo *Visualizing Data using t-SNE*, dos autores Laurens van der Maaten e Geoffrey Hinton, a t-SNE é uma redução de dimensionalidade não linear técnica adequada para incorporar dados de alta dimensão para visualização em um espaço de baixa dimensão de duas ou três dimensões, baseada no trabalho *Stochastic Neighbor Embedding* originalmente desenvolvido por Sam Roweis e Geoffrey Hinton [91]. A t-SNE foi usada para testar se o modelo proposto tinha sucesso em agrupar os pixels das imagens por semelhança. Isso garantiria que as imagens tivessem uma reconstrução de maior qualidade, uma vez que o modelo saberia identificar a posição dos pixels e as características que os faziam semelhantes. Dessa forma, o método seria capaz de reconstruir

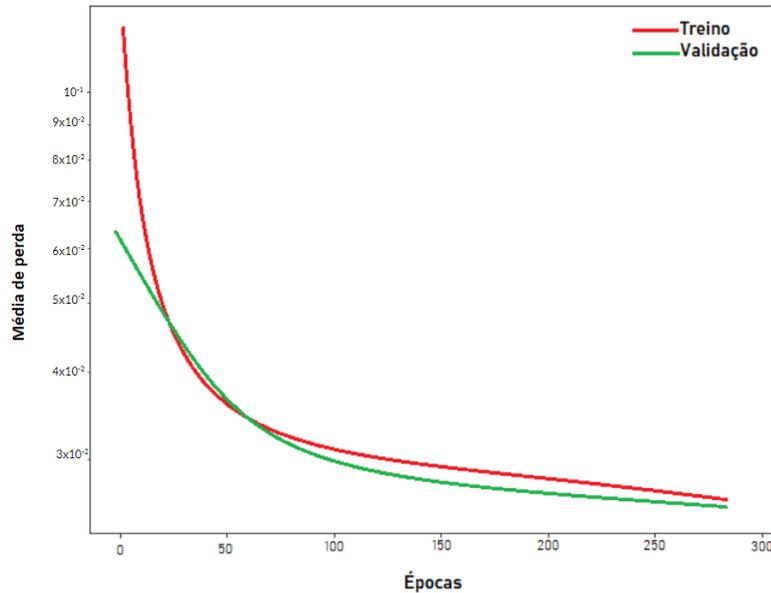


Figura 4.9: Gráfico representando a média da função de perda nos processo de treinamento e validação pelo número de épocas.

as imagens com sucesso, pois saberia identificar em quais posições era mais plausível colocar um determinado pixel. Fazendo uma analogia simples, imagine que quiséssemos transformar a imagem original em um quebra-cabeças utilizando a reconstrução para processar a imagem. Para tanto, a imagem original teria que ser representada por seus pixels. Os pixels seriam as peças do quebra-cabeças. A t-SNE busca saber se o modelo proposto era capaz de identificar quais peças (pixels) encaixavam perfeitamente ou quais peças (pixels) estavam os cantos da imagem, por exemplo.

Como as imagens utilizadas nos testes do método proposto eram de tamanho 512X512, cada imagem possuía cerca de 262.144 pixels. Assim, a t-SNE permite uma visualização dos dados de forma mais clara. O gráfico da Figura 4.10, foi gerado através de um código desenvolvido por Roberto Stelling [92]. Os pixels foram divididos em 10 novas classes devido a alta quantidade de pixels em cada imagem. Podemos ver que ele distingue um dígito do outro, embora existam algumas exceções com pontos que se enquadram em outras categorias. A t-SNE construiu um mapa em que a separação entre as classes é quase perfeita. Uma inspeção mais detalhada do mapa t-SNE revela que grande parte da estrutura local dos dados foi capturada com sucesso, revelando a classe natural dos 262.144 pixels presentes em cada uma das imagens reconstruídas e os distribuindo de forma a representa-los por suas características comuns, dando ao modelo proposto comprovação de sua competência na tratativa das imagens, ou seja, o modelo proposto conseguiu identificar cada um dos pixels das imagens durante o processo de reconstrução e teve sucesso em agrupa-los por semelhança quando as imagens estavam sendo reconstruídas, propiciando uma reconstrução sucinta. Usando mais uma vez a analogia do quebra-cabeças, a t-SNE mostrou que o método desenvolvido conseguiu, dadas todas as peças (pixels) disponíveis, montar o quebra-cabeças (imagem reconstruída) com sucesso, independente de qual imagem do banco de dados estivesse sendo

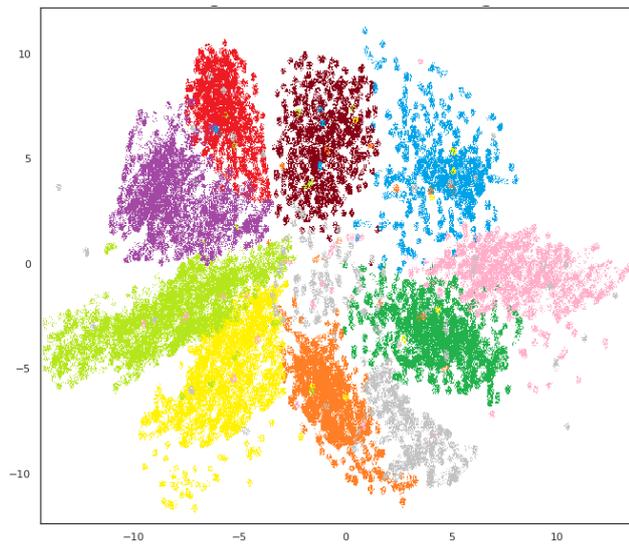


Figura 4.10: Gráfico da métrica T-SNE com 10 classes de agrupamento por semelhança para os pixels das imagens de tamanho 512x512.

reconstruída.

Decorre da discussão anterior que existe uma conexão entre a literatura disponível e o que a experiência prática mostra. Os resultados dos testes apresentaram um modelo que teve sucesso em recuperar as imagens 512x512 (que são consideradas de tamanho grande) de ressonância magnética, preservando a maior parte das características das imagens originais. Os resultados obtidos mostram, ainda, que o modelo conseguiu incorporar os três tipos de regularização testados, tendo apresentado resultados plausíveis e de qualidade visual perceptível para todos eles.

Capítulo 5

Análise Final

A reconstrução de imagens de ressonância magnética aplicando técnicas de aprendizado profundo e ferramentas de CS é um campo promissor que pode contribuir para um diagnóstico médico mais preciso. Nesta dissertação, procuramos realizar uma revisão sucinta dos fundamentos teóricos que envolvem a formação de imagens de ressonância e a sua degradação de forma a apresentar um modelo que pudesse atender a demanda de refinamento na reconstrução de imagens de ressonância magnética. Também era de interesse apresentar um modelo análogo com o que vem sendo desenvolvido pela comunidade acadêmica frente ao uso de redes neurais convolucionais. Esta dissertação trabalhou com dois tipos de arquiteturas de redes neurais: autoencoders e U-Net como modelos codificador-decodificador, mostrando que, enquanto os autoencoders não alcançaram resultados adequados para a reconstrução das IRMs, a U-Net apresentou resultados satisfatórios. Um fator importante nesta proposta é o uso da regularização como ferramenta de otimização das imagens de ressonância. Tal configuração aumentou a probabilidade das imagens de saída terem uma boa qualidade.

