UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

MATEUS BALTAZAR DE ALMEIDA

Comparações de Modelos Neurais em Redução de Ruído de Imagens de Tomografia Computadorizada

Recife

2024

MATEUS BALTAZAR DE ALMEIDA

Comparações de Modelos Neurais em Redução de Ruído de Imagens de Tomografia Computadorizada

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador: Prof. Dr. Tsang Ing Ren

Coorientador: Prof. Dr. Luis Filipe Alves Pereira

Recife

2024

Catalogação na fonte
Bibliotecária: Mônica Uchôa, CRB4-1010

A447c Almeida, Mateus Baltazar de

Comparações de modelos neurais em redução de ruído de imagens de tomografia computadorizada / Mateus Baltazar de Almeida. – 2024.

62 f.: il., fig.; tab.

Orientador: Tsang Ing Ren.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Informática. Programa de Pós-graduação em Ciência da Computação. Recife, 2024.

Inclui referências e anexos.

1. Redução de ruído. 2. Tomografia computadorizada. 3. Avaliação holística. I. Ren, Tsang Ing (orientador). II. Título.

006.31 CDD (23. ed.) UFPE - CCEN 2024 – 76

MATEUS BALTAZAR DE ALMEIDA

COMPARAÇÕES DE MODELOS NEURAIS EM REDUÇÃO DE RUÍDO DE IMAGENS DE TOMOGRAFIA COMPUTADORIZADA

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de concentração: Inteligência Computacional.

Aprovado em: 27 de março de 2024.

BANCA EXAMINADORA

_____

Prof. Dr. Tsang Ing Ren (Orientador)
Centro de Informática - UFPE

_____

Prof. Dr. Cleber Zanchettin (Examinador Interno)
Centro de Informática - UFPE

_____

Prof. Dr. Alceu de Souza Britto Junior (Examinador Externo)
Programa de PósGraduação em Informática - PUC/PR

Dedico o presente trabalho ao meu tio, padrinho e amigo Junior Mário Baltazar de Oliveira. Que me orientou e incentivou a prosseguir nos momentos mais difíceis, me mostrando que para tudo o que eu pretender realizar, **só vou conseguir se tentar**. Das brincadeiras aos puxões de orelha, tudo o que ele me fez, servirá como ensinamento. Um exemplo de amigo, profissional e humano, pelo qual sempre me inspirei para seguir em frente. Gostaria de poder agradecê-lo mais uma vez, pela grande parte do que me faz ser o que sou *(in memorian)*.

AGRADECIMENTOS

Agradeço, primeiramente, aos meus pais, Marisvalda Baltazar e Paulo César, e ao meu irmão, Adalberto Baltazar, por me proporcionarem um ótimo apoio para que eu conseguisse ultrapassar todo o percurso do mestrado. Em especial, à minha mãe, que, apesar do árduo trajeto que tem enfrentado ultimamente, está sempre presente e incentivando minhas decisões.

Aos meus orientadores, Prof. Dr. Tsang Ing Ren, Prof. Dr. Luis Filipe e Prof. Dr. George Darmiton, pela atenção, compreensão e, principalmente, pelas ótimas orientações (não só no âmbito da pesquisa). Agradeço também ao meu grande amigo, Gersonilo Oliveira, por sempre me apoiar e torcer pelo meu êxito nessa jornada.

Ao Centro de Informática (CIn) da Universidade Federal de Pernambuco (UFPE), por me conceder uma excelente infraestrutura, e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte com a bolsa de estudo de Mestrado.

E, por fim, a todos os demais familiares, colegas de curso e amigos, que direta ou indiretamente contribuíram para a conclusão desta etapa.

RESUMO

Reduzir a exposição de pacientes à radiação ionizante em exames de Tomografia Computadorizada é primordial para minimizar os riscos oncogênicos, desafiando o desenvolvimento de técnicas que viabilizem imagens claras mesmo com doses baixas de radiação. A remoção eficaz de ruído das imagens médicas mediante modelos neurais surge como alternativa da recuperação da qualidade de imagens produzidas com menor corrente elétrica nos tubos de raio-X. Este estudo centra-se na comparação equitativa de diferentes arquiteturas de redes neurais nessa aplicação, confrontando a adequação das avaliações tradicionais, ancoradas em análises pontuais que ignoram variações experimentais, permitindo conclusões superficiais e tendenciosas, além de abordar a inconsistência entre métricas quantitativas e qualidade visual. Com ênfase nos desafios inerentes, a metodologia adotada perpassou por um controle rigoroso de hiperparâmetros, tal como o *batch size* e *seeds*, para garimpar nuances no desempenho dos modelos sob variações experimentais. As experimentações, sob um meticuloso ambiente de testes, buscaram garantir replicabilidade e entender cuidadosamente o papel das funções de perda na preservação de detalhes diagnósticos. O Efeito MSE, um artefato visual induzido por funções de perda pixel-wise, foi um dos delineamentos qualitativos observados, desafiando a confiabilidade das métricas tradicionais. A pesquisa evidencia a complexidade subjacente nas análises de desempenho, postulando o design de um *framework* que confere maior confiabilidade aos resultados, além de apontar para a necessidade de práticas de avaliação mais robustas e holísticas no campo da reconstrução de imagens médicas.

Palavras-chave: redução de ruído; tomografia computadorizada; avaliação holística.

ABSTRACT

Reducing patients' exposure to ionizing radiation in Computed Tomography scans is crucial to minimize oncogenic risks, challenging the development of techniques that enable clear images even with low radiation doses. The effective removal of noise from medical images through neural models emerges as a solution to restore the quality of images produced with lower current in the X-ray tubes. This study focuses on the equitable comparison of different neural network architectures in this application, confronting the adequacy of traditional evaluations, anchored in punctual analyses that ignore experimental variations, allowing superficial and biased conclusions, as well as addressing the inconsistency between quantitative metrics and visual quality. With an emphasis on inherent challenges, the adopted methodology passed through a rigorous control of hyperparameters, such as batch size and seeds, to glean nuances in the performance of models under experimental variations. The experiments, under a meticulous testing environment, sought to ensure replicability and carefully understand the role of loss functions in preserving diagnostic details. The MSE Effect, a visual artifact induced by pixel-wise loss functions, was one of the qualitative delineations observed, challenging the reliability of traditional metrics. The research highlights the underlying complexity in performance analyses, proposing the design of a framework that provides greater reliability to the results, as well as pointing to the need for more robust and holistic evaluation practices in the field of medical image reconstruction.

Keywords: holistic evaluation; noise reduction; computed tomography.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A Tomografia Computadorizada (CT, do inglês *Computed Tomography*) é reconhecida como uma ferramenta diagnóstica essencial na medicina moderna, proporcionando imagens detalhadas das estruturas internas do corpo humano, que são cruciais para o diagnóstico preciso e o planejamento do tratamento de diversas condições clínicas. No entanto, o processo de aquisição dessas imagens envolve a exposição do paciente à radiação ionizante, o que traz consigo riscos potenciais à saúde, incluindo o aumento do risco de câncer (KRILLE et al., 2010). Diante disso, a redução da dose de radiação utilizada em procedimentos de CT surgiu como uma estratégia vital para minimizar esses riscos, garantindo a segurança do paciente. Contudo, a captura de imagens de CT com baixa dosagem de radiação (LDCT, do inglês *Low Dose Computed Tomography*) tende a resultar em imagens apresentando ruído e artefatos significativamente aumentados, comprometendo a nitidez e a clareza das mesmas. Esta compensação entre redução de radiação e qualidade da imagem sublinha a necessidade de desenvolver técnicas de redução de ruído eficazes (QIN et al., 2019), para otimizar tanto a segurança quanto a eficácia diagnóstica dos exames de CT .

O ruído em imagens de LDCT, especialmente o ruído de Poisson que é intrinsecamente ligado ao processo de captura de imagem baseado em radiação, compromete a qualidade das imagens ao obscurecer detalhes finos e importantes, quando comparado com captura da imagem de CT com dosagem normal de radiação (NDCT, do inglês *Normal Dose Computed Tomography*). Este ruído não apenas deteriora a qualidade visual geral das imagens, como também pode prejudicar a acurácia diagnóstica (CHEN et al., 2017). O desafio, portanto, reside em desenvolver métodos capazes de recuperar imagens nítidas e detalhadas a partir dessas capturas ruidosas, preservando detalhes importantes sem necessitar o aumento da dose de radiação. A meta é, assim, garantir a obtenção de imagens de alta qualidade diagnóstica, mantendo a segurança do paciente em conformidade com o princípio ALARA (As Low As Reasonably Achievable) (BRENNER; HALL, 2007) da radioproteção (CHEN et al., 2017).

Neste cenário, as abordagens atuais de remoção de ruído em imagens LDCT estão cada vez mais voltadas para a utilização de modelos neurais avançados, dada a sua capacidade de aprender padrões complexos e de adaptar-se a uma ampla gama de cenários com ruído (CHEN et al., 2017; LI et al., 2020). Esses modelos, que incluem variedades como a U-Net (RONNEBERGER; FISCHER; BROX, 2015), a *Residual Encoder-Decoder Convolutional Neural*

Network (RED-CNN) (CHEN et al., 2017), a *Wasserstein Generative Adversarial Neural Network* (WGAN) (YANG et al., 2018) e a *Self-Attention Convolutional Neural Network* (SACNN) (LI et al., 2020), têm demonstrado um potencial significativo para aprimorar eficazmente a qualidade visual das imagens LDCT. Essas tecnologias representam uma evolução notável em relação às técnicas tradicionais de processamento de imagens, direcionando o campo para soluções mais eficientes e adaptativas no desafio de aprimoramento de imagens médicas diante da necessária redução de exposição à radiação (CHEN et al., 2017; LI et al., 2020).

No cerne da eficácia dos modelos neurais para remoção de ruído em imagens LDCT, residem as funções de perda escolhidas para o treinamento dessas redes, desempenhando um papel crítico ao guiar o processo de aprendizado do modelo, influenciando diretamente a qualidade das imagens reconstruídas. Redes como RED-CNN (CHEN et al., 2017) e U-Net (RONNEBERGER; FISCHER; BROX, 2015) podem ser treinadas utilizando uma variedade de funções de perda, tais como o MSE (do inglês, *Mean Squared Error*) (WANG; BOVIK, 2009), Charbonnier (CHARBONNIER et al., 1997), e funções de perda perceptuais (JOHNSON; ALAHI; FEI-FEI, 2016). A escolha da função de perda pode determinar a capacidade do modelo em preservar detalhes importantes nas imagens, ao mesmo tempo em que reduz o ruído. Entretanto, alguns desses métodos podem resultar no Efeito MSE: um fenômeno onde a reconstrução fidedigna de detalhes minuciosos é comprometida, levando a imagens suavizadas que carecem de nitidez e precisão na representação de texturas e bordas (YANG et al., 2018; LI et al., 2020; YU et al., 2017; WOLTERINK et al., 2017), além de implicar em valores maiores de métricas na análise quantitativa (YANG et al., 2018; LI et al., 2020). Esse efeito é particularmente notável em modelos treinados com a função de otimização que prioriza a minimização do erro pixel a pixel (YANG et al., 2018; LI et al., 2020).

A comparação entre diferentes modelos neurais é comumente desafiada pela variabilidade inerente ao processo de treinamento e validação dessas redes. Avaliações baseadas em execuções unitárias podem levar a conclusões enganosas, devido a uma ampla gama de variáveis experimentais, como a escolha do *batch size*, a *seed* para inicialização de pesos e ordenação dos dados apresentados, a taxa de aprendizagem e as versões de softwares e hardwares utilizados. Esses fatores podem influenciar significativamente o desempenho do modelo, introduzindo um grau de imprevisibilidade nos resultados que dificulta a realização de comparações diretas e objetivas entre diferentes abordagens. Assim, a confiabilidade dos resultados pode ser comprometida, sujeita a certas configurações experimentais ou até predisposta ao viés de interpretação dos autores, ressaltando a complexidade de estabelecer avaliações definitivas

sobre a superioridade de um modelo em detrimento de outros.

Nesse contexto, a necessidade de adotar métodos de comparação mais sistemáticos e abrangentes torna-se evidente. Conduzindo múltiplas execuções e variando hiperparâmetros cruciais como *batch size* e *seed*, é possível construir uma nuvem de resultados para cada modelo, englobando uma diversidade de cenários experimentais que oferecem uma visão mais ampla do seu desempenho. Esta abordagem permite não apenas a identificação da faixa de desempenho esperada sob diferentes condições, mas também proporciona uma base mais sólida e confiável para a comparação entre modelos. Tal estratégia minimiza a influência de variabilidades individuais e artefatos experimentais, facilitando uma avaliação mais justa e equitativa da eficácia dos modelos neurais em realizar a remoção de ruído em imagens LDCT.

1.1 OBJETIVOS

Este trabalho tem como objetivo principal investigar a aplicabilidade de modelos neurais de ponta na tarefa de redução de ruído em imagens LDCT. A pesquisa busca ir além das tradicionais análises baseadas em execuções únicas, adotando uma abordagem que considera uma ampla variedade de resultados gerados a partir da variação de hiperparâmetros cruciais como *batch size* e *seed*. Essa metodologia permite uma avaliação detalhada e representativa do desempenho dos modelos. Além disso, o estudo visa também entender o impacto das diferentes funções de perda na qualidade visual final das imagens, focando em aspectos como a presença do Efeito MSE.

Portanto, os objetivos específicos são:

- Avaliar a eficiência de distintas arquiteturas de redes neurais, incluindo RED-CNN, U-Net, WGAN e SACNN, na remoção de ruído em imagens LDCT, considerando a influência de variáveis experimentais chave, como, *batch size* e *seed*. Esta análise se propõe a gerar uma nuvem de resultados por modelo, abarcando uma gama de condições experimentais para definir uma visão abrangente e precisa de sua capacidade e consistência;
- Determinar o papel que ajustes nos hiperparâmetros exercem sobre as métricas de desempenho (PSNR, SSIM e NRMSE), visando realçar a necessidade de múltiplas repetições experimentais na realização de comparações equitativas entre diferentes modelos;
- Investigar o efeito de diversas funções de perda, tais como MSE, Charbonnier e percep-

tual, no processo de remoção de ruído de imagens LDCT. Será dada uma atenção especial ao entendimento de como essas funções influenciam a manifestação do Efeito MSE e a preservação de detalhes importantes nas imagens, combinando análises quantitativas e observações qualitativas para observar as nuances introduzidas por cada abordagem;

- Desenvolver e aplicar uma metodologia integrada de avaliação que corrobora tanto métricas quantitativas quanto análises visuais qualitativas. Este procedimento visa fornecer um cenário de análise mais completo, que supere as limitações de estudos baseados em execuções únicas, facilitando uma compreensão aprofundada e crítica sobre a real eficácia dos modelos na tarefa proposta.

1.2 ORGANIZAÇÃO DO DOCUMENTO

Os demais capítulos deste documento, estão organizados da seguinte maneira:

- No Capítulo 2, introduzimos os fundamentos teóricos e conceituais relacionados à remoção de ruído em imagens de tomografia computadorizada de baixa radiação utilizando modelos neurais profundos.

- No Capítulo 3, expomos a metodologia adotada no estudo, incluindo a seleção dos modelos neurais, a definição dos experimentos e a estratégia de avaliação.

- No Capítulo 4, apresentamos os resultados obtidos a partir da aplicação da metodologia delineada, destacando a análise quantitativa e observações qualitativas dos modelos e suas variações experimentais.

- No Capítulo 5, descrevemos as conclusões finais e considerações sobre os resultados alcançados.

2 FUNDAMENTAÇÃO TEÓRICA

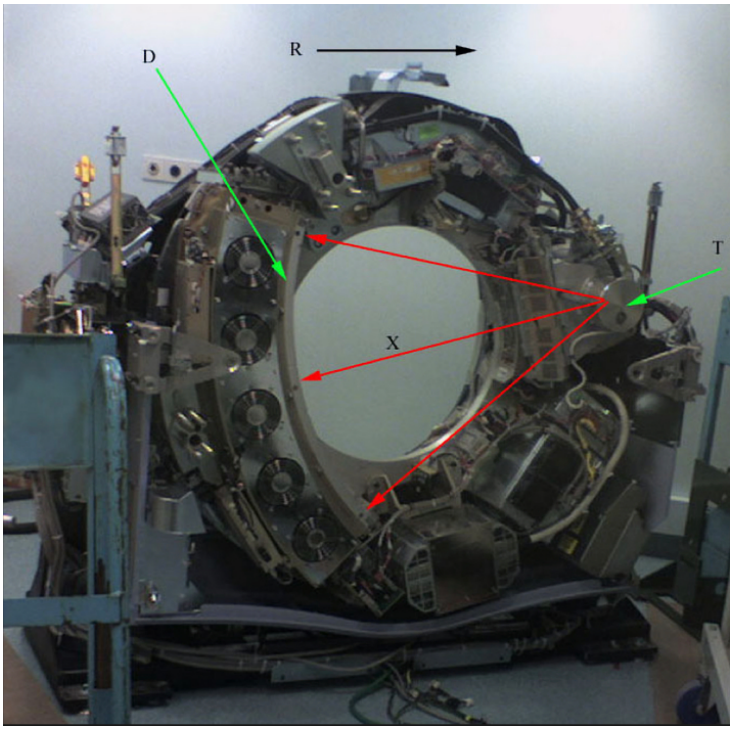
Neste capítulo, apresentaremos os conceitos fundamentais da tomografia computadorizada, desde a técnica de captura de imagens até a importância da redução da dosagem de radiação para a segurança do paciente. Abordaremos o desafio da remoção de ruído em imagens de tomografia computadorizada, ressaltando a necessidade de soluções baseadas em modelos neurais para essa finalidade. Procederemos com uma análise detalhada de variados modelos neurais, incluindo RED-CNN, U-Net, WGAN e SACNN, além de examinarmos diversas funções de perda, suas possíveis combinações e a categorização dessas funções em dois grupos distintos: pixel-wise e não pixel-wise. Concluiremos com uma discussão sobre dois hiperparâmetros (*batch size* e *seed*) fundamentais, que foram amplamente aplicados nos experimentos conduzidos para este estudo.

