



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

Carlos Antônio Alves Júnior

**Combinação de Redes Neurais Convolucionais para reconstrução de Tomografias  
Computadorizadas com Baixas Dosagens de Radiação**

Recife

2023

Carlos Antônio Alves Júnior

**Combinação de Redes Neurais Convolucionais para reconstrução de Tomografias Computadorizadas com Baixas Dosagens de Radiação**

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

**Área de concentração:** Inteligência Computacional.

**Orientador:** Tsang Ing Ren

**Coorientador:** Luis Filipe Alves Pereira

Recife

2023

Catálogo na fonte  
Bibliotecária Nataly Soares Leite Moro, CRB4-1722

A474c Alves Júnior, Carlos Antônio  
Combinação de redes neurais convolucionais para reconstrução de tomografias computadorizadas com baixas dosagens de radiação / Carlos Antônio Alves Júnior – 2023.  
60 f.: il., fig., tab.

Orientador: Tsang Ing Ren.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2023.  
Inclui referências.

1. Inteligência computacional. 2. Tomografias computadorizadas. 3. Redes Neurais Convolucionais. I. Tsang Ing Ren (orientador). II. Título

006.31

CDD (23. ed.)

UFPE - CCEN 2023 – 142

**Carlos Antônio Alves Júnior**

**“Combinação de Redes Neurais Convolucionais para reconstrução de Tomografias Computadorizadas com Baixas Dosagens de Radiação”**

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

Aprovado em: 9 de agosto de 2023.

**BANCA EXAMINADORA**

---

Prof. Carlos Alexandre Barros de Mello  
Centro de Informática/UFPE

---

Prof. Dr. Hae Yong Kim  
Departamento de Engenharia de Sistemas Eletrônicos/USP

---

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

Dedico essa dissertação à minha mãe, Maria Sueli de Melo, mulher guerreira que apesar de todas as dificuldades e injustiças jogadas em seu caminho, sempre fez o melhor possível para que seus filhos não precisassem passar pelo mesmo caminho espinhoso.

## **AGRADECIMENTOS**

O presente trabalho foi desenvolvido com o apoio da Fundação de Amparo a Ciência e Tecnologia de Pernambuco (FACEPE) através da concessão de bolsa de estudos de Mestrado.

## RESUMO

Riscos relacionados a exposição excessiva de pacientes à radiação durante a aquisição de tomografias computadorizadas são motivo de preocupação crescente na comunidade médica. Tomografias obtidas com menores quantidade de radiação ou redução de projeções poderiam aliviar o problema, mas resultam em artefatos, ruídos e distorções na imagem reconstruída. Sendo assim, diversos métodos baseados na aplicação de Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) para recuperação das tomografias obtidas com baixa dosagem (LDCTs, do inglês *Low-dose Computed Tomography*), apresentaram bons resultados. Métodos baseados em CNNs 3D obtiveram resultados ainda mais promissores, visto que exploram melhor as relações entre os pixels das visões ortogonais dos volumes CT. CNNs 3D possuem, no entanto, um custo computacional maior associado. Além disso, são mais sensíveis ao problema do desbalanceamento de classes, que consiste na aprendizagem enviesada em favor de uma classe de pixels mais abundante. Ao mesmo tempo, métodos baseados em CNNs 2D não exploram tão bem as relações entre pixels das visões ortogonais, visto que processam cada fatia do volume tomográfico separadamente, o que leva a uma demanda computacional menor. Neste trabalho, é apresentado um método que busca explorar as relações entre os pixels dos eixos Axial, Coronal e Sagittal dos volumes LDCT utilizando-se de um ensemble de quatro CNNs 2D, onde, três delas processam os eixos ortogonais do volume LDCT de forma separada, e a quarta faz a fusão das saídas das três redes anteriores. Dessa forma, resultados tão bons quanto ou superiores aos oriundos do uso de CNNs 3D, podem ser obtidos sem os custos computacionais associados as mesmas. Para os experimentos, foram utilizados duas bases de dados, uma simulando tomografias computadorizadas com projeções-esparsas, e a outra com dados reais de tomografias computadorizadas obtidas com redução direta de radiação. O método proposto apresentou resultados majoritariamente superiores aos modelos 2D e 3D.

