Avaliação de Modelos de Segmentação para Remoção de Ruídos em Tomografias Computadorizadas

Natalia Camillo do Carmo

Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2025

Natalia Camillo do Carmo

Avaliação de Modelos de Segmentação para Remoção de Ruídos em Tomografias Computadorizadas

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: André Ricardo Backes

Uberlândia
2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

C287 Carmo, Natalia Camillo do, 1998-
2025 Avaliação de Modelos de Segmentação para Remoção de Ruídos em Tomografias Computadorizadas [recurso eletrônico] / Natalia Camillo do Carmo. - 2025.

Orientador: André Ricardo Backes.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.di.2025.330
Inclui bibliografia.
Inclui ilustrações.

1. Computação. I. Backes, André Ricardo,1981-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. III. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Computação
Av. João Naves de Ávila, 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação, 09/2025, PPGCO				
Data:	10 de junho de 2025	Hora de início:	09:00	Hora de encerramento:	11:50
Matrícula do Discente:	12222CCP018				
Nome do Discente:	Natalia Camillo do Carmo				
Título do Trabalho:	Avaliação de Modelos de Segmentação para Remoção de Ruídos em Tomografias Computadorizadas				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	----------				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Alessandra Aparecida Paulino - FACOM/UFU, Dalcimar Casanova DAINF/UTFPR e André Ricardo Backes - DC/UFSCar orientador da candidata.

Os examinadores participaram desde as seguintes localidades: André Ricardo Backes - São Carlos/SP e Dalcimar Casanova - Pato Branco/PR. Os outros membros da banca e o aluno participaram da cidade de Uberlândia.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. André Ricardo Backes, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(à) Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir ao(à) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(à) candidato(a):

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente

ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **André Ricardo Backes**, **Usuário Externo**, em 10/06/2025, às 16:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Alessandra Aparecida Paulino**, **Professor(a) do Magistério Superior**, em 10/06/2025, às 16:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Dalcimar Casanova**, **Usuário Externo**, em 11/06/2025, às 14:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6384942** e o código CRC **722E5228**.

Referência: Processo nº 23117.037234/2025-67 SEI nº 6384942

Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.

Agradecimentos

Agradeço ao meu orientador, Prof. Dr. André Backes, pela compreensão, apoio e ensinamentos valiosos que levarei para a vida. À minha família, minha mãe Angélia, meu pai Frederico e minha irmã Letícia, pelo incentivo e encorajamento incondicionais; e aos meus amigos, especialmente a Anna Luiza, pelo suporte ao longo desta jornada. Agradeço também à CAPES pelo apoio financeiro e à Universidade Federal de Uberlândia (UFU) pela acolhida e infraestrutura disponibilizada para a realização deste trabalho.

"Nada na vida deve ser temido, somente compreendido. Agora é hora de compreender mais para temer menos."

(Marie Curie)

Resumo

A tomografia computadorizada de baixa dosagem (LDCT) reduz a exposição à radiação ionizante, mas gera imagens com ruído e perda de qualidade. Este trabalho analisa o uso de redes neurais convolucionais de segmentação para remoção de ruído em imagens LDCT, com o objetivo de preservar detalhes estruturais e relevantes. Foram avaliadas quatro arquiteturas: U-Net, LinkNet, PSPNet e MA-Net. Essas redes foram combinadas com diferentes *backbones* (ResNet50, VGG16, DenseNet121, Inceptionv4, MobileNetV2 e EfficientNet-B2). Utilizou-se o conjunto de dados AAPM 2016 da Mayo Clinic. Os resultados demonstram que a combinação U-Net + Inceptionv4 obteve o melhor desempenho, com PSNR de 48,797 e SSIM de 0,977, superando métodos tradicionais. As análises evidenciam que o uso de mecanismos de atenção e *backbones* mais avançados pode aprimorar a recuperação de detalhes em imagens ruidosas.

Palavras-chave: LDCT. Remoção de ruídos. Redes de segmentação. Mecanismos de atenção.

Abstract

Low-dose computed tomography (LDCT) reduces exposure to ionizing radiation, but generates noisy images with poor quality. This work analyzes the use of convolutional neural networks for segmentation to remove noise from LDCT images, aiming to preserve structural details and relevant information. Four architectures were evaluated: U-Net, LinkNet, PSPNet, and MA-Net. These networks were combined with different backbones (ResNet50, VGG16, DenseNet121, Inceptionv4, MobileNetV2, and EfficientNet-B2). The AAPM 2016 dataset from the Mayo Clinic was used. The results demonstrate that the combination of U-Net + Inceptionv4 obtained the best performance, with a PSNR of 48.797 and SSIM of 0.977, outperforming traditional methods. The analyses show that the use of more advanced attention mechanisms and backbones can improve detail recovery in noisy images.

Keywords: LDCT. Noise removal. Segmentation networks. Attention mechanisms.

Lista de ilustrações

Lista de tabelas

Sumário

Capítulo 1

Introdução

A tomografia computadorizada (TC) é um exame de imagem amplamente utilizada na medicina por sua capacidade de gerar imagens transversais detalhadas do corpo humano, permitindo diferenciar com precisão tecidos, músculos, vasos e órgãos internos. Sua sensibilidade é superior à de exames convencionais de raios X e também apresenta um custo mais acessível do que métodos como a ressonância magnética (MRI). Por isso, é comum em acompanhamentos clínicos, especialmente em pacientes com doenças como o câncer.(KROEKER et al., 2011).

No entanto, a TC utiliza radiação ionizante, um tipo de energia que pode danificar o DNA das células. Isso acontece por que a radiação de baixa ionização, que é o caso dos raios X em TC, causam danos leves e espalhados pela célula porém percorrem distâncias mais longas no corpo (BRENNER; HALL, 2007; RYAN, 2012). Sendo assim a repetida exposição, mesmo em doses consideradas baixas, pode levar a efeitos cumulativos a longo prazo, como mutações celulares e, em casos mais graves, câncer. Estudos sugerem que cerca de 5% dos casos anuais de câncer podem estar relacionados à exposição à radiação, inclusive aquela proveniente de exames diagnósticos. (SMITH-BINDMAN et al., 2025).

Em (SILVA, 2013), são apresentadas estatísticas sobre o impacto da exposição acumulada à radiação. Nos Estados Unidos, cerca de 29 mil casos de câncer são estimados como decorrentes de 72 milhões de tomografias realizadas. No Brasil, esses números são menores devido à menor quantidade de equipamentos disponíveis. No entanto, a tendência é de crescimento na aquisição desses aparelhos, dada a relação custo-benefício do exame em comparação com outras modalidades como a MRI, o que aumentará o número de pacientes expostos à TC.

Para aliviar os riscos associados à radiação, uma solução é o uso da tomografia computadorizada de baixa dose, conhecida como *Low-Dose CT* (LDCT). Ao reduzir a dose de radiação emitida, o exame se torna mais seguro. Mas essa redução compromete a qualidade da imagem, já que menos fótons são captados e o sinal é mais fraco. O computador não consegue distinguir entre o que é uma estrutura real do que é um ruído ou artefato, prejudicando a análise por parte dos especialistas (MOEN et al., 2021).

A melhoria da qualidade dessas imagens LDCT, sem aumentar a exposição do paciente, tornou-se um área de estudo (WANG et al., 2021; KANG; MIN; YE, 2017). Métodos clássicos de redução de ruído como filtros espaciais (mediano, bilateral, Gaussiano) e reconstruções

iterativas como o *Filtered Back Projection* (FBP) apresentam limitações em preservar detalhes finos da imagem. Métodos mais recentes, como o *Statistical Iterative Reconstruction* (SIR), podem introduzir artefatos específicos e têm alto custo computacional (KAUR; DONG, 2023).

Com o avanço do aprendizado profundo (*Deep Learning*, DL), surgiram novas possibilidades. Redes neurais convolucionais (CNNs) têm demonstrado excelente desempenho em tarefas de reconstrução e melhoria de imagens médicas. (KULATHILAKE et al., 2021). Modelos de segmentação baseados em CNNs têm sido propostos como alternativa para a remoção de ruído, pois suas estruturas de *encoder-decoder* permitem aprender representações detalhadas da imagem original (CHEN et al., 2017a), favorecendo uma reconstrução mais precisa.

Dessa forma, o trabalho propõe avaliar diferentes modelos de segmentação, combinados com diversos *backbones* (codificadores) pré-treinados, com o objetivo de realizar a remoção de ruído em imagens LDCT. A meta é identificar quais arquiteturas oferecem os melhores resultados tanto em métricas quantitativas (como PSNR e SSIM), quanto na preservação visual da estrutura da imagem.

1.1 Motivação

Os modelos de aprendizado profundo têm apresentado forte capacidade de generalização e adaptação à presença de diferentes tipos de ruído. Isso é particularmente importante em imagens LDCT, cujas variações são amplas e complexas. Redes convolucionais, por exemplo, conseguem extrair características relevantes da imagem original, possibilitando uma reconstrução mais fiel mesmo com ruído (KULATHILAKE et al., 2021).

Além disso, mecanismos de atenção integrados a redes convolucionais, como no modelo *Attention U-Net*, demonstraram capacidade de focar automaticamente em regiões relevantes da imagem, suprimindo informações irrelevantes e melhorando a precisão em tarefas médicas complexas (LI et al., 2020).

Dentro desse panorama, a avaliação das arquiteturas das redes, com ou sem atenção, torna-se fundamental para compreender como diferentes modelos afetam o desempenho em tarefas de remoção de ruído em LDCT.

1.2 Objetivos

O objetivo principal do trabalho é avaliar diferentes modelos de segmentação com *backbones* na tarefa de remoção de ruído em imagens LDCT. Os objetivos específicos do trabalho são:

- ❏ Utilizar *transfer learning* na implementação dos modelos e seus *backbones*;
- ❏ Comparar os resultados com trabalhos da literatura utilizando métricas PSNR e SSIM;
- ❏ Investigar a correlação entre essas métricas e a qualidade visual percebida das imagens reconstruídas.

1.3 Hipótese

A principal hipótese deste trabalho é que as redes convolucionais voltadas para segmentação apresentam um bom potencial para lidar de forma eficiente com a tarefa de remoção de ruídos em imagens LDCT, especialmente devido à sua capacidade de capturar padrões espaciais e relacionamentos locais em diferentes escalas. As redes contêm etapas de extração de atributos, *backbones*, que têm um impacto fundamental no resultado da imagem sem ruído.

1.4 Contribuições

As principais contribuições do trabalho são:

- ❏ Avaliação experimental de diferentes combinações entre arquiteturas de segmentação e *backbones* para remoção de ruído em LDCT;
- ❏ Exploração do uso de mecanismos de atenção na tarefa de remoção de ruído (*denoising*);
- ❏ Análise crítica sobre o uso das métricas PSNR e SSIM em reconstrução de imagem, com destaque para suas limitações;
- ❏ Proposição de combinações específicas que obtiveram os melhores resultados em LDCT.

1.5 Organização da Dissertação

A estrutura desta dissertação está organizada em seis capítulos. O Capítulo 1 apresenta a introdução do problema, sua motivação, objetivos, hipóteses e contribuições. O Capítulo 2 traz a revisão da literatura sobre tomografia, técnicas de remoção de ruído e redes neurais convolucionais. O Capítulo 3 discute os trabalhos relacionados. No Capítulo 4, detalha-se a proposta metodológica, incluindo dados, pré-processamento, treinamento e métricas utilizadas. O Capítulo 5 apresenta a análise e discussão dos resultados experimentais. Por fim, o Capítulo 6 traz as conclusões e sugestões para trabalhos futuros.

Capítulo 2

Revisão da Literatura

2.1 Tomografia computadorizada

A TC é uma técnica de imagem médica que gera cortes transversais do corpo, permitindo uma visualização detalhada e em alta resolução. Ao empregar feixes de raios X emitidos em múltiplos ângulos, a TC permite uma reconstrução volumétrica do corpo por meio de algoritmos matemáticos que estimam os coeficientes de atenuação dos tecidos (MAKEN; GUPTA, 2022; KAK; SLANEY, 2001). A Figura 1 mostra o processo resumido de como acontece a aquisição das imagens.

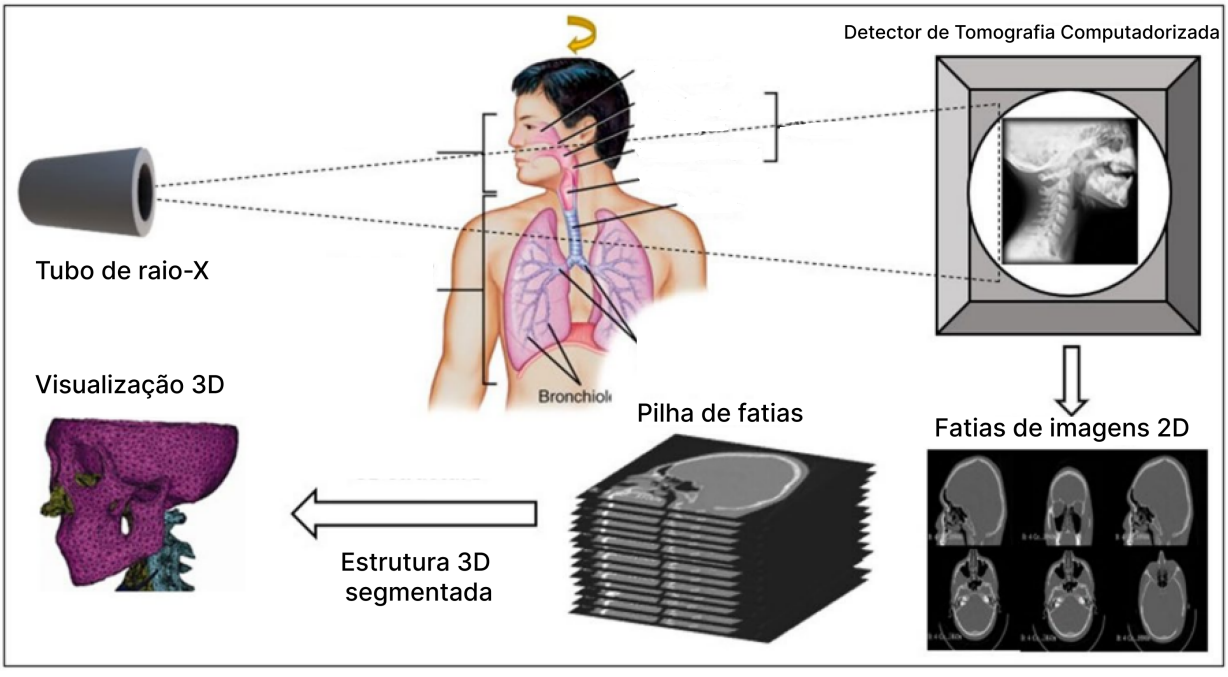


Figura 1 – Imagem ilustrativa do funcionamento de uma TC (MAKEN; GUPTA, 2022).

A obtenção de imagens de qualidade por meio da TC depende de diversos fatores técnicos, incluindo a corrente do tubo (mA), a voltagem (kV), o tempo de rotação e o tipo de colimação. Alterações nesses parâmetros afetam diretamente a dose de radiação e a resolução da imagem (MCCOLLOUGH; BRUESEWITZ; JR, 2006). Portanto, há um equilíbrio necessário entre a

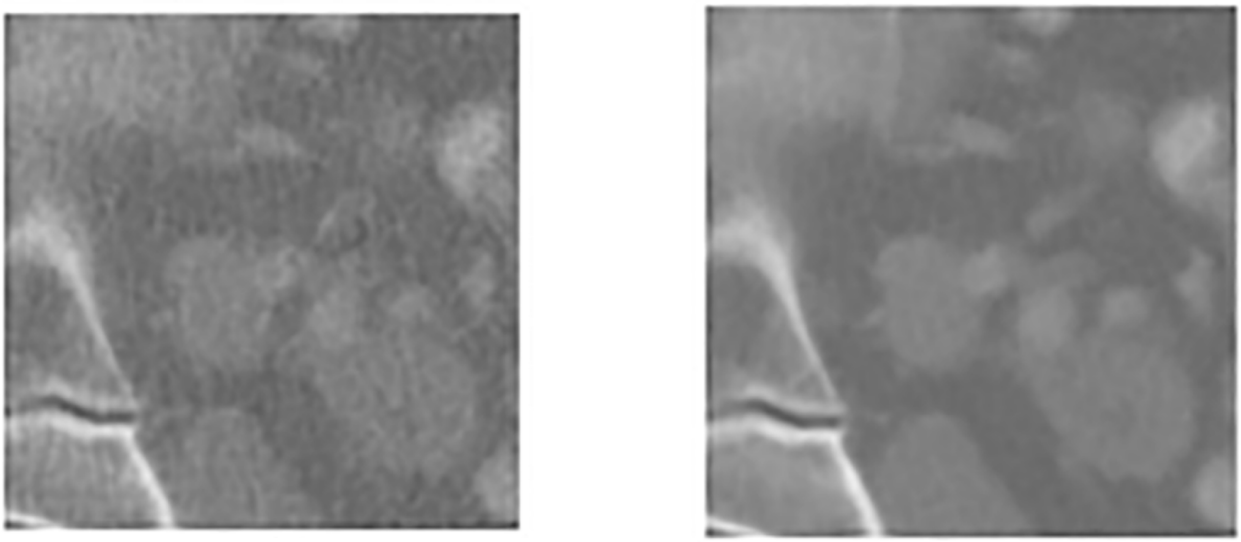
nitidez da imagem e a exposição do paciente à radiação ionizante.