2.1 TOMOGRAFIA COMPUTADORIZADA (CT)

A Tomografia Computadorizada é uma tecnologia fundamental para o diagnóstico por imagens na área médica. Através de um exame de CT, é possível gerar um modelo tridimensional que contém a representação das estruturas internas de um corpo. Tal modelo é reconstruído a partir das medidas registradas por centenas de feixes de raio-X, irradiados em direção ao paciente a partir de ângulos ao longo de uma revolução completa de 360 graus, gerando assim a imagem sinograma. Diferentes composições de tecidos absorvem diferentes níveis de radiação, funcionando como uma espécie de barreira entre o emissor e a placa de captura, sendo assim os feixes que transpassam diferentes órgãos irão apresentar níveis diferentes de energia ao emergir do corpo. Dois exemplos de tecidos com níveis de absorção de raio-X drasticamente diferentes são os ossos e os pulmões, que absorvem muita e pouca radiação respectivamente. As máquinas de CT possuem uma espécie de anel giratório que circunda o paciente, onde em um ponto está posicionado o emissor de radiação, e na outra extremidade do anel referente ao emissor existe uma placa que captura a radiação emergida do corpo, como mostra a Figura 1.

Contudo, as dosagens de radiação ionizante (raio-X) utilizadas nos diagnósticos e tratamentos que utilizam CT, agridem de forma intensificada a saúde dos pacientes, podendo induzi-los ao câncer e outros problemas (KRILLE et al., 2010). Caracterizando um cenário ainda mais grave para os pacientes que necessitam realizar exames de CT recorrentemente.

Figura 1 – Máquina de Tomografia Computadorizada sem carcaça. R: direção de rotação do anel. T: emissor de radiação. X: direção que os feixes de raio-X irão seguir. D: placa de captura de raios-X.



Fonte: Acervo *radiopaedia.org*

Com o objetivo de reduzir os danos à saúde dos pacientes, a captura de imagens de CT com baixa dosagem de radiação é uma abordagem amplamente estudada. Ao reduzir a corrente elétrica nos tubos emissores de raio-X diminui-se a intensidade da radiação gerada, ocasionando a geração de imagens LDCT. No entanto, essas novas imagens possuem ruídos advindos da captura de baixas quantidades de fótons por pixel da imagem final. Dessa maneira, surge o problema de realizar a recuperação da imagem LDCT removendo os ruídos, transformando-a numa aproximação do que seria a imagem de CT capturada com a dosagem normal de radiação.

A priori, os algoritmos construídos para realizar a aproximação de imagens NDCT a partir de imagens LDCT funcionam de forma iterativa (MURPHY et al., 2021), no entanto, devido às peculiaridades de possuírem um alto grau de complexidade de configuração e exigirem a necessidade de um longo tempo de execução, o método de reconstrução *Filtered Back-Projection* (FBP) (NETT, 2021) é utilizado em conjunto com métodos de recuperação de imagens LDCT baseados em aprendizagem de máquina, satisfazendo os dois principais empecilhos relatados anteriormente, agregando ainda mais potencial de melhorar as aproximações às imagens NDCT. A relação entre os métodos iterativos e o papel dos métodos baseados em aprendizagem de máquina para recuperação de imagens médicas ruidosas é melhor discutido em (QIN

et al., 2019).

Diante dos desafios impostos pela necessidade de imagens de CT de alta qualidade, e os riscos à saúde associados à exposição à radiação, a aplicação de modelos neurais surge como uma solução inovadora e eficaz. Esses modelos aproveitam o aprendizado profundo para aprimorar imagens LDCT tornando-as comparáveis com as imagens NDCT, ao mesmo tempo em que mantêm a segurança do paciente, e reduz o tempo e os recursos necessários para a reconstrução das imagens.

2.2 MODELOS NEURAIS

Nesta seção, serão apresentados diferentes métodos de aprendizagem de máquina com abordagens distintas para a reconstrução de imagens ruidosas de CT. Cada modelo possui sua própria arquitetura e estratégias específicas para lidar com a reconstrução da imagem. Essa diversidade de abordagens busca investigar a qualidade das imagens sintetizadas por modelos que almejam alcançar níveis de fidelidade comparáveis às imagens obtidas com dosagens normais de radiação.

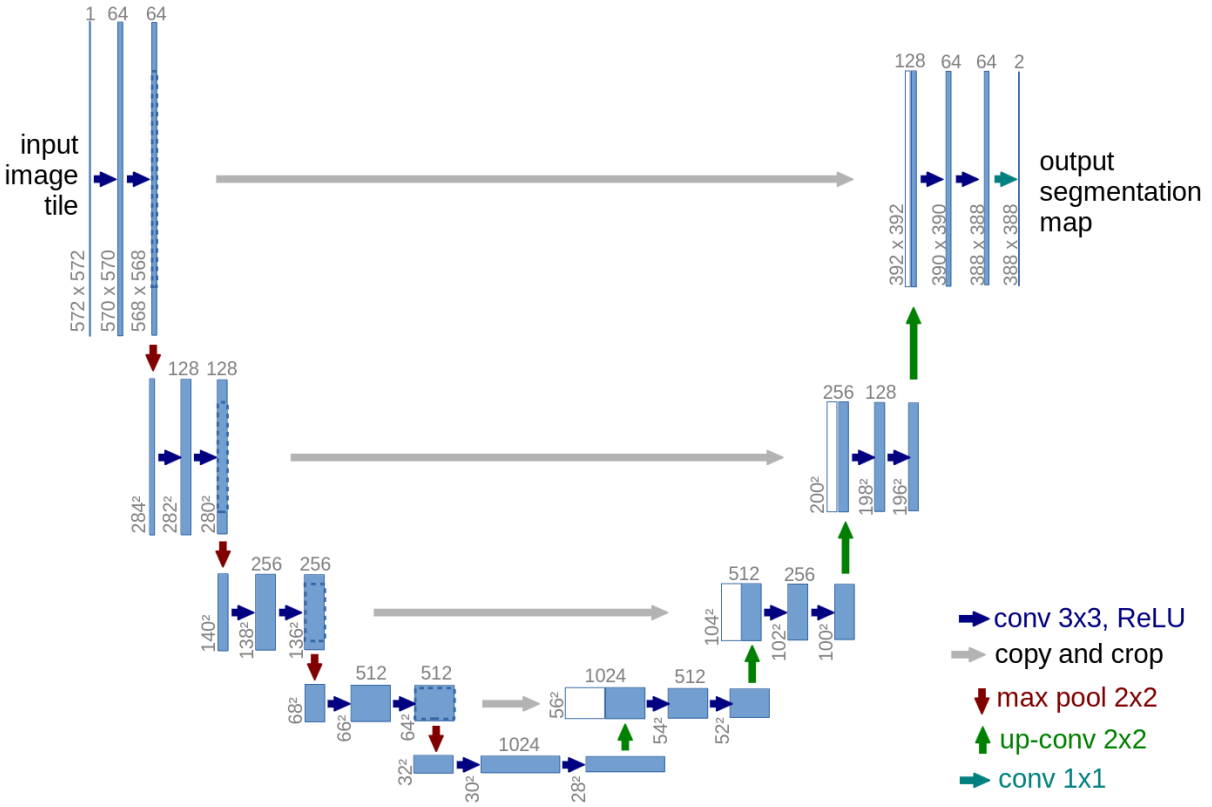
2.2.1 U-Net

A U-Net (RONNEBERGER; FISCHER; BROX, 2015), apresentada inicialmente em 2015, é um modelo revolucionário no campo das redes neurais convolucionais, destacando-se notavelmente nas áreas de segmentação e reconstrução de imagens. Esta arquitetura é facilmente reconhecida por sua estrutura peculiar, que se assemelha à letra "U", uma característica que advém da configuração simétrica de suas partes de codificação (*encoder*) e decodificação (*decoder*). Durante a fase de codificação, a rede realiza um processo de compressão progressiva da imagem, identificando e retendo aspectos fundamentais como bordas e texturas. Em contrapartida, a fase de decodificação trabalha na expansão desses dados condensados, reconstruindo a imagem com um enfoque particular nos detalhes mais sutis.

Um elemento-chave da U-Net é o seu uso inovador de *skip-connections* entre o *encoder* e o *decoder*, permitindo que informações detalhadas sejam transmitidas diretamente entre as duas partes da rede em diferentes níveis de profundidade. Esta abordagem facilita a fusão de características essenciais em várias escalas, crucial para a reconstrução detalhada da imagem. Essas *skip-connections* trabalham pela concatenação de canais, garantindo que detalhes finos não

se percam durante o processo de decodificação, um aspecto vital para a precisão e a qualidade da imagem reconstruída. A Figura 2 ilustra sua arquitetura, com codificador, decodificador e *skip-connections*. Para essa arquitetura serão exploradas três funções de perdas: MSE (*Mean Squared Error*), Charbonnier e perceptual; detalhadamente apresentadas nas seções 2.3.1 e 2.3.2.

Figura 2 – Arquitetura do modelo neural U-Net (RONNEBERGER; FISCHER; BROX, 2015), originalmente desenvolvida para segmentação semântica. É possível identificar seu formato em "U", o *encoder* constituído de convoluções normais e operações de *pooling* condensando o dado de entrada, o *decoder* formado por convoluções normais e transpostas, e as *skip-connections*, na ilustração representadas por "*copy and crop*", representam a cópia da saída de cada bloco convolucional do *encoder* e um recorte central desses dados para haver compatibilidade ao concatenar com os dados do *decoder*.



Fonte: (RONNEBERGER; FISCHER; BROX, 2015)

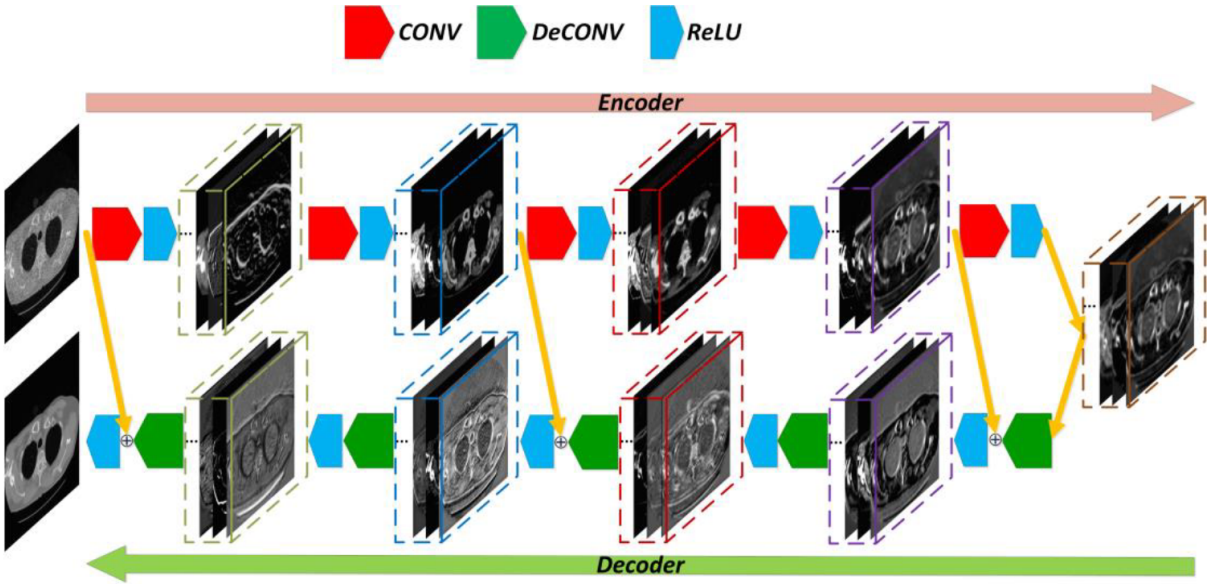
A capacidade da U-Net de manipular e ajustar características em diversas escalas espaciais a torna particularmente eficaz na melhoria das imagens LDCT, realçando a identificação de padrões e estruturas complexas sem comprometer os detalhes cruciais, graças à integração de *skip-connections*. Sua versatilidade e eficiência na captura e reconstrução de nuances sutis fazem dela uma escolha prezada para o aprimoramento de imagens médicas, que frequentemente requerem um alto nível de detalhamento e precisão.

2.2.2 *Residual Encoder-Decoder Convolutional Neural Network* (RED-CNN)

Desde sua publicação em 2017, a *Residual Encoder-Decoder Convolutional Neural Network* (RED-CNN) (CHEN et al., 2017), tem sido considerada uma arquitetura inovadora no campo da redução de ruído em imagens LDCT. Este modelo é construído sobre o conceito de redes neurais residuais, estruturas que se destacam por permitir que os sinais de entrada sejam transmitidos diretamente para camadas posteriores por meio de conexões residuais. Essas estruturas facilitam o treinamento de redes mais profundas ao mitigar o problema do desaparecimento do gradiente.

Na arquitetura da RED-CNN, blocos de camadas realizam previsões residuais em relação à entrada, em vez de executar diretamente a reconstrução desejada, permitindo que o modelo aprenda a discrepância entre a saída esperada e a imagem de entrada. Em contraste com as *skip-connections*, apresentadas na U-net em 2.2.1, que concatenam as saídas de camadas distintas, a RED-CNN adota operações de soma por meio de suas conexões residuais. A Figura 3 ilustra sua arquitetura, composta pelo *encoder*, *decoder* e conexões residuais. Assim como a U-Net, essa modelo será explorado com três funções de perdas: MSE (*Mean Squared Error*), Charbonnier e perceptual; apresentadas nas seções 2.3.1 e 2.3.2.

Figura 3 – Arquitetura do modelo neural RED-CNN (CHEN et al., 2017). É possível identificar seu *encoder* constituído de convoluções normais, o *decoder* formado por convoluções transpostas, e as conexões residuais, ilustradas por setas laranjas, representando a cópia da saída de cada bloco convolucional do *encoder* para serem somadas com os dados do *decoder*. Sendo válido destacar a primeira conexão residual, que afetará diretamente a saída do *decoder* por meio da soma.



Fonte: (CHEN et al., 2017)

A combinação de conexões residuais com a arquitetura *encoder-decoder* confere à RED-

CNN eficácia notável na remoção de ruído de imagens LDCT, mantendo a integridade dos detalhes importantes. Esta característica é especialmente valiosa em contextos clínicos, onde a precisão e a qualidade da imagem são primordiais, oferecendo uma solução robusta e confiável para aprimoramento de imagem.

2.2.3 *Wasserstein Generative Adversarial Neural Network* (WGAN)

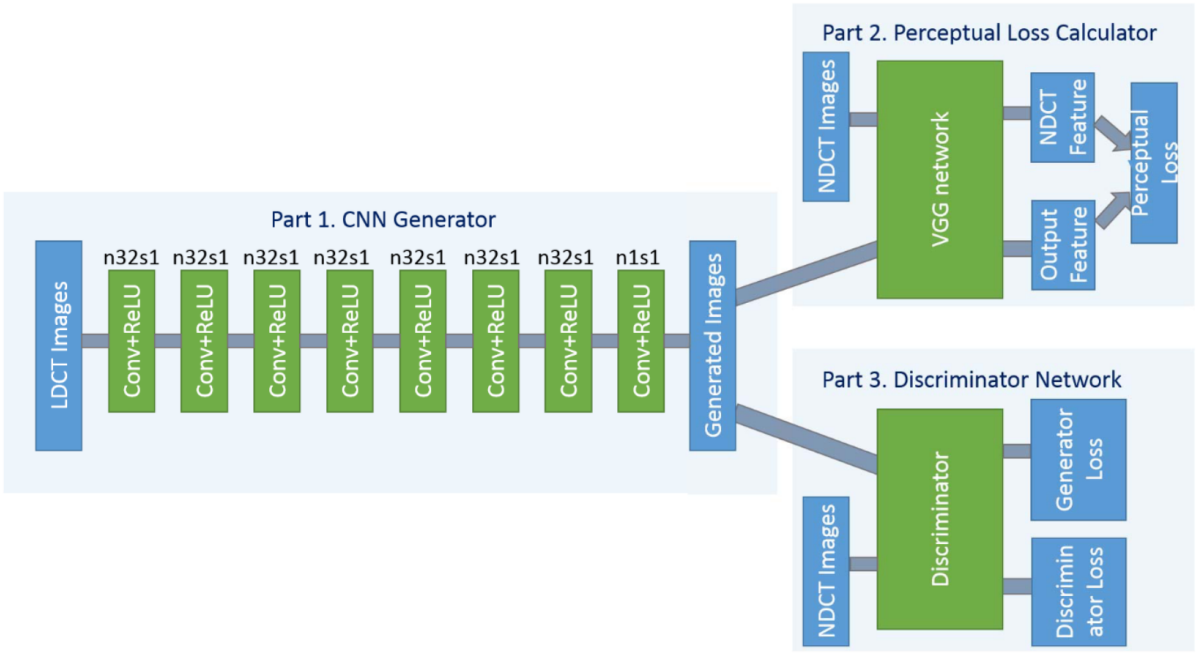
As Redes Generativas Adversárias (GANs) (GOODFELLOW et al., 2014), surgiram como um paradigma revolucionário na geração de dados sintéticos, particularmente em aplicações de imagens. Utilizando uma estrutura composta por duas redes concorrentes, um gerador e um discriminador, as GANs aprendem a produzir dados indistinguíveis dos reais, enquanto simultaneamente aprimoram a capacidade do discriminador em diferenciá-los. Esta abordagem encontrou aplicação na problemática de remoção de ruído em imagens LDCT, como evidenciado em (WOLTERINK et al., 2017), demonstrando a capacidade das GANs de sintetizar imagens de alta qualidade a partir de entradas ruidosas.