**Palavras-chave:** combinação; tomografias computadorizadas; CNN; baixa dosagem.

## ABSTRACT

Risks related to the excessive exposure of patients to radiation during the acquisition of computerized tomography's (CT) are a growing concern over the medical community. CT scans obtained with radiation levels reduced or sparse projections may relieve the problem, but results in noisier volumes. Several methods based on the use of Convolutional Neural Networks (CNNs) for low-dose CT scans recovery (LDCTs, *Low-dose Computed Tomography*), have shown good results. Methods based on 3D CNNs resulted in even more promising outputs, since they better explore the intra-voxial relationships between the orthogonal CT volumes views. 3D CNNs, however, have a higher computational cost associated with them, along with being more sensitive to the class imbalance problem, which consists in a biased learning in favor of a more abundant pixel class, while 2D CNNs based methods, are unable to explore the intra-voxial relationships on the orthogonal axis efficiently, since each slice from the CT volume are processed separately, which lead to a lower computational demand. In this work, a method is presented that seeks to explore the intra-voxial relationships of the Axial, Coronal and Sagittal axes of the LDCT volumes using an ensemble of four 2D CNNs, where three of them process the orthogonal axes of the LDCT volume separately, and the fourth merges the outputs of the previous three networks. In this way, results that are as good or superior to those referring to the use of 3D CNNs, can be obtained without the computational costs associated with them. For the experiments, two datasets were used, one simulating computed tomography scans with sparse-projection and the other, with real data from computed tomography scans obtained with direct radiation reduction. The proposed method showed results that were mostly superior to the 2D and 3D models.

**Keywords:** ensemble; computed tomography; CNN; low-dose.

## LISTA DE FIGURAS

- Figura 1 – Diferentes visões de uma mesma tomografia. (a) representa a visão/eixo sagital, (b) Axial e (c) Coronal. . . . . 17
- Figura 2 – Diferentes visões de um mesmo volume NDCT (dosagem normal). O círculo vermelho aponta como um conjunto de pixels se complementa nas diferentes visões. . . . . 20
- Figura 3 – Uma mesma fatia do volume reconstruída com diferentes quantidade de projeções (10, 30, 60, 90 e quantidade padrão do Qure.ai (720)). . . . . 21
- Figura 4 – Exemplo de CNN com 2 camadas escondidas. A CNN conta ainda com 2 operações de *max pooling* . . . . . 22
- Figura 5 – Exemplo de aplicação do *kernel* (1, 0, -1) com *strides* de tamanho 1 e 2. Nota-se que com *Stride* = 2 o resultado é duas vezes menor que *Stride* = 1. 24
- Figura 6 – A Figura abaixo mostra a arquitetura da REDCNN, proposta em (Chen; Zhang et al., 2017). A parte de cima mostra o processo de codificação da arquitetura, enquanto a parte de baixo mostra o processo inverso, de decodificação. . . . . 26
- Figura 7 – A Figura abaixo mostra a visão geral do método proposto em (YANG et al., 2018). Na "Part 1" observa-se a arquitetura da rede geradora, enquanto nas "Part 2 e 3" lista-se o funcionamento da *perceptual loss* e da rede discriminadora respectivamente. . . . . 27
- Figura 8 – A Figura abaixo mostra a visão geral do método proposto em (HUANG et al., 2020). Em (a), observa-se o fluxo de entrada e processamento das imagens, enquanto (b) apresenta a arquitetura da CNN primária. É importante notar que a primeira CNN é chamada de RCNN(texture), pois essa tem como objetivo fazer a limpeza inicial de forma a extrair a textura principal do CT. 28