É importante destacar que a TC não é apenas um método anatômico, mas também funcional, sendo utilizada, por exemplo, na angiotomografia para avaliação de vasos, ou na TC de perfusão para análise de tecidos cerebrais. A flexibilidade desse tipo de exame é uma das razões pela sua ampla adoção na medicina moderna (MIER; MIER, 2015).

Em situações como detecção precoce de câncer o uso repetido de TC demanda alternativas com menor dose, o que reforça a relevância de abordagens como a LDCT (GHOLIZADEH-ANSARI; ALIREZAIE; BABYN, 2020).

2.2 Tomografia de Baixa Dose (LDCT)

A LDCT tem o objetivo de reduzir a radiação emitida durante o exame, tornando-o mais seguro para o paciente. (MCCOLLOUGH et al., 2009). A estratégia envolve a diminuição da dose administrada por meio de ajustes na corrente do tubo, tempo de exposição e técnicas de reconstrução, preservando ao máximo a acurácia diagnóstica. Essa abordagem tornou-se importante em contextos de rastreamento e observação, como no diagnóstico precoce de câncer de pulmão em populações de risco (TEAM, 2011).



(a) (b)

Figura 2 – LDCT (a) em comparação com uma tomografia de dose normal (NDCT) (b).

Contudo, a redução de dose provoca efeitos colaterais importantes na qualidade da imagem, como pode ser observado na Figura 2 que compara uma TC normal e uma LDCT. Os principais problemas estão relacionados ao aumento do ruído quântico, que é inerente à redução do número de fótons detectados (BARRETT; KEAT, 2004). Esse ruído afeta principalmente regiões de baixa densidade e pode comprometer a visualização de detalhes sutis, como pequenas lesões ou alterações de textura. Artefatos estruturais, como granularidade e borramento, também podem surgir, prejudicando a interpretação clínica (KALENDER, 2011).

Para contornar essas limitações, diversas abordagens foram exploradas. Métodos de reconstrução iterativa e técnicas de filtragem foram amplamente utilizados para compensar a perda

de qualidade. No entanto, com a evolução da inteligência artificial, surgiram modelos baseados em aprendizado profundo capazes de modelar e remover o ruído de forma mais eficaz. Esses métodos aprendem a relação entre imagens de baixa e alta dose a partir de exemplos anotados, apresentando desempenho superior às técnicas tradicionais (CHEN et al., 2017a).

O interesse por LDCT impulsionou o desenvolvimento de bases de dados públicas, como a AAPM *Low Dose Grand Challenge*, que viabilizam pesquisas comparativas entre diferentes métodos de reconstrução e *denoising*. A integração de LDCT com métodos de aprendizado profundo representa uma frente de inovação na radiologia moderna (YANG et al., 2018)

2.3 Métodos Clássicos de Remoção de Ruído

Técnicas como os filtros Gaussiano, mediano e bilateral são amplamente usadas para reduzir ruídos em imagens (GONZALEZ; WOODS, 2009). Mesmo sendo simples, essas abordagens apresentam limitações. O principal problema é o comprometimento das bordas anatômicas, que podem ser suavizadas excessivamente, dificultando a visualização de estruturas finas (HAO; ZHOU; GUO, 2020). Isso é particularmente prejudicial em imagens médicas, nas quais detalhes sutis podem indicar informações relevantes.

Por outro lado, métodos de reconstrução iterativa, como o *Filtered Back Projection* (FBP), o *Algebraic Reconstruction Technique* (ART) e o *Statistical Iterative Reconstruction* (SIR), tentam modelar a aquisição física da imagem e iterativamente ajustar a reconstrução para reduzir o erro entre os dados reais e a estimativa computacional (KAK; SLANEY, 2001; BEISTER; KOLDITZ; KALENDER, 2012). Embora mais robustos, os métodos iterativos exigem maior poder computacional e tempo de processamento.

O desempenho desses métodos ainda é limitado quando a dose de radiação é muito baixa, pois a informação capturada pode não ser suficiente para uma reconstrução precisa (WANG, 2016). Essas limitações abriram espaço para o uso de métodos baseados em aprendizado de máquina, que conseguem lidar melhor com padrões complexos de ruído e preservação de detalhes estruturais (CHEN et al., 2017a; SHAN et al., 2018).

2.4 Redes Neurais Convolucionais

Com o avanço do aprendizado profundo, surgiram novas formas de lidar com ruído em imagens. As redes neurais convolucionais, *Convolutional Neural Networks*, (CNNs) se destacam por conseguirem extrair padrões e características relevantes automaticamente. Isso permite que modelos baseados nessas redes reconstruam imagens com maior precisão e menos perda de informação.

CNNs são redes neurais artificiais projetadas para processar dados com estrutura de grade, como imagens. Elas se diferenciam das redes neurais tradicionais por utilizarem operações de convolução, que extraem automaticamente padrões relevantes de imagens, como bordas, tex-

turas e formas, sem necessidade de intervenção manual (GOODFELLOW; BENGIO; COURVILLE, 2016).

A arquitetura de uma CNN, como ilustrado na Figura 3, é composta por camadas organizadas em três dimensões: altura, largura e profundidade, sendo essa última relacionada aos canais de cor ou aos filtros aprendidos, e não ao número total de camadas da rede. Cada camada transforma o volume de entrada em um novo volume de saída, aprendendo representações cada vez mais abstratas da imagem original.

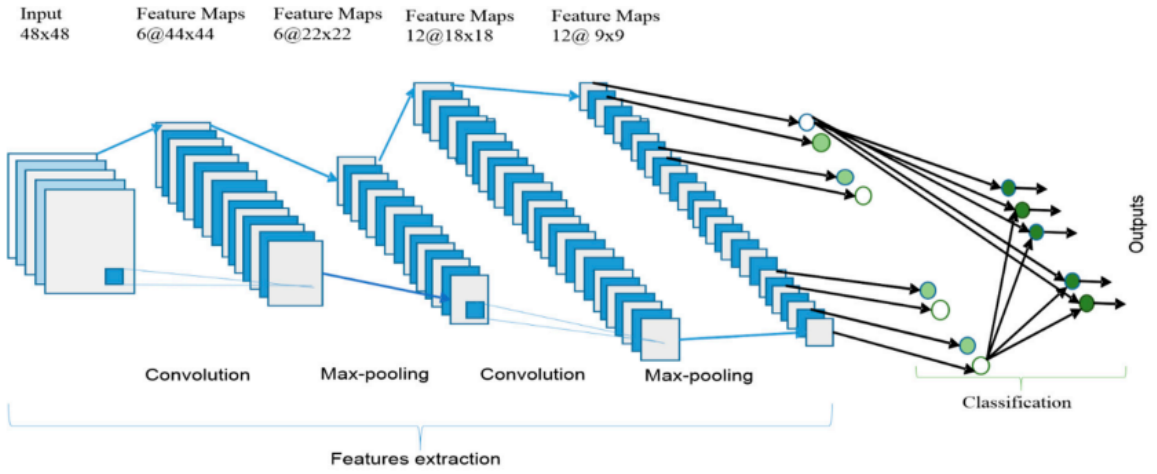


Figura 3 – Ilustração de uma CNN generalizada com suas camadas (ALOM et al., 2019).

A primeira camada essencial é a convolucional, que aplica filtros deslizantes sobre a imagem para gerar os chamados mapas de características (*feature maps*). Cada filtro é treinado para detectar um padrão específico, e seus pesos são ajustados durante o treinamento. Após a convolução, os resultados passam por uma função de ativação, geralmente ReLU, que introduz não-linearidade ao modelo, permitindo-o aprender relações mais complexas (VOULODIMOS et al., 2018).

A seguir, camadas de *pooling* são utilizadas para reduzir a dimensão espacial dos mapas de características, resumindo a informação e reduzindo a complexidade computacional. Essa redução também contribui para a generalização do modelo, pois diminui a chance de overfitting (VOULODIMOS et al., 2018). O tipo mais comum é o *max pooling*, que preserva os valores mais representativos em regiões locais.

O empilhamento de múltiplas camadas de convolução e *pooling* permite que a CNN aprenda representações hierárquicas: das estruturas mais simples nas primeiras camadas até padrões mais complexos nas camadas profundas. Essa capacidade torna as CNNs especialmente eficazes em tarefas de visão computacional, como classificação, detecção, segmentação e, no caso deste trabalho, remoção de ruído em imagens de tomografia de baixa dose (LDCT).

Durante o treinamento, a escolha dos hiperparâmetros como o tamanho dos filtros e o número de camadas influencia diretamente a capacidade da rede em capturar detalhes relevantes sem comprometer a estrutura da imagem original (GOODFELLOW; BENGIO; COURVILLE,

2016). O ajuste adequado desses parâmetros é crucial para que o modelo consiga distinguir entre ruído e sinal útil.

2.5 Modelos de Segmentação

Modelos de segmentação baseados em aprendizado profundo têm se mostrado eficazes em diversas tarefas de análise de imagem, especialmente na separação de regiões relevantes em contextos clínicos (RONNEBERGER; FISCHER; BROX, 2015; LITJENS et al., 2017). A segmentação divide uma imagem em áreas com características semelhantes, permitindo que algoritmos foquem em estruturas específicas, como tecidos, órgãos ou regiões ruidosas (LIU et al., 2021). Essa capacidade é particularmente útil na remoção de ruído em imagens LDCT, pois facilita a diferenciação entre detalhes anatômicos e artefatos indesejados.

Os modelos e suas arquiteturas foram implementados a partir da proposta de Iakubovskii (2019). O termo *backbone* refere-se, neste contexto, à porção codificadora (*encoder*) da arquitetura *encoder-decoder*, responsável pela extração de atributos em múltiplas resoluções espaciais (HE et al., 2016). Os *backbones* empregados foram previamente treinados em grandes bases de dados, como o ImageNet, permitindo uma extração de características robusta desde as primeiras camadas. Essa etapa é crucial, pois fornece à rede a capacidade de identificar e preservar padrões estruturais importantes, mesmo em imagens comprometidas por ruído.

A s arquitetura de um modelo de segmentação, como mostrado na Figura 4, possuí duas partes principais: o *encoder*, responsável por extrair características em múltiplas resoluções espaciais, e o *decoder*, que reconstrói uma versão refinada da imagem ou gera a máscara de segmentação. Essa máscara permite que a rede identifique regiões específicas da imagem onde o ruído deve ser atenuado (HESAMIAN et al., 2019).

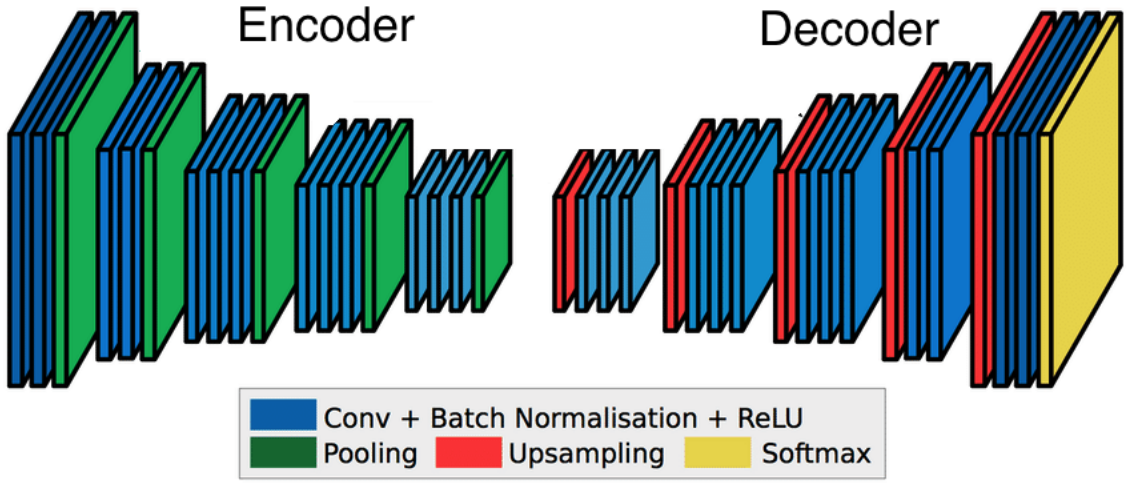


Figura 4 – Arquitetura *encoder-decoder* típica de modelos de segmentação (SYED; MORRIS, 2019).

A seguir, são apresentadas as arquiteturas utilizadas na pesquisa, como também os *backbones*

testados em combinação com cada uma delas. Essa abordagem permitiu avaliar como diferentes estratégias de extração e reconstrução de atributos impactam a eficácia na remoção de ruído.

2.5.1 U-Net

U-Net é uma arquitetura de CNN que inicialmente tinha o intuito de focar na segmentação de imagens biomédicas. Criada em 2015, a arquitetura é caracterizada por sua estrutura em forma de U, a qual lhe dá o nome e é mostrada na Figura 5. A U-Net consiste em um caminho de contração e um caminho expansivo.

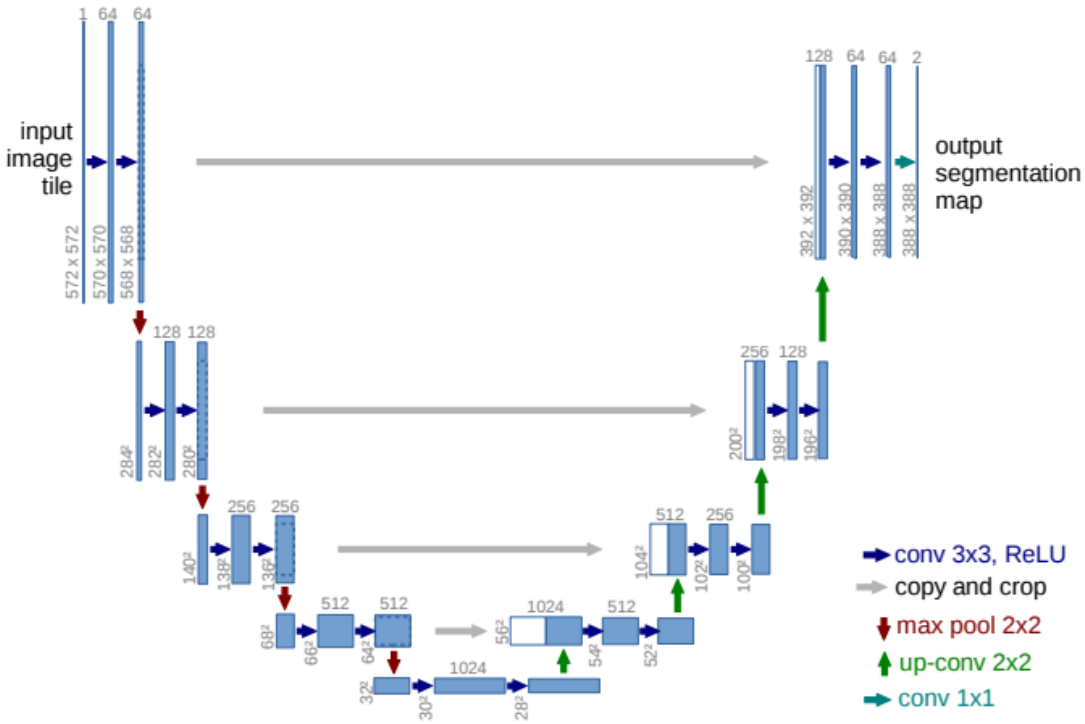


Figura 5 – Arquitetura U-Net (RONNEBERGER; FISCHER; BROX, 2015).