Nesse contexto, a *Wasserstein Generative Adversarial Network* (WGAN) (ARJOVSKY; CHINTALA; BOTTOU, 2017) representa uma evolução notável como uma das variantes das GANs, introduzindo uma métrica de distância mais estável e confiável para o treinamento do modelo. Especialmente na remoção de ruído em imagens LDCT (YANG et al., 2018), a WGAN evidenciou uma capacidade superior em gerar reconstruções de alta fidelidade, mitigando desafios comuns relacionados ao colapso do modelo e proporcionando uma convergência mais confiável e regular. Com o objetivo de reduzir o tempo computacional necessário para execução da restrição Lipschitz incorporada na WGAN, foi desenvolvido a implantação de uma penalidade no gradiente tornando o modelo mais eficiente (GULRAJANI et al., 2017).

A Figura 4 ilustra a arquitetura do modelo neural WGAN, proposto em (YANG et al., 2018), apresentando tanto a estruturação do modelo gerador como também o fluxo de execução dos dados sintetizados para cálculo do valor de perda pelo componente perceptual e discriminador, que serão detalhados na seções, 2.3.2 e 2.3.3 respectivamente, além da descrição da combinação dessas duas funções de perda em 2.3.4. A Figura 5 ilustra a arquitetura do modelo adversário, responsável por discriminar amostras sintetizadas e amostras NDCT.

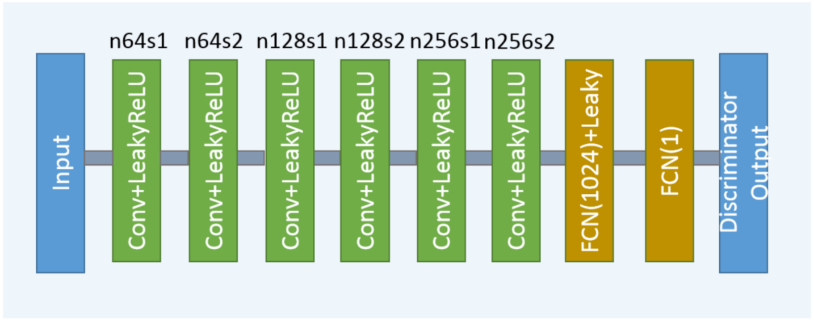
A utilização de modelos adversários, como a WGAN e suas variações, mostra grande potencial na reconstrução de imagens LDCT, pois permitem sintetizações precisas e de qualidade, essenciais para aplicações clínicas que exigem imagens médicas com alta fidelidade em detalhes.

Figura 4 – Arquitetura do modelo neural WGAN (YANG et al., 2018). É possível identificar, na *Part 1.* o modelo convolucional gerador, composto por convoluções normais, sendo a imagem sintetizada por esse gerador utilizada para cálculo da perda pela *Part 2.*, uma rede *encoder* VGG (SIMONYAN; ZISSERMAN, 2015) utilizada como função de perda perceptual, e pela *Part 3.* um modelo discriminador, apresentado na Figura 5. Ambas as funções de perdas estão melhor descritas na seção subsequente 2.3.



Fonte: (YANG et al., 2018)

Figura 5 – Arquitetura do modelo neural discriminador da WGAN (YANG et al., 2018). Responsável por discriminar imagens sintetizadas pelo modelo gerador apresentado na Figura 4 e imagens NDCT. Constituído por camadas convolucionais comuns e camadas completamente conectadas.



Fonte: (YANG et al., 2018)

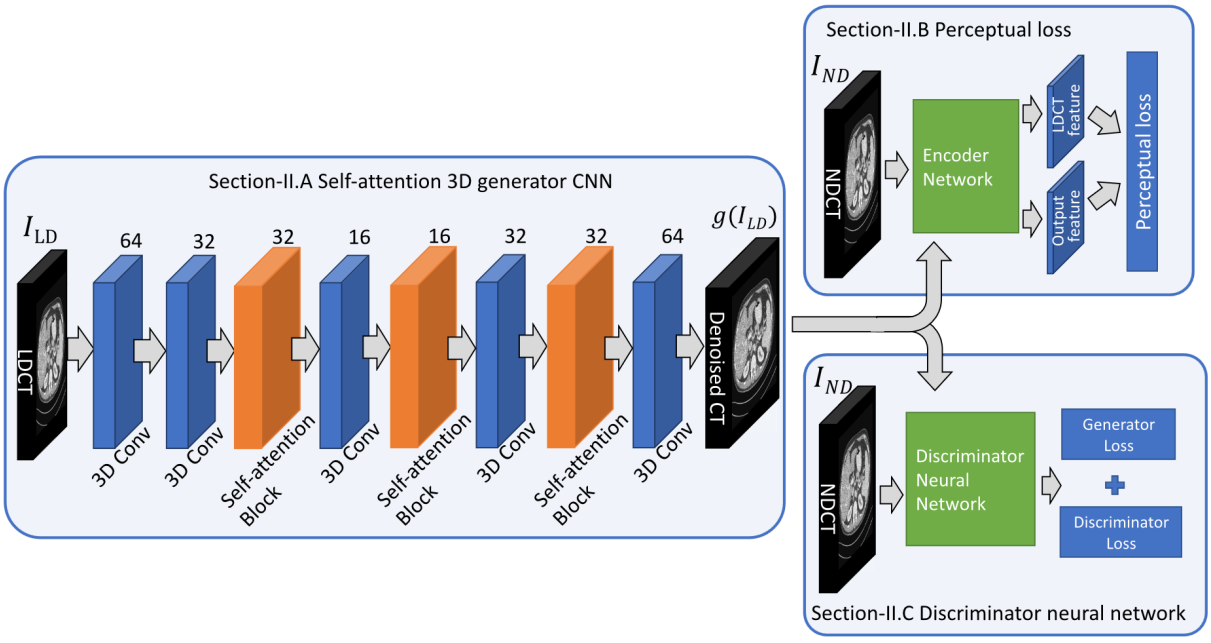
2.2.4 *Self-Attention Convolutional Neural Network* (SACNN)

A *Self-Attention Convolutional Neural Network* (SACNN) (LI et al., 2020), uma variante avançada das GANs, integra o mecanismo de autoatenção (*self-attention*) na sua arquitetura, especificamente na porção geradora da rede. O mecanismo *self-attention* confere à SACNN a habilidade de capturar dependências contextuais de longo alcance por toda a imagem, essencial para a reconstrução detalhada das imagens.

O módulo de atenção na SACNN calcula pesos de atenção com base na relação entre cada pixel da imagem com todos os outros, realizando a correlação entre linhas e colunas da imagem. Esse processo permite que o modelo priorize, durante a fase de geração, regiões da imagem que são mais informativas, melhorando o foco em áreas críticas que contribuem significativamente para a fidelidade da imagem reconstruída. A atenção para regiões relevantes da imagem possibilita um aprimoramento específico das características que mais impactam a qualidade visual da imagem resultante.

A arquitetura do SACNN, apresentada em (LI et al., 2020), é exibida na Figura 6, que não apenas mostra a estrutura do modelo gerador, mas também ilustra como os dados sintetizados passam pelo processo de avaliação de perda utilizando componentes perceptuais e discriminadores, detalhadamente descritos nas seções 2.3.2 e 2.3.3, respectivamente. Além disso, a seção 2.3.4 discute a integração dessas funções de perda. A SACNN compartilha o mesmo discriminador da WGAN aplicada ao contexto de imagens LDCT (YANG et al., 2018), já apresentado na seção 2.2.3 pela Figura 5.

Figura 6 – Arquitetura do modelo neural SACNN (LI et al., 2020). É possível identificar, na *Section-II.A* o modelo convolucional gerador, composto por convoluções e blocos de *self-attention*, sendo a imagem sintetizada por esse gerador utilizada para cálculo da perda pela *Section-II.B*, uma rede *encoder* utilizada como função de perda perceptual, e pela *Section-II.C* um modelo discriminador, sendo o mesmo utilizado pela WGAN descrita em 2.2.3 e já apresentado pela Figura 5. Ambas as funções de perdas estão melhor descritas na seção subsequente 2.3.



Fonte: (LI et al., 2020)

A introdução do módulo *self-attention* altera sua capacidade gerativa, o que se traduz na habilidade aprimorada de distinguir informações relevantes e de mitigar ruídos dispersos pela

imagem. Embora possua requisitos computacionais elevados devido a operação *self-attention*, refletindo em um alto consumo de memória e tempo de execução, esse refinamento constante das representações internas da SACNN consolida sua posição como uma ferramenta robusta para a reconstrução de imagens LDCT.

2.3 DEFINIÇÃO DAS FUNÇÕES DE PERDA

A escolha criteriosa das funções de perda é um dos aspectos fundamentais na estruturação de experimentos envolvendo o treinamento de modelos neurais, especialmente quando se trata da remoção de ruído de imagens. Em nosso estudo, a diversificação das funções de perda permite explorar as capacidades intrínsecas de diferentes arquiteturas de redes neurais, como RED-CNN (CHEN et al., 2017) e U-Net (RONNEBERGER; FISCHER; BROX, 2015), bem como modelos avançados como WGAN (YANG et al., 2018) e SACNN (LI et al., 2020), cada um com suas especificidades e adaptabilidade a abordagens de perda. Enquanto as arquiteturas RED-CNN e U-Net foram submetidas a treinamentos com funções de perda do tipo MSE (*Mean Squared Error*) (WANG; BOVIK, 2009), Charbonnier (CHARBONNIER et al., 1997), e perceptual (JOHNSON; ALAHI; FEI-FEI, 2016), os modelos WGAN e SACNN foram investigados através de suas funções de perda intrínsecas, oferecendo uma análise compreensiva sobre a eficácia de cada método na refinada tarefa de remoção de ruído em imagens de tomografia computadorizada.

2.3.1 Funções de Perda *Mean Squared Error* (MSE) e Charbonnier

No cerne do treinamento de nossos modelos neurais, as funções de perda MSE (*Mean Squared Error*) (WANG; BOVIK, 2009) e Charbonnier (CHARBONNIER et al., 1997) funcionam como fundamentos para a aprendizagem supervisionada, priorizando a minimização da discrepância entre as imagens reconstruídas e as de referência. A função de perda MSE é amplamente adotada devido à sua simplicidade e eficácia na quantificação do erro pixel a pixel entre a saída do modelo (y) e o alvo esperado ($\hat{y}$), sendo expressa pela Equação 2.1. Por outro lado, a função de perda Charbonnier, uma variante robusta e diferenciável da função $L1$, confere propriedades suavizantes e uma tendência menor a ser influenciada por outliers, o que a torna particularmente adequada para tarefas de recuperação de imagens. Esta função é formulada pela Equação 2.2, onde ϵ é um termo de suavização pequeno e constante, visando evitar a

descontinuidade no ponto zero. A inclusão desse termo ϵ na Charbonnier permite um equilíbrio eficaz entre sensibilidade aos detalhes e robustez contra variações abruptas, essencial para a melhoria da performance na remoção de ruído de imagens.

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2.1}$$

$$L_{Charb} = \sqrt{(y_i - \hat{y}_i)^2 + \epsilon^2} \tag{2.2}$$

2.3.2 Funções de Perda Perceptual

A função de perda perceptual (JOHNSON; ALAHI; FEI-FEI, 2016), diferentemente das abordagens tradicionais baseadas na exatidão pixel a pixel, incide sobre a preservação das características perceptuais essenciais das imagens. Este tipo de perda é especialmente relevante no contexto de remoção de ruído de imagens de tomografia computadorizada, onde a manutenção de detalhes precisos é de suma importância. A perda perceptual utiliza modelos pré-treinados de redes neurais profundas, como a VGG-19 (SIMONYAN; ZISSERMAN, 2015), para extrair características de alto nível das imagens comparadas, mirando em uma representação mais fidedigna do espaço de características humanamente perceptíveis (JOHNSON; ALAHI; FEI-FEI, 2016). Neste trabalho foi utilizado camada de saída "block5_conv4" da VGG-19, representado como F_{VGG} na Equação 2.3, que avalia a distância entre os mapas de características de uma imagem reconstruída e a correspondente imagem de referência.

$$L_{VGG} = ||F_{VGG}(y) - F_{VGG}(\hat{y})||^2 \tag{2.3}$$

2.3.3 Funções de Perda Adversária

A função de perda adversária constitui o núcleo das Redes Generativas Adversárias (GANs) (GOODFELLOW et al., 2014), uma arquitetura composta por duas redes em competição: um gerador (G) que sintetiza imagens, e um discriminador (D) que avalia a autenticidade das imagens sintetizadas $G(x)$ em comparação às imagens reais $\hat{y}$. Este modelo é baseado no conceito de teoria dos jogos, onde G busca maximizar a probabilidade de D cometer erros, tornando as imagens sintetizadas indistinguíveis das reais. Em termos matemáticos, a função

de perda adversária pode ser expressa pela Equação 2.4, que compreende tanto a minimização do erro de classificação pelo discriminador quanto a maximização deste erro pelo gerador. Esta dualidade incentiva o gerador a aperfeiçoar a geração de imagens, resultando em reconstruções com propriedades visuais mais realistas e diminuição dos artefatos.

$$\min_G \max_D V(D,G) = \mathbb{E}[\log D(\hat{y})] + \mathbb{E}[\log(1 - D(G(x)))] \tag{2.4}$$

A função $\log D(\hat{y})$ mede o quão eficientemente o discriminador reconhece as imagens reais como autênticas, e $\log(1-D(G(x)))$ avalia o equívoco do discriminador em relação às imagens geradas, estimulando assim o gerador a se tornar cada vez mais eficaz na criação de amostras visualmente indistinguíveis das reais.

2.3.4 Combinação de Funções de Perda

A combinação de funções de perda em modelos de aprendizado profundo apresenta-se como uma estratégia eficaz para atingir objetivos complementares durante o treinamento de redes neurais. Este processo de hibridização envolve a fusão de duas ou mais funções de perda, cada qual com seu propósito específico, regulada por parâmetros (λ), que definem a contribuição relativa de cada componente no cálculo da perda total. Neste trabalho, adotou-se essa abordagem tanto para harmonizar as características das imagens geradas quanto para guiar o início do treinamento. Em particular, a combinação da função de perda perceptual com MSE foi estrategicamente utilizada para guiar o início do treinamento dos modelos, buscando o equilíbrio entre a precisão pixel a pixel e a preservação de aspectos perceptuais críticos. A equação simplificada para este caso pode ser representada como:

$$L_{VGG\_total} = L_{MSE} + \lambda_1 L_{VGG} \tag{2.5}$$

onde L_{VGG_total} representa a função de perda total, L_{MSE} se refere à perda média quadrada, e L_{VGG} é a perda perceptual calculada com base em características extraídas da VGG-19. O parâmetro λ_1 controla a contribuição relativa da perda perceptual na função de perda total. Durante os experimentos, adotou-se $\lambda_1 = 1$ para L_{VGG_total}, inicialmente priorizando L_{MSE} devido aos seus resultados iniciais mais significativos, especialmente quando as imagens geradas ainda estão distantes do padrão de referência (*groundtruth*). Conforme o modelo avança no aprendizado, a importância de L_{VGG} aumenta gradualmente em L_{VGG_total} devido

a célere redução de L_{MSE}, oferecendo uma representação mais relevante das características perceptuais das imagens reconstruídas.

Para os modelos WGAN (2.2.3) e SACNN (2.2.4), explorou-se a combinação da função de perda adversária com a perceptual. Essa junção foi direcionada para aprimorar a qualidade visual das imagens reconstruídas, encorajando uma maior fidelidade detalhista e textural, fundamental em aplicações clínicas. A relação para essa combinação pode ser simplificada na forma:

$$L_{ADV\_total} = L_{ADV} + \lambda_2 L_{VGG\_total} \tag{2.6}$$

onde, L_{ADV} aplica-se à perda oriunda da disputa entre o gerador e o discriminador em arquiteturas GAN, e L_{VGG_total} continua a representar a perda perceptual enquanto λ_2 regula sua contribuição para a perda total L_{ADV_total}. Neste estudo dotou-se $\lambda_2 = 10$ para garantir uma equalização entre os termos L_{VGG_total} e L_{ADV} ao decorrer do treinamento. Esta abordagem direciona o treinamento de forma que valorize tanto a autenticidade das representações geradas e guie o início do treinamento pela componente L_{MSE} (pertencente a L_{VGG_total}) quanto a preservação inerente de detalhes visuais relevantes, impulsionando assim o desempenho global dos modelos.

2.3.5 Funções de Perda pixel-wise e não pixel-wise

No contexto deste trabalho, as funções de perda utilizadas para treinar os modelos neurais foram agrupadas em duas categorias principais: pixel-wise e não pixel-wise. Esta classificação é fundamental para entender o impacto de cada abordagem nas características visuais das imagens reconstruídas e na eficácia geral dos modelos na remoção de ruído de imagens de tomografia computadorizada.

As funções de perda pixel-wise, incluindo o MSE e Charbonnier, apresentados em 2.3.1, conduzem a otimização com base na discrepância pixel a pixel entre a imagem gerada pelo modelo e a imagem de referência, avaliando a precisão de forma localizada, como é possível perceber nas Equações 2.1 e 2.2. Esse método de comparação direta oferece uma métrica clara e quantificável de erro, facilitando o treinamento de redes neurais focadas em minimizar discrepâncias específicas identificadas em cada posição do pixel. Apesar de eficientes em termos métricos quantitativos, as funções de perda pixel-wise podem não capturar integralmente a

qualidade perceptual e as características essenciais das imagens médicas.

Por outro lado, as funções de perda não pixel-wise, visam superar as limitações das abordagens pixel-wise ao focar na comparação de características perceptuais de alto nível entre a imagem reconstruída e a original. Este enfoque direciona os modelos a aprenderem a reconstruir imagens assemelhando-se mais fielmente à percepção visual humana, priorizando a reconstrução de texturas, contornos, e outros detalhes de alta frequência.