- Figura 9 – O método proposto em (KANG et al., 2019) consiste em utilizar duas redes geradoras,  $G_{AB}$  e  $G_{BA}$ , em conjunto com dois discriminadores adversariais,  $D_A$  e  $D_B$ , para realizar a recuperação de imagens de tomografia computadorizada de baixa dose (LDCT) por meio de imagens de tomografia computadorizada de dose normal (NDCT). A figura abaixo apresenta uma visão geral do método, onde  $A$  e  $B$  denotam os domínios LDCT e NDCT, respectivamente. Após o processamento de cada imagem pelas redes geradoras, a qualidade dos resultados é avaliada pelos discriminadores correspondentes. Uma das características interessantes do método é a utilização de três funções de perda diferentes: adversarial, cíclica e identidade. . . . . 29
- Figura 10 – A Figura abaixo mostra a mudança feita por (GU; YE, 2021) no método proposto em (KANG et al., 2019). A mudança mais óbvia é o uso de apenas uma rede geradora  $G$  ao invés de duas ( $G_{AB}$  e  $G_{BA}$ ). Além disso, é possível ver as camadas AdaIN  $F$ . . . . . 30
- Figura 11 – Arquitetura da rede geradora com as camadas AdaIN. A arquitetura original era uma U-net padrão onde, as camadas de normalização de *batch* foram substituídas pelas camadas AdaIN. . . . . 30
- Figura 12 – Na figura é possível observar o esquema geral do método proposto por (IBRAGIMOV et al., 2017), que utilizou múltiplas visões no processo de segmentação da veia porta em imagens CT. O método consistiu na aplicação de três redes neurais convolucionais, cada uma processando uma visão separadamente e criando três visualizações melhoradas da veia porta. Para criar a segmentação final, as saídas das três redes foram combinadas por meio de uma média. Essa abordagem permitiu uma melhoria significativa nos resultados de segmentação da veia porta em comparação com métodos que utilizam apenas uma única visão. . . . . 31
- Figura 13 – Esquema geral do uso das múltiplas visões por (MORTAZI et al., 2017). Novamente, múltiplas CNNs são utilizadas no processamento de cada eixo e, para o resultado final, é utilizado uma estratégia de fusão que, medindo a confiabilidade de cada CNN, busca maximizar a fidelidade da segmentação. 31

Figura 14 – A Figura abaixo mostra a visão geral do método proposto em (KUSHIBAR et al., 2017). Em (A) é possível ver a arquitetura da CNN 2.5D, que possui diferentes entradas para cada visão e para as informações espaciais do Atlas. As visões são, inicialmente, processadas pela mesma organização estrutural (apresentada em (B)) e posteriormente, são utilizadas em conjunto com a informação espacial do Atlas na obtenção da segmentação final. A camada final ( <i>softmax</i> ) possui a probabilidade de cada estrutura subcortical (ao todo, são 15 possíveis estruturas), e a segmentação é feita com base na estrutura de maior probabilidade. . . . .	32
Figura 15 – O esquema para treinamento do método em duas etapas: (a) e (b). Na etapa (a), Uma partição $p_1$ do conjunto de treino $T$ é usado pra treinar a primeira fase de combinação sendo as redes referentes a sagital (1), axial (2), e coronal (3). Como resultado, um conjunto de modelos $\{m_1, m_2, m_3\}$ é gerado. Em (b), uma partição $p_2$ de $T$ é processado pelos modelos treinados anteriormente, gerando tuplas $((\hat{v}_k^{m_1}, \hat{v}_k^{m_2}, \hat{v}_k^{m_3}), v_k)$ ; usando conjuntos de tuplas, o modelo final (fusão) $m_4$ é treinado. Em (c), os modelos treinados anteriormente são organizados de forma a computar $\hat{v}^{m_4} = m_4(\hat{v}^{m_1}, \hat{v}^{m_2}, \hat{v}^{m_3})$ . . . . .	36
Figura 16 – Alguns exemplos de tomografias encontradas na base de dados Qure.ai CQ500. . . . .	38
Figura 17 – Alguns exemplos de tomografias encontradas na base de dados Mayo-Challenge. . . . .	39
Figura 18 – Simulações LDCT da base de dados Qure.ai CQ500. Cada imagem representa (respectivamente, da esquerda pra direita, de cima para baixo) reconstruções com 10, 30, 60 e 90 projeções. . . . .	41
Figura 19 – Descrição da arquitetura da U-net utilizada neste trabalho. U-net 2D e 3D diferem-se pelo uso de convoluções 2D e 3D. . . . .	42
Figura 20 – Descrição da arquitetura da RED-CNN utilizada neste trabalho. RED-CNN 2D e 3D diferem-se pelo uso de convoluções 2D e 3D. . . . .	42
Figura 21 – Descrição da arquitetura da GRC-GAN utilizada neste trabalho. GRC-GAN 2D e 3D diferem-se pelo uso de convoluções 2D e 3D. Fonte (ALMEIDA et al., 2021) (Imagem cedida pelo autor). . . . .	43