Mostramos, através de testes com imagens de ressonância magnética do cérebro humano, que os autoencoders não são indicados para a reconstrução de imagens pois, em seu processo de manipulação das imagens, muitos detalhes importantes da fisiologia do tecido humano são perdidos, acarretando possíveis falhas na detecção de um tumor, por exemplo. Já no caso da U-Net, utilizando o mesmo conjunto de imagens, obtivemos resultados melhores. A reconstrução das imagens gerou uma saída com qualidade visual satisfatória, preservando as principais características e detalhes do tecido humano. Os testes foram consistentes não só para o caso de reconstrução direta como também para os casos de reconstrução cruzada, salientados pelos processos de regularização L1, L2 e TV e definição de um padrão de amostragem adequado ao modelo proposto. A U-Net foi capaz de obter um alto nível de precisão para um problema relativamente complexo, mesmo para uma taxa de amostragem muito pequena. Especialmente, conseguimos melhorar a reconstrução L2 rec+L2-unet de trabalhos anteriores.

O resultados foram comprovados não só pela inspeção visual, mas também pelo uso de

métricas de avaliação de desempenho, em especial, a PSNR e o SSIM. Nos valores PSNR, foi observado que a reconstrução de imagens apresentou simetria e baixos níveis de dispersão dos dados. Embora os valores de PSNR das reconstruções testadas tenham apresentado maior distinção, essas diferenças não afetaram de forma grave o desempenho do modelo. Os valores de SSIM, por outro lado, apresentaram resultados ainda mais satisfatórios, uma vez que quase não foram encontradas discrepâncias na avaliação do SSIM. Por conseguinte, o método demonstra ter potencial para reconstrução de imagens de ressonância magnética não só do cérebro humano, mas também imagens referentes de outras partes da anatomia humana.

Além do uso da inspeção visual e das métricas de avaliação de desempenho, a veracidade do modelo proposto foi sustentado pelos valores encontrados para a função de perda dos dados ao longo das épocas constituídos de uma decrescimento exponencial, sendo muito próximos de zero. Este decrescimento mostra que o modelo obteve sucesso em reconstruir as imagens conservando suas características visuais mais significativas. O uso da t-SNE também foi de grande valia para a avaliação do método proposto porque indicou que os dados reconstruídos apresentaram capacidade de serem incorporados em espaços de de baixa dimensão e semelhança consolidada, uma vez que os dados foram agrupados por equivalência com sucesso pela t-SNE. O manejo destas funções encerra a análise do modelo proposto não apenas confirmando os valores numéricos da PSNR e SSIM mas também confirmando que a escolha da U-Net para a reconstrução de imagens de ressonância magnética foi acertada e adequada para o problema exposto.

5.1 Estudos Futuros

Ao realizarmos os testes com a configuração básica proposta, o padrão de amostragem escolhido foi de 20 linhas radiais. Diante disso, um estudo que pode ser realizado é a escolha de padrões de amostragem com números maiores de linhas radiais como 40, 60, 80 ou 100, por exemplo. Este estudo pode ser feito a partir das interações entre os processos de regularização L1, L2 e TV e novos padrões de amostragem. Essa análise futura pode obter melhores resultados para o modelo L2 e uma diminuição do efeito “pintura” para o modelo TV, levando em consideração a precisão. Outros estudos podem ser feitos escolhendo outros conjuntos de imagens de ressonância magnética que contenham outros órgãos ou tecidos humanos a fim de testar a resiliência do modelo quando submetido a características fisiológicas diferentes. Uma outra contribuição que pode gerar um estudo futuro consistente é experimentar outras arquiteturas de redes neurais artificiais para a reconstrução das imagens.

Referências Bibliográficas

- [1] F. Posses, “Ressonância magnética (rmn): o que é, tipos e como funciona,” in *Radiologia - STAR Telerradiologia*, p. 1, 2022.
- [2] H. Jia, W. Cai, H. Huang, and Y. Xia, “H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries-6th International Workshop, Springer*, pp. 58–68, 2021.
- [3] R. R. Borges, K. C. Iarosz, A. M. Batista, I. L. Caldas, F. S. Borges, and E. L. Lameu, “Sincronização de disparos em redes neuronais com plasticidade sináptica,” *Revista Brasileira de Ensino de Física*, vol. 37, pp. 2310–1, 2015.
- [4] S. Favoretto *et al.*, “Máquinas de aprendizado extremo aplicadas à identificação de pessoas através de eletrocardiograma (ecg),” Master’s thesis, Universidade Tecnológica Federal do Paraná, 2016.
- [5] H. Lee and Y.-P. P. Chen, “Image based computer aided diagnosis system for cancer detection,” *Expert Systems with Applications*, vol. 42, no. 12, pp. 5356–5365, 2015.
- [6] M. A. Ponti and G. B. P. Da Costa, “Como funciona o deep learning,” *arXiv preprint arXiv:1806.07908*, 2018.
- [7] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, “Stacked convolutional auto-encoders for hierarchical feature extraction,” in *International conference on artificial neural networks*, pp. 52–59, Springer, 2011.
- [8] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [9] F. Hibraj, M. Pelillo, S. Salzo, and M. Pontil, “Efficient tensor kernel methods for sparse regression,” *arXiv preprint arXiv:2003.10482*, 2020.
- [10] K. Krupa and M. Bekiesińska-Figatowska, “Artifacts in magnetic resonance imaging,” *Polish journal of radiology*, vol. 80, p. 93, 2015.

- [11] L. Schettini, “Estudo dos artefatos da imagem diagnóstica gerada em aparelhos ressonância magnética,”
- [12] S. A. Mirowitz, “Mr imaging artifacts: Challenges and solutions,” *Magnetic resonance imaging clinics of North America*, vol. 7, no. 4, pp. 717–732, 1999.
- [13] A. Thomaz, F. Mylene, M. Cristiano, J. Lima, A. Krylov, and Y. Ding, “Undersampled magnetic resonance image reconstructions based on a combination of u-nets and l1, l2, and tv optimizations,” *IEEE International Conference on Imaging Systems and Techniques*, vol. 1, pp. 1–6, 2022.
- [14] J. C. Ye, “Compressed sensing MRI: a review from signal processing perspective,” *BMC Biomed. Engineering*, vol. 1, no. 1, pp. 1–17, 2019.
- [15] L. Feng, “Golden-angle radial mri: Basics, advances, and applications,” *Journal of Magnetic Resonance Imaging*, 2022.
- [16] M. de Oliveira Santos, “Estimativa 2018: incidência de câncer no brasil,” *Revista Brasileira de Cancerologia, INCA*, vol. 64, no. 1, pp. 119–120, 2018.
- [17] C. Thomson and D. Forman, “Cancer survival in england and the influence of early diagnosis: what can we learn from recent eurocare results?,” *British journal of cancer*, vol. 101, no. 2, pp. S102–S109, 2009.
- [18] W. H. Organization *et al.*, *World Cancer Report: Cancer Research for Cancer Development*. IARC, 2020.
- [19] J. H. Thrall, X. Li, Q. Li, C. Cruz, S. Do, K. Dreyer, and J. Brink, “Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 504–508, 2018.
- [20] R. N. Bracewell, “Strip integration in radioastronomy,” *Australian Journal of Physics*, vol. 9, no. 2, pp. 198–217, 1956.
- [21] NobelPrize.org., “Godfrey n. hounsfield – facts,” in *Nobel Prize Outreach AB 2022.*, p. 1, 2022.
- [22] M. Yuen and H. R. Wu, “Reconstruction artifacts in digital video compression,” in *Proc. SPIE Digital Video Compression: Algorithms and Technologies*, vol. 2419, (San Jose, CA), pp. 455–65, SPIE, 1995.
- [23] D. C. Peters, F. R. Korosec, T. M. Grist, W. F. Block, J. E. Holden, K. K. Vigen, and C. A. Mistretta, “Undersampled projection reconstruction applied to mr angiography,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 43, no. 1, pp. 91–101, 2000.