A função de perda perceptual, apresentada em 2.3.2 aproveita representações internas de redes neurais pré-treinadas, como a VGG-19 (SIMONYAN; ZISSERMAN, 2015), para avaliar a semelhança entre as características extraídas de ambos, imagens geradas e de referência, permitindo uma avaliação da semelhança em um espaço de características mais abstrato. Além da perda perceptual, as funções de perda adversárias, melhor descrita em 2.3.3, também se enquadram nesta categoria, uma vez que utilizam um modelo discriminador treinado para diferenciar imagens mesmo com uma alta fidelidade visual, enriquecendo a preservação de detalhes e a qualidade perceptual geral.

2.4 HIPERPARÂMETROS *BATCH SIZE* E *SEED*

Ao abordarmos as replicações de experimentos em redes neurais, nos deparamos frequentemente com hiperparâmetros selecionados de modo facultativo, muitas das vezes porque não há as especificações descritas nos trabalhos referenciais. Dois hiperparâmetros exemplificativos dessa problemática são o *batch size* e a *seed*, que frequentemente são ajustados sem um critério rigoroso, incluindo variações substanciais que podem comprometer a consistência dos resultados entre diferentes replicações de um mesmo experimento, afetando a confiabilidade dos resultados obtidos.

O *batch size*, contempla o número de amostras de dados processadas antes da atualização dos parâmetros do modelo em cada iteração. Valores maiores proporcionam gradientes de erro mais estáveis devido à variedade de amostras contidas no lote, o que contribui para uma otimização mais eficaz. Por outro lado, valores menores podem resultar em uma convergência mais lenta e até mesmo irregular, em função da reduzida diversidade de amostras presentes no lote. Já a *seed*, ou semente, é essencial para a geração de números pseudoaleatórios, utilizada na inicialização dos pesos da rede neural e em todas as operações que requerem aleatoriedade. A escolha de diferentes *seeds* pode induzir a variações significativas durante o treinamento, exercendo influência direta sobre o desempenho e a capacidade de generalização do modelo.

Neste trabalho, daremos um enfoque especial ao impacto do *batch size* e da *seed* na replicação dos experimentos, ilustrando como esses hiperparâmetros podem contribuir para as divergências observadas quando diferentes pesquisadores tentam reproduzir resultados prévios. Além destes, é importante destacar que outros fatores como métodos de inicialização dos pesos dos modelos, a taxa de aprendizagem (*learning rate*), a ordem de apresentação dos dados, bem como divergências em versões de softwares e hardwares, influenciam igualmente a consistência de resultados entre replicações de estudos. Todos esses elementos compõem o conjunto de variáveis que serão consideradas minuciosamente no decorrer desta investigação, para garantir uma análise mais precisa e controlada dos modelos neurais empregados na redução de ruído em imagens de tomografia computadorizada.

3 METODOLOGIA

Neste capítulo, adentramos na estrutura metodológica detalhada para condução dos experimentos neste estudo. Inicialmente, abordamos a base de dados e o pré-processamento das imagens. A seguir, configurações experimentais e definição dos hiperparâmetros são exploradas para garantir uma abordagem abrangente e rigorosa. No âmbito da implementação dos experimentos, abordamos o ambiente computacional, a reprodutibilidade dos resultados, e os procedimentos de treinamento e validação dos modelos. Esta metodologia integrada visa fornecer um arcabouço robusto para a avaliação quantitativa e observação qualitativa dos modelos neurais, a fim de fornecer conclusões sobre a remoção de ruído em imagens de tomografia computadorizada.

3.1 BASE DE DADOS

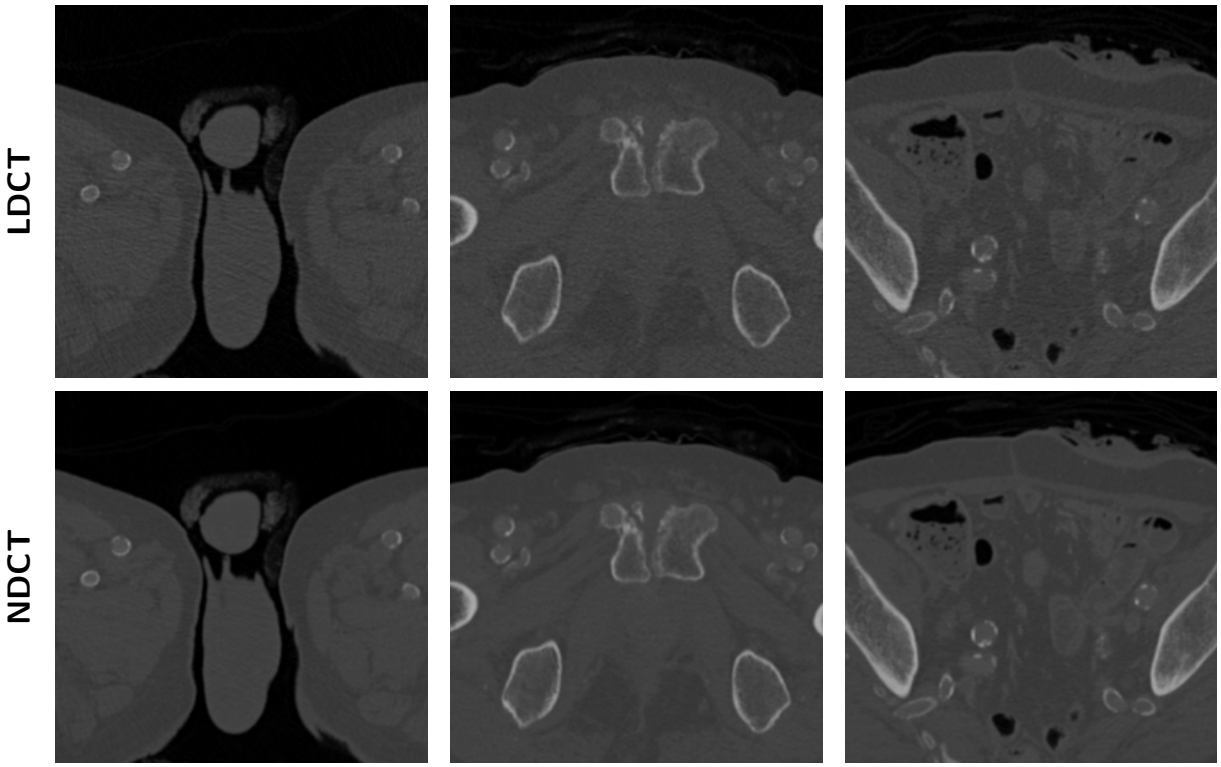
3.1.1 *Mayo Challenge Dataset*

O *Low Dose CT Grand Challenge* (AMERICAN ASSOCIATION OF PHYSICISTS IN MEDICINE (AAPM), 2016), organizado pela *American Association of Physicists in Medicine* (AAPM) em 2016, foi uma iniciativa destinada a avaliar quantitativamente o desempenho de técnicas de redução de ruído e reconstrução iterativa em conjuntos de imagens de CT de baixa dosagem. O objetivo principal para os participantes era minimizar o ruído nas imagens, visando otimizar a detecção de lesões hepáticas. Para simular condições de dose reduzida, inseriu-se ruído de Poisson nos dados de projeção, simulando uma imagem reconstruída com 25% da dose original. Os dados de projeção incluíam casos de 10 pacientes, abrangendo lesões sutis e típicas, além de casos sem lesões, para testar a eficácia das técnicas de *denoising* em uma variedade de condições. Por esses motivos, a base de dados *Mayo Challenge* é compatível com os interesses do presente estudo. Na Figura 7 mostramos alguns exemplos que compõem a base de dados, exemplificando a versão com ruído (LDCT), que representa a captura com uma baixa dosagem de radiação, e sem ruído (NDCT), representando a captura com a dosagem normal de radiação. A diferença entre esses dois tipos de aquisição e suas vantagens e desvantagens foi melhor abordada em 2.1.

O ruído de Poisson é uma característica inerente ao processo de detecção de fótons em imagens de CT. Este tipo de ruído é proporcional à raiz quadrada do número de eventos

contados, o que significa que quanto maior o sinal (ou dose), menor a influência relativa do ruído de Poisson sobre a qualidade da imagem. No contexto da competição, a inserção desse ruído foi utilizada para simular imagens de CT de baixa dosagem, replicando as condições encontradas em exames com redução de exposição à radiação. Esse método permite avaliar e aprimorar técnicas de redução de ruído e reconstrução de imagens, mantendo a segurança do paciente através da minimização da exposição à radiação, como foi apresentado em 2.1.

Figura 7 – Exemplos de imagens que compõem o *dataset* da competição *Low Dose CT Grand Challenge*, após realizar um recorte central (256x256) para melhor visualização. A primeira linha representa a imagem gerada a partir de projeções corrompidas com ruído de Poisson, representando a captura da tomografia com uma baixa dosagem de radiação (LDCT). Enquanto que a segunda linha representa a versão das mesmas imagens sem o ruído, sendo este o resultado da captura da tomografia com a dosagem normal de radiação (NDCT). O processo de captura das imagens de CT está melhor descrito em 2.1.



Fonte: Imagem produzida pelo autor.

Para a realização dos experimentos, foi selecionado o subconjunto *1mm B30* pertencente à base de dados *Mayo Challenge*. Especificamente, foram escolhidas amostras de 7 pacientes (*L067, L310, L286, L333, L291, L109* e *L143*) para o conjunto de treinamento, 1 paciente (*L192*) para a validação e 2 pacientes (*L096 e L506*) para o conjunto de teste. Essa seleção estratégica dos pacientes garantiu a representatividade e diversidade necessárias para avaliar a eficácia dos modelos.

3.1.2 Pré-processamento de Dados

No que tange ao pré-processamento dos dados utilizados neste estudo, uma atenção especial foi dada para assegurar a integridade e a padronização dos dados, facilitando assim a execução dos experimentos. Felizmente, a base de dados selecionada já se encontrava normalizada no intervalo de 0 a 1, um requisito preliminar para a adequada alimentação dos modelos neurais escolhidos. Porém, considerando as limitações de memória gráfica inerentes ao treinamento de redes convolucionais profundas e a necessidade de empregar valores maiores de *batch size*, optou-se por subdividir as imagens iniciais, que originalmente apresentavam uma resolução de 512x512, em *patches* menores de 128x128 pixels. Esta abordagem permitiu uma multiplicação efetiva do número de amostras disponíveis para treinamento, criando assim 16 quadrantes distintos por imagem.

Adicionalmente, com o objetivo de otimizar o desempenho computacional e mitigar o significativo gargalo de tempo associado ao carregamento de dados durante os treinamentos, foi empregada a ferramenta FFCV (LECLERC et al., 2023). Através do uso de técnicas de paralelização e pré-carregamento, essa ferramenta propiciou uma redução substancial no tempo total de treinamento, excedendo uma economia de 50% em comparação com abordagens tradicionais de carregamento de dados, garantindo assim uma maior eficiência e agilidade na fase experimental deste trabalho.

3.2 METODOLOGIA DE EXPERIMENTAÇÃO

3.2.1 Justificativa da Escolha das Arquiteturas de Redes Neurais

A escolha das arquiteturas de redes neurais para este estudo foi pautada por modelos do estado-da-arte dentro do campo de processamento de imagens de tomografia computadorizada. Foi dada ênfase às capacidades distintas dos modelos em manter características, que conferem na respectiva imagem capturada com dosagem normal de radiação, ao mesmo tempo que buscam atenuar os artefatos intrínsecos ao processo de captura de imagem com menor exposição à radiação.

Para representar esse espectro, foram incorporadas arquiteturas consagradas e inovadoras como o RED-CNN (CHEN et al., 2017), U-Net (RONNEBERGER; FISCHER; BROX, 2015), WGAN (YANG et al., 2018) e SACNN (LI et al., 2020), melhor apresentadas em 2.2.2, 2.2.1, 2.2.3 e 2.2.4,

respectivamente. Esses modelos foram selecionados tendo em vista não só suas contribuições metodológicas para a área de aprendizagem profunda, mas também pela compatibilidade com uma variedade de funções de perda, para diversificação dos experimentos. Por exemplo, RED-CNN e U-Net possuem a flexibilidade necessária para operar com diferentes funções de perda, tais como MSE (WANG; BOVIK, 2009), Charbonnier (CHARBONNIER et al., 1997) e Perceptual (JOHNSON; ALAHI; FEI-FEI, 2016), como melhor abordado na seção subsequente. Ao incorporar essas arquiteturas ao nosso conjunto de análise, almejamos estabelecer uma base sólida para comparações significativas e clareza sobre as melhores práticas neste campo essencial da inteligência artificial aplicada à remoção de ruído de imagens de CT.

3.2.2 Otimização e Hiperparâmetros

Na configuração dos experimentos para o treinamento dos modelos neurais abordados neste estudo, foi adotado o algoritmo de otimização Adam, amplamente utilizado devido à sua eficiência. Os hiperparâmetros específicos do Adam foram estabelecidos como $\beta_1 = 0,9$ e $\beta_2 = 0,999$, configurações recomendadas por padrão que demonstram um bom equilíbrio entre a aceleração da convergência inicial e a estabilização da convergência nos estágios posteriores do treinamento. A taxa de aprendizagem, outro componente crítico na eficácia do processo de otimização, foi fixada em $1e-4$ para todos os modelos, incluindo os adversários. Por fim, o número de épocas selecionado foi 30 para todas as execuções experimentais, que se revelou suficientemente amplo para permitir que as redes atingissem um estado de convergência, ao mesmo tempo que garantiu uniformidade nas condições de treinamento entre os distintos modelos ao escolher a última época treinada para avaliação dos mesmos.

Essa padronização de hiperparâmetros contribui para uma comparação equitativa do desempenho dos modelos, centrando o foco nas diferenças arquitetônicas e as funções de perda específicas sem discrepâncias advindas de métodos de otimização. A seguir, no detalhamento dos conjuntos experimentais, será explorada a implementação de configurações variadas através da manipulação de *seed* e *batch size*.

3.2.3 Detalhamento das Conjuntos Experimentais

Para assegurar a robustez e a consistência dos resultados obtidos, foi imprescindível a criação de um esquema experimental diversificado. A estruturação deste esquema se deu

através da implementação sistemática de diferentes configurações experimentais, moduladas pela variação de dois eixos de configuração cruciais: *batch size* e *seed*, melhor descritos em 2.4.

A seleção dos valores de *seed* (10 e 11) foi uma decisão estratégica tomada para investigar a sensibilidade dos modelos em relação à inicialização aleatória, tendo em vista que diferentes inicializações podem potencialmente levar a trajetórias de aprendizado distintas, e, consequentemente, a resultados divergentes. A inclusão de uma terceira *seed* (12) foi prevista como um mecanismo de contingência, para assegurar a obtenção de um volume de dados satisfatório para análise, especialmente nos casos em que determinadas configurações de treinamento resultam em aprendizado inadequado, identificado por um modelo incapaz de aprender a tarefa proposta com a configuração vigente.

Em relação ao *batch size*, os valores determinados foram 128, 64, 32, 16 e 8. Esta gama de tamanhos oferece um amplo espectro para análise, desde lotes maiores, que proporcionam um gradiente de erro mais estável, mas exigem maior capacidade computacional, até lotes menores, que podem aumentar a variabilidade do treinamento, mas são menos demandantes em termos de memória gráfica. A SACNN, devido à sua expressiva utilização de memória gráfica provocada pelo mecanismo de atenção, como já antecipado em 2.2.4, exigiu um ajuste no *batch size*. A relação adotada para a SACNN foi: 16, 12, 8, 4 e 2. Este ajuste visou assegurar a viabilidade do treinamento desta arquitetura em hardware disponível. Além disso, foi necessário reduzir a complexidade da U-Net dividindo o número de *kernels* de suas camadas internas por 4, devido à incapacidade do modelo, em sua configuração padrão (RONNEBERGER; FISCHER; BROX, 2015), de aprender ao utilizar funções de perda pixel-wise. Mesmo após 10 execuções distintas (com variação de 5 valores de *batch size* e 2 valores de *seed*), o modelo não convergia até que essa redução fosse feita.

A implementação dessas configurações, variando tanto o *seed* quanto o *batch size*, criou um conjunto amostral compreensivo de 10 execuções por modelo, proporcionando uma base de dados rica e diversificada para análise. A inclusão de uma terceira *seed* (12), empregada estrategicamente para assegurar a obtenção de pelo menos 8 execuções exitosas por modelo, foi um mecanismo adicionado para garantir que cada modelo tivesse representação suficiente em nossas análises. Esse esquema experimental assegura o rigor na avaliação e na compreensão dos resultados, favorecendo uma interpretação abrangente sobre o desempenho das arquiteturas neurais.

Neste contexto, as combinações entre modelos, descritos em 2.2, e funções de perda,

descritas em 2.3, surgem como um forte alicerce para a compreensão profunda dos efeitos distintos que cada parâmetro pode exercer sobre a qualidade final das imagens reconstruídas. Especificamente, o emprego de arquiteturas como RED-CNN e U-Net, com as funções de perda MSE, Charbonnier e Perceptual, representa uma tentativa de abordar a problemática sob variadas perspectivas, divergindo tanto na abordagem de otimização pixel-wise quanto na avaliação baseada em características perceptuais. Ademais, a inclusão dos modelos WGAN e SACNN, introduz um espectro ainda mais amplo de análises possíveis. Formando assim 8 combinações, esta sistemática prevê entender o comportamento individual de cada abordagem, construindo assim um estudo abrangente e holístico sobre a remoção de ruído em imagens de CT, crucial para aplicações clínicas futuras.