O caminho de contração, também chamado de caminho de codificação, captura o contexto da imagem de entrada usando uma série de camadas convolucionais e de *pool* máximo para reduzir a amostragem das dimensões espaciais. Ou seja, há a aplicação repetida de convoluções, seguidas por uma função de ativação (ReLU, nesse caso) e uma operação de agrupamento para *downsampling* (RONNEBERGER; FISCHER; BROX, 2015). Sendo assim, podemos dizer que a rede "contrai" as imagens originais. O caminho expansivo, ou de decodificação, usa camadas de *upsampling* e convoluções transpostas. A resolução espacial dos mapas de atributos é aumentada durante a fase de *upsampling*, permitindo que a rede reconstitua um mapa de segmentação denso. E esse mapa de segmentação vai ter as mesmas dimensões espaciais da imagem de entrada, com isso expandindo as imagens contraídas.

Conexões de *skip* são usadas na U-Net para conectar camadas correspondentes dos caminhos de codificação e decodificação. A rede coleta dados locais e globais justamente por conta desses links criados. Isso permite preservar detalhes espaciais essenciais e melhorar a acurácia da

segmentação, especialmente em regiões com ruído ou contraste baixo. Os mapas de atributos do caminho de codificação são concatenados com os do caminho de decodificação, possibilitando a reconstrução detalhada da imagem segmentada (RONNEBERGER; FISCHER; BROX, 2015).

A escolha da U-Net neste trabalho deve-se ao fato de que essa arquitetura foi originalmente desenvolvida para segmentação de imagens biomédicas, o que a torna naturalmente adequada para lidar com imagens médicas como as de tomografia computadorizada. Sua estrutura permite preservar e recuperar detalhes finos da imagem, mesmo após múltiplas operações de *downsampling* por conta das suas conexões de salto. Essa característica é essencial em tarefas que envolvem ruído, pois favorece a reconstrução precisa das regiões anatômicas de interesse.

2.5.2 LinkNet

A LinkNet é uma arquitetura de rede neural convolucional projetada especificamente para realizar segmentação semântica de forma eficiente, com baixo custo computacional. Sua concepção visa aplicações em tempo real, como veículos autônomos ou sistemas médicos com restrições de hardware, nos quais desempenho e leveza da rede são fatores críticos. A arquitetura foi introduzida por Chaurasia e Culurciello (2017) e desde então vem sendo utilizada em tarefas de segmentação de imagens com diferentes contextos.

A estrutura da LinkNet é composta por duas principais partes: o *backbone*, que realiza a extração progressiva das características da imagem, e o *decoder*, que reconstrói o mapa segmentado. Esse processo segue a ideia de reduzir a resolução da imagem para capturar atributos globais e depois restaurá-la utilizando os atributos aprendidos, garantindo que as informações espaciais relevantes sejam mantidas (CHAURASIA; CULURCIELLO, 2017).

Diferente da U-Net, que utiliza *skip connections* entre camadas simétricas do *backbone* e do *decoder*, a LinkNet introduz um método alternativo que emprega conexões residuais internas. Cada bloco da rede, tanto no *backbone* quanto no *decoder*, inclui uma conexão residual direta entre sua entrada e saída, como é mostrado na Figura 6. Isso significa que a entrada do bloco é somada à sua saída após passar pelas transformações convolucionais, técnica herdada da arquitetura ResNet. Essa abordagem facilita o treinamento da rede, como também melhora a propagação do gradiente e permite maior profundidade com menor risco de perda de informações (CHAURASIA; CULURCIELLO, 2017).

A ausência de conexões tradicionais de salto entre diferentes estágios da rede faz com que a LinkNet mantenha uma estrutura mais enxuta e, ao mesmo tempo, eficiente. Ao adotar um número reduzido de parâmetros em comparação com outras arquiteturas, a LinkNet demonstra ser eficaz na manutenção de detalhes estruturais (CHAURASIA; CULURCIELLO, 2017)

A LinkNet foi incluída neste estudo por seu equilíbrio entre desempenho e custo computacional. Por ser uma arquitetura leve, com menor quantidade de parâmetros em comparação com outras redes de segmentação, ela é particularmente indicada para aplicações em tempo real ou contextos com limitação de recursos computacionais. Além disso, o uso de conexões residuais internas facilita o treinamento de redes mais profundas e reduz perdas de informação.

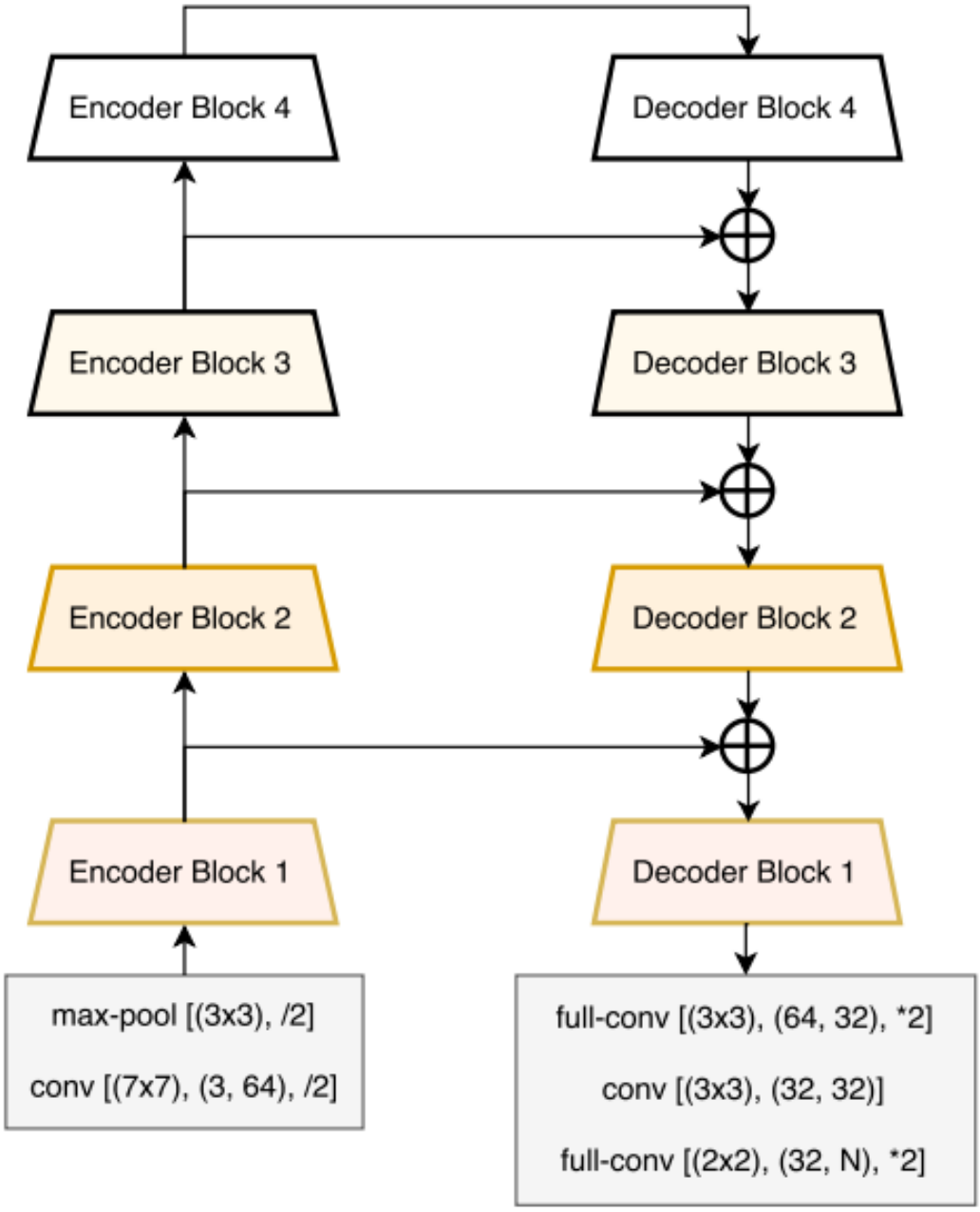


Figura 6 – Arquitetura LinkNet (CHAURASIA; CULURCIELLO, 2017).

2.5.3 PSPNet

A *Pyramid Scene Parsing Network* (PSPNet), proposta por (ZHANG et al., 2017), é uma arquitetura de segmentação semântica que se destaca pelo uso do módulo de *pooling* em pirâmide para capturar informações de contexto em múltiplas escalas. Essa estratégia torna o modelo altamente eficiente para compreender a cena global, mesmo quando os objetos possuem tamanhos variados ou aparecem em posições incomuns na imagem.

O modelo baseia-se na ResNet como *backbone*, responsável por extrair um mapa de características a partir da imagem de entrada. A característica principal da PSPNet está no seu módulo de *parsing* em pirâmide, que realiza agrupamentos adaptativos com janelas de diferentes tamanhos. Isso permite que o modelo processe a imagem em várias escalas simultaneamente, garantindo uma integração entre informações locais (detalhes) e globais (contexto de cena).

A arquitetura completa é composta por um caminho de codificação baseado na ResNet, seguido pelo módulo piramidal e por uma fase de reconstrução que utiliza camadas de convolução e *upsampling* para gerar o mapa final de segmentação. Como mostrado na Figura 7, a região central do modelo corresponde ao módulo de*pooling* em pirâmide, o qual desempenha papel crucial no ganho de robustez da arquitetura.

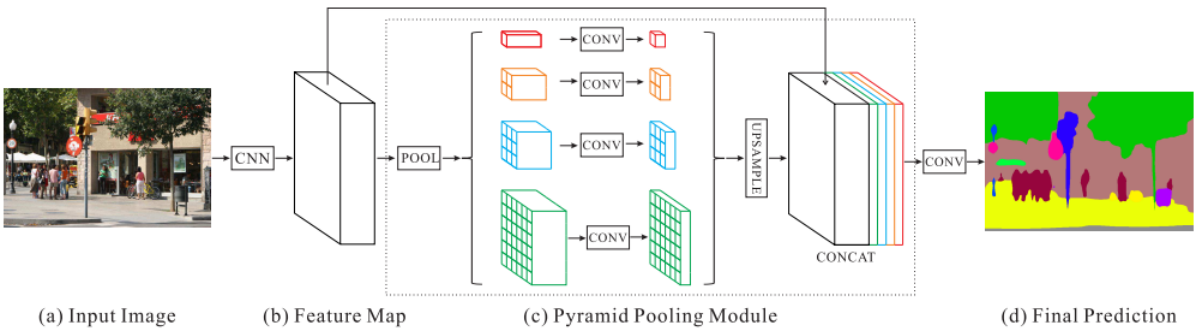


Figura 7 – Arquitetura PSPNet (ZHANG et al., 2017).

A inclusão da PSPNet neste trabalho foi motivada por seu desempenho superior em *benchmarks* de segmentação semântica, conforme evidenciado por estudos como (ZHU et al., 2021). Sua principal contribuição está na capacidade de capturar informações contextuais em múltiplas escalas por meio do módulo de *pooling* em pirâmide, o que permite integrar informações globais e locais de forma eficaz. Como a rede tinha resultados bons em tarefas de segmentação em cenas complexas, foi contemplado testar em imagens médicas.

2.5.4 MA-Net

A Multi-scale Attention Network (MA-Net), desenvolvida por Fan et al. (2020), foi projetada para refinar a extração de características por meio de mecanismos de atenção aplicados tanto no espaço quanto nos canais da imagem. A proposta central da arquitetura é integrar informações

locais e contextuais globais, ajustando dinamicamente a importância das diferentes regiões e filtros da imagem durante o processo de segmentação.

A MA-Net é composta por dois módulos principais: o *Position-wise Attention Block* (PAB), responsável por aplicar atenção espacial, e o *Multi-scale Fusion Attention Block* (MFAB), que implementa atenção por canal. O PAB foca em destacar as regiões da imagem com maior relevância para a tarefa de segmentação, enquanto o MFAB permite que o modelo aprenda quais canais contêm as informações mais discriminativas. Essa separação clara entre atenção espacial e por canal facilita a especialização dos filtros da rede e amplia a capacidade do modelo de identificar detalhes relevantes mesmo em imagens com ruído intenso.

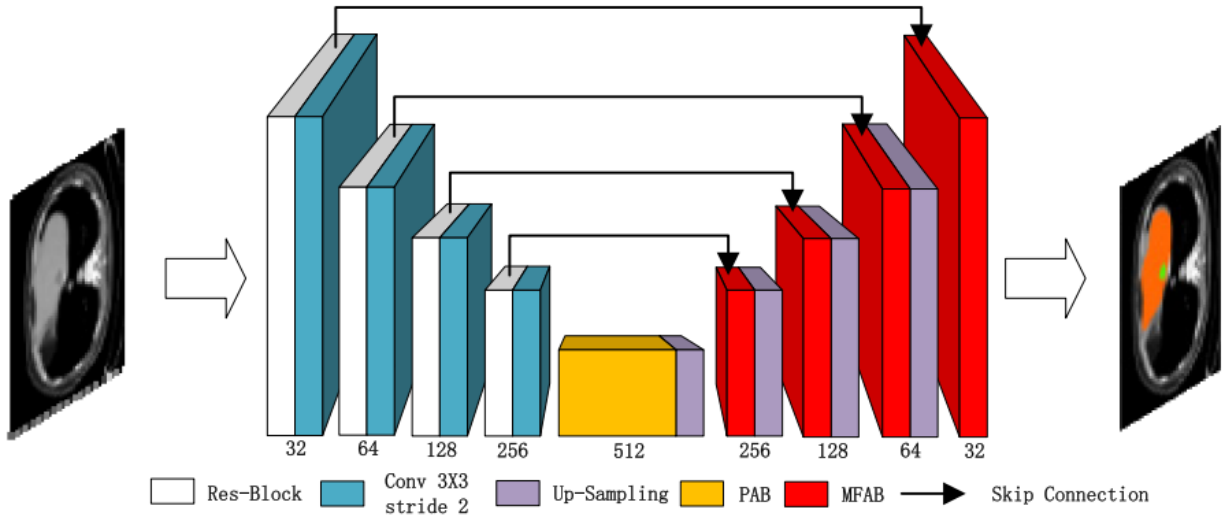


Figura 8 – Arquitetura MA-Net (FAN et al., 2020).

Como representado na Figura 8, a arquitetura combina esses dois blocos de maneira que os mapas de atenção gerados refinam as ativações da rede ao longo do processo de codificação e decodificação. Isso permite à rede destacar estruturas anatômicas significativas e suprimir regiões ruidosas ou com baixa informação.

A motivação para o uso da MA-Net neste trabalho está na sua habilidade de tratar eficientemente contextos desafiadores como o das imagens LDCT. Essas imagens apresentam variações de ruído em diferentes escalas, e a capacidade da MA-Net de processar múltiplos níveis de detalhe por meio de atenção torna-a uma candidata promissora. A adaptação da rede às particularidades do conteúdo da imagem, ajustando o foco espacial e de canais, contribui diretamente para uma reconstrução mais limpa da imagem segmentada.

Além disso, a presença de mecanismos de atenção permite que a MA-Net opere de forma seletiva, aprendendo a distinguir padrões estruturais dos ruídos que os degradam. Essa seletividade é particularmente importante em cenários clínicos, onde detalhes finos como bordas de órgãos ou lesões podem ser facilmente mascarados por artefatos.

2.5.5 Backbones

Para aprimorar a capacidade de extração de atributos das imagens, foram utilizados diferentes modelos de *backbones* pré-treinados como base para as arquiteturas de segmentação. Eles foram treinados previamente no conjunto ImageNet e fornecem representações robustas de imagens em diferentes níveis de abstração.

O ResNet50, por exemplo, possui 50 camadas e é conhecido por suas conexões residuais que permitem o treinamento de redes profundas ao evitar o problema do desaparecimento do gradiente (HE et al., 2016). Essa característica faz com que o modelo mantenha informações de baixo e alto nível, sendo útil para preservar estruturas relevantes em imagens ruidosas.

Já o VGG-16 apresenta uma arquitetura sequencial e simples, com convoluções de tamanho 3×3, que favorecem a captação de detalhes locais. Embora tenha uma quantidade maior de parâmetros, sua eficiência na extração de padrões hierárquicos a torna uma escolha recorrente em tarefas de segmentação, especialmente em contextos onde a estrutura das imagens precisa ser bem preservada (YU et al., 2020).