- [24] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, “Deep learning for undersampled MRI reconstruction,” *Physics in Medicine & Biology*, vol. 63, no. 13, p. 135007, 2018.
- [25] C. B. Paschal and H. D. Morris, “K-space in the clinic,” *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 19, no. 2, pp. 145–159, 2004.
- [26] M. J. Ehrhardt and M. M. Betcke, “Multicontrast MRI reconstruction with structure-guided total variation,” *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 1084–1106, 2016.
- [27] K. T. Block, M. Uecker, and J. Frahm, “Undersampled radial mri with multiple coils. iterative image reconstruction using a total variation constraint,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086–1098, 2007.
- [28] S. Mizusawa, Y. Sei, R. Orihara, and A. Ohsuga, “Computed tomography image reconstruction using stacked U-Net,” *Computerized Medical Imaging and Graphics*, vol. 90, p. 101920, 2021.
- [29] Y. Chen, C.-B. Schönlieb, P. Liò, T. Leiner, P. L. Dragotti, G. Wang, D. Rueckert, D. Firmin, and G. Yang, “Ai-based reconstruction for fast mri—a systematic review and meta-analysis,” *Proceedings of the IEEE*, vol. 110, no. 2, pp. 224–245, 2022.
- [30] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated mri data,” *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
- [32] J. Zheng and L. Peng, “An autoencoder-based image reconstruction for electrical capacitance tomography,” *IEEE Sensors Journal*, vol. 18, no. 13, pp. 5464–5474, 2018.
- [33] E. Amaro Júnior and H. Yamashita, “Aspectos básicos de tomografia computadorizada e ressonância magnética,” *Brazilian Journal of Psychiatry*, vol. 23, pp. 2–3, 2001.
- [34] L. C. A. Madureira, C. S. Oliveira, C. Seixas, V. De Nardi, R. P. C. Araújo, and C. Alves, “Importância da imagem por ressonância magnética nos estudos dos processos interativos dos órgãos e sistemas,” *Revista de Ciências Médicas e Biológicas*, vol. 9, pp. 13–19, 2010.
- [35] R. H. Hunt and K. M. Thomas, “Magnetic resonance imaging methods in developmental science: a primer,” *Development and Psychopathology*, vol. 20, no. 4, pp. 1029–1051, 2008.

- [36] A. A. Mazzola, “Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional,” *Revista Brasileira de Física Médica*, vol. 3, no. 1, pp. 117–129, 2009.
- [37] S. M. Arcuri and P. K. McGuire, “Ressonância magnética funcional e sua contribuição para o estudo da cognição em esquizofrenia,” *Brazilian Journal of Psychiatry*, vol. 23, pp. 38–41, 2001.
- [38] A. C. Silva, *Algoritmos para Diagnóstico Assistido de nódulos Pulmonares Solitários EM Imagens de Tomografia Computadorizada*. PhD thesis, Pontifical Catholic University of Rio de Janeiro, Brazil, 2004.
- [39] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [40] T. Plevaya, R. Ravodin, and A. Filchenkov, “Skin Lesion Primary Morphology Classification With End-To-End Deep Learning Network,” in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 247–250, IEEE, feb 2019.
- [41] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [42] G. Zhao and M. Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.
- [43] M. L. Giger, “Machine learning in medical imaging,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, 2018.
- [44] T. Escovedo and A. Koshiyama, *Introdução a Data Science: Algoritmos de Machine Learning e métodos de análise*. Casa do Código, 2020.
- [45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [46] J. Gomes, “Reconstrução de imagens de tomografia por impedância elétrica usando máquinas de aprendizado extremo e algoritmos de retroprojeção,” pp. 1–110, Universidade Federal de Pernambuco, Departamento de Engenharia Biomédica, 2019.
- [47] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, “Deep learning in medical imaging: general overview,” *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017.
- [48] H. Matsumoto, Y. Terao, A. Yugeta, H. Fukuda, M. Emoto, T. Furubayashi, T. Okano, R. Hanajima, and Y. Ugawa, “Where do neurologists look when viewing brain CT

- images? an eye-tracking study involving stroke cases,” *PLoS ONE*, vol. 6, p. e28928, dec 2011.
- [49] V. Flovik, “How to use machine learning for anomaly detection and condition monitoring,” in *Concrete use case for machine learning and statistical analysis*, Towards Data Science Inc, 2018.
- [50] D. RAVI, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Yang, “Deep learning for health informatics,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 4–21, Jan 2017.
- [51] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [52] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, “Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network,” *BMC medical research methodology*, vol. 18, no. 1, p. 24, 2018.
- [53] J. Jiménez, M. Skalic, G. Martínez-Rosell, and G. De Fabritiis, “K deep: Protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks,” *Journal of chemical information and modeling*, vol. 58, no. 2, pp. 287–296, 2018.
- [54] J. Erly, M. Ervik, R. Dikshit, S. Eser, C. Mathers, *et al.*, “Cancer incidence and mortality worldwide: Iarc cancer base no. 11,” *GLOBOCAN 2012 v1*, 2012.
- [55] G. . R. F. Collaborators *et al.*, “Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the global burden of disease study 2015,” *Lancet (London, England)*, vol. 388, no. 10053, p. 1659, 2016.
- [56] J. E. R. de Queiroz and H. M. Gomes, “Introdução ao processamento digital de imagens,” *Rita*, vol. 13, no. 2, pp. 11–42, 2006.
- [57] A. Singh *et al.*, “Review of brain tumor detection from mri images,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 3997–4000, IEEE, 2016.
- [58] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [59] A. C. Chang, “Intelligence-based medicine,” *Artificial Intelligence and Human Cognition in Clinical Medicine and Healthcare*, pp. 397–412, 2020.