3.3 IMPLEMENTAÇÃO DOS EXPERIMENTOS

3.3.1 Ambiente Computacional: Hardware e Software

Os experimentos foram conduzidos em um sistema equipado com uma placa gráfica Nvidia RTX 2080 Ti e um processador Intel i9-9900k. Quanto ao software, o ambiente operacional adotado foi o Linux Ubuntu 22.04. As principais ferramentas utilizadas incluíram Driver Nvidia v535.129.03, CUDA v11.2, cuDNN v8, Python v3.10.12, TensorFlow v2.11.0 e FFCV v0.0.3.

3.3.2 Reprodutibilidade

A reprodutibilidade dos experimentos é um componente essencial para garantir a confiabilidade e a validade dos resultados obtidos. Uma das estratégias adotadas para assegurar a consistência nos experimentos realizados foi manter as mesmas amostras e a mesma ordem de apresentação dos dados em todas as execuções. Vale ressaltar que a variação dos valores das *seeds* não afetou a ordem de apresentação dos dados, mas sim influenciou a construção dos modelos, que para a inicialização dos seus pesos, foi utilizada a função *Glorot Uniform* [1] em conjunto com as *seeds* específicas do experimento, enquanto os *bias* foram inicializados com zeros.

O esforço dedicado à reprodutibilidade dos experimentos foi priorizado, buscando resolver questões de não determinismo nos *frameworks* utilizados. Uma abordagem adotada foi a desa-

[1] <https://www.tensorflow.org/api_docs/python/tf/keras/initializers/GlorotUniform>

tivação de otimizações específicas que realizam operações não determinísticas devido à ordem e precisão das execuções. A desativação dessas otimizações, embora tenha estendido o tempo de treinamento em pelo menos o dobro, permitiu que em múltiplas execuções de um mesmo experimento os mesmos resultados fossem gerados no mesmo hardware. Vale ressaltar que diferentes hardwares ainda podem gerar resultados diferentes, mas essa abordagem contribuiu significativamente para garantir a consistência e a replicabilidade dos resultados obtidos, essenciais para a validação e interpretação adequada dos experimentos realizados. (TENSORFLOW, 2024)

3.4 AVALIAÇÃO QUANTITATIVA

3.4.1 Descrição e Fundamentação das Métricas Utilizadas

Na avaliação da performance de modelos neurais destinados à redução de ruído em imagens de CT, a aplicação de métricas quantitativas assume um papel fundamental. Entre as métricas predominantes, destacam-se a *Peak Signal-to-Noise Ratio* (PSNR), *Structural Similarity Index* (SSIM) e a *Root Normalize Mean Square Error* (NRMSE). A seguir, discutiremos cada uma dessas métricas, relacionando as predições de imagens NDCT (y) em comparação às suas respectivas imagens reais NDCT ($\hat{y}$) para todos os N pixels que compõem as imagens.

- A *Normalized Root Mean Squared Error* (NRMSE) quantifica a disparidade entre dois sinais baseando-se no quadrado da subtração entre eles, logo, busca-se minimizar esse valor. Mais precisamente, é a normalização da raiz quadrada da média do erro quadrático entre os pixels das imagens y e $\hat{y}$:

$$NRMSE(y, \hat{y}) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\hat{y}_{max} - \hat{y}_{min}}, \tag{3.1}$$

- A *Peak-to-Signal Noise Ratio* (PSNR) mensura a fidelidade do sinal com respeito ao nível de ruído, logo, busca-se maximizar esse valor. Calculada pela divisão entre o quadrado do maior valor possível que um pixel da imagem $\hat{y}$ pode possuir, sobre o erro quadrático médio entre os pixels das imagens y e $\hat{y}$:

$$PSNR(y, \hat{y}) = 10 \cdot \log_{10}\left(\frac{MAX_{\hat{y}}^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}\right). \tag{3.2}$$

- A *Structural Similarity Index Measure* (SSIM) (WANG et al., 2004) atua qualificando a degradação da informação estrutural entre dois sinais, logo, busca-se maximizar esse valor. Para isso, essa métrica realiza a avaliação entre as imagens y e $\hat{y}$ relacionada às características de luminescência (l), contraste (c) e estrutura (s):

$$
\begin{aligned}
SSIM(y,\hat{y}) &= [l(y,\hat{y})^{\alpha} \cdot c(y,\hat{y})^{\beta} \cdot s(y,\hat{y})^{\gamma}], \\
l(y,\hat{y}) &= \frac{2\mu_y\mu_{\hat{y}} + c_1}{\mu_y^2 + \mu_{\hat{y}}^2 + c_1}, \\
c(y,\hat{y}) &= \frac{2\sigma_y\sigma_{\hat{y}} + c_1}{\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2}, \\
s(y,\hat{y}) &= \frac{\sigma_{y\hat{y}} + c_3}{\sigma_y\sigma_y + c_3},
\end{aligned}
\tag{3.3}
$$

  onde μ representa a média do sinal y ou $\hat{y}$, σ^2 é a variância dessa média e $\sigma_{y\hat{y}}$ é a covariância. As variáveis c_1, c_2 e c_3 tem o papel de estabilizar as divisões, e α, β e γ são pesos usualmente utilizados como 1.

Em resumo, as métricas PSNR, SSIM e NRMSE são influenciadas por modelos treinados com funções de perda do tipo pixel-wise, como destaca a literatura (YANG et al., 2018; LI et al., 2020). Modelos otimizados com MSE, por exemplo, tendem a apresentar altos valores de PSNR e SSIM, e baixos valores de NRMSE devido à minimização direta do erro pixel a pixel. É evidente que as métricas PSNR e NRMSE focam na semelhança pixel a pixel, no entanto, a SSIM, apesar de avaliar aspectos de estrutura e luminância, ainda favorece imagens geradas por modelos treinados com funções de perda pixel-wise, pois ela se concentra na diferença média entre as imagens (LI et al., 2020). Portanto, é essencial considerar que altas pontuações nessas métricas podem ser um reflexo do treinamento de modelos com funções de perda do tipo pixel-wise, como será discutido a seguir ao abordar o Efeito MSE.

3.4.2 Protocolo de Validação e Análise Estatística

Neste estudo, adotou-se um protocolo de validação planejado para avaliar o desempenho dos modelos de redução de ruído das imagens. Fundamental para a integridade e fidelidade das análises, o protocolo estabelece que a avaliação das métricas de desempenho será realizada exclusivamente no término do processo de treinamento, mais precisamente na 30ª e última época. Esta abordagem garante que os modelos tenham tido oportunidade ampla para

convergir, fornecendo um quadro de referência consistente para uma comparação justa entre os diversos experimentos.

Adicionalmente, faz-se importante ressaltar a exclusão das execuções nas quais os modelos não demonstram qualquer aprendizado, uma medida crucial para assegurar a precisão da análise. Tal decisão, embasada nos critérios delineados em 3.2.3, que demanda pelo menos 8 treinamentos exitosos, visa eliminar dados que poderiam distorcer os resultados, mantendo o foco nas configurações de treinamento que efetivamente contribuíram para a melhoria do desempenho da rede, representando o cenário de êxito na tentativa de execução do modelo replicado.

Para a análise estatística dos resultados obtidos, empregou-se o Teste de Friedman, um método não paramétrico concebido para avaliar diferenças entre múltiplas distribuições. A escolha desse teste é particularmente pertinente para nossa pesquisa, tendo em vista sua capacidade de identificar desvios significativos no desempenho entre os modelos sob condições experimentais variáveis, sem inferir suposições acerca da distribuição normal dos dados. O intuito primordial deste teste em nosso estudo é verificar a equivalência estatística entre as métricas dos modelos, permitindo um discernimento mais acurado sobre a eficácia relativa das diferentes arquiteturas neurais e funções de perda adotadas.

Na aplicação do Teste de Friedman, cada grupo experimental é formado por sua combinação de arquitetura, função de perda, *batch size* e *seed*, como detalhado em 3.2.3. A hipótese nula do teste sugere que todos os modelos exibem um desempenho equivalente, enquanto a rejeição dessa hipótese indicaria variações estatisticamente significativas, justificando uma investigação mais detalhada para identificar as configurações que destacam-se positiva ou negativamente.

3.4.3 Comparação de Desempenho: Tabelas de Métricas e Boxplots

A abordagem tradicional de comparar o desempenho de modelos de inteligência artificial, particularmente no campo da redução de ruído em imagens médicas, tem sido predominantemente realizada por meio de tabelas métricas baseadas em execuções únicas de modelos. Essa metodologia, embora possa proporcionar uma visão do desempenho do modelo, falha em capturar a variabilidade inerente aos processos de treinamento de aprendizado profundo, incluindo as influências de inicializações aleatórias, como o valor de *seed* e funções de inicialização dos pesos do modelos neurais, e a seleção de hiperparâmetros e configurações experimentais, como

o *batch size*, taxa de aprendizagem, versão se *software* e *hardware*. Consequentemente, este método pode levar a conclusões potencialmente enganosas sobre a eficácia de modelos diferentes ou configurações de treinamento, ocultando discrepâncias que surgem devido a estes e outros fatores estocásticos.

Por outro lado, considerar uma amostra variada de execuções que manipulam cuidadosamente parâmetros básicos como *seed* e *batch size*, emerge como uma estratégia mais robusta e equitativa para a avaliação de desempenho de modelos. A utilização de boxplots para representação dos resultados, por sua natureza, oferecem uma visão rica sobre a distribuição dos dados, proporcionando um entendimento mais amplo e detalhado da performance dos modelos em diferentes cenários e condições, destacando a mediana, os quartis e possíveis valores discrepantes (*outliers*).

Adotar uma abordagem mais abrangedora e metódica, tal como a exploração de variações experimentais de um modelo e utilização de boxplots em complemento a análises métricas tradicionais, é de suma importância para uma avaliação acurada dos modelos de aprendizado de máquina. Isso não apenas demonstra a consistência ou variabilidade do desempenho de um modelo sob diversas condições de execução, como também permite uma comparação mais justa entre diferentes arquiteturas e configurações. A rigidez e a variabilidade do aprendizado profundo requerem uma consideração cuidadosa para garantir que as conclusões sobre a eficácia dos modelos sejam fundamentadas em uma base estatística sólida e representativa da real capacidade dos modelos.

Portanto, a transição para uma avaliação de desempenho baseada em uma nuvem de resultados associados à capacidade de um modelo, constitui um avanço importante na pesquisa em aprendizado de máquina. Esse avanço não apenas melhora nossa confiança nas conclusões extraídas dos experimentos, mas também fornece esclarecimentos detalhados que são cruciais para o avanço e a aplicação efetiva das tecnologias de inteligência artificial no domínio das imagens médicas e além.

3.5 OBSERVAÇÃO QUALITATIVA

3.5.1 Definição do Efeito MSE

O Efeito MSE, comumente observado em modelos de aprendizado profundo treinados com funções de perda focadas na precisão pixel a pixel (YANG et al., 2018; LI et al., 2020; YU et al.,

2017; WOLTERINK et al., 2017), leva a uma consequência indesejada de borramento nas imagens geradas. Este fenômeno resulta na perda de detalhes finos, particularmente nos componentes de alta frequência, como textura e bordas precisas dos objetos, levando a uma característica de uniformização indesejável na superfície dos tecidos presentes nas imagens médicas (LI et al., 2020). Tal efeito não só compromete a fidelidade visual das imagens reconstruídas, mas também pode mascarar ou alterar características vitais para a interpretação e diagnóstico médico preciso. (YANG et al., 2018)

A formação do Efeito MSE está intrinsecamente ligada ao critério de otimização adotado pelas funções de perda pixel-wise. Conforme apontado em pesquisas (YANG et al., 2018; LI et al., 2020), ao minimizar as diferenças diretas entre os pixels da imagem gerada e da imagem de referência, essas funções inadvertidamente favorecem soluções que nivelam variações sutis de intensidade, essenciais para a representação de texturas e bordas. Esse comprometimento se traduz na produção de imagens "suavizadas", onde o contraste entre diferentes estruturas teciduais pode ser significativamente reduzido, levando a uma perda de informação crítica.

Investigações teóricas e experimentais (YU et al., 2017) elucidam que a manifestação do Efeito MSE origina-se na busca dos modelos treinados por soluções que estritamente atendam aos critérios do MSE. Essas soluções, embora otimamente alinhadas ao objetivo quantitativo de minimização do erro médio quadrático, frequentemente não refletem a complexidade e a integridade perceptual dos dados reais. A aplicação de funções de perda perceptual surge como um contraponto estratégico a esse problema, promovendo o aprendizado de características cruciais, que melhor replicam a percepção visual humana e mantêm a integridade dos detalhes mais sutis, além de induzir a uma menor incidência de artefatos irregulares nos resultados, como melhor apresentado na seção subsequente (LI et al., 2020). Entretanto, mesmo com uma reconstrução mais fiel à percepção humana, métricas qualitativas podem ser menores quando utiliza-se funções de perda perceptuais (YANG et al., 2018; LI et al., 2020), como melhor descrito em 3.4.1.

3.5.2 Importância da Qualidade Perceptual em Imagens Médicas

A consideração das funções de perda perceptuais apresenta-se não somente como uma abordagem técnica inovadora, mas também como um requisito fundamental para elevar tanto a qualidade quanto a aplicabilidade clínica das imagens médicas produzidas por técnicas de aprendizado profundo. A implementação dessas funções de perda é estratégica no combate

ao Efeito MSE, com o objetivo explícito de mitigar suas manifestações indesejadas e, simultaneamente, garantir a manutenção de nuances significativas nas imagens. Tal abordagem é corroborada por estudos (YANG et al., 2018; LI et al., 2020; WOLTERINK et al., 2017; YU et al., 2017), os quais não apenas confirmam a viabilidade e eficácia dessas funções na superação dos desafios impostos pelo Efeito MSE, mas também destacam a importância vital de adotar estratégias que priorizem a fidelidade perceptual. Essas estratégias asseguram a preservação de características imprescindíveis para a interpretação diagnóstica e análise médica, reforçando a premissa de que a qualidade das imagens médicas não depende apenas de métricas quantitativas, mas também da riqueza de detalhes cruciais (tratados pelos trabalhos como "realistas").

Esta necessidade de manter informações cruciais torna a escolha da função de perda um aspecto decisivo na determinação da utilidade das imagens reconstruídas. A eficácia das funções de perda perceptuais, realça o potencial desses métodos no fortalecimento da precisão diagnóstica através da melhoria perceptual das imagens. A pesquisa conduzida em (WOLTERINK et al., 2017) enfatiza um aspecto notável: um gerador de GAN, orientado unicamente por perda adversarial, manifestou habilidade para atenuar de forma significativa o ruído em imagens CT cardíacas de baixa dosagem, conservando a fidelidade dos valores de tecidos, sustentando a validade das informações diagnósticas, além de reforçar a visão de que a perda perceptual promove a retenção de características essenciais da imagem.

Embora métricas quantitativas como PSNR e SSIM possam fornecer uma avaliação objetiva da fidelidade de reconstrução das imagens, existe uma discrepância notável entre essas avaliações e a qualidade visual percebida, particularmente no contexto de imagens médicas. Tal desconexão é crítica na aplicação de técnicas avançadas de reconstrução de imagem, onde a preferência das métricas quantitativas por soluções que minimizem o erro pixel a pixel pode comprometer a utilidade das imagens geradas. Este dilema é abordado em (YANG et al., 2018), indicando que, embora PSNR e SSIM sejam indicadores úteis, eles podem não ser inteiramente suficientes para avaliar a complexidade das reconstruções em imagens de tomografia computadorizada de baixa dosagem, reforçando a importância de adotar abordagens que priorizem a qualidade perceptual e a preservação de informações vitais.

4 RESULTADOS

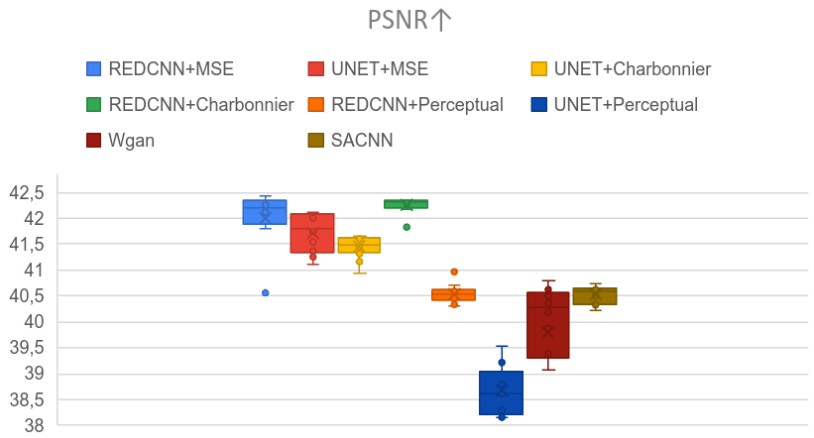
Neste capítulo, desdobram-se os resultados advindos da aplicação rigorosa da metodologia delineada no capítulo anterior, usando métricas como PSNR, SSIM e NRMSE. Os resultados das análises quantitativa e qualitativa realçam tanto as peculiaridades dos modelos com funções de perda não pixel-wise quanto o impacto visível do Efeito MSE nas imagens geradas por funções de perda pixel-wise. Ademais, a investigação dos efeitos causados por variáveis de hiperparâmetros enfatizam a complexidade e os desafios enfrentados, como a realização de uma comparação justa entre modelos neurais.

4.1 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS QUANTITATIVOS

Neste segmento do capítulo, concentramo-nos na exposição e discussão dos dados quantitativos colhidos durante a fase de experimentação, com especial atenção aos resultados de treinamento dos modelos de acordo com as métricas de PSNR, SSIM e NRMSE. Importante salientar que estas métricas foram selecionadas por sua relevância e capacidade de fornecer percepções detalhadas sobre a qualidade de reconstrução das imagens processadas pelos modelos. A apresentação desses resultados é realizada mediante o uso de gráficos boxplot, que fornecem uma rica visão comparativa entre os dois grupos de funções de perda: pixel-wise e não pixel-wise.