A DenseNet121 se destaca por suas conexões densas, onde cada camada recebe como entrada as saídas de todas as camadas anteriores. Isso promove um reaproveitamento das características extraídas e melhora o fluxo de informações, resultando em representações mais completas com menor risco de perda de detalhes relevantes (HUANG et al., 2017).

O Inceptionv4 combina a arquitetura Inception, que processa múltiplas escalas simultaneamente, com conexões residuais. Essa união permite à rede analisar padrões em diferentes níveis de granularidade, facilitando a separação de ruído e estrutura em imagens médicas com variação de textura (SZEGEDY et al., 2017).

O MobileNetv2 foi projetado para ambientes com restrições computacionais. Utiliza convoluções separáveis em profundidade e blocos residuais invertidos, reduzindo o número de operações necessárias sem comprometer significativamente a precisão. Essa economia de recursos o torna adequado para aplicações móveis ou clínicas com hardware limitado (SANDLER et al., 2018).

Por fim, o EfficientNetB2 adota a técnica de escalonamento composto, ajustando proporcionalmente profundidade, largura e resolução da rede. Isso garante um bom equilíbrio entre desempenho e eficiência computacional, sendo particularmente útil em cenários que exigem baixo consumo de memória, mas alta precisão (TAN; LE, 2019).

O uso combinado desses *backbones* com diferentes modelos de segmentação permitiu uma análise abrangente sobre como as extrações iniciais de atributos afetam o resultado final da remoção de ruído, sendo uma das principais contribuições do trabalho.

Capítulo 3

Trabalhos Relacionados

Este capítulo apresenta um levantamento dos principais trabalhos relacionados à remoção de ruído em imagens médicas, com ênfase em tomografias computadorizadas de baixa dosagem (LDCT). São discutidas abordagens baseadas em redes neurais convolucionais (CNN), redes com mecanismos de atenção (AN) e *Generative Adversarial Network* (GAN), destacando seus desempenhos em termos quantitativos (PSNR, SSIM), conjuntos de dados utilizados, estratégias de treinamento e limitações. Ao final, uma comparação com o modelo proposto do trabalho é apresentada.

Modelos baseados em CNN têm sido amplamente utilizados na remoção de ruídos em imagens médicas. Chen et al. (2017a) propuseram o RED-CNN, uma rede convolucional com arquitetura de *autoencoder* e conexões residuais, voltada para a atenuação de ruído em imagens LDCT. Em seus experimentos, o modelo obteve PSNR variando entre 38,99 e 39,20 e SSIM entre 0,933 e 0,934, destacando-se como um dos melhores entre os métodos comparados na época, tanto em dados simulados quanto clínicos. A rede mostrou bom desempenho geral, especialmente na preservação de bordas e texturas finas, mas pode ser limitada por sua dependência exclusiva da função de perda baseada em erro quadrático médio (MSE), que nem sempre representa bem a percepção visual (CHEN et al., 2017a).

Outras abordagens, como a de (ZHANG et al., 2021), propuseram o uso de arquiteturas inspiradas em redes Inception e GoogleNet, com extração de multi-características. O modelo mostrou bons resultados quantitativos, mas exigiu alto custo computacional e apresentou riscos de overfitting em função do grande número de mapas de características (GOODFELLOW et al., 2014).

Técnicas baseadas em GAN também têm sido exploradas. Tran, Nguyen e Arai (2020) propuseram um modelo em duas etapas utilizando GANs para simular ruído realístico em imagens brutas e posteriormente treinar uma CNN para removê-lo. Embora o modelo tenha apresentado ganhos na generalização, ele depende de metadados e estimativas do nível de ruído, o que pode limitar sua aplicação clínica (TRAN; NGUYEN; ARAI, 2020).

Mecanismos de atenção foram empregados em trabalhos como (TIAN et al., 2020), que utilizaram atenção espacial para refinar a restauração de imagens CT. Apesar da melhora visual observada, o modelo não obteve desempenho competitivo nas métricas PSNR e SSIM.

Zhang et al. (2021), por sua vez, apresentaram uma evolução utilizando múltiplos módulos de atenção, alcançando melhor preservação estrutural e maior qualidade quantitativa.

Wang et al. (2023) propuseram o modelo DADN, baseado em aprendizado por transferência e atenção dual, treinado com os dados do AAPM 2016 — o mesmo utilizado neste trabalho. O modelo demonstrou resultados superiores em comparação com abordagens anteriores, especialmente por sua capacidade de adaptação ao domínio e preservação de detalhes finos.

Tabela 1 – Comparação entre trabalhos relacionados

Autor	**Modelo**	**Tipo Imagem**	**Base**	**PSNR**	**SSIM**	**Treino**
Chen et al. (2017)	RED-CNN (CNN + Res. AE)	LDCT simulada (ROI)	AAPM 2016	38,99–39,20	0,933–0,934	Sup.
Zhang et al. (2021)	CNN tipo Inception	CT abd.	Base própria	–	–	Sup.
Tran et al. (2020)	GAN + CNN	CT c/ ruído real	Simulada	35,5	0,920	Semi-sup.
Tian et al. (2020)	Atenção espacial (AN)	CT simulada	Simulada	34,7	0,901	Sup.
Zhang et al. (2023)	Atenção múltipla (AN)	LDCT simulada	AAPM 2016	39,1	0,947	Sup.
Wang et al. (2023)	DADN	LDCT simulada	AAPM 2016	45,53	0,973	Sup.
Este trabalho	U-Net+Inceptionv4	LDCT simulada abd.	AAPM 2016	**48,797**	**0,977**	Sup.

Nota: Sup. = supervisionado; Semi-sup. = semi-supervisionado; LDCT = tomografia computadorizada de baixa dose (simulada); Res. AE = residual autoencoder; abd. = abdominal; c/ = com; ROI = região de interesse.

A Tabela 1 resume as principais características dos trabalhos analisados. Nota-se que as abordagens que incorporam atenção e adaptação ao domínio obtiveram melhores resultados quantitativos. Ainda assim, há lacunas na literatura quanto à avaliação sistemática do impacto de diferentes codificadores e arquiteturas de segmentação na tarefa de remoção de ruído em LDCT.

Em comparação com os trabalhos discutidos, o trabalho se diferencia ao realizar uma análise abrangente da combinação entre diferentes arquiteturas de segmentação com *backbones* variados, com e sem mecanismos de atenção. Os experimentos foram conduzidos utilizando exclusivamente dados reais do conjunto AAPM 2016, com avaliação quantitativa. Isso proporciona uma comparação justa e detalhada com os métodos do estado da arte, contribuindo para o avanço na análise das pesquisas de reconstrução de imagens médicas com redes de segmentação.

A partir dessa análise, conclui-se que há espaço para aprimorar o desempenho das redes de remoção de ruído por meio da escolha adequada de *backbone* e arquitetura, com potencial de obter melhor preservação estrutural e desempenho geral, conforme será demonstrado no Capítulo 4.

Capítulo 4

Metodologia

Neste capítulo, é apresentada a metodologia adotada para a remoção de ruído em imagens LDCT utilizando modelos de segmentação. São descritos o conjunto de dados utilizado, as etapas de pré-processamento, os modelos treinados, os parâmetros empregados durante o treinamento, e as métricas de avaliação aplicadas para mensurar a qualidade das imagens reconstruídas.

4.1 Conjunto de dados

Foi utilizado o conjunto de imagens AAPM 2016 *Low-dose CT Grand Challenge*, da contribuição de *Mayo Clinic* e organizado pela *American Association of Physicists in Medicine* (AAPM) [1].

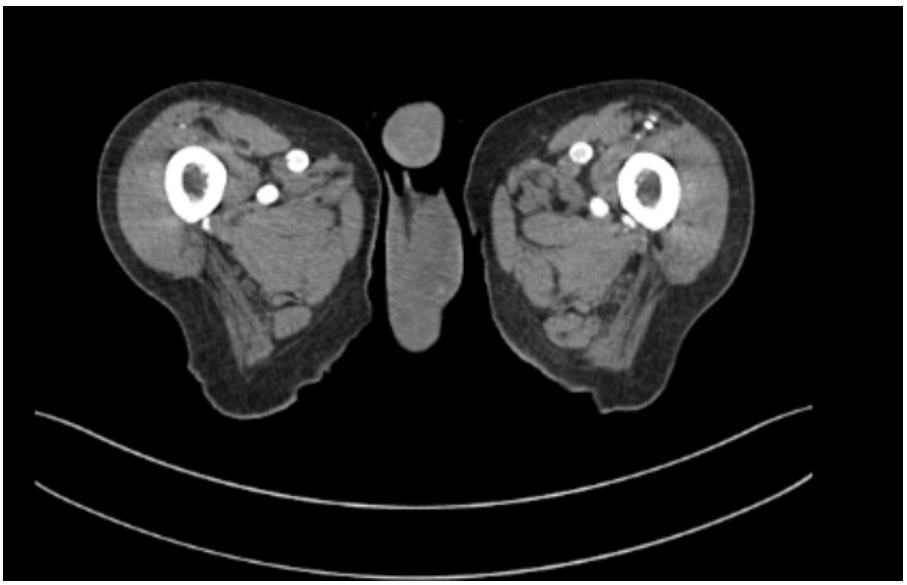
O conjunto de dados utilizado neste trabalho é composto por 30 exames de tomografia computadorizada abdominal com contraste, realizados na fase venosa portal (também conhecida como fase hepática), utilizando o equipamento Siemens SOMATOM Flash. Para cada paciente, estão disponíveis imagens adquiridas em dois níveis de dose de radiação: uma com dose completa (padrão), denominada NDCT (Normal Dose CT) e outra com um quarto da dose, referida como LDCT (Low Dose CT). Uma observação a ser feita é que as imagens LDCT não foram adquiridas diretamente, mas sim simuladas a partir das imagens NDCT por meio da adição de ruído de Poisson (MCCOLLOUGH et al., 2017).

No total, o conjunto contém 2.378 cortes axiais distribuídos entre os 30 pacientes, sendo que para cada fatia NDCT existe uma correspondente LDCT simulada. A Figura 9 ilustra visualmente a diferença entre uma fatia de imagem com dose normal e sua versão simulada com dose reduzida.

As imagens foram reconstruídas com diferentes espessuras de fatia e *kernels*. Para este trabalho, optou-se por utilizar cortes com espessura de 3 mm, valor comumente adotado na literatura por equilibrar nível de detalhe anatômico e custo computacional.

Também foram selecionadas imagens reconstruídas com o *kernel* B30, classificado como de suavização média. Essa escolha proporciona um bom equilíbrio entre preservação de estruturas,

[1] Low-dose CT Grand Challenge



(a) Dose completa (NDCT)

(b) Dose de um quarto (LDCT)

Figura 9 – Imagem de dose completa e quarto de dose do paciente “L506”.

controle de ruído e contraste, sendo amplamente utilizada em estudos semelhantes. A padronização desses parâmetros visa garantir reprodutibilidade e facilitar comparações com trabalhos correlatos.

As imagens estão organizadas ao longo do eixo axial (Z), de forma que cada fatia representa uma seção transversal do corpo humano, disposta sequencialmente da parte superior à inferior. Essa organização permite um emparelhamento direto entre os cortes equivalentes das versões NDCT e LDCT, o que é fundamental para o treinamento supervisionado dos modelos.

4.2 Remoção de ruído usando DL

A proposta deste trabalho é aplicar arquiteturas de segmentação baseadas em redes neurais convolucionais (CNNs) para a tarefa de remoção de ruído em imagens LDCT. Diferente de abordagens tradicionais que utilizam redes específicas para remoção de ruído, este estudo avalia se modelos de segmentação, que originalmente foram desenvolvidos para segmentação semântica, podem ser eficazes nessa tarefa, aproveitando sua capacidade de identificar padrões nas imagens e reconstruir as regiões afetadas pelo ruído.

Foram utilizados modelos como U-Net, LinkNet, PSPNet e MA-Net, combinados com diferentes *backbones* pré-treinados. A expectativa é que essas redes, ao aprenderem padrões espaciais complexos, consigam separar de forma eficiente o ruído das estruturas anatômicas relevantes, gerando imagens com melhor qualidade visual e diagnóstica.

4.2.1 Pré-processamento

O pré-processamento das imagens envolveu três etapas principais: conversão para unidades Hounsfield (HU), normalização dos valores de intensidade e aplicação de técnicas de aumento de dados. Primeiramente, as imagens foram convertidas para a escala de HU, que é amplamente utilizada na área médica por expressar de forma padronizada a densidade dos tecidos. Essa conversão é essencial para garantir que os modelos de rede aprendam a partir de informações clínicas significativas, e não apenas de intensidades relativas dos *pixels* ((KAK; SLANEY, 2001)).

Em seguida, as imagens foram normalizadas para o intervalo [0,1]. Essa normalização é uma prática comum no treinamento de redes neurais profundas, pois facilita a convergência dos algoritmos de otimização, melhora a estabilidade numérica e reduz o risco de explosão ou desaparecimento do gradiente (GOODFELLOW; BENGIO; COURVILLE, 2016).

Por fim, foram aplicadas técnicas de aumento de dados com o objetivo de expandir a diversidade do conjunto de treinamento sem a necessidade de adquirir novas imagens. As transformações incluíram:

- Aplicação de um desfoque gaussiano com limite de intensidade entre 1 e 3, utilizado para suavizar a imagem de forma leve (probabilidade de 0,4);
- Efeito de desfoque por movimento da câmera, com limite de desfoque 3 e probabilidade de 0,2, que auxilia na regularização do modelo;
- Variações aleatórias leves de brilho e contraste, simulando pequenas variações nas condições de iluminação;
- Adição de ruído gaussiano à imagem, com variância entre 5,0 e 15,0, para ajudar o modelo a lidar com diferentes tipos de ruído.

A aplicação dessas técnicas é justificada pela necessidade de tornar o modelo mais robusto a pequenas variações naturais nas imagens médicas, como mudanças na posição do paciente, iluminação ou ruído simulado (SHORTEN; KHOSHGOFTAAR, 2019).

Além disso, optou-se por não utilizar aumentos agressivos, como deformações elásticas ou ruídos adicionados artificialmente, pois essas alterações podem interferir na estrutura anatômica da imagem e confundir o modelo durante o aprendizado, especialmente em tarefas que dependem da preservação de detalhes finos, como é o caso da remoção de ruído em LDCT (GOCERI, 2023; SHAH et al., 2021).

Essas etapas de pré-processamento, portanto, visam garantir que o modelo aprenda de forma eficiente e generalizável, equilibrando diversidade no treinamento com a preservação da integridade das informações clínicas relevantes.

4.2.2 Parâmetros e Configurações do Modelo

Os modelos foram implementados utilizando a biblioteca *segmentation-models-pytorch* (SMP), que oferece arquiteturas eficientes para segmentação e regressão de imagens, com *backbones* pré-treinados no ImageNet, otimizando a capacidade de generalização mesmo com conjuntos de dados limitados. Essa biblioteca permite acessar facilmente diferentes*backbones* com variados tamanhos e complexidades, influenciando diretamente o número de parâmetros treináveis. Na tabela 2 são mostradas as estimativas aproximadas de parâmetros para cada combinação. Por exemplo, modelos como o VGG16 possuem cerca de 138 milhões de parâmetros, enquanto arquiteturas mais compactas, como MobileNetV2, possuem aproximadamente 3,5 milhões, o que impacta o tempo de treinamento e a capacidade de evitar *overfitting* (IAKUBOVSKII, 2019).

As configurações de treinamento adotadas neste trabalho foram cuidadosamente selecionadas com o objetivo de garantir a convergência eficiente do modelo, evitando o overfitting e respeitando as particularidades do problema. O número de épocas foi definido em 50, considerado suficiente para que o modelo convergisse de forma estável sem incorrer em sobreajuste, conforme indicado em estudos anteriores (GOODFELLOW; BENGIO; COURVILLE, 2016).

O tamanho do *batch* utilizado foi 4, o que representa um compromisso entre o uso eficiente da memória da GPU e a estabilidade nas atualizações dos gradientes, um fator especialmente relevante ao se trabalhar com imagens médicas de alta resolução. A função de perda escolhida foi o erro quadrático médio (*Mean Squared Error*, MSE), apropriada para problemas de regressão contínua como a remoção de ruído, pois penaliza mais fortemente os erros maiores, promovendo predições mais precisas (LU, 2010).