- [60] G. Zaharchuk, E. Gong, M. Wintermark, D. Rubin, and C. Langlotz, “Deep learning in neuroradiology,” *American Journal of Neuroradiology*, vol. 39, no. 10, pp. 1776–1784, 2018.
- [61] G. Y. G. Y. HOSAKI and D. F. RIBEIRO, “Deep learning: ensinando a aprender,” 2021.
- [62] D. R. Araújo, J. F. Martins-Filho, and C. J. Bastos-Filho, “Aprendizagem nao-supervisionada para avaliação de métricas topológicas de redes ópticas,”
- [63] R. N. Bryan, “Introduction to the science of medical imaging.,” in *Cambridge University Press*, pp. 1–20, Cambridge University, 2010.
- [64] L. Shao, D. Wu, and X. Li, “Learning deep and wide: A spectral method for learning deep networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 12, pp. 2303–2308, 2014.
- [65] W. Luo, J. Li, J. Yang, W. Xu, and J. Zhang, “Convolutional sparse autoencoders for image classification,” *IEEE transactions on neural networks and learning systems*, vol. 29, no. 7, pp. 3289–3294, 2017.
- [66] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [67] A. C. Spolti *et al.*, “Classificação de vias através de imagens aéreas usando deep learning,” 2018.
- [68] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, “Unet 3+: A full-scale connected u-net for medical image segmentation,” in *ICASSP 2020-2020 IEEE Intern. Conf. on Acoustics, Speech and Signal Proces.*, pp. 1055–1059, IEEE, 2020.
- [69] R. T. H. H. Q. Z. Y. I. X. H. Y.-W. C. H. Huang, L. Lin and J. Wu, “QUEST: Unet 3+: A full-scale connected u-net for medical image segmentation,” *IEEE Intern. Conf. on Acoustics, Speech and Signal Proces.*, vol. 1, no. 1055–1059, 2020.
- [70] A. Spolti, V. C. Guizilini, C. C. Mendes, M. D. Croce, A. R. de Geus, H. C. de Oliveira, A. R. Backes, and J. R. Souza, “Application of u-net and auto-encoder to the road/non-road classification of aerial imagery in urban environments.,” in *VISIGRAPP (4: VISAPP)*, pp. 607–614, 2020.
- [71] C. M. Verdun, “Compressive sensing,” *Federal University of Rio de Janeiro, Rio de Janeiro*, 2016.
- [72] A. L. Pilastri, “Compressive sensing in medical imaging,” *Programa Doutoral em Engenharia Informática. Faculdade de Engenharia da Universidade do Porto*, 2015.

- [73] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, “Compressive sensing: From theory to applications, a survey,” *Journal of Communications and networks*, vol. 15, no. 5, pp. 443–456, 2013.
- [74] R. d. F. Góis, *Investigação do uso de imagens mamográficas adquiridas em diferentes digitalizadores no desempenho do processamento de imagens para detecção de microcalcificações*. PhD thesis, Universidade de São Paulo, 2017.
- [75] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid mr imaging,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [76] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [77] C. G. Graff and E. Y. Sidky, “Compressive sensing in medical imaging,” *Applied optics*, vol. 54, no. 8, pp. C23–C44, 2015.
- [78] M. Fornasier and H. Rauhut, “Compressive sensing,” *Handbook of mathematical methods in imaging*, vol. 1, pp. 187–229, 2015.
- [79] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, vol. 2. Springer, 2010.
- [80] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [81] A. Hore and D. Ziou, “Image quality metrics: PSNR vs. SSIM,” in *2010 20th international conference on pattern recognition*, pp. 2366–2369, IEEE, 2010.
- [82] D. S. Turaga, C. Yingwei, and J. Caviedes, “No reference psnr estimation for compressed pictures,” in *Proc. IEEE International Conference on Image Processing*, vol. 3, pp. 61–4 vol, 2002.
- [83] M. Yousufi, M. Amir, U. Javed, M. Tayyib, S. Abdullah, H. Ullah, I. M. Qureshi, K. S. Alimgeer, M. W. Akram, and K. B. Khan, “Application of compressive sensing to ultrasound images: A review,” *BioMed Research International*, vol. 2019, 2019.
- [84] E. M. Bellon, E. M. Haacke, P. E. Coleman, D. C. Sacco, D. A. Steiger, and R. E. Gangarosa, “Mr artifacts: a review,” *American Journal of Roentgenology*, vol. 147, no. 6, pp. 1271–1281, 1986.
- [85] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, “f-anogan: Fast unsupervised anomaly detection with generative adversarial networks,” *Medical image analysis*, vol. 54, pp. 30–44, 2019.

- [86] L. Gondara, “Medical image denoising using convolutional denoising autoencoders,” in *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, pp. 241–246, IEEE, 2016.
- [87] K. G. Helmer, J. L. Ambite, J. Ames, *et al.*, “Enabling collaborative research using the biomedical informatics research network (BIRN),” *Journal of the American Medical Informatics Association*, vol. 18, no. 4, pp. 416–422, 2011.
- [88] E. Anello, “Convolutional autoencoder in pytorch on mnist dataset,” in *Data Series*, p. 1, 2021.
- [89] V. V. Estrela, H. A. Magalhães, and O. Saotome, “Total variation applications in computer vision,” in *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing*, pp. 41–64, IGI Global, 2016.
- [90] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [91] L. Van der Maaten and G. Hinton, “Visualizing data using t-sne.,” *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [92] R. Stelling, “Abrindo a caixa preta do t-sne,” in *Observable HQ*, p. 1, 2018.

Undersampled Magnetic Resonance Image Reconstructions Based on a Combination of U-Nets and L1, L2, and TV Optimizations

Aline A.S. Thomaz*, Jonathan A.S. Lima⁺,
Cristiano J. Miosso⁺, and Mylène C.Q. Farias*
Department of Electrical Engineering*
Biomedical Engineering Graduate Program⁺
University of Brasília (UnB)
Brasília, Brazil

Andrey S. Krylov
Faculty of Computational
Mathematics and Cybernetics,
Lomonosov Moscow State University
Moscow, Russia

Yong Ding
College of Information Science
and Electronic Engineering,
Zhejiang University
Zhejiang, China

Abstract—Magnetic Resonance Imaging (MRI) plays a major role in the diagnosis of several diseases. However, the acquisition of measurements in the k -space domain, which is the basis for image reconstruction, takes a long time compared to other imaging modalities and is comparatively costly. In this context, undersampled MRI reconstruction is an approach for decreasing the acquisition duration and the exam’s final cost. With this objective, both Compressed Sensing (CS) and Deep Learning (DL) provide techniques for generating good quality MRI images from undersampled measurements. In this paper, we combine CS and DL methods in order to investigate the potential increase in image quality over each isolated approach. We use reconstructions from highly undersampled MRI signals using two CS approaches, the L1 and total minimizations, as inputs to a U-Net. We also use, for comparison, the reconstructions from the same undersampled signals using L2 minimization, and also test them as inputs to a U-net. The goal is to evaluate whether the U-Net can improve the results of the CS reconstructions after learning from degraded and original image pairs. Our experimental results suggest that the combination of L1 or TV minimization with U-Nets can improve reconstruction, in terms of objective image quality, over each technique used in isolation.