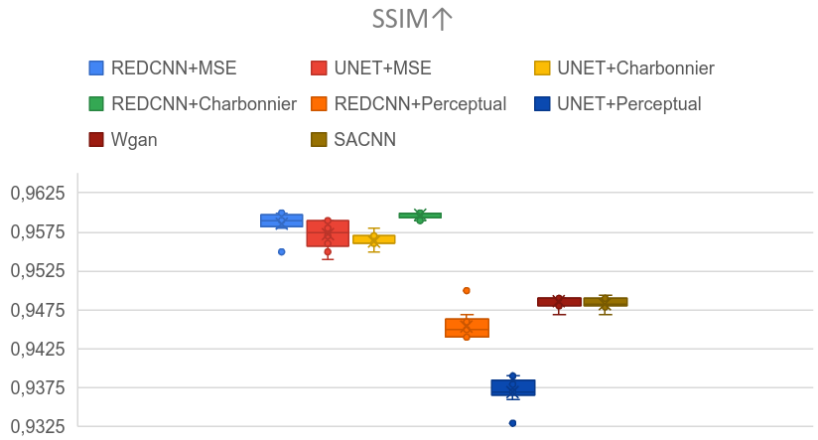
A partir dos boxplots apresentados nas Figuras 8, 9 e 10, e como complemento, os valores de desvio padrão expostos na Tabela 1 das três métricas (PSNR, SSIM e NRMSE) dos conjuntos experimentais descritos em 3.2.3 e expostos mais detalhadamente no Anexo A, é possível observar os resultados quantitativos globais dos experimentos. Desse modo, um padrão comum emerge claramente na análise dos boxplots, evidenciando a diferença de desempenho entre os modelos treinados com funções de perda pixel-wise e aqueles treinados com funções de perda não pixel-wise (estas funções estão detalhadas em 2.3.5). Essa diferença notável não apenas sublinha a influência significativa que a escolha da função de perda pode ter no resultado final das imagens reconstruídas, mas também destaca a particularidade de cada métrica na captura desses efeitos, como destacado em 3.4.1.

Figura 8 – Gráfico boxplot para a métrica PSNR.



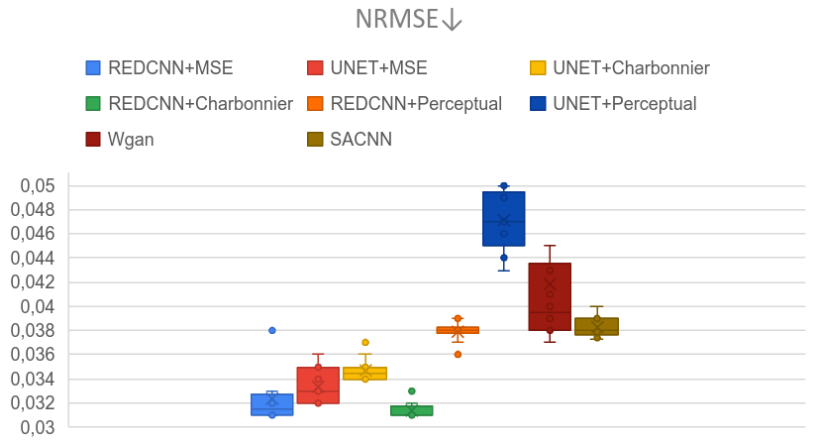
Fonte: Imagem produzida pelo autor.

Figura 9 – Gráfico boxplot para a métrica SSIM.



Fonte: Imagem produzida pelo autor.

Figura 10 – Gráfico boxplot para a métrica NRMSE.



Fonte: Imagem produzida pelo autor.

Tabela 1 – Desvio padrão (STD, do inglês *standard deviation*) das métricas PSNR, SSIM e NRMSE para as execuções de cada combinação (modelo + função de perda) testada. Complementando os resultados apresentados nos boxplots destas mesmas métricas nas Figuras 8, 9 e 10.

Modelo + *Loss*	**STD PSNR**	**STD SSIM**	**STD NRMSE**
RED-CNN + MSE	0,5747	0,0014	0,0022
U-Net + MSE	0,3577	0,0016	0,0014
U-Net + Charbonnier	0,2162	0,0007	0,0009
RED-CNN + Charbonnier	0,1656	0,0004	0,0006
RED-CNN + Perceptual	0,1782	0,0014	0,0008
U-Net + Perceptual	0,4571	0,0016	0,0024
WGAN	1,1648	0,0006	0,0062
SACNN	0,1705	0,0006	0,0008

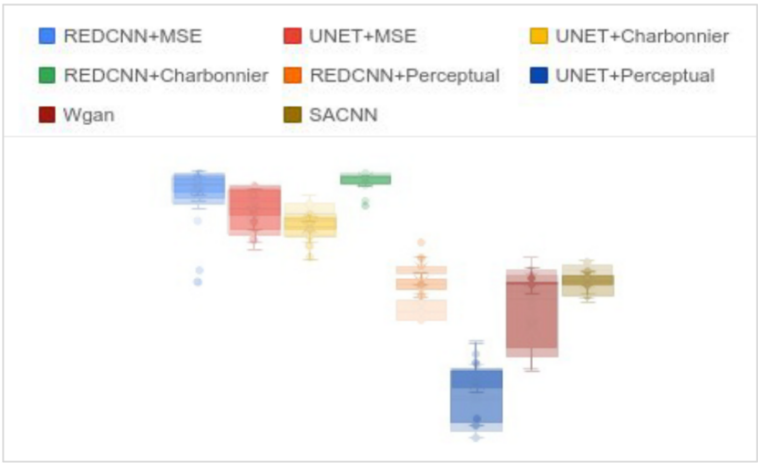
Os modelos treinados com funções de perda pixel-wise, em geral, apresentaram melhores métricas (maiores valores de PSNR e SSIM e menores valores de NRMSE), demonstrando conformidade com o que se esperaria, pois, assim como tais funções, as métricas em questão utilizam a diferença direta entre os pixels das imagens, principalmente PSNR e NRMSE, como apresentado em 3.4.1. Contudo, a distinção aparente em métricas entre o grupo pixel-wise e não pixel-wise, não captura toda a história. O desempenho superior para o grupo pixel-wise, conforme indicado pelos valores dessas métricas, não necessariamente se traduz em uma maior qualidade visual (WOLTERINK et al., 2017; YANG et al., 2018; LI et al., 2020; YU et al., 2017), abrindo espaço para a discussão que segue nos tópicos relacionados à interpretação dos resultados qualitativos.

Por outro lado, os modelos que fazem uso de funções de perda não pixel-wise, apesar de apresentarem uma certa disparidade com valores inferiores de métricas em comparação com o grupo pixel-wise, ressaltam uma abordagem diferente na reconstrução de imagens. Esses modelos, através de uma ótica menos focada na precisão pixel a pixel, possibilitam uma ênfase maior na manutenção de características estruturais e texturais das imagens (WOLTERINK et al., 2017; YANG et al., 2018; LI et al., 2020; YU et al., 2017), o que pode não ser plenamente refletido pelas métricas quantitativas PSNR, SSIM e NRMSE.

Uma análise da sobreposição dos gráficos boxplot das três métricas selecionadas (após espelhamento horizontal do gráfico da NRMSE), apresentado na Figura 11, revela uma conclusão notável: *independentemente da métrica específica em foco, observa-se uma constância marcante na diferença de desempenho entre os modelos treinados com funções de perda pixel-wise e aqueles treinados com funções de perda não pixel-wise.* Essa uniformidade nos resultados,

visualmente representada pela quase total sobreposição dos boxplots para cada modelo específico, sublinha a consistência na maneira como as funções de perda influenciam o desempenho dos modelos baseados nas métricas individuais.

Figura 11 – Sobreposição dos gráficos boxplot, esmaecidos, das três métricas observadas PSNR, SSIM e NRMSE (após espelhamento horizontal dessa última).



Fonte: Imagem produzida pelo autor.

Prosseguindo com o desdobramento dos resultados quantitativos obtidos nesta pesquisa, a atenção agora volta-se para a explanação dos resultados advindos do Teste de Friedman, detalhado em 3.4.2. Este teste, realizado sobre a equivalência entre modelos com funções de perda não pixel-wise, reveste-se de especial importância, tendo em vista a averiguação de uma possível uniformidade de desempenho entre as diferentes configurações desses modelos. A métrica PSNR foi escolhida como a variável de interesse para este teste estatístico, fundamentada na sua proeminência na medição do nível de ruído de imagens, e na observação de uma sobreposição quase completa nos boxplots desta e das demais métricas (apresentada na Figura 11).

Importa recordar que o Teste de Friedman é empregado para comparar três ou mais grupos emparelhados, fundamentado numa análise de ranks e não pressupondo normalidade dos dados. No contexto desta pesquisa, o teste foi aplicado para investigar se a performance, conforme quantificada pela métrica PSNR entre as diversas configurações de treinamento que empregam funções de perda não pixel-wise, divergia de maneira estatisticamente significativa.

Os resultados desse teste elucidaram pontos essenciais para a compreensão do impacto das funções de perda não pixel-wise. Com um valor estatístico ~14.6 e um p-valor de ~0.0022, as evidências iniciais suplantaram a Hipótese Nula (H0), sugerindo diferenças significativas entre ao menos um par de grupos analisados. Esta evidência embasou uma inspeção mais minuciosa

das configurações, onde uma peculiaridade da configuração U-Net, associada à função de perda perceptual, distinguiu-se substancialmente de maneira negativa das outras, conforme apresenta a Figura 8.

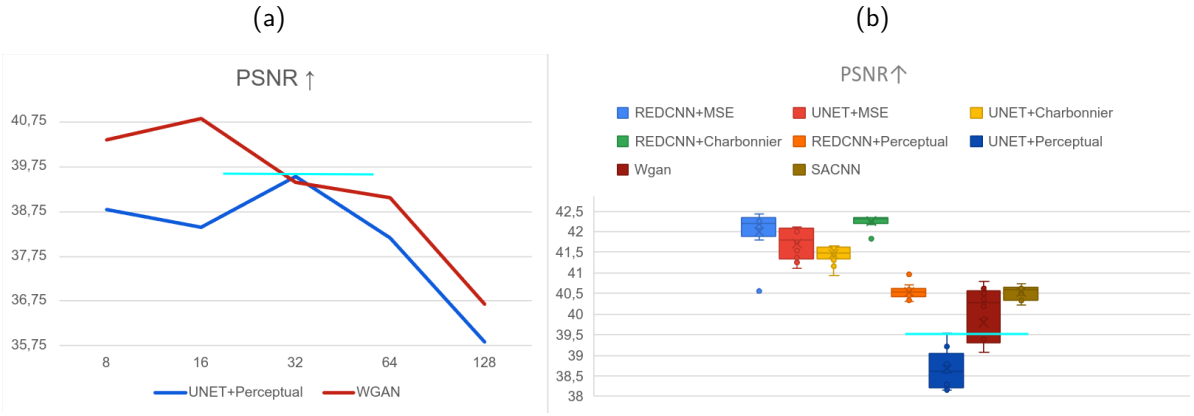
A exclusão estratégica da U-Net com função de perda perceptual do agrupamento em análise conduziu a uma nova interpretação dos resultados, revelando um valor estatístico de ~0.66 e um p-valor de ~0.72. Esta reavaliação colocou as diferenças de performance entre os modelos restantes dentro da margem da variabilidade estatística, não mais sugerindo a existência de diferenças significativas sob a lupa da métrica PSNR. Esta constatação reforça a noção de que, descontando-se uma configuração atípica, os modelos (RED-CNN, WGAN e SACNN) que adotam funções de perda não pixel-wise mantêm um desempenho quantitativamente comparável entre si.

Este resultado nos entrega duas conclusões importantes. *Primeiramente, destaca que a maioria dos modelos que usam funções de perda não pixel-wise apresenta uma performance muito similar*. Isso desafia a ideia de que um modelo é significativamente melhor que outro simplesmente com base em análises isoladas, como as comparações realizadas em (CHEN et al., 2017; WOLTERINK et al., 2017; YANG et al., 2018; LI et al., 2020). Em particular, após a exclusão da configuração da U-Net em combinação com a função de perda perceptual, os modelos restantes, RED-CNN empregando função de perda perceptual, WGAN, e SACNN, demonstraram equivalência em suas performances. Esta observação é importante considerando que estes modelos, como melhor descrito em 3.2.1, são distintos nas suas arquiteturas e complexidade computacional, o que indica que avanços e inovações em modelos neurais podem ser adaptáveis entre si, contanto que apropriadamente calibrados com funções de perda adequadas.

Em seguida, a análise então redireciona o foco específico para a configuração empregando a U-Net associada à função de perda perceptual, a qual se diferencia consideravelmente em desempenho, conforme evidenciado pelo teste estatístico e identificação visual nos gráficos das Figuras 8, 9 e 10, em comparação com as demais configurações. Uma inspeção mais detalhada, utilizando por exemplo, gráficos comparativos sob certas condições específicas, como variações no *batch size* e escolha de um único valor de *seed*, revela que esta configuração possui a capacidade de alcançar, e até mesmo superar, o desempenho da WGAN, como apresentado na Figura 12. Entretanto, uma avaliação de espectro mais amplo, tal qual a conduzida neste estudo, elucidou uma predominância quantitativa em termos de métricas por parte da WGAN. Esta observação ressalta a importância de uma análise compreensiva que vá além de avaliações pontuais, para compreender verdadeiramente a eficácia relativa dos modelos em ambientes

experimentais rigorosamente controlados.

Figura 12 – Gráficos comparativos entre modelo U-Net com função de perda perceptual e WGAN. Evidenciando em (a) uma comparação pontual onde a U-Net supera a WGAN em PSNR em uma configuração específica de *batch size* (32). Enquanto que em (b) os modelos são comparados de maneira holística, revelando superioridade distante por parte da WGAN em PSNR. Em ambos os gráficos, está presente uma linha de cor ciano, que destaca um valor da métrica PSNR análogo em ambos os gráficos.



Fonte: Imagem produzida pelo autor.

Em conclusão, os resultados quantitativos apresentados neste segmento nos conduzem a uma reflexão crucial sobre a importância de adotar uma abordagem holística e metódica, como melhor abordado em 3.4.3, para impulsionar o avanço no campo da remoção de ruído em imagens médicas por meio da inteligência artificial. A constatação da equivalência de desempenho entre modelos diversos ressalta a necessidade, ao escolher ou desenvolver uma arquitetura para a redução de ruído em imagens de CT, de considerar não apenas a inovação ou a complexidade computacional dos modelos, mas também, como veremos detalhadamente no tópico seguinte, a aplicação adequada das funções de perda alinhadas com as exigências específicas da tarefa.

4.2 INTERPRETAÇÃO DOS RESULTADOS QUALITATIVOS

A interpretação das observações qualitativas, melhor descrita em 3.5, emerge como um pilar fundamental na avaliação de modelos de remoção de ruído em imagens de tomografia computadorizada, complementando rigorosamente as métricas quantitativas previamente discutidas. Enquanto a análise quantitativa fornece uma visão objetiva sobre a precisão na reconstrução de imagens, as observações qualitativas se concentram em aspectos visuais essenciais que transcendem os números, tais como a fidelidade de texturas, o contraste entre os tecidos e a preservação de detalhes diagnósticos cruciais. Esta dimensão da análise é de

particular importância no contexto médico, onde a capacidade de discernir sutilezas visuais nas imagens pode diretamente influenciar a precisão diagnóstica e, por extensão, os procedimentos clínicos adotados.

Para uma representação visual abrangente dos resultados obtidos das observações qualitativas, referenciamos o conjunto de imagens das Figuras 13 e 14. Este quadro é compilado a partir das execuções de maior êxito de cada configuração neural estudada, apresentadas na Tabela 2, escolhidas com base na métrica PSNR, que é comumente utilizadas como um indicativo da fidelidade global da imagem restaurada em comparação com o original, como melhor apresentada em 3.4.1. O conjunto de imagens é estruturado de tal maneira que cada linha apresenta o resultado de uma configuração diferente, iniciando com a imagem obtida com baixa dosagem de radiação e culminando com a imagem de alta dosagem, que serve como nosso *groundtruth*. Adicionalmente, algumas colunas são incluídas para demonstrar a variabilidade dos resultados em diferentes cenários tomográficos.