Como otimizador, optou-se pelo Adam, dada sua robustez e capacidade de ajustar automaticamente as taxas de aprendizado para cada parâmetro, o que acelera a convergência em relação a métodos mais tradicionais, como o Stochastic Gradient Descent (SGD) (SUN et al., 2024). A taxa de aprendizado inicial foi fixada em 1×10^{-4}, um valor conservador que ajuda a evitar grandes oscilações na atualização dos pesos durante as etapas iniciais do treinamento.

Para melhorar ainda mais a convergência, foi empregado o *scheduler* ReduceLROnPlateau, configurado com um parâmetro de *patience* de 5 épocas e um fator de decaimento de 0,5. Esse mecanismo reduz automaticamente a taxa de aprendizado quando o desempenho na validação se estabiliza, contribuindo para escapar de mínimos locais e refinar o ajuste final do modelo (SMITH, 2017).

Tabela 2 – Número total de parâmetros (em milhões) para as diferentes combinações de modelos e *backbones*.

Model	**VGG16**	**Inceptionv4**	**DenseNet121**	**ResNet50**	**EfficientNet-B3**	**MobileNet_v2**
U-Net	23M	48M	13M	32M	10M	6M
LinkNet	16M	46M	10M	31M	7M	4M
PSPNet	15M	41M	7M	24M	7M	2M
MA-Net	27M	114M	44M	147M	13M	48M

4.2.3 Treinamento

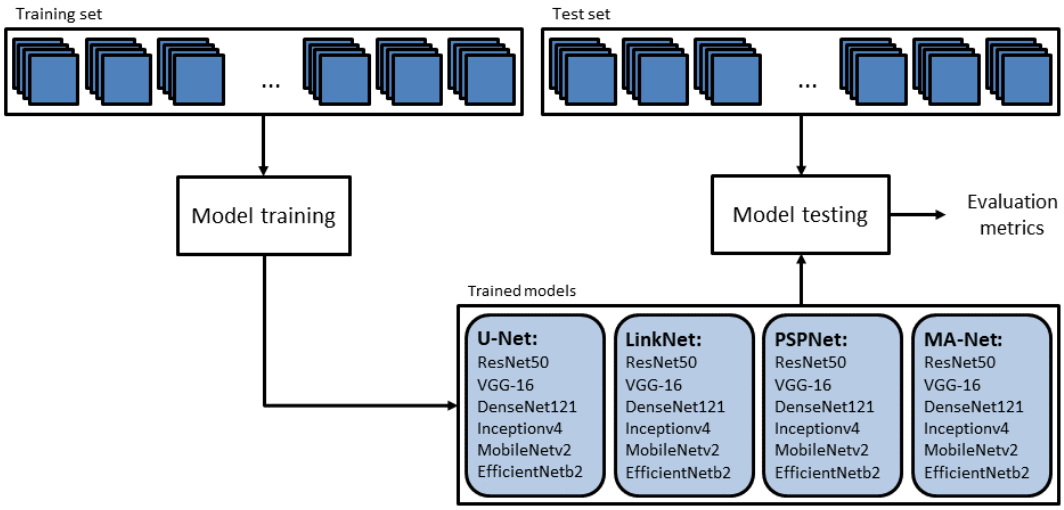


Figura 10 – Fluxo de trabalho do design do experimento, incluindo divisão do conjunto de dados, treinamento do modelo e avaliação.

A Figura 10 apresenta um resumo do processo experimental adotado neste trabalho. O conjunto de dados foi dividido pelo método holdout, destinando 80% para treinamento e 20% para teste. Essa estratégia é amplamente utilizada em contextos com volume limitado de dados, como em aplicações médicas, devido à sua simplicidade, eficiência computacional e capacidade de manter a independência entre treinamento e avaliação final (KOHAVI et al., 1995).

Para garantir o correto pareamento entre imagens LDCT e NDCT, foram implementadas classes personalizadas do tipo *NumpyDataset*, que carregam e sincronizam os dados de forma eficiente durante o treinamento. Devido ao tamanho e complexidade das imagens como previamente mencionado, o *batch size* foi fixado em 4, promovendo um equilíbrio entre o uso da memória da GPU e a estabilidade na atualização dos parâmetros do modelo a cada iteração.

Cada arquitetura de rede foi combinada com seis diferentes *backbones*, totalizando 24 combinações distintas avaliadas. Os *backbones* foram carregados com pesos pré-treinados no ImageNet, o que favorece a extração de características visuais robustas e melhora a capacidade de generalização dos modelos em tarefas de remoção de ruído, conforme discutido em Wang et al. (2021).

Durante o treinamento, o conjunto de treinamento foi subdividido para incluir um conjunto de validação interna, utilizado para monitorar o desempenho do modelo e ajustar hiperparâmetros, como a taxa de aprendizado, através de técnicas como *early stopping* e *learning rate scheduling*. Esta validação interna é fundamental para prevenir *overfitting*, garantindo que o modelo não aprenda características específicas do conjunto de treinamento que não generalizam para novos dados. O conjunto de teste permaneceu completamente separado e foi utilizado exclusivamente para a avaliação final da capacidade de generalização dos modelos, assegurando uma análise imparcial e confiável dos resultados (GOODFELLOW; BENGIO; COURVILLE, 2016). Para isso, métricas quantitativas específicas foram calculadas para avaliar a qualidade das imagens reconstruídas, incluindo medidas como PSNR e SSIM.

4.3 Métricas de avaliação

Para comparar o desempenho de um modelo de remoção ruído, ou avaliar a melhora observada após a redução de ruído em um conjunto de dados específico, é necessário ter uma métrica de qualidade quantitativa. Existem duas medidas que são bastante usadas na literatura, as clássicas PSNR e SSIM.

4.3.1 PSNR

O *Peak Signal to Noise Ratio* (PSNR) é uma métrica que quantifica a razão entre o sinal máximo possível de uma imagem e o ruído presente na imagem reconstruída, expressa em decibéis. Para isso, utiliza-se o erro quadrático médio (MSE) entre a imagem original limpa s e a imagem ruidosa ou reconstruída x, definido por:

$$MSE(x, s) = \frac{1}{n} \sum_{i=1}^{n} (x_i - s_i)^2$$

A partir do MSE, o PSNR é calculado como:

$$PSNR(x, s) = 10 \log_{10} \frac{R^2}{MSE(x, s)}$$

onde R representa o valor máximo possível do pixel na imagem (por exemplo, 255 para imagens de 8 bits). Um valor alto de PSNR indica uma boa correspondência entre x e s, ou seja, baixo nível de ruído, enquanto valores baixos indicam maior discrepância entre as imagens (CHOW; PARAMESRAN, 2016).

O PSNR é amplamente utilizado devido à sua simplicidade e facilidade de cálculo, sendo um padrão para avaliar quantitativamente a precisão de métodos de restauração de imagem

(BUADES; COLL; MOREL, 2005). No entanto, ele mede apenas diferenças pixel a pixel, sem considerar aspectos perceptuais, o que pode resultar em alta pontuação mesmo quando distorções visuais perceptíveis estão presentes (WANG; BOVIK, 2006).

4.3.2 SSIM

O *Structural Similarity Index* (SSIM) foi desenvolvido para superar limitações de métricas tradicionais, avaliando a similaridade estrutural entre duas imagens x e s por meio de componentes de luminância, contraste e estrutura local, conforme a fórmula:

$$SSIM(x, s) = \frac{(2\mu_x\mu_s + c_1)(2\sigma_{xs} + c_2)}{(\mu_x^2 + \mu_s^2 + c_1)(\sigma_x^2 + \sigma_s^2 + c_2)}$$

onde μ_x e μ_s são as médias das intensidades dos pixels das imagens x e s, representando o brilho; σ_x^2 e σ_s^2 são as variâncias, que indicam o contraste; e σ_{xs} é a covariância que captura a correlação estrutural entre as imagens. As constantes c_1 e c_2 são incluídas para evitar divisões por zero e garantir estabilidade numérica (WANG et al., 2004).

O SSIM varia entre 0 e 1, sendo 1 a correspondência estrutural perfeita entre as imagens. Essa métrica é considerada mais próxima da percepção humana, pois avalia a preservação de padrões visuais essenciais, além de refletir diferenças em luminância e contraste (WANG; BOVIK, 2006). Por essa razão, o SSIM é amplamente utilizado para avaliar qualidade visual em imagens médicas, onde a preservação da estrutura anatômica é fundamental para diagnósticos confiáveis (ZHU; WANG; LIU, 2007).

Apesar de sua maior fidelidade perceptual, o SSIM possui algumas limitações, como maior complexidade computacional em relação ao PSNR e sensibilidade a parâmetros do cálculo local, como o tamanho da janela de análise (WANG et al., 2004).

Capítulo 5

Análise e discussão dos resultados

Este capítulo apresenta e discute os resultados obtidos a partir dos experimentos realizados com redes de segmentação aplicadas à remoção de ruído em imagens LDCT. A análise compreende aspectos quantitativos, perceptuais e comparativos, fornecendo uma avaliação ampla sobre o desempenho das arquiteturas testadas.

Foram realizados 24 experimentos, combinando cada uma das 4 arquiteturas de redes neurais U-Net, PSPNet, LinkNet e MA-Net com 6 modelos de *backbones* diferentes: VGG16, DenseNet121, Efficient-B2, InceptionV4, MobileNet_v2 e ResNet50 [1]. Essa combinação total resulta em 24 configurações experimentais distintas.

O treinamento e validação foram feitos em um ambiente Colab Pro, no qual a média de tempo gasto foi entre 2 e 4 horas por modelo. Para avaliar os modelos, foram utilizadas métricas quantitativas especificadas na Seção 4.3.

5.1 Análise Quantitativa

Esta seção apresenta a análise quantitativa dos resultados obtidos nos experimentos, utilizando as métricas PSNR e SSIM.

As tabelas 3 e 4 a seguir apresentam os valores médios e desvios padrão de cada combinação testada. Os modelos com melhores desempenhos em cada rede são destacados em negrito.

	LinkNet	PSPNet	U-Net	MA-Net
VGG16	**48,400 ± 0,623**	32,327 ± 0,376	**47,959 ± 0,606**	**48,609 ± 0,654**
Inceptionv4	27,824 ± 15,632	32,427 ± 0,411	**48,797 ± 0,672**	**48,651 ± 0,656**
ResNet50	**44,823 ± 4,255**	32,019 ± 1,055	48,138 ± 0,746	29,747 ± 13,803
DenseNet121	25,349 ± 21,704	32,428 ± 0,396	**48,671 ± 0,471**	35,817 ± 17,324
EfficientNet-b2	**45,672 ± 0,517**	32,434 ± 0,388	46,564 ± 0,404	36,877 ± 9,726
MobileNet_v2	44,668 ± 0,894	32,442 ± 0,380	46,717 ± 0,537	39,362 ± 7,867

Tabela 3 – Resultados obtidos para o PSNR para as diferentes combinações de redes e *backbones*

A arquitetura U-Net, independentemente do *backbone* utilizado, teve os melhores desempenhos médios tanto em PSNR quanto em SSIM. A combinação U-Net com InceptionV4 foi a

[1] Segmentation Models Pytorch

	LinkNet	PSPNet	U-Net	MA-Net
VGG16	**0,974 ± 0,005**	0,799 ± 0,020	**0,975 ± 0,005**	**0,973 ± 0,005**
Inceptionv4	**0,974 ± 0,005**	0,803 ± 0,020	**0,977 ± 0,005**	**0,974 ± 0,005**
ResNet50	**0,973 ± 0,006**	0,800 ± 0,023	**0,976 ± 0,005**	0,956 ± 0,022
DenseNet121	0,931 ± 0,051	0,798 ± 0,020	**0,976 ± 0,003**	0,952 ± 0,042
EfficientNet-b2	0,972 ± 0,006	0,798 ± 0,020	0,918 ± 0,011	**0,966 ± 0,006**
MobileNet_v2	0,968 ± 0,007	0,788 ± 0,019	0,959 ± 0,007	0,965 ± 0,010

Tabela 4 – Resultados obtidos para o SSIM para as diferentes combinações de redes e *backbones*.

que mais se destacou, alcançando PSNR de 48,797 e SSIM de 0,977, seguida pelas combinações U-Net com DenseNet121, ResNet50 e VGG16. Esses modelos mantiveram valores de PSNR acima de 48 e SSIM igual ou superior a 0,975, sugerindo uma excelente capacidade de reconstruir imagens com alta fidelidade. O bom desempenho está diretamente relacionado à robustez dos *backbones* utilizados, que possibilitaram uma extração de atributos eficiente, e ao fato de a U-Net possuir uma estrutura que favorece a retenção de detalhes finos, principalmente em imagens médicas.

As combinações com piores resultados foram as baseadas na arquitetura PSPNet, com destaque para os experimentos com DenseNet121, ResNet50 e VGG16, que tiveram PSNR em torno de 32 e SSIM entre 0,79 e 0,80. Mesmo utilizando *backbones* mais potentes, a PSPNet não apresentou ganhos consistentes. A principal hipótese é que a estratégia de *pyramid pooling* dessa arquitetura, voltada para captar o contexto global da imagem, não favorece a preservação de microestruturas importantes nas tarefas de remoção de ruído, que exigem mais atenção ao nível local. Isso também se refletiu nas combinações com MobileNet_v2, EfficientNet-B2 e InceptionV4, que, apesar de suas particularidades, não conseguiram melhorar significativamente os resultados dentro dessa arquitetura.

Outro ponto importante é a relação entre o desempenho dos modelos e a quantidade de parâmetros em cada combinação, como mostrado na Tabela 2. Em geral, as melhores médias de PSNR e SSIM foram alcançadas por modelos com número moderado a alto de parâmetros, como U-Net com InceptionV4 (48M) e DenseNet121 (13M). Porém, nem sempre um maior número de parâmetros significa melhor desempenho. Um exemplo disso é a combinação MA-Net com ResNet50, que apesar de ter cerca de 147 milhões de parâmetros, teve desempenho inconsistente, sugerindo que o simples aumento da complexidade não garante bons resultados se a arquitetura não for adequada à tarefa.

Ao contrário, modelos mais leves como os que utilizam o *backbone* MobileNet_v2, mesmo com apenas 2 a 6 milhões de parâmetros, apresentaram desempenho competitivo em algumas combinações. Por exemplo o modelo LinkNet alcançou um SSIM de 0,968, e com U-Net, um PSNR de 46,717, ambos resultados acima da média geral. Esses casos mostram que arquiteturas eficientes, quando bem ajustadas e combinadas com uma estrutura adequada (como a U-Net), conseguem gerar boas predições mesmo com limitações de capacidade. Isso destaca que o número de parâmetros é apenas um dos fatores que influenciam no desempenho e que o equilíbrio entre arquitetura, função de perda, otimização e capacidade de generalização é mais

determinante do que apenas o tamanho do modelo.

O desempenho dos modelos também foi influenciado por aspectos do treinamento. O uso do otimizador Adam com taxa de aprendizado inicial de 1×10^{-4} permitiu uma convergência rápida sem oscilações indesejadas, especialmente importante para redes mais profundas. O *scheduler* ReduceLROnPlateau foi fundamental para ajustar automaticamente a taxa de aprendizado nos momentos em que o desempenho na validação deixava de subir, permitindo que os modelos refinassem seus pesos com mais precisão nas etapas finais do treinamento.

Para complementar, os *boxplots* das Figuras 11 e 12 ilustram a distribuição dos resultados, facilitando a análise de variabilidade e desempenho por modelo. É possível observar maior consistência de desempenho em algumas combinações, como U-Net+InceptionV4, e alta variabilidade em outras, como PSPNet com DenseNet121.