Index Terms—Magnetic resonance imaging, Compressed sensing, Deep learning, U-net

I. INTRODUCTION

Magnetic resonance imaging (MRI) is a medical imaging technique that is used to acquire images of internal organs and tissues, thus revealing the anatomy and the physiological processes of the body. It can provide images with higher pixel contrast compared to most other medical imaging techniques, by emphasizing tissue properties related to T1 and T2 relaxation times [1]. In order to do so, MRI scanners use strong magnetic fields, magnetic field gradients, and radio waves. Contrarily to Computational Tomography (CT) techniques, it does not require the use of X-rays or other form of ionizing radiation, thus providing an important advantage for children, pregnant women, and patients who require repeated imaging procedures. MRI techniques produce medical images that have a high spatial resolution and a wide range of soft tissue contrasts, and is thus

This work was supported by the CNPq-BRICS funding and by the University of Brasília (UnB).

able to depict the body anatomy in great detail. Also, MRI techniques are able to generate images in more planes and can reconstruct organs and anatomical regions in 3 dimensions. In summary, MRI is one of the most important and non-invasive methods for visualizing internal biological tissues [2].

However, the MRI acquisition process is quite slow, as it requires the equipment to synchronize the measurement procedure in thousands of k -space locations with severe spatial gradient field changes. One of the approaches for decreasing the acquisition time is to undersample the measurements, and Compressed Sensing (CS) is an advanced signal processing approach for computing signals based on undersampled linear measurements. [3]–[7]. In fact, CS corresponds to an advanced theory and set of algorithms for reconstructing signals from a number of linear measurements that is below that established by the Shannon-Nyquist criterion, by exploring the fact that such signal has a sparse representation in known transformed domain. CS has been widely used in MRI [4] to reconstruct the desired image from an acquired compressed (undersampled) set of measurements in the k -space.

Since MRI is a non-invasive method that allows the visualization of internal biological tissues and the differentiation between different types of tissues, it has become an important tool for diagnosing diseases [2]. But, although MRI techniques provide a high degree of details, diagnosis can be difficult and time-consuming, and lower quality images can lead to diagnosis variations and visual fatigue. Also, if problems happen during acquisition, the image may present an even lower quality, with a poor contrast and the presence of degradation that prevents medical decisions. In these cases, image processing techniques are used to improve the image quality and reduce the time required for accurate assessment, therapy planning, and tumor progression tracking. In summary, diagnosis based on MRI analysis is a time-consuming process that depends heavily on the expert’s proficiency and the image quality.

While CS has conquered relative success in the MRI reconstruction in the last decade, machine learning techniques have been presenting great popularization in the last few years. Hyun *et al.* [8] proposed an undersampled MRI reconstruction

method using zero-fill reconstruction as input of a U-net [9] deep learning (DL) architecture from Cartesian undersampling measurements.

In this paper, we further advance Hyun *et al.* work. First, we also explore CS reconstruction as input for the U-net. Specifically, we test using ℓ_2 , ℓ_1 , and TV Optimizations (CS) as the inputs for the U-Net. Our hypothesis is that inputting better-quality images to the U-Net would make it easier for the U-Net to get to the original high-quality image. In this proposed approach, we train 3 U-Nets, one for each optimization method. In the sequence, we use a pseudo-radial measurement trajectory, that is more suitable for CS reconstruction, a type of trajectory that present better incoherence and usually present better quality image for fewer measurements in CS reconstruction. The proposed methods are adequate for reconstructing severely undersampled images (under 10% of the original resolution) in the k -space, which leads to a better image quality.

The remaining of the paper is organized as follows. Section II presents the fundamentals of the MRI reconstruction optimization process. Section III details the proposed method. Sections IV and V details the experimental setup and results, respectively. Finally, Section VI presents our conclusions.

II. COMPRESSED SENSING MRI RECONSTRUCTION

Compressed sensing is an strategy used to acquire and reconstruct MRIs, which exploit spatial and temporal correlations using irregular undersampling schemes and a nonlinear reconstruction that uses the data sparsity in a suitable transform domain. Let x be an MRI image sampled in the k -space. The k -space measurements are

$$y = S(\mathcal{F}(x)), \quad (1)$$

where $\mathcal{F}(x)$ is the discrete 2D Fourier of the x image and S is the sampling pattern function. From this process, we obtain the samples y in the discrete k -space.

Since using Cartesian k -space trajectories with random undersampling patterns often creates aliasing artifacts [8], we use a sampling pattern composed of a golden angle radial sampling of the k -space [10]. In general, radial sampling is a commonly chosen trajectory for compressed sensing (undersampled) MRI reconstructions because it is less susceptible to motion artifacts caused by a higher sampling density in the lower frequencies [5]. Also, the undersampling of this technique does not affect the contrast of the reconstructed MRI [11]. In the particular case of the golden angle radial sampling, the angle of the radial lines is increased continuously by 111.25 degrees, which provides a coverage of the k -space that is more uniform than what is obtained with the evenly sampled radial trajectory. The golden angle radial sampling also provides a high temporal incoherence and a Point Spread Function (PSF) that has a smaller side lobe. This leads to a better compressed sensing reconstruction.

Given the measurements y , the reconstruction process consists of solving an optimization reconstruction problem to recover an image \hat{x} that is as close as possible to the original image x . In this work, we consider 3 types of reconstructions. First, the

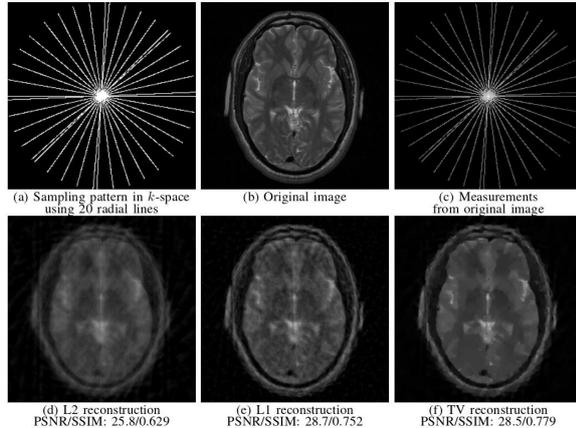


Fig. 1. Measurement process in the k -space and image reconstruction results.

ℓ_2 reconstruction, which generates an image with the minimum norm in the k -space:

$$\hat{x} = \operatorname{argmin} \|y\|, \text{ subject to } y = S(\mathcal{F}(x)). \quad (2)$$

The solution for the ℓ_2 reconstruction is the pseudo-inverse of the operator $S(\mathcal{F}(\cdot))$ of y . This solution is simple, but usually does not provide a correct solution for undersampled signals.

The second type of reconstruction is an ℓ_1 norm minimization, which is given by the following equation:

$$\hat{x} = \operatorname{argmin} \|y\|_1, \text{ subject to } y = S(\mathcal{F}(x)). \quad (3)$$

Under a restriction of the data sparsity and given enough measurements, the compressed sensing theory states that the ℓ_1 solution is the the correct one [12]. If we suppose the images are sparse, which is a common assumption for MRIs [6], no sparsifying transform is required. However, even if the images are not sparse, the ℓ_1 resulting images usually present better image quality than correspondent ℓ_2 reconstruction.