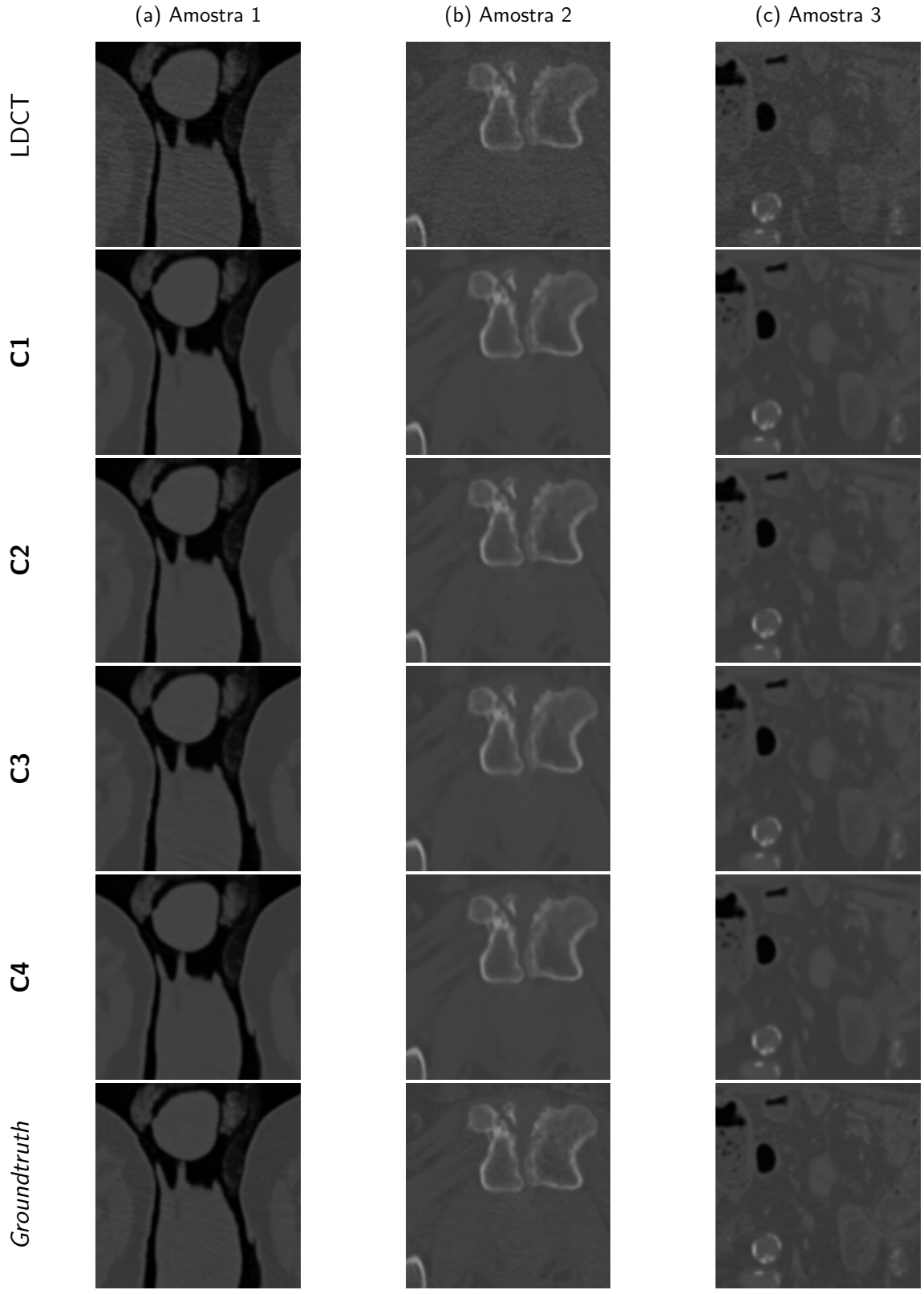
Tabela 2 – Configurações de *batch size* e *seed* das execuções com maior PSNR de cada modelo neural experimentado, devidamente apresentados em 3.2.3. Para cada configuração dos modelos também está sendo apresentado o valor de PSNR atingido.

Modelo + *Loss*	***Batch Size***	***Seed***	**PSNR ↑**
RED-CNN + MSE	8	10	42,418
U-Net + MSE	16	12	42,127
U-Net + Charbonnier	32	11	41,657
RED-CNN + Charbonnier	16	11	42,349
RED-CNN + Perceptual	128	10	40,964
U-Net + Perceptual	32	11	39,541
WGAN	16	11	40,808
SACNN	8 (32)	11	40,735

A observação das Figuras 13 e 14 revela uma valiosa percepção sobre o equilíbrio entre redução de ruído e preservação de detalhes, elemento-chave na busca por soluções adequadas em radiologia diagnóstica, como abordado por (WOLTERINK et al., 2017; YANG et al., 2018; LI et al., 2020) em suas análises qualitativas. Ao explorar visualmente essas reconstruções, somos capazes de discernir as nuances que distinguem as abordagens de redução de ruído, não apenas em termos de suavização de ruído, mas também na sua capacidade de manter ou mesmo realçar detalhes críticos essenciais para interpretações médicas precisas.

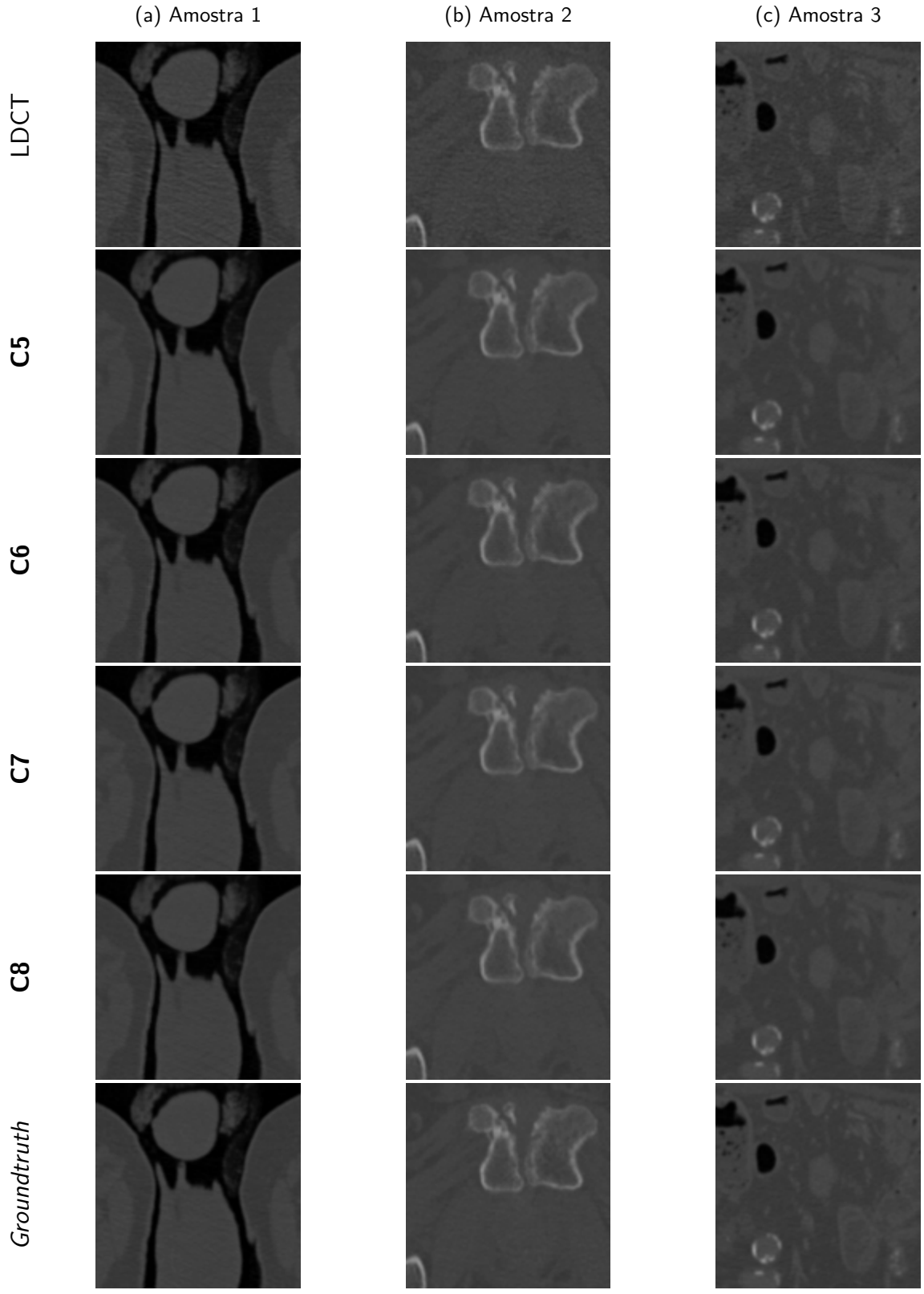
A investigação qualitativa das imagens geradas pela configuração com melhor PSNR de cada grupo (pixel-wise e não pixel-wise), revela esclarecimentos distintos sobre o Efeito MSE,

Figura 13 – Imagens sintetizadas por modelos pixel-wise, descrito em 2.3, com recorte central para melhor visualização, acompanhadas da imagem de baixa dosagem (LDCT) e *groundtruth*. Cada modelo representa sua melhor configuração de *seed* e *batch size* baseado na mátrica PSNR (↑), conforme apresentado na Tabela 2. No quadro de figuras abaixo os modelos estão elencados como: **C1**: RED-CNN + MSE; **C2**: U-Net + MSE; **C3**: U-Net + Charbonnier, **C4**: RED-CNN + Charbonnier.



Fonte: Imagem produzida pelo autor.

Figura 14 – Imagens sintetizadas por modelos **não** pixel-wise, descrito em 2.3, com recorte central para melhor visualização, acompanhadas da imagem de baixa dosagem (LDCT) e *groundtruth*. Cada modelo representa sua melhor configuração de *seed* e *batch size* baseado na mátrica PSNR (↑), conforme apresentado na Tabela 2. No quadro de figuras abaixo os modelos estão elencados como: **C5**: RED-CNN + Perceptual; **C6**: U-Net + Perceptual; **C7**: WGAN, **C8**: SACNN.



Fonte: Imagem produzida pelo autor.

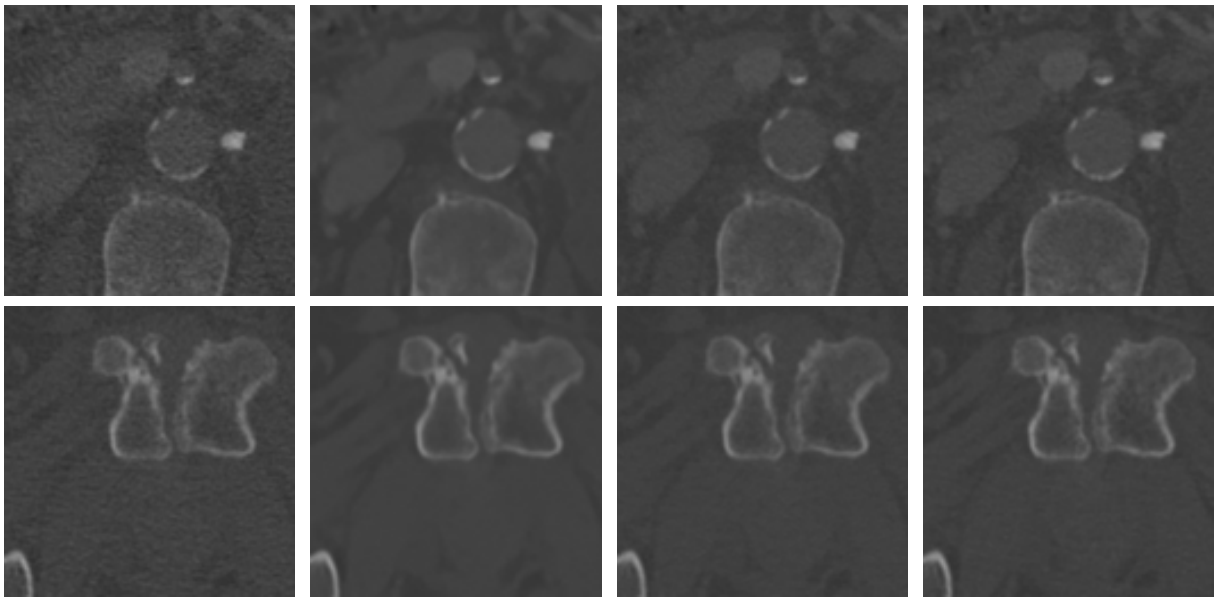
descrito em 3.5.1. As configurações comparadas, RED-CNN com função de perda MSE (pixel-wise) e RED-CNN com função de perda perceptual (não pixel-wise), com *batch size* e *seeds* especificados na Tabela 2. Apesar do alto desempenho quantitativo do RED-CNN, guiado pela minimização do erro quadrático médio, um exame mais atento das imagens resultantes desvenda o borramento característico associado ao Efeito MSE, como apresentado na Figura 15. Esse efeito se manifesta na forma de uma aparente simplificação dos detalhes de alta frequência da imagem, onde a textura e a nitidez sofrem, evidenciando uma homogeneização dos tecidos. Portanto, mesmo atingindo valores elevados de PSNR, o RED-CNN + MSE demonstra limitações na preservação de qualidades visuais, ilustrando uma falha comum das funções de perda pixel-wise em reproduzir a complexidade natural dos tecidos (WOLTERINK et al., 2017; YANG et al., 2018; LI et al., 2020; YU et al., 2017).

Por outro lado, a configuração RED-CNN + Perceptual, que não se baseia na conformidade pixel a pixel mas em aproximar a distribuição dos dados de treinamento por meio da sua rede adversária, oferece uma abordagem alternativa. Ainda observando a Figura 15, ao comparar as imagens reconstruídas utilizando o RED-CNN + Perceptual com as produzidas pelo RED-CNN + MSE, nota-se uma preservação superior da textura e dos detalhes finos, ou seja, das altas frequências. A intensidade do contraste e a clareza dos contornos nos tecidos são mais evidentes, aproximando-se da qualidade visual encontrada nas imagens de alta dosagem (groundtruth), indicando a potencialidade das funções de perda não pixel-wise em manter ou mesmo realçar características cruciais. Esta diferença aponta diretamente para a desconexão entre as métricas tradicionais de avaliação de desempenho e a qualidade visual percebida, desafiando a premissa de que melhorias métricas correspondem diretamente a uma reconstrução mais confiável das imagens.

A presença do Efeito MSE, apresentado em 3.5.1, e a subsequente perda detalhes de alta frequência, como a textura, em modelos treinados com funções de perda pixel-wise, lançam luz sobre uma questão crítica: a escolha da função de perda não é apenas um detalhe técnico, mas sim um fator determinante na utilidade clínica das imagens reconstruídas, como apresentado em 3.5.2. As comparações lado a lado das imagens geradas por RED-CNN + MSE e RED-CNN + Perceptual, junto às versões de baixa e alta dosagem de radiação, enfatizam a necessidade de equilibrar métricas quantitativas e investigações qualitativas na avaliação de modelos de remoção de ruído em imagens de tomografia computadorizada, como já destacado em 3.5.2. O reconhecimento do Efeito MSE, destacado em 3.5.1, reforça a importância de direcionar o desenvolvimento de modelos não apenas para a otimização de métricas estatísticas, mas

Figura 15 – Comparação entre imagens sintetizadas por modelos treinados com loss pixel-wise e não pixel-wise, acompanhadas da imagem de baixa dosagem (LDCT) e *groundtruth*. É observável em *(b)* a presença do Efeito MSE, melhor descrito em 3.5.1.

(a) LDCT (b) RED-CNN + MSE (c) RED-CNN + Percep. (d) *Groundtruth*



Fonte: Imagem produzida pelo autor.

também para a preservação da integridade visual das imagens médicas, garantindo assim sua relevância e aplicabilidade no cenário clínico.

4.3 DESAFIOS NA AVALIAÇÃO QUANTITATIVA E NECESSIDADE DE INSPEÇÃO CLÍNICA

A identificação do Efeito MSE, apontado em 3.5.1, e a constatação da desconexão entre métricas convencionais e a qualidade visual perceptual apontam para uma falha sistêmica no campo da redução de ruído em imagens de tomografia computadorizada: a lacuna existente na disponibilidade de métricas quantitativas que corretamente capturem a relevância clínica e a fidelidade visual das imagens reconstruídas. Embora métricas como PSNR, SSIM e NRMSE ofereçam visões valiosas sobre a precisão e o desempenho técnico dos modelos, elas pecam ao não contemplar integralmente as nuances perceptuais cruciais na reconstrução das imagens com características fidedignas, como melhor abordado em 3.5.1 e 3.5.2. Outras métricas, como *Visual Information Fidelity* (VIF) (SHEIKH; BOVIK, 2006; WANG; BOVIK, 2009) e *Information Fidelity Criterion* (IFC) (SHEIKH; BOVIK; VECIANA, 2005), utilizadas por exemplo em (LI et al., 2020), poderiam trazer novas dimensões de análise ao considerar aspectos mais complexos da

percepção visual. No entanto, a eficácia dessas métricas alternativas na captura da qualidade perceptual, em um contexto tão especializado quanto a interpretação de imagens médicas de CT, ainda permanece pouco explorada neste estudo.

A consequência da ausência de métricas robustas e especializadas é a inevitável necessidade de avaliações médicas manuais para determinar a utilidade diagnóstica das imagens reconstruídas. No entanto, essa abordagem esbarra em desafios práticos significativos, principalmente devido à complexidade envolvida na organização de painéis de especialistas médicos, que devem avaliar a vasta gama de imagens geradas por diferentes configurações de modelos. Essa exigência se mostra inviável em muitos cenários de pesquisa, pela demanda de tempo dos especialistas e recursos. Mesmo que essa análise especializada seja ideal e altamente desejável para determinar a aplicabilidade clínica das imagens, a insuficiência de métricas quantitativas apropriadas deixa um vácuo que impede uma avaliação abrangente e objetiva da qualidade das imagens médicas produzidas por técnicas computacionais avançadas, como as empregadas na remoção de ruído. Portanto, identificar ou desenvolver métricas que possam efetivamente ser a ponte entre a precisão técnica e a aplicabilidade clínica das imagens se apresenta como uma diretriz crucial para futuras pesquisas na área.

4.4 OBSERVAÇÕES SOBRE O DESEMPENHO RELATIVO DOS MODELOS EM CONDIÇÕES VARIÁVEIS

A análise detalhada do desempenho relativo dos modelos sob a influência de hiperparâmetros variáveis (como *seed* e *batch size*), conforme apresentado nas Figuras 8, 9 e 10, pelas Tabelas 3 e 4 e mais abrangentemente pelo Anexo A, revela uma camada de complexidade que desafia as percepções tradicionais formadas a partir de avaliações baseadas em execuções isoladas, como exemplificado pela Figura 12. A adoção de um conjunto amostral expandido de execuções permite não apenas uma compreensão mais precisa do comportamento dos modelos em diferentes condições, como também ilumina a natureza intrinsecamente variável da aprendizagem de máquina. Essa abordagem desvenda um espectro mais amplo de resultados que podem confrontar pressupostos estabelecidos através de testes únicos, demonstrando que variações aparentemente menores nos hiperparâmetros podem produzir discrepâncias significativas no desempenho dos modelos. Esta observação é crucial, visto que embasa a argumentação de que avaliações mais holísticas e repetitivas são essenciais para uma avaliação justa e representativa da capacidade dos modelos.