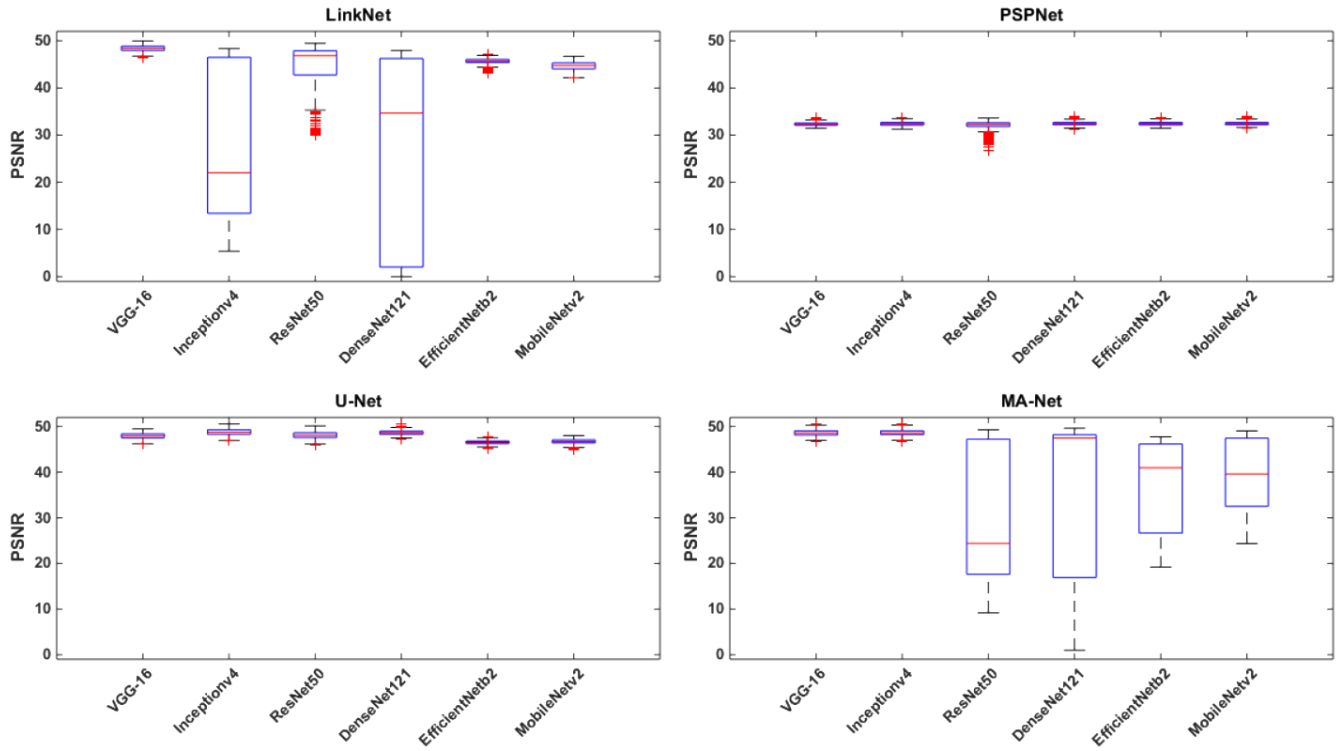


Figura 11 – Distribuição dos valores de PSNR por combinação Rede-*Backbone*.

O *boxplot* do PSNR da Figura 11 mostra que as redes U-Net e PSPNet apresentam menor variabilidade de valores, independentemente do *backbone* utilizado. A maioria de seus valores está próxima da mediana, com os resultados da U-Net sendo substancialmente superiores aos da PSPNet. Por outro lado, as redes LinkNet e MA-Net apresentam grande variabilidade dependendo do*backbone* utilizado, sendo esta última a que apresenta a maior discrepância de valores. Enquanto a LinkNet apresenta resultados ruins para apenas dois *backbone* e alta presença de *outliers* para o ResNet50, a MA-Net apresenta resultados robustos para apenas dois *backbones*, VGG-16 e Inceptionv4. Para os demais *backbones*, seus resultados estão dispersos em uma ampla faixa de valores e com medianas inferiores às de outras redes, como U-Net e LinkNet.

Em relação ao SSIM, a variabilidade dos resultados é menor, independentemente da rede

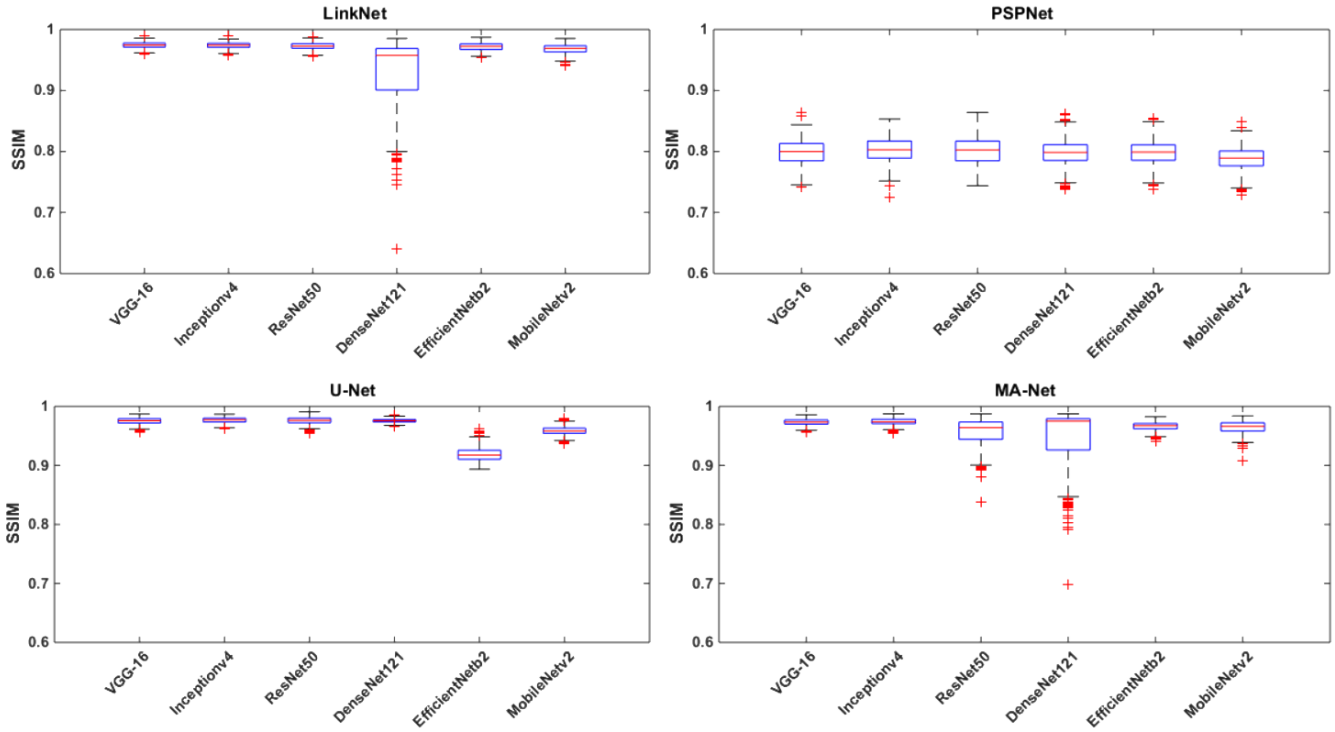


Figura 12 – Distribuição dos valores de SSIM por combinação Rede-*Backbone*.

e do *backbone* utilizado, conforme demonstrado na distribuição do *boxplot* na Figura 12. Com poucas exceções, a maioria dos valores obtidos está próxima da mediana. A PSPNet foi a que apresentou as maiores variações nos valores e os piores resultados de SSIM. Em relação ao *backbone*, a DenseNet121 apresentou a maior variabilidade, independentemente da rede utilizada. No geral, a rede U-Net apresentou os melhores resultados, sendo superada por outras redes apenas quando combinada com *backbones* rasos, como EfficientNetb2 e MobileNetv2.

Essas diferenças entre PSNR e SSIM são explicadas pela natureza dessas métricas. O PSNR verifica o erro absoluto entre os pixels e nem sempre se correlaciona com a qualidade percebida (por exemplo, é amplamente afetado pelo desfoque). Ao contrário do PSNR, o SSIM foi projetado para refletir a percepção visual humana, o que o torna mais eficaz na avaliação da qualidade perceptual das imagens.

5.2 Análise Qualitativa

Além da avaliação numérica, é fundamental considerar a percepção visual da qualidade das imagens reconstruídas. Esta análise qualitativa visa observar, de forma subjetiva, como cada modelo se comporta na remoção de ruído e na preservação de estruturas anatômicas relevantes.

A Figura 13 apresenta exemplos visuais representativos das reconstruções realizadas pelos diferentes modelos. Observa-se que as combinações com U-Net e MA-Net tendem a preservar melhor os detalhes anatômicos, com maior nitidez nas bordas e contraste estrutural. O modelo MA-Net com InceptionV4 destaca-se por apresentar regiões menos ruidosas e transições suaves entre tecidos. No entanto, a PSPNet demonstrou imagens mais borradas e com perda de

textura, indicando que seu foco em contexto global compromete a reconstituição de detalhes locais. Já a LinkNet mostrou desempenho intermediário, com bons contornos, mas contraste .

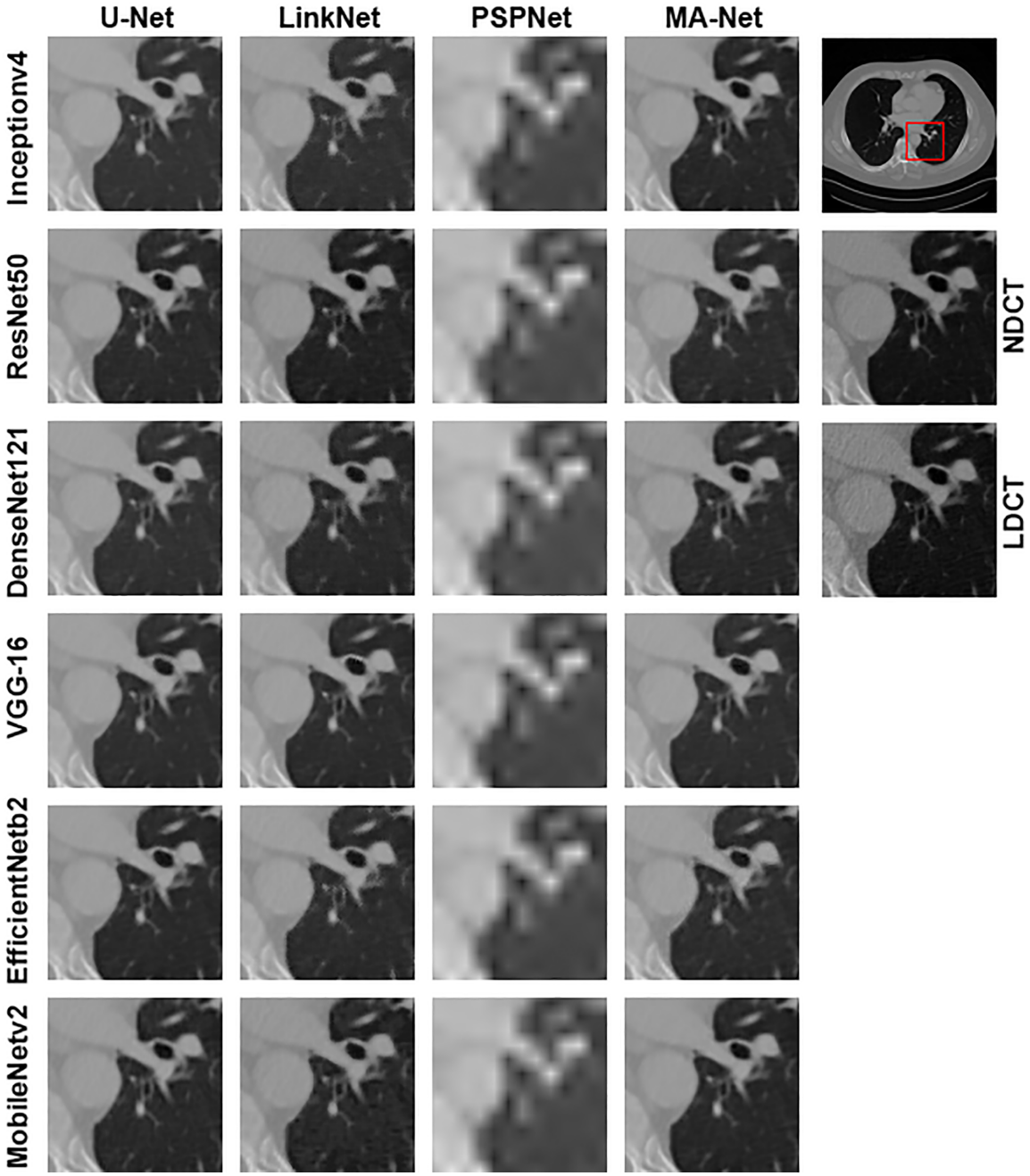


Figura 13 – Exemplos visuais das reconstruções realizadas pelas diferentes combinações de redes e *backbones*

A Figura 14 apresenta um segundo conjunto de reconstruções. A combinação MA-Net com EfficientNet-B2 produz imagens visualmente mais escuras, o que pode comprometer a visibilidade de estruturas de menor intensidade. Esse escurecimento excessivo pode ser consequência da interação entre os mecanismos de atenção e o *backbone* usado. Por outro lado, a LinkNet, especialmente com *backbones* leves como MobileNetV2 ou ResNet50, apresenta bons resultados visuais, mantendo definição razoável das bordas e contraste entre tecidos.

Durante o processo de reconstrução de imagens ruidosas, é comum observar perdas de brilho, contraste e detalhes finos em relação à imagem original. Isso ocorre porque os modelos

de remoção de ruído, ao tentarem "limpar" a imagem, podem suavizar excessivamente certas regiões ou eliminar informações de alta frequência que são confundidas com ruído. Em modelos mais simplificados, esse efeito é ainda mais acentuado, resultando em imagens que parecem opacas ou lavadas, com textura e luminosidade alteradas. Além disso, arquiteturas que utilizam funções de ativação como ReLU tendem a zerar partes das ativações , o que pode reduzir os níveis de brilho ou apagar detalhes em regiões menos intensas da imagem original (GLOROT; BENGIO, 2010).

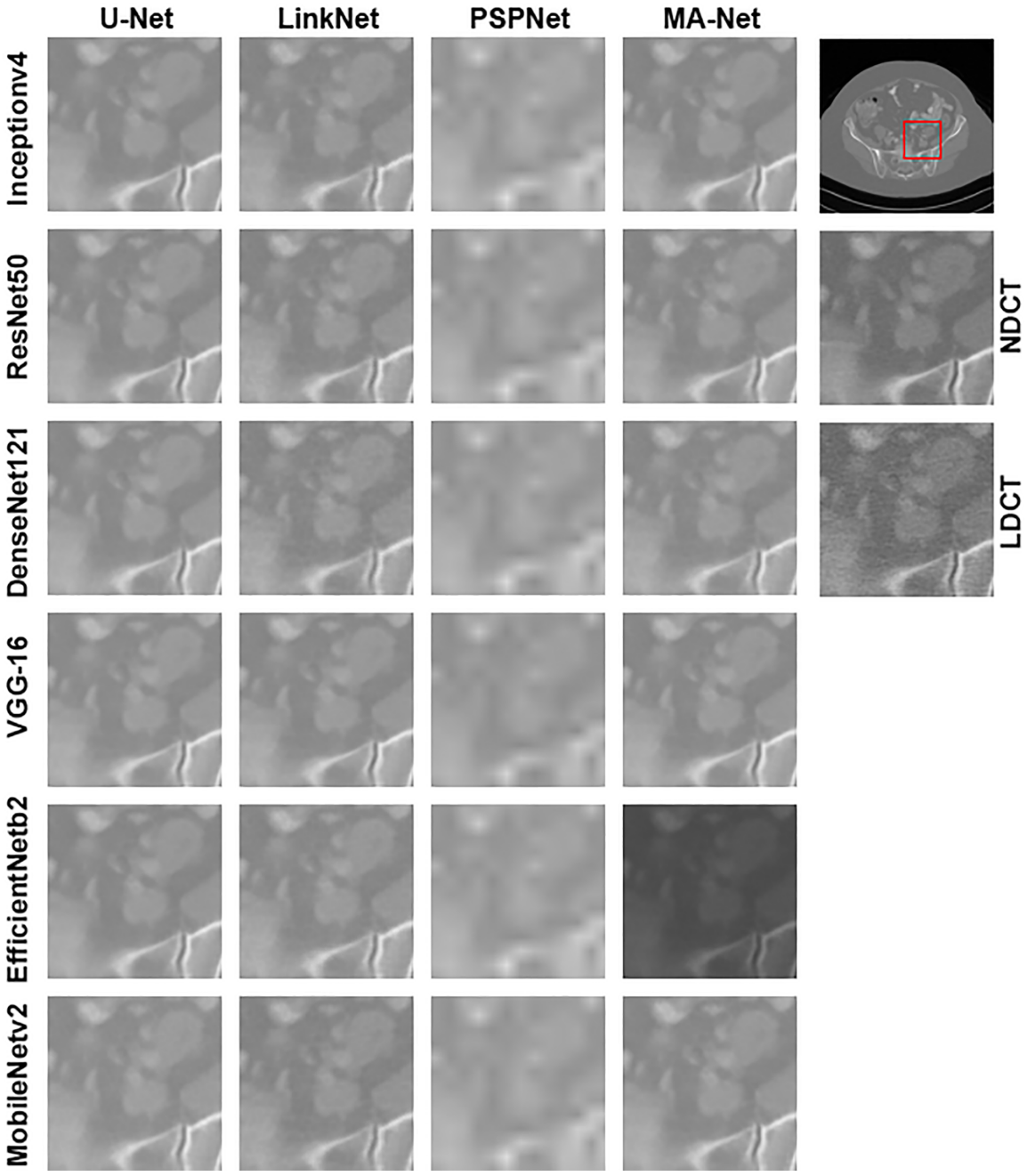


Figura 14 – Exemplos visuais adicionais das reconstruções por diferentes arquiteturas.