The third type of reconstruction is the Total Variation (TV) minimization [13], which is given by the following equation:

$$\hat{x} = \operatorname{argmin} \|\nabla y\|, \text{ subject to } y = S(\mathcal{F}(x)), \quad (4)$$

where the $\|\nabla \cdot\|$ is the horizontal and vertical finite difference norm ℓ_2 [7], [13]. This approach has been largely employed for MRI reconstruction, because it reinforces piece-wise constant regions of the images, emphasizing tissues with constant properties [14]–[17]. However, over high undersampling ratio, TV generated images can present a visual artifact similar to a “painting” effect.

To perform the tests in this work, we used only 20 pseudo radial lines and a 256×256 rectangular grid. This corresponds to 5,436 samples taken from a 256×256 grid, corresponding to a 8.365% sampling ratio. It is worth emphasizing that this is a very small number of radial lines, which generates highly undersampled measurements that are a challenge to the compressed sensing reconstruction process. To illustrate this, Figure 1 (a) shows the radial sampling pattern of the k -space, Figure 1 (b) shoes the original image from which the samples are taken, and Figure 1 (c) shows the corresponding sampled image

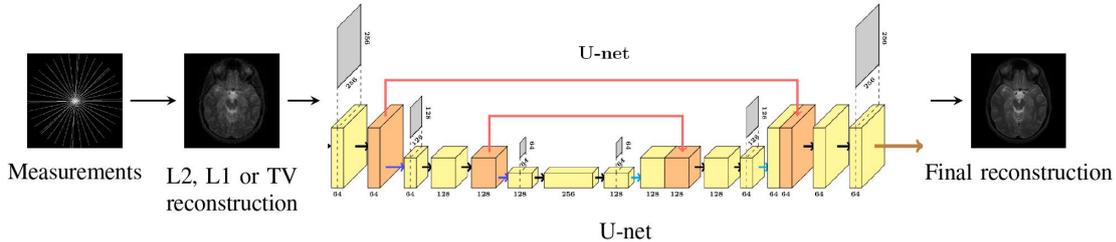


Fig. 2. Complete architecture of our model. Figure adapted from [8].

containing the magnitude of the measurements taken. Figures 1 (d), (e), (f) shows, respectively, the ℓ_2 (L2), ℓ_1 (L1), and TV reconstructions from the measurements in Figure 1 (c). The ℓ_1 and TV minimization reconstructions were performed using the IAFNNESTA minimization algorithms [16]¹.

III. PROPOSED METHODOLOGY

In this work, our goal is to use machine learning technique to improve the quality of images reconstructed with the norm optimization methods described in previous section. More specifically, we aim to train neural networks that take the reconstructed images and output images that have a better overall quality. For this goal, we have chosen the Convolution Neural Network (CNN) [18] known as U-net [9], which is an encoder-decoder CNN (no fully connected layers) that represents the input data in a low-dimensional space that preserves high frequency content. To avoid vanishing gradients, the U-net architecture connects the first intermediate layers to the decoder part in order [8], [18].

In medical imaging, the U-net has been successfully used for image segmentation [18]–[20] and image reconstruction [8], [21], [22]. For example, Hyun *et al.* used the U-net for under-sampled MRI reconstruction, using as inputs ℓ_2 reconstructions obtained from k -space horizontal lines.

In this work, we first reconstruct the images from under-sampled measurements, using radial measurements and reconstruction methods presented in previous section. Next, we train three U-Nets with similar architecture proposed by Hyun *et al.*, with the goal to minimize the difference between the CS reconstructed images for each optimization method and the original images. We hypothesize that the Unet can learn to get rid of the visual degradation commonly presented in the undersampled optimization reconstruction. Figure 2 shows the block-diagram of the proposed methodology.

For the training procedure, we simulate the measurement process, reconstruct with an optimization method (L2, L1 or TV) and use the result as the input images for a training procedure with the U-net. The goal is to train the network to become as close as possible to the original image, taking into consideration that in production we only have access to the measurements. In this case, we reconstruct the image with the same optimization method and forward it by the network to produce the final

image. In this work, we perform a comprehensive analysis of the proposed method, by selecting a fixed sampling trajectory and testing different minimization reconstruction techniques.

IV. EXPERIMENTAL SETUP

We implement the U-net model using the Pytorch learning framework². To run our simulations, we use 4,020 brain MRI images taken from the BIRN dataset [23]. We split the dataset in 80% for training, 10% for validation and 10% for testing. For the training process, we use the Mean Square Error (MSE) as the loss function and a batch size of 8. We use the Adam optimizer with a constant 0.0001 learning rate and a 0.00001 weight decay. We train the model for 300 epochs, selecting the parameters of the epoch with the smallest MSE value.

As mentioned earlier, we use as inputs 3 types of reconstructions (L2, L1, and TV), repeating the exact training process for each one of these input images and generating 3 models:

- **L2-Unet:** model trained with images obtained with the L2 reconstruction process;
- **L1-Unet:** model trained with images obtained with the L1 reconstruction process;
- **TV-Unet:** model trained with images obtained with the TV reconstruction process.

For the image reconstruction simulations, we used 20 radial lines of image measurements, which is a very low number of radial lines. Finally, as performance metrics, we use two quality metrics: the Peak Signal to Noise Ratio (PSNR) and the structural similarity index measure (SSIM) [24]. These metrics are computed between the original and final output images, with higher values indicating a better image qualities.

V. EXPERIMENTAL RESULTS

Figure 3 shows 5 sample reconstruction results obtained from the complete test dataset, with each line corresponding to the results of one (out of 5) sample MRI content. The first column shows the 5 sample originals taken from the dataset. The second, third, and fourth columns show the corresponding reconstruction results obtained with the the U-net architecture when the inputs are, respectively, L2, L1, and TV reconstructed images. The PSNR and SSIM values for each reconstructed images are shown below the corresponding image. Note that

¹Available for download at <https://github.com/JonathanAlis/IAFNNESTA>.