Durante a fase experimental, certos casos foram observados onde variações nos hiperparâmetros, melhor apresentados em 3.2.3, não apenas influenciaram o desempenho dos modelos, mas também inviabilizaram um aprendizado adequado, resultando, em todas as épocas treinadas, na produção de imagens completamente pretas, como as configurações expostas nas Tabelas 3 e 4 pelas células sem valor de métrica. Essas ocorrências, identificadas ao variar a *seed* e o *batch size*, enfatizam o impacto crítico que esses parâmetros podem exercer sobre o processo de treinamento e a importância de sua cuidadosa seleção e controle. Confrontados, principalmente pelo grupo piexel-wise, com a incapacidade de aprendizado manifestada nesses episódios, foi imperativo expandir o espectro de *seeds* utilizadas para incluir valores adicionais, pela incorporação da *seed* 12, para assegurar um conjunto amostral mínimo de 8 execuções por configuração, evidente na Tabela 3. Essa estratégia não só reitera a variabilidade inerente aos experimentos de aprendizado de máquina como também destaca a necessidade de adaptabilidade na condução de experimentações robustas, capazes de transcender limitações pontuais e obter revelações genuínas sobre a performance e confiabilidade dos modelos analisados.

Tabela 3 – Resultados em PSNR (↑) dos conjuntos experimentais dos modelos treinados com função de perda pixel-wise. Células sem valores representam a incapacidade do modelo em aprender com a configuração específica, resultando apenas em imagens pretas. A descrição dos experimentos estão apresentadas em 3.2.3. Na tabela, os modelos estão apresentados como: **C1**: RED-CNN + MSE; **C2**: U-Net + MSE; **C3**: U-Net + Charbonnier, **C4**: RED-CNN + Charbonnier.

Seed	***Batch Size***	**C1**	**C2**	**C3**	**C4**
	128	40,552	41,112	41,159	
	64		41,371	41,381	42,165
10	32	42,165	41,85	41,393	42,348
	16	42,346	42,01	41,55	42,314
	8	42,418	42,09	41,642	42,34
	128				41,833
	64				42,287
11	32			41,657	42,325
	16				42,349
	8			41,612	
	128		41,256	40,935	
	64	41,798	41,543	41,313	
12	32	42,122	41,751	41,634	
	16	42,253	42,127	41,442	
	8	42,324	42,09	41,637	

Tabela 4 – Resultados em PSNR (↑) dos conjuntos experimentais dos modelos treinados com função de perda **não** pixel-wise. Células sem valores representam a incapacidade do modelo em aprender com a configuração específica, resultando apenas em imagens pretas. A descrição dos experimentos estão apresentadas em 3.2.3. Na tabela, os modelos estão apresentados como: **C5**: RED-CNN + Perceptual; **C6**: U-Net + Perceptual; **C7**: WGAN, **C8**: SACNN. Sendo válido recordar, como melhor apresentado em 3.2.3, este último é executado com valores de *batch size* diferentes: 16 (128), 12 (64), 8 (32), 4 (16) e 2 (8).

Seed	***Batch Size***	**C5**	**C6**	**C7**	**C8**
	128	40,964	38,284	40,624	40,578
	64	40,707	38,154	40,182	40,494
10	32	40,551	38,61	39,88	40,625
	16	40,459	39,212	40,535	40,321
	8	40,535	38,899	40,499	40,223
	128	40,335		36,691	40,323
	64	40,589	38,16	39,059	40,735
11	32	40,491	39,541	39,389	40,735
	16	40,539	38,385	40,808	40,629
	8	40,301	38,797	40,356	40,592

4.5 SUPERAÇÃO DE OBSTÁCULOS NA AVALIAÇÃO HOLÍSTICA DE MODELOS NEURAIS

No campo dinâmico da inteligência artificial aplicada à medicina, a realização de um conjunto de treinamentos com configurações experimentais distintas para um mesmo modelo neural confronta-se, intrinsecamente, com complexidades que vão além das enfrentadas em treinamentos com configurações únicas. Esta extensão do escopo experimental abarca não apenas a diversidade dos dados e a variação dos hiperparâmetros, mas se depara também com o desafio crucial das operações não determinísticas (TENSORFLOW, 2024) inerente a muitos *frameworks* de aprendizado de máquina, como melhor abordado em 3.3.2. Tal ambiente exige uma metodologia que não só contemple a multiplicidade de condições experimentais, mas que também assegure a replicabilidade e fidelidade dos resultados, permitindo uma avaliação precisa do impacto das variáveis em estudo.

Diante deste contexto, o desenvolvimento de um *framework* robusto é apresentado como solução central aos desafios mencionados. Este sistema é desenhado para facilitar a condução holística de experimentos, fornecendo uma plataforma coerente que permite a manipulação controlada dos hiperparâmetros, o gerenciamento de operações não determinísticas, e a garantia de padrões replicáveis de experimentação. A sua arquitetura flexível serve como a espinha dorsal para estudos experimentais em uma gama variada de aplicações, indo além da remoção

de ruído em imagens médicas, para incluir um espectro abrangente de problemas que podem ser explorados com diferentes modelos neurais e métodos de avaliação.

5 CONCLUSÃO

A jornada empreendida neste trabalho revela que comparações de modelos neurais derivadas de execuções isoladas ou configurações únicas podem ocultar a verdadeira capacidade de um modelo neural, bem como sua generalizabilidade em aplicações práticas, sobretudo na área crítica da remoção de ruído em imagens de tomografia. Comprovou-se que variações nos hiperparâmetros, frequentemente escolhidos sem rigor nas replicações, como o *seed* e o *batch size*, carregam um impacto considerável no desempenho dos modelos. Tais observações, detalhadamente desenvolvidas no Capítulo 4, evidenciaram a essencialidade de conduzir múltiplas execuções sob variadas configurações de treinamento. Esta abordagem não apenas corrobora a robustez do modelo, mas também oferece uma visão mais fidedigna e confiável sobre sua performance, navegando para além das limitações impostas por testagens singulares e proporcionando uma compreensão abrangente do comportamento dos modelos.

Referindo-se às funções de perda empregadas, um padrão emergiu: modelos treinados com funções de perda pixel-wise, apesar de suas vantagens em termos de métricas quantitativas, tais como PSNR e SSIM, paradoxalmente podem conduzir à perda de detalhes finos e texturas, sendo esta uma consequência direta do Efeito MSE. Em contrapartida, estratégias de treinamento que se apoiam em funções de perda não pixel-wise, ainda que possam exibir métricas quantitativas menores, revelam uma capacidade notável de gerar imagens que retêm as qualidades perceptuais críticas às demandas clínicas, preservando os detalhes diagnósticos vitais. Este discernimento não unicamente sublinha a importância de considerar o equilíbrio entre as métricas objetivas e a qualidade visual na seleção da função de perda, como também realça o potencial dessas funções em satisfazer as necessidades clínicas ao manter os detalhes diagnósticos fundamentais (WOLTERINK et al., 2017; YU et al., 2017).

A investigação minuciosa do desempenho de modelos neurais para remoção de ruído em imagens de tomografia computadorizada evidenciou uma lacuna substancial entre as métricas quantitativas convencionais e as observações visuais percebidas. Métricas como PSNR e SSIM, apesar de sua utilidade incontestável em fornecer avaliações quantitativas comparativas, falham frequentemente em refletir a complexidade e a relevância clínica das imagens reconstruídas, como abordam (YANG et al., 2018; LI et al., 2020). A presença do Efeito MSE ressalta, portanto, a exigência de métricas mais especializadas capazes de avaliar a qualidade perceptual das imagens médicas de forma mais alinhada às necessidades clínicas. A inclusão

de métricas com ênfase nas características psico-visuais humanas (LI et al., 2020), como a *Visual Information Fidelity* (VIF) (SHEIKH; BOVIK, 2006; WANG; BOVIK, 2009) e *Information Fidelity Criterion* (IFC) (SHEIKH; BOVIK; VECIANA, 2005), apresenta-se como uma alternativa valiosa para suprir essa demanda. Ainda assim, a constatação de que a utilidade diagnóstica das imagens sintetizadas requer avaliações médicas manuais e especializadas, reflete sobre a complexidade do julgamento da qualidade das imagens médicas, sugerindo que a integração entre métricas especializadas e análises qualitativas detalhadas é essencial na avaliação das imagens reconstruídas por modelos de aprendizado profundo.

Este estudo, ao abraçar uma análise abrangente dos modelos convolucionais para remoção de ruído em imagens de tomografia computadorizada, iluminou as nuances críticas associadas à avaliação de tais modelos. A necessidade de equilibrar a precisão métrica com a relevância clínica das imagens sintetizadas sublinha uma área de considerável complexidade e interesse, demandando futuras investigações. Em sua essência, este trabalho argumenta por uma abordagem multidimensional na avaliação da eficácia dos modelos neurais, instigando uma perspectiva que integre análises quantitativas e observações qualitativas com testes específicos para melhores práticas no avanço de tecnologias de remoção de ruído. A busca por modelos que não só performem satisfatoriamente em métricas estabelecidas mas também preservem a integridade visual e diagnóstica das imagens demonstra ser um objetivo distinto e significativo no desenvolvimento de tecnologias auxiliadoras na área da saúde.

REFERÊNCIAS

AMERICAN ASSOCIATION OF PHYSICISTS IN MEDICINE (AAPM). *Low Dose CT Grand Challenge*. 2016. <https://www.aapm.org/GrandChallenge/LowDoseCT/>. Acesso em: 08 mar. 2024.

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. *Wasserstein GAN*. 2017. <https://arxiv.org/abs/1701.07875>. Acesso em: 18 mar. 2024.

BRENNER, D. J.; HALL, E. J. Computed tomography — an increasing source of radiation exposure. *New England Journal of Medicine*, v. 357, n. 22, p. 2277–2284, 2007. PMID: 18046031. Disponível em: <https://doi.org/10.1056/NEJMra072149>.

CHARBONNIER, P.; BLANC-FERAUD, L.; AUBERT, G.; BARLAUD, M. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, v. 6, n. 2, p. 298–311, 1997.

CHEN, H.; ZHANG, Y.; KALRA, M. K.; LIN, F.; CHEN, Y.; LIAO, P.; ZHOU, J.; WANG, G. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, Institute of Electrical and Electronics Engineers (IEEE), v. 36, n. 12, p. 2524–2535, dez. 2017. ISSN 1558-254X. Disponível em: <http://dx.doi.org/10.1109/TMI.2017.2715284>.

GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. In: GHAHRAMANI, Z.; WELLING, M.; CORTES, C.; LAWRENCE, N.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. v. 27. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>.

GULRAJANI, I.; AHMED, F.; ARJOVSKY, M.; DUMOULIN, V.; COURVILLE, A. C. Improved training of wasserstein gans. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/892c3b1c6dccd52936e27cbd0ff683d6-Paper.pdf>.

JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016. p. 694–711. ISBN 978-3-319-46475-6.

KRILLE, L.; HAMMER, G. P.; MERZENICH, H.; ZEEB, H. Systematic review on physician's knowledge about radiation doses and radiation risks of computed tomography. *European journal of radiology*, Elsevier, v. 76, n. 1, p. 36–41, 2010.

LECLERC, G.; ILYAS, A.; ENGSTROM, L.; PARK, S. M.; SALMAN, H.; MADRY, A. *FFCV: Accelerating Training by Removing Data Bottlenecks*. 2023. <https://arxiv.org/abs/2306.12517>. Acesso em: 18 mar. 2024.

LI, M.; HSU, W.; XIE, X.; CONG, J.; GAO, W. Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. *IEEE Transactions on Medical Imaging*, v. 39, n. 7, p. 2289–2301, 2020.

MURPHY, A.; GAJERA, J.; LUIJKX, T.; HACKING, C.; QAQISH, N. *Iterative reconstruction (CT)*. 2021. <https://radiopaedia.org/articles/iterative-reconstruction-ct?lang=us>. Acesso em: 14 mar. 2024.

NETT, B. *Filtered Back-Projection (FBP) Illustrated Guide for Radiologic Technologists*. 2021. <https://howradiologyworks.com/filtered-backprojection-fbp-illustrated-guide-for-radiologic-technologists/>. Acesso em: 14 mar. 2024.

QIN, C.; SCHLEMPER, J.; CABALLERO, J.; PRICE, A. N.; HAJNAL, J. V.; RUECKERT, D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, v. 38, n. 1, p. 280–290, 2019.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4.

SHEIKH, H.; BOVIK, A. Image information and visual quality. *IEEE Transactions on Image Processing*, v. 15, n. 2, p. 430–444, 2006.

SHEIKH, H.; BOVIK, A.; VECIANA, G. de. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, v. 14, n. 12, p. 2117–2128, 2005.

SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. <https://arxiv.org/abs/1409.1556>. Acesso em: 18 mar. 2024.

TENSORFLOW. *tf.config.experimental.enable_op_determinism*. 2024. <https://www.tensorflow.org/api_docs/python/tf/config/experimental/enable_op_determinism>. Acesso em: 08 mar. 2024.

WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, v. 26, n. 1, p. 98–117, 2009.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

WOLTERINK, J. M.; LEINER, T.; VIERGEVER, M. A.; IšGUM, I. Generative adversarial networks for noise reduction in low-dose ct. *IEEE Transactions on Medical Imaging*, v. 36, n. 12, p. 2536–2545, 2017.

YANG, Q.; YAN, P.; ZHANG, Y.; YU, H.; SHI, Y.; MOU, X.; KALRA, M. K.; ZHANG, Y.; SUN, L.; WANG, G. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, v. 37, n. 6, p. 1348–1357, 2018.

YU, S.; DONG, H.; YANG, G.; SLABAUGH, G.; DRAGOTTI, P. L.; YE, X.; LIU, F.; ARRIDGE, S.; KEEGAN, J.; FIRMIN, D.; GUO, Y. *Deep De-Aliasing for Fast Compressive Sensing MRI*. 2017. <https://arxiv.org/abs/1705.07137>. Acesso em: 18 mar. 2024.

ANEXO A – CONFIGURAÇÕES E MÉTRICAS EXPERIMENTAIS

PSNR | Batches 8, 16, 32, 64, 128 | Method+Loss | Last Epoch (30)

Seed	BatchSize	REDCNN+Perceptual	REDCNN+MSE	UNET+MSE	UNET+Charbonnier	UNET+Perceptual	REDCNN+Charbonnier	Wgan	SACNN *
10	128	40,964	40,552	41,112	41,159	38,284		40,624	40,578
	64	40,707		41,371	41,381	38,154	42,165	40,182	40,494
	32	40,551	42,165	41,85	41,393	38,61	42,348	39,88	40,625
	16	40,459	42,346	42,01	41,55	39,212	42,314	40,535	40,321
	8	40,535	42,418	42,09	41,642	38,899	42,34	40,499	40,223
11	128	40,335				41,833		36,691	40,323
	64	40,589			38,16		42,287	39,059	40,735
	32	40,491			41,657	39,541	42,325	39,389	40,735
	16	40,539				38,385	42,349	40,808	40,629
	8	40,301			41,612	38,797		40,356	40,592
12	128			41,256	40,935				
	64		41,798	41,543	41,313				
	32		42,122	41,751	41,634				
	16		42,253	42,127	41,442				
	8		42,324	42,09	41,637				

** SACNN was trained with batch sizes 16, 12, 8, 4 and 2*

SSIM | Batches 8, 16, 32, 64, 128 | Method+Loss | Last Epoch (30)

Seed	BatchSize	REDCNN+Perceptual	REDCNN+MSE	UNET+MSE	UNET+Charbonnier	UNET+Perceptual	REDCNN+Charbonnier	Wgan	SACNN *
10	128	0,95	0,955	0,954	0,956	0,937		0,949	0,949
	64	0,947		0,956	0,956	0,939	0,959	0,949	0,949
	32	0,945	0,959	0,958	0,956	0,937	0,96	0,949	0,948
	16	0,944	0,96	0,958	0,956	0,937	0,96	0,948	0,948
	8	0,945	0,96	0,959	0,957	0,937	0,96	0,949	0,947
11	128	0,946				0,936	0,959	0,947	0,9479
	64	0,945				0,939	0,96	0,948	0,9486
	32	0,944			0,958	0,938	0,96	0,949	0,9493
	16	0,944			0,956	0,933	0,96	0,949	0,9483
	8	0,944						0,949	0,9483
12	128			0,955	0,955				
	64		0,958	0,957	0,957				
	32		0,959	0,957	0,957				
	16		0,959	0,959	0,956				
	8		0,959	0,959	0,956				

** SACNN was trained with batch sizes 16, 12, 8, 4 and 2*

NRMSR | Batches 8, 16, 32, 64, 128 | Method+Loss | Last Epoch (30)

Seed	BatchSize	REDCNN+Perceptual	REDCNN+MSE	UNET+MSE	UNET+Charbonnier	UNET+Perceptual	REDCNN+Charbonnier	Wgan	SACNN *
10	128	0,036	0,038	0,036	0,036	0,049		0,038	0,038
	64	0,037		0,035	0,035	0,05	0,032	0,04	0,038
	32	0,038	0,032	0,033	0,035	0,047	0,031	0,041	0,038
	16	0,038	0,031	0,032	0,034	0,044	0,031	0,038	0,039
	8	0,038	0,031	0,032	0,034	0,046	0,031	0,038	0,04
11	128	0,039					0,033	0,059	0,039
	64	0,038				0,05	0,031	0,045	0,0374
	32	0,038			0,034	0,043	0,031	0,043	0,0373
	16	0,038				0,049	0,031	0,037	0,0377
	8	0,039			0,034	0,046		0,039	0,0379
12	128			0,035	0,037				
	64		0,033	0,034	0,035				
	32		0,032	0,033	0,034				
	16		0,031	0,032	0,035				
	8		0,031	0,032	0,034				

** SACNN was trained with batch sizes 16, 12, 8, 4 and 2*