A Figura 15 mostra uma visualização ampliada das melhores combinações de rede e *backbone*. Notam-se microestruturas e bordas finas preservadas, principalmente na U-Net com InceptionV4. Mesmo em regiões de transição complexa, como interfaces entre órgãos, os con-

tornos permanecem nítidos e bem definidos. Isso sugere que a rede conseguiu manter uma boa separação entre ruído e informação anatômica real.

A MA-Net com InceptionV4 também apresentou bom desempenho nas imagens ampliadas, embora com leve tendência à suavização em áreas intermediárias. Algumas variações tonais mais sutis ficam menos evidentes em comparação com a U-Net, mas ainda assim, os contornos principais são preservados com precisão.

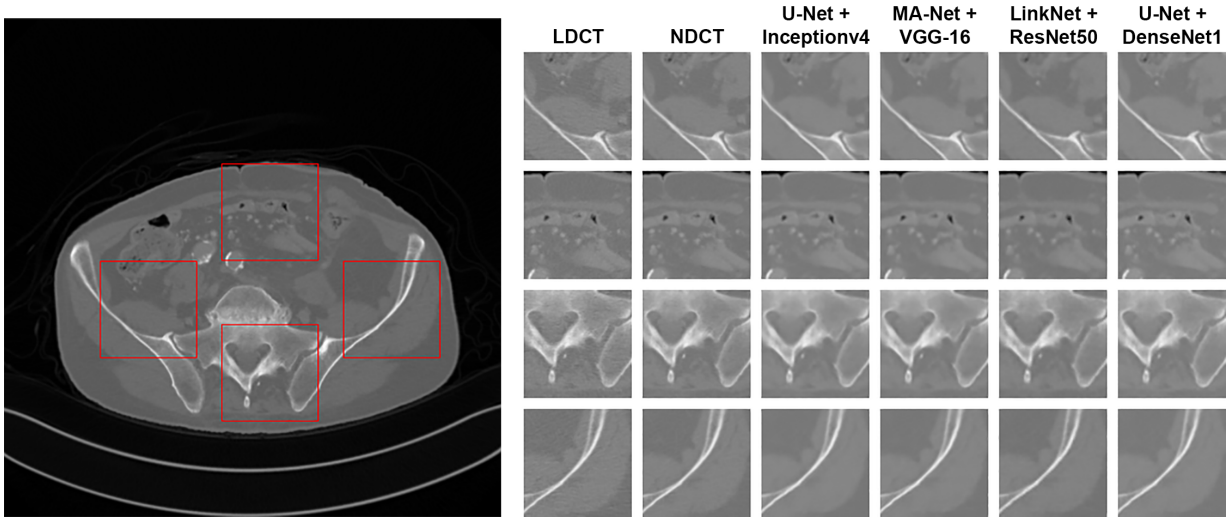


Figura 15 – Visualização ampliada das melhores combinações de rede e backbone.

Já a Figura 16 apresenta ampliações das piores combinações. Em PSPNet com DenseNet121, as regiões anatômicas aparecem borradas e com baixa definição, dificultando a distinção entre tecidos. As bordas tornam-se imprecisas e há perda de textura significativa. No entanto, a LinkNet com DenseNet, apesar de apresentar contraste mais suave, consegue manter um nível visual aceitável de estrutura, com contornos razoavelmente preservados e sem introduzir artefatos evidentes.

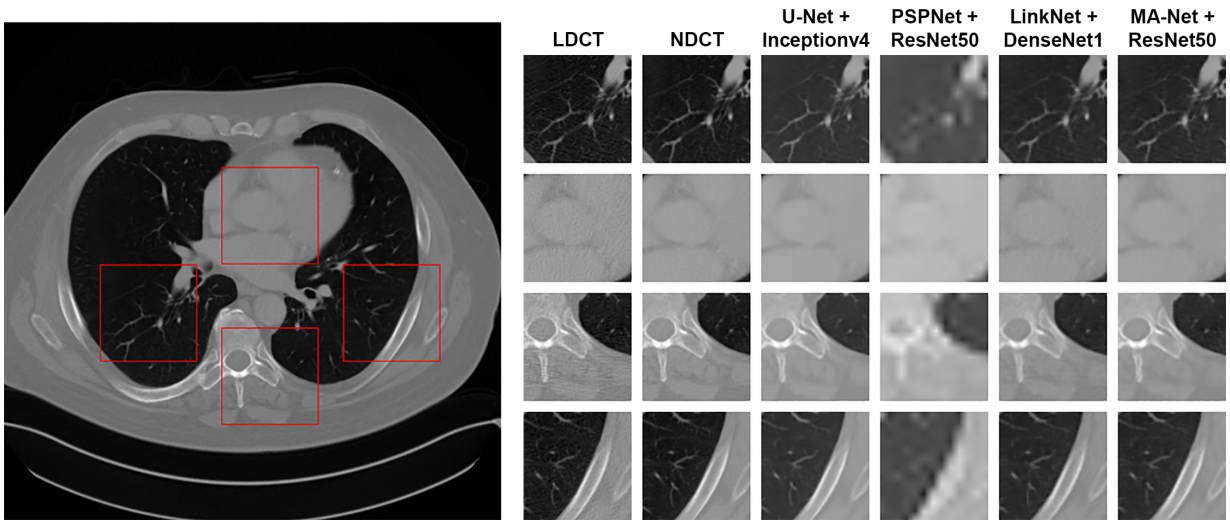


Figura 16 – Visualização ampliada da melhor combinação (U-Net+InceptionV4) comparada às piores combinações.

Comparando visualmente as ampliações das melhores e piores arquiteturas (Figuras 15 e

16), percebe-se que a fidelidade estrutural é o fator mais impactado nas reconstruções ruins. Modelos mais eficazes preservam não só os contornos, mas também a variação tonal interna das estruturas. Em contraste, os piores modelos tendem a achatar a imagem, reduzindo contraste local e apagando variações sutis essenciais para diagnóstico.

A análise ampliada evidenciou que pequenos defeitos, como desfoque e perda de contraste, tornam-se mais visíveis sob inspeção detalhada. Tais aspectos reforçam a importância de incluir avaliações qualitativas em estudos com imagens m

5.3 Comparação com a Literatura

A Tabela 5 apresenta os resultados comparativos de modelos usados no mesmo conjunto de imagens LDCT com o intuito de remoção de ruído.

Modelo	Média PSNR	Média SSIM
LDCT (Baseline)	40,3427	0,923
RED-CNN (CHEN et al., 2017a)	45,1813	0,971
ED-CNN (CHEN et al., 2017b)	44,8748	0,969
WGAN-VGG (YANG et al., 2018)	43,3313	0,963
WGAN-RAM (LI et al., 2021a)	44,5934	0,966
MAPNN-D5 (SHAN et al., 2019)	44,7677	0,969
DADN (WANG et al., 2023)	45,5286	0,973
U-Net+Inceptionv4	**48,797**	**0,977**

Tabela 5 – Resultados comparativos.

A Tabela 5 resume os resultados de modelos relevantes na literatura utilizando a base AAPM 2016. Modelos como RED-CNN, ED-CNN, WGAN e DADN alcançaram bons resultados, com PSNR entre 44 e 45. A abordagem proposta, especificamente a combinação U-Net com InceptionV4, superou esses métodos, alcançando 48,797 de PSNR e 0,977 de SSIM.

Apesar de serem métodos consolidados, esses modelos apresentam limitações em suas arquiteturas. O RED-CNN, por exemplo, é baseado em *autoencoders* com conexões residuais, mas não utiliza estratégias que integram informações em diferentes escalas nem atenção espacial, o que dificulta a preservação de detalhes finos da imagem (CHEN et al., 2017a). Outro modelo, como o ED-CNN emprega convoluções dilatadas, que ajudam a capturar padrões mais amplos, mas podem gerar artefatos, especialmente em áreas de baixo contraste (GHOLIZADEH-ANSARI; ALIREZAIE; BABYN, 2020).

Modelos baseados em GAN, como o WGAN, focam em gerar imagens visualmente mais naturais com o auxílio de discriminadores. No entanto, essa estratégia muitas vezes torna o treinamento instável e sensível ao ruído de entrada (LI et al., 2021b). E por por mais realistas que pareçam, essas imagens podem não preservar fielmente estruturas anatômicas importantes para o diagnóstico, o que limita sua aplicabilidade clínica (YANG et al., 2018).

O MAPNN-D5, usa uma rede neural multimodal com cinco níveis de atenção, o que ajuda a equilibrar a redução de ruído com a preservação de texturas. Porém, seu desempenho ainda

fica abaixo da proposta deste trabalho, provavelmente por não usar *backbones* pré-treinados nem técnicas de análise em múltiplas escalas de forma tão eficaz.

Já o modelo DADN combina aprendizado por atenção dual com transferência de domínio. Embora seus resultados sejam bons, esse modelo exige mais ajustes de hiperparâmetros e depende de mecanismos extras de adaptação entre domínios, o que pode dificultar sua adoção em ambientes clínicos com maior diversidade anatômica (WANG et al., 2023).

Em contraste, redes de segmentação como a U-Net, combinadas com *backbones* avançados como o InceptionV4, demonstraram maior versatilidade e eficiência na tarefa de reconstrução. Sua arquitetura *encoder-decoder* é naturalmente adaptável a diferentes escalas, e os pesos pré-treinados ampliam a capacidade do modelo de generalizar com menos dados rotulados.

5.4 Análise Final

A análise dos resultados confirma que as redes segmentadoras com diferentes *backbones* têm um bom potencial para lidar com a remoção de ruídos em imagens LDCT. *Backbones* como Inceptionv4 e VGG16 demonstraram alta eficácia, indicando que a escolha de um *backbone* com boa capacidade de extração de características é essencial. A MA-Net, por exemplo, foi a arquitetura que mais se beneficiou de *enconders* bem projetados, como VGG16 e Inceptionv4, devido ao seu uso de mecanismos de atenção que destacam regiões importantes na reconstrução da imagem.

As métricas PSNR e SSIM, embora amplamente utilizadas, mostram limitações ao não capturarem plenamente a qualidade visual das imagens. Como a hipótese sugere, altos valores dessas métricas nem sempre garantem que as imagens reconstruídas estejam livres de artefatos perceptíveis.

Um argumento sobre a acurácia desses índices implica que por serem consideradas apenas medidas baseadas em erros de pixel ou na similaridade estrutural podem não detectar artefatos sutis ou inconsistências que afetam a experiência visual do observador (MITTAL; MOORTHY; BOVIK, 2012). Assim, há uma necessidade de complementar a análise quantitativa com avaliações qualitativas ou subjetivas para obter uma compreensão mais abrangente da qualidade visual, uma vez que a percepção humana envolve fatores complexos que vão além dos números fornecidos por PSNR e SSIM.

Capítulo 6

Conclusão

O trabalho teve como objetivo investigar o uso de modelos de segmentação para melhorar a qualidade de imagens LDCT, que costumam apresentar bastante ruído. Foram avaliadas 24 combinações diferentes de arquiteturas de segmentação (U-Net, PSPNet, LinkNet e MA-Net) com *backbones* pré-treinados (VGG16, ResNet50, DenseNet121, Inceptionv4, MobileNetV2 e EfficientNetB2).

Os resultados mostraram que a combinação entre U-Net e Inceptionv4 se destacou, alcançando os melhores valores de PSNR 48,797 e SSIM 0,977. Isso indica que essa configuração conseguiu não só remover o ruído de forma eficaz, mas também preservar bem as estruturas anatômicas das imagens.

Outros modelos, como a MA-Net, que usa mecanismos de atenção, também apresentaram resultados bastante competitivos, especialmente quando combinados com *backbones* mais robustos. No entanto, seu desempenho variou mais de acordo com o *backbone* utilizado, o que mostra que não existe uma única solução ideal, a combinação entre arquitetura e codificador é essencial para um bom resultado.

Além dos resultados quantitativos, foi possível observar que métricas como PSNR e SSIM nem sempre refletem a percepção visual da qualidade da imagem. Em alguns casos, mesmo com bons valores, ainda era possível ver artefatos ou perda de contraste em regiões importantes. Isso reforça a ideia de que avaliações qualitativas também são essenciais em estudos dessa natureza.

De forma geral, o trabalho mostrou que usar modelos de segmentação, que originalmente foram pensados para tarefas de segmentação e classificação de objetos, pode ser uma alternativa promissora para melhorar imagens médicas ruidosas. Além disso, o uso de *backbone* pré-treinados e mecanismos de atenção demonstrou impacto positivo na qualidade final das imagens reconstruídas.

6.1 Principais Contribuições

A contribuição deste trabalho para a literatura reside na demonstração de que a combinação de arquiteturas de segmentação com *backbones* avançados pode superar métodos tradicionais e até mesmo alguns dos modelos mais recentes. Em particular, a integração da U-Net com o

backbone InceptionV4 mostrou resultados expressivos, com PSNR e SSIM superiores aos alcançados por modelos estabelecidos como o RED-CNN e o DADN. Esse resultado reforça a ideia de que a capacidade de extrair características ricas e detalhadas, proporcionada pelo InceptionV4, aliada à robustez da estrutura U-Net, é determinante para a restauração de imagens LDCT, evidenciando novas direções para a pesquisa em reconstrução de imagens médicas.

Além disso, o estudo contribui para o estado da arte ao evidenciar que, embora métricas quantitativas como PSNR e SSIM sejam importantes indicadores de desempenho, elas não capturam integralmente a qualidade visual percebida. A superioridade da combinação U-Net+InceptionV4, que não só alcançou elevados índices quantitativos, mas também demonstrou uma reconstrução visualmente consistente, propõe um novo paradigma para avaliação de modelos de remoção de ruído. Essa abordagem estimula a repensar os métodos de avaliação, integrando análises qualitativas e subjetivas para complementar os resultados numéricos.

Por fim, o trabalho estabelece um referencial comparativo robusto entre diversas arquiteturas e backbones, o que contribui significativamente para a literatura ao oferecer *insights* detalhados sobre as vantagens e limitações de cada combinação. Ao demonstrar que modelos com *backbones* eficientes e, especialmente, aqueles incorporando mecanismos de atenção, como o MA-Net, podem melhorar substancialmente a qualidade das imagens reconstruídas, este estudo abre caminho para futuras pesquisas. Tais investigações poderão explorar de forma mais aprofundada a integração de módulos de atenção e estratégias de decodificação, consolidando um novo patamar no desenvolvimento de soluções para a melhoria de imagens de baixa dosagem em tomografias computadorizadas.

6.2 Trabalhos Futuros

A partir dos resultados encontrados neste estudo, várias possibilidades interessantes surgem para dar continuidade à pesquisa. A seguir, estão algumas direções que podem ser exploradas nos próximos trabalhos:

- ❏ Testar os modelos em exames reais: a base AAPM 2016 é muito útil para experimentos controlados, mas trabalhar com dados reais de hospitais, com ruídos vindos diretamente dos equipamentos, seria um passo importante para aproximar os modelos da prática clínica.

- ❏ Explorar outras regiões do corpo e tipos de exame: este trabalho focou em tomografias abdominais, mas vale investigar se os modelos também funcionam bem em imagens do tórax, crânio, ou até em outros exames como ressonância magnética. Cada caso pode trazer desafios diferentes, e seria interessante entender essas variações.