²Available at <https://pytorch.org/>

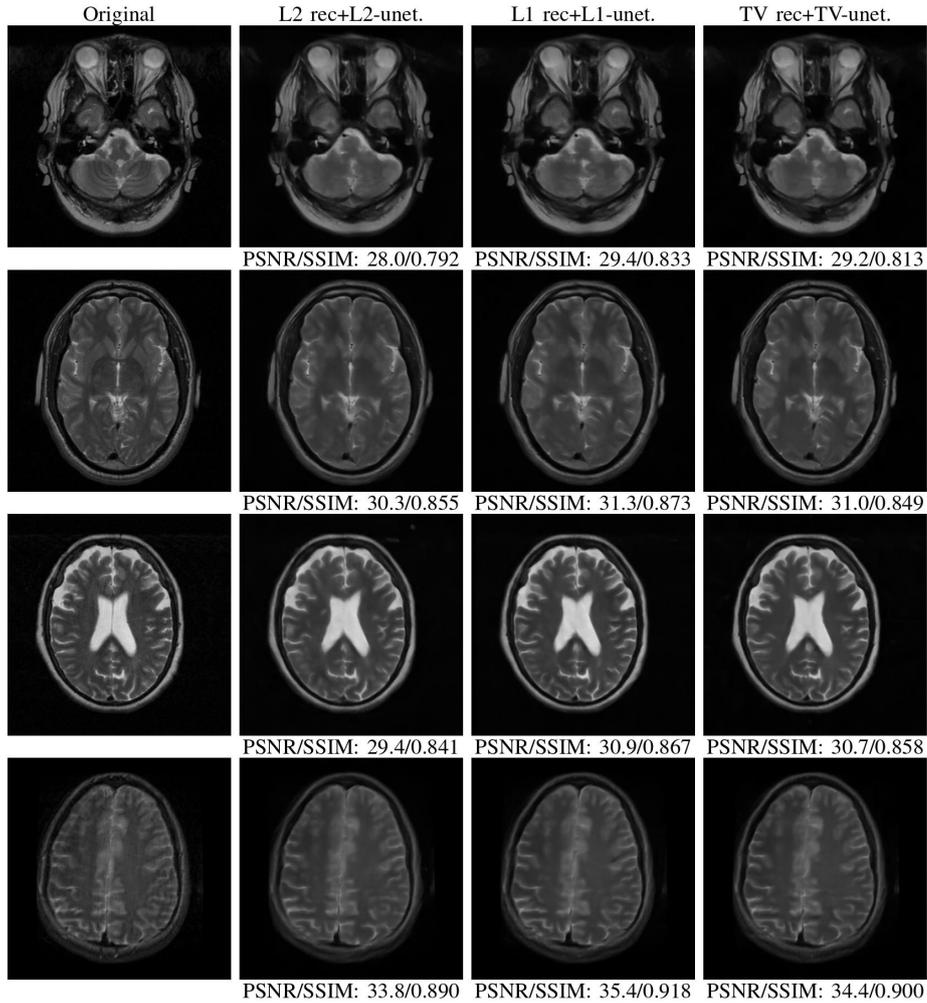


Fig. 3. Image reconstruction examples from the test dataset.

TABLE I
AVERAGE PSNR/SSIM OF THE UNET OUTPUTS COMPARED TO THE ORIGINAL IMAGES. THE INPUTS ARE THE L2, L1 AND TV RECONSTRUCTIONS, AND COLUMN INPUT INDICATE THEIR PSNR/SSIM COMPARED TO THE ORIGINAL.

	Input	L2-unet	L1-unet	TV-unet
L2 rec	27.9/0.672	32.1/0.865	28.7/0.755	28.8/0.768
L1 rec	31.2/0.799	26.7/0.793	33.5/0.891	31.4/0.837
TV rec	31.5/0.832	24.6/0.786	31.2/0.854	33.3/0.877

all three models were able to obtain images with a good visual quality. Also, the images obtained with the 3 models are similar to the original image and to each other. This is supported by the reported PSNR and SSIM values.

More specifically, in terms of the quality metrics, the Unets significantly improved the visual quality of the recon-

structed images, specially for reconstruction methods that initially presented poor performance. The L2 reconstruction presented an average PSNR/SSIM of 27.9/0.672, but after applying the trained L2-Unet the average PSNR/SSIM increased to 32.1/0.865. The L1 reconstruction presented initially an average PSNR/SSIM of 31.2/0.799, which increased to 33.5/0.891 after using the L1-unet. Finally, the TV reconstruction had an average PSNR/SSIM of 31.5/0.832, which was increased to 33.3/0.877 after using the TV-Unet. Notice that, the bigger quality jump occur for the L2 reconstruction. However, as expected, TV-Unet and L1-Unet presented higher final quality scores. With reference to the SSIM values, we observe that the performance of the L2 reconstruction is slightly inferior than the performance of the L1 and TV reconstructions, which have very close values. These results are an evidence that using Unets improve the quality of the reconstructed images.

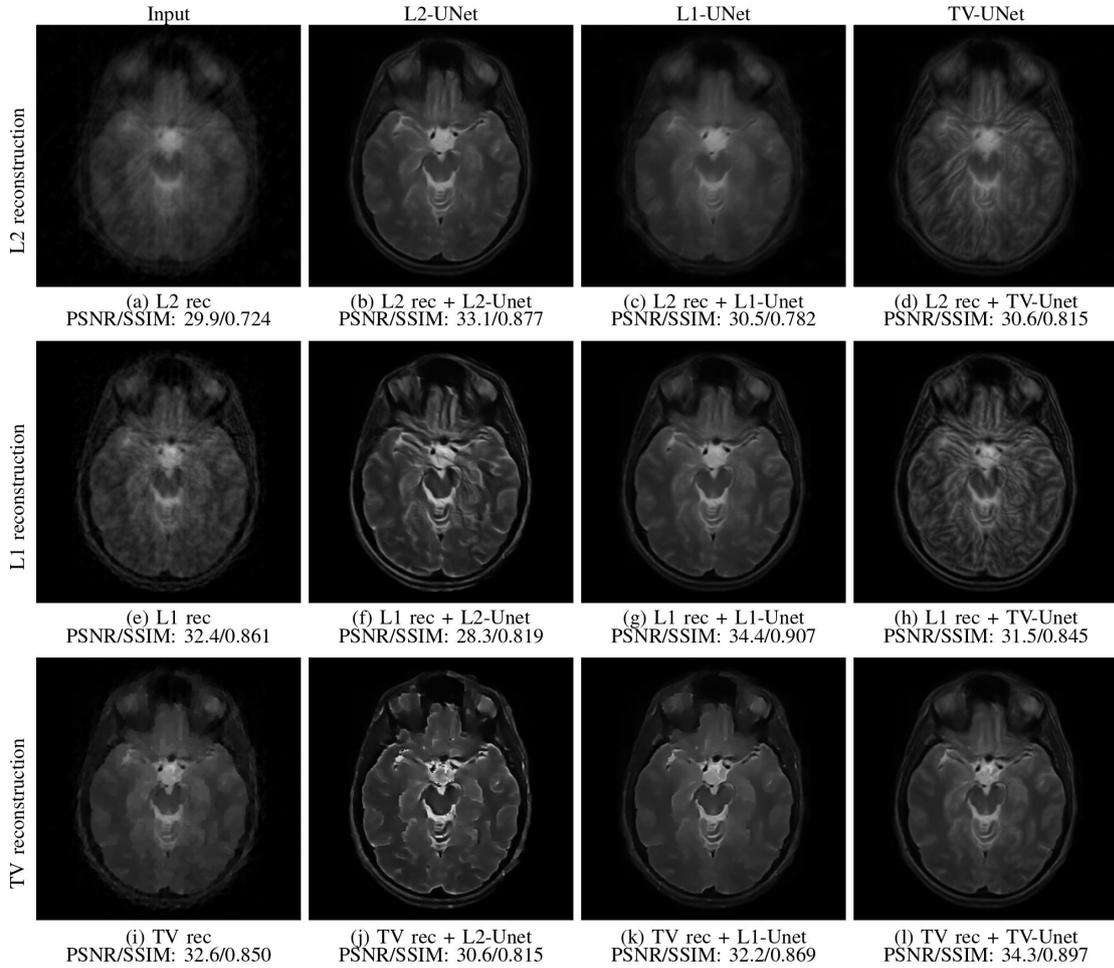


Fig. 4. Average PSNR/SSIM of the U-Net outputs compared to the original images. The inputs are the L2, L1 and TV reconstructions, and column Input indicate their PSNR/SSIM compared to the original.

We also analyze how different trained Unet models transform the images generated by different optimization methods, so we perform a cross-image reconstruction comparison test. In this test, the cross comparison is formed by all combinations of optimization methods and U-nets: L2, L1 and TV optimizations; and not using Unet, L2-UNet, L1-UNet and TV-UNet; a total of 12 combinations. Table V shows the average PSNR and SSIM values for this test, while Figure 4 shows an example image. Rows in this table correspond to optimization reconstruction methods, and columns correspond to Unet models, with the exception of the column input, for the sake of comparison. Notice that best results occur for matching types: L2 optimization reconstruction + L2-UNet, L1 optimization reconstruction + L1-UNet and TV optimization reconstruction + TV-UNet. L1 and TV based methods presented better results than L2 based.

In Figure 4 we can also see how the Unets transform the images. The Unets were trained to remove the kind of degra-

ation presented in images generated by the correspondent optimization reconstruction. Using Unet for different type of optimization generated intriguing results. For instance, L2-UNet is able to deblur the image, because L2 images are very blurry, so the L2-UNet generated strong edges on L1 and TV generated images (Fig. 4(f), (j)). Notice also that L1-UNet and TV-UNet generated blurred images when the input is the blurred L2 reconstruction (Fig. 4(c), (d)), unlike the L2-UNet. Other example is that TV-UNet generated images with brain structures that were not presented in the original. TV-Unets learns to That happens because the TV reconstruction gets rid of the . TV-UNet use to decrease the 'painting aesthetic' that the TV reconstruction presents, what generates the odd strong edges on L2 optimization reconstruction.

Figure 5 shows the PSNR and SSIM distribution plots for all samples of the test set. Not only the average and median values is higher for models of matching types, but they also are

distributed in smaller range, showing consistent better results.

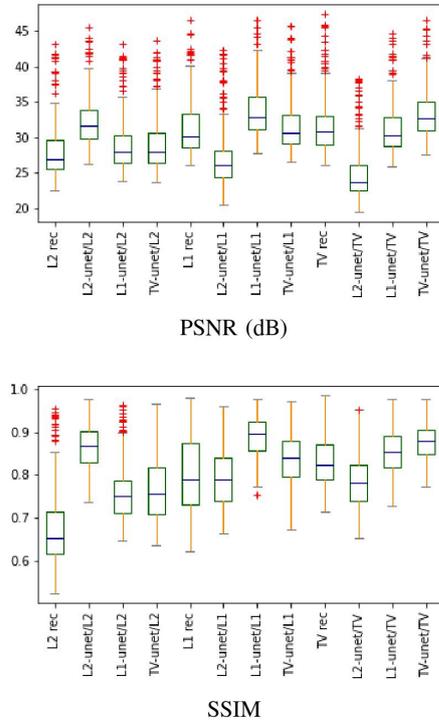


Fig. 5. PSNR (dB) and SSIM values for the cross comparison.

VI. CONCLUSIONS AND FUTURE WORKS

In this work, we combine common optimization reconstruction techniques for undersampled MRI with the U-Net trained for the reconstruction inputs. The U-Net was able to obtain a high level of accuracy for a relatively complex problem, even for a very small sampling ratio. Specially, we were able to improve the L2 rec+L2-unet reconstruction from previous works.

However, another study that can be done is from the interactions between the L1, L2 and TV models using a greater number of radial lines such as 40, 60, 80 or 100, for example, taking into account precision, specificity and sensitivity, hoping to obtain better results for the L2 model and a decrease in the "paint" effect for the TV model. The results obtained will also be useful for future studies, improvements in methodology, testing images of other datasets and other parts of the body, and eventually in practical applications in hospitals and MRI machines.

REFERENCES

[1] R. N. Bryan, *Introduction to the Science of Medical Imaging*. Cambridge University Press, 2009.
 [2] M. T. Vlaardingerbroek and J. A. Boer, *Magnetic resonance imaging: theory and practice*. Springer Science & Business Media, 2013.

[3] J. Lima, C. J. Miosso, M. Farias, and R. von Borries, "Evaluation of Different Types of Filters in Magnetic Resonance Imaging Using Compressive Sensing with Pre-Filtering," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2018, pp. 5575–5578. [Online]. Available: <https://ieeexplore.ieee.org/document/8513618/>

[4] J. C. Ye, "Compressed sensing MRI: a review from signal processing perspective," *BMC Biomed. Engineering*, vol. 1, no. 1, pp. 1–17, 2019.

[5] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[7] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[8] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Physics in Medicine & Biology*, vol. 63, no. 13, p. 135007, 2018.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, "An optimal radial profile order based on the golden ratio for time-resolved MRI," *IEEE transactions on medical imaging*, vol. 26, no. 1, pp. 68–76, 2006.

[11] K. Scheffler and J. Hennig, "Reduced circular field-of-view imaging," *Magnetic resonance in medicine*, vol. 40, no. 3, pp. 474–480, 1998.

[12] E. J. Candès and Y. Plan, "Near-ideal model selection by l_1 minimization," *The Annals of Statistics*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[13] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.

[14] V. N. Varghees, M. S. Manikandan, and R. Gini, "Adaptive MRI image denoising using total-variation and local noise estimation," in *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012)*. IEEE, 2012, pp. 506–511.

[15] C. Li, W. Yin, and Y. Zhang, "User's guide for TV-L3: TV minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 46-47, p. 4, 2009.

[16] J. A. Lima, F. B. da Silva, R. von Borries, C. J. Miosso, and M. C. Farias, "Isotropic and anisotropic filtering norm-minimization: A generalization of the TV and TGV minimizations using NESTA," *Signal Processing: Image Communication*, vol. 85, p. 115856, 2020.

[17] M. J. Ehrhardt and M. M. Betcke, "Multicontrast MRI reconstruction with structure-guided total variation," *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 1084–1106, 2016.

[18] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[19] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "Nas-unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44 247–44 257, 2019.

[20] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected u-net for medical image segmentation," in *ICASSP 2020-2020 IEEE Intern. Conf. on Acoustics, Speech and Signal Proces.* IEEE, 2020, pp. 1055–1059.

[21] S. Mizusawa, Y. Sei, R. Orihara, and A. Ohsuga, "Computed tomography image reconstruction using stacked U-Net," *Computerized Medical Imaging and Graphics*, vol. 90, p. 101920, 2021.

[22] T. Wang, X. Xu, J. Xiong, Q. Jia, H. Yuan, M. Huang, J. Zhuang, and Y. Shi, "ICA-UNet: ICA inspired statistical UNet for real-time 3d cardiac cine MRI segmentation," in *Intern. conf. on med. image comput. and computer-assisted intervention*. Springer, 2020, pp. 447–457.

[23] K. G. Helmer, J. L. Ambite, J. Ames *et al.*, "Enabling collaborative research using the biomedical informatics research network (BIRN)," *Journal of the American Medical Informatics Association*, vol. 18, no. 4, pp. 416–422, 2011.

[24] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.