- ❏ Usar métricas que se aproximem mais da percepção humana: embora PSNR e SSIM sejam métricas consolidadas, elas nem sempre refletem bem o que o olho humano percebe. Outras abordagens, como o *Learned Perceptual Image Patch Similarity* (LPIPS), ou até

avaliações com especialistas da área médica, podem trazer uma visão mais completa da qualidade das imagens.

- ❏ Experimentar novas arquiteturas, como Transformers: modelos baseados em atenção, como os Transformers, vêm ganhando espaço na área de visão computacional. Vale a pena investigar se eles podem oferecer vantagens também na tarefa de remoção de ruído em imagens médicas.

- ❏ Otimizar o uso de recursos computacionais: muitas vezes não se tem acesso a máquinas potentes para treinar os modelos. Explorar técnicas como treinamento distribuído, uso de *mixed precision* ou acumulação de gradiente pode ajudar a reduzir o custo computacional e tornar o processo mais acessível.

- ❏ Investigar o uso de aprendizado com poucos ou nenhum rótulo: como nem todo hospital tem acesso a bases de dados bem anotadas, testar modelos que aprendem de forma auto-supervisionada ou semi-supervisionada pode tornar a tecnologia mais viável em cenários com poucos recursos.

Referências

ALOM, M. Z. et al. A state-of-the-art survey on deep learning theory and architectures. **Electronics**, Multidisciplinary Digital Publishing Institute, v. 8, n. 3, p. 292, 2019. <https://doi.org/10.3390/electronics8030292>.

BARRETT, J. F.; KEAT, N. Artifacts in ct: recognition and avoidance. **Radiographics**, Radiological Society of North America, v. 24, n. 6, p. 1679–1691, 2004. <https://doi.org/10.1148/rg.246045065>.

BEISTER, M.; KOLDITZ, D.; KALENDER, W. A. Iterative reconstruction methods in x-ray ct. **Physica medica**, Elsevier, v. 28, n. 2, p. 94–108, 2012. <https://doi.org/10.1016/j.ejmp.2012.01.003>.

BRENNER, D. J.; HALL, E. J. Computed tomography—an increasing source of radiation exposure. **New England journal of medicine**, Mass Medical Soc, v. 357, n. 22, p. 2277–2284, 2007. <https://doi.org/10.1056/nejmra072149>.

BUADES, A.; COLL, B.; MOREL, J.-M. A non-local algorithm for image denoising. In: IEEE. **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**. [S.l.], 2005. v. 2, p. 60–65. <https://ieeexplore.ieee.org/document/1467314>.

CHAURASIA, A.; CULURCIELLO, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: IEEE. **2017 IEEE visual communications and image processing (VCIP)**. [S.l.], 2017. p. 1–4. <https://doi.org/10.1109/vcip.2017.8305148>.

CHEN, H. et al. Low-dose ct with a residual encoder-decoder convolutional neural network. **IEEE transactions on medical imaging**, IEEE, v. 36, n. 12, p. 2524–2535, 2017. <https://doi.org/10.1109/tmi.2017.2715284>.

_____. Low-dose ct via convolutional neural network. **Biomedical optics express**, Optical Society of America, v. 8, n. 2, p. 679–694, 2017. <https://doi.org/10.1364/boe.8.000679>.

CHOW, L. S.; PARAMESRAN, R. Review of medical image quality assessment. **Biomedical signal processing and control**, Elsevier, v. 27, p. 145–154, 2016. <https://doi.org/10.1016/j.bspc.2016.02.006>.

FAN, T. et al. Ma-net: A multi-scale attention network for liver and tumor segmentation. **IEEE Access**, IEEE, v. 8, p. 179656–179665, 2020. <https://doi.org/10.1109/access.2020.3025372>.

GHOLIZADEH-ANSARI, M.; ALIREZAIE, J.; BABYN, P. Deep learning for low-dose ct denoising using perceptual loss and edge detection layer. **Journal of Digital Imaging**, v. 33, n. 2, p. 504–515, 2020. <https://link.springer.com/article/10.1007/s10278-019-00274-4>.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: **Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)**. [S.l.: s.n.], 2010. (Proceedings of Machine Learning Research, v. 9), p. 249–256. <https://proceedings.mlr.press/v9/glorot10a.html>.

GOCERI, E. Medical image data augmentation: Techniques, comparisons and interpretations. **Artificial Intelligence Review**, v. 56, n. 11, p. 12561–12605, 2023. <https://link.springer.com/article/10.1007/s10462-023-10453-z>.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3rd. ed. Upper Saddle River, NJ: Pearson Education, 2009.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Massachusetts: MIT Press, 2016. <http://www.deeplearningbook.org>.

GOODFELLOW, I. et al. Generative adversarial nets. **Advances in neural information processing systems**, v. 27, 2014. <https://doi.org/10.1145/3422622>.

HAO, S.; ZHOU, Y.; GUO, Y. A brief survey on semantic segmentation with deep learning. **Neurocomputing**, Elsevier, v. 406, p. 302–321, 2020. <https://doi.org/10.1016/j.neucom.2019.11.118>.

HE, K. et al. Deep residual learning for image recognition. In: IEEE. **Proceedings of the IEEE conference on computer vision and pattern recognition**. Beijing, 2016. p. 770–778. <https://doi.org/10.1109/cvpr.2016.90>.

HESAMIAN, M. H. et al. Deep learning techniques for medical image segmentation: achievements and challenges. **Journal of digital imaging**, Springer, v. 32, p. 582–596, 2019. <https://doi.org/10.1007/s10278-019-00227-x>.

HUANG, G. et al. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 4700–4708. <https://doi.org/10.1109/cvpr.2017.243>.

IAKUBOVSKII, D. **Segmentation Models PyTorch**. 2019. <https://github.com/qubvel/segmentation_models.pytorch>. GitHub repository. Disponível em: <https://github.com/qubvel/segmentation_models.pytorch>.

KAK, A. C.; SLANEY, M. **Principles of computerized tomographic imaging**. [S.l.]: SIAM, 2001. <https://doi.org/10.1137/1.9780898719277>.

KALENDER, W. A. **Computed tomography: fundamentals, system technology, image quality, applications**. [S.l.]: John Wiley & Sons, 2011.

KANG, E.; MIN, J.; YE, J. C. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. **Medical physics**, Wiley Online Library, v. 44, n. 10, p. e360–e375, 2017. <https://doi.org/10.1002/mp.12344>.

KAUR, A.; DONG, G. A complete review on image denoising techniques for medical images. **Neural Processing Letters**, Springer, v. 55, n. 6, p. 7807–7850, 2023. <https://doi.org/10.1007/s11063-023-11286-1>.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. **Ijcai**. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

KROEKER, K. I. et al. Patients with ibd are exposed to high levels of ionizing radiation through ct scan diagnostic imaging: a five-year study. **Journal of clinical gastroenterology**, LWW, v. 45, n. 1, p. 34–39, 2011. <https://doi.org/10.1097/mcg.0b013e3181e5d1c5>.

KULATHILAKE, K. S. H. et al. A review on deep learning approaches for low-dose computed tomography restoration. **Complex & Intelligent Systems**, Springer, p. 1–33, 2021. <https://doi.org/10.1007/s40747-021-00405-x>.

LI, C. et al. Attention unet: A nested attention-aware u-net for liver ct image segmentation. In: IEEE. **Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP)**. [S.l.], 2020. p. 345–349. <https://doi.org/10.1109/ICIP40778.2020.9190761>.

LI, M. et al. Incorporation of residual attention modules into two neural networks for low-dose ct denoising. **Medical Physics**, Wiley Online Library, v. 48, n. 6, p. 2973–2990, 2021. <https://doi.org/10.1002/mp.14856>.

LI, Z. et al. Low-dose ct image denoising with improving wgan and hybrid loss function. **Computational and Mathematical Methods in Medicine**, Wiley Online Library, v. 2021, n. 1, p. 2973108, 2021. <https://doi.org/10.1155/2021/2973108>.

LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical image analysis**, Elsevier, v. 42, p. 60–88, 2017. <https://doi.org/10.1016/j.media.2017.07.005>.

LIU, X. et al. A review of deep-learning-based medical image segmentation methods. **Sustainability**, MDPI, v. 13, n. 3, p. 1224, 2021. <https://doi.org/10.3390/su13031224>.

LU, Z. J. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Oxford University Press, 2010. <https://doi.org/10.1111/j.1467-985x.2010.00646_6.x>.

MAKEN, P.; GUPTA, A. 2d-to-3d: A review for computational 3d image reconstruction from x-ray images. **Archives of Computational Methods in Engineering**, Springer, p. 1–30, 2022. <https://doi.org/10.1007/s11831-022-09790-z>.

MCCOLLOUGH, C. H. et al. Low-dose ct for the detection and classification of metastatic liver lesions: Results of the 2016 low dose ct grand challenge. **Medical Physics**, v. 44, n. 10, p. e339–e352, 2017. <https://doi.org/10.1002/mp.12345>.

MCCOLLOUGH, C. H.; BRUESEWITZ, M. R.; JR, J. M. K. Ct dose reduction and dose management tools: overview of available options. **Radiographics**, Radiological Society of North America, v. 26, n. 2, p. 503–512, 2006. <https://doi.org/10.1148/rg.262055138>.

MCCOLLOUGH, C. H. et al. Strategies for reducing radiation dose in ct. **Radiologic Clinics of North America**, v. 47, n. 1, p. 27, 2009. <https://doi.org/10.1016/j.rcl.2008.10.006>.

MIER, W.; MIER, D. Advantages in functional imaging of the brain. **Frontiers in human neuroscience**, Frontiers Media SA, v. 9, p. 249, 2015. <https://doi.org/10.3389/fnhum.2015.00249>.

MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C. No-reference image quality assessment in the spatial domain. **IEEE Transactions on Image Processing**, IEEE, v. 21, n. 12, p. 4695–4708, 2012. <https://doi.org/10.1109/tip.2012.2214050>.

MOEN, T. R. et al. Low-dose ct image and projection dataset. **Medical Physics**, v. 48, n. 2, p. 902–911, 2021. <https://doi.org/10.1002/mp.14594>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18**. [S.l.], 2015. p. 234–241. <https://doi.org/10.1007/978-3-319-24574-4_28>.

RYAN, J. L. Ionizing radiation: the good, the bad, and the ugly. **Journal of Investigative Dermatology**, Elsevier, v. 132, n. 3, p. 985–993, 2012. <https://doi.org/10.1038/jid.2011.411>.

SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2018. p. 4510–4520. <https://doi.org/10.1109/cvpr.2018.00474>.

SHAH, Z. H. et al. Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. **Photonics Research**, OSA, v. 9, n. 5, p. B168–B181, 2021. <https://doi.org/10.1364/prj.416437>.

SHAN, H. et al. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction. **Nature Machine Intelligence**, Nature Publishing Group UK London, v. 1, n. 6, p. 269–276, 2019. <https://doi.org/10.1038/s42256-019-0057-9>.

_____. 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network. **IEEE transactions on medical imaging**, IEEE, v. 37, n. 6, p. 1522–1534, 2018. <https://doi.org/10.1109/tmi.2018.2832217>.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of big data**, Springer, v. 6, n. 1, p. 1–48, 2019. <https://doi.org/10.1186/s40537-019-0197-0>.

SILVA, R. C. F. da. Tomografia computadorizada e risco de câncer. **Rede Câncer, Instituto Nacional de Câncer (INCA), Rio de Janeiro**, v. 21, p. 38–40, 2013. <https://ninho.inca.gov.br/jspui/handle/123456789/15095>.

SMITH-BINDMAN, R. et al. Projected lifetime cancer risks from current computed tomography imaging. **JAMA internal medicine**, 2025. <https://doi.org/10.1001/jamainternmed.2025.0505>.

SMITH, L. N. Cyclical learning rates for training neural networks. In: IEEE. **2017 IEEE winter conference on applications of computer vision (WACV)**. [S.l.], 2017. p. 464–472. <https://doi.org/10.1109/wacv.2017.58>.

SUN, H. et al. An improved medical image classification algorithm based on adam optimizer. **Mathematics**, MDPI, v. 12, n. 16, p. 2509, 2024. <https://doi.org/10.3390/math12162509>.

SYED, A.; MORRIS, B. T. Sseg-lstm: Semantic scene segmentation for trajectory prediction. In: IEEE. **2019 IEEE Intelligent Vehicles Symposium (IV)**. [S.l.], 2019. p. 2504–2509. <https://doi.org/10.1109/ivs.2019.8813801>.

SZEGEDY, C. et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: **Proceedings of the AAAI conference on artificial intelligence**. [S.l.: s.n.], 2017. v. 31, n. 1. <https://doi.org/10.1609/aaai.v31i1.11231>.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **Proceedings of the 36th International Conference on Machine Learning**. [S.l.: s.n.], 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. <https://proceedings.mlr.press/v97/tan19a.html>.

TEAM, N. L. S. T. R. Reduced lung-cancer mortality with low-dose computed tomographic screening. **New England Journal of Medicine**, Mass Medical Soc, v. 365, n. 5, p. 395–409, 2011. <https://doi.org/10.1056/nejmoa1102873>.

TIAN, C. et al. Attention-guided cnn for image denoising. **Neural Networks**, Elsevier, v. 124, p. 117–129, 2020. <https://doi.org/10.1016/j.neunet.2019.12.024>.

TRAN, L. D.; NGUYEN, S. M.; ARAI, M. Gan-based noise model for denoising real images. In: **Proceedings of the Asian Conference on Computer Vision**. [S.l.: s.n.], 2020. <https://doi.org/10.1007/978-3-030-69538-5_34>.

VOULODIMOS, A. et al. Deep learning for computer vision: A brief review. **Computational intelligence and neuroscience**, Hindawi, v. 2018, 2018. <https://doi.org/10.1155/2018/7068349>.

WANG, G. A perspective on deep imaging. **IEEE access**, IEEE, v. 4, p. 8914–8924, 2016. <https://doi.org/10.1109/access.2016.2624938>.

WANG, G. et al. Deep tomographic image reconstruction: Yesterday, today, and tomorrow—editorial for the 2nd special issue "machine learning for image reconstruction". **IEEE transactions on medical imaging**, IEEE, v. 40, n. 11, p. 2956–2964, 2021. <https://doi.org/10.1109/tmi.2021.3115547>.

WANG, J. et al. Domain-adaptive denoising network for low-dose ct via noise estimation and transfer learning. **Medical Physics**, v. 50, n. 1, p. 74–88, 2023. <https://doi.org/10.1002/mp.15952>.

WANG, Z.; BOVIK, A. C. **Modern Image Quality Assessment**. San Rafael, CA: Morgan & Claypool Publishers, 2006. v. 2. (Synthesis Lectures on Image, Video, and Multimedia Processing, v. 2). <https://doi.org/10.1007/978-3-031-02238-8>.

WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. **IEEE transactions on image processing**, IEEE, v. 13, n. 4, p. 600–612, 2004. <https://doi.org/10.1109/tip.2003.819861>.

YANG, Q. et al. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. **IEEE transactions on medical imaging**, IEEE, v. 37, n. 6, p. 1348–1357, 2018. <https://doi.org/10.1109/tmi.2018.2827462>.

YU, Z. et al. A comparative analysis of visual encoding models based on classification and segmentation task-driven cnns. **Computational and Mathematical Methods in Medicine**, Wiley Online Library, v. 2020, n. 1, p. 5408942, 2020. <https://doi.org/10.1155/2020/5408942>.

ZHANG, J. et al. Cnn and multi-feature extraction based denoising of ct images. **Biomedical Signal Processing and Control**, Elsevier, v. 67, p. 102545, 2021. <https://doi.org/10.1016/j.bspc.2021.102545>.

ZHANG, K. et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. **IEEE transactions on image processing**, IEEE, v. 26, n. 7, p. 3142–3155, 2017. <https://doi.org/10.1109/tip.2017.2662206>.

ZHU, L.; WANG, G.; LIU, Y. Image quality evaluation method based on structural similarity. In: SPIE. **MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications**. [S.l.], 2007. v. 6790, p. 1439–1448. <https://doi.org/10.1117/12.774817>.

ZHU, X. et al. Coronary angiography image segmentation based on pspnet. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 200, p. 105897, 2021. <https://doi.org/10.1016/j.cmpb.2020.105897>.