Figura 22 – Resultados gerados pela RED-CNN 2D, RED-CNN 3D, e o método proposto aplicado a RED-CNN para quatro fatias LDCT reconstruídas de 10, 30, 60, e 90 projeções X-ray. . . . .	46
Figura 23 – Resultados gerados pela U-net 2D, U-net 3D, e o método proposto aplicado a U-net para quatro fatias LDCT reconstruídas de 10, 30, 60, and 90 X-ray projeções X-ray. . . . .	47
Figura 24 – Mapas de calor indicando as zonas com maior erro nos resultados da RED-CNN. . . . .	48
Figura 25 – Mapas de calor indicando as zonas com maior erro nos resultados da U-net.	49
Figura 26 – Resultados gerados pela RED-CNN 2D, RED-CNN 3D, e o método proposto aplicado a RED-CNN para a recuperação de uma fatia obtida com redução de radiação. . . . .	50
Figura 27 – Zoom das três reconstruções baseadas na RED-CNN. A área circulada em vermelho mostra a região com maior distinção entre as três figuras. . . . .	51
Figura 28 – Resultados gerados pela GRC-GAN 2D, GRC-GAN 3D, e o método proposto aplicado a GRC-GAN para a recuperação de uma fatia obtida com redução de radiação. . . . .	51
Figura 29 – Zoom das três fatias recuperadas utilizando métodos baseados na GRC-GAN.	52

## LISTA DE TABELAS

- Tabela 1 – SSIM, PSNR and RMSE (Média $\pm$  desvio padrão) obtidos usando os métodos baseados na RED-CNN para supressão de ruídos e artefatos em LDCTs reconstruídos com projeções-esparsas. Resultados referentes ao conjunto de teste. Os melhores resultados estão destacados em negrito. . . . . 48
- Tabela 2 – SSIM, PSNR and RMSE (Média $\pm$  desvio padrão) obtidos usando os métodos baseados na U-net para supressão de ruídos e artefatos em LDCTs reconstruídos com projeções-esparsas. Resultados referentes ao conjunto de teste. Os melhores resultados estão destacados em negrito. . . . . 49
- Tabela 3 – SSIM, PSNR e NRMSE (Média $\pm$  desvio padrão) obtidos usando os métodos baseados na RED-CNN e GRCGAN para recuperação de LDCTs obtidos com redução direta de radiação. Resultados referentes ao conjunto de teste. Os melhores resultados estão destacados em negrito. . . . . 52

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	OBJETIVOS	17
1.2	ORGANIZAÇÃO DO TRABALHO	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>19</b>
2.1	TOMOGRAFIAS COMPUTADORIZADAS	19
<b>2.1.1</b>	<b>Dosagem Normal de Radiação</b>	<b>19</b>
<b>2.1.2</b>	<b>Dosagem Reduzida de Radiação</b>	<b>20</b>
2.2	REDES NEURAS CONVOLUCIONAIS	21
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>25</b>
3.1	CNNS EM LOW-DOSE	25
3.2	MÚLTIPLAS CNNS EM LOW-DOSE	27
<b>3.2.1</b>	<b>Métodos que utilizam as múltiplas visões</b>	<b>29</b>
<b>4</b>	<b>MÉTODO PROPOSTO</b>	<b>34</b>
<b>5</b>	<b>AMBIENTE EXPERIMENTAL</b>	<b>37</b>
5.1	DATASETS	37
<b>5.1.1</b>	<b>Qure.ai</b>	<b>37</b>
<b>5.1.2</b>	<b>Mayo-Challenge</b>	<b>38</b>
<b>5.1.3</b>	<b>Simulação da aquisição em Low-Dose</b>	<b>40</b>
5.2	ARQUITETURAS	41
5.3	AMBIENTE DE TESTES	44
<b>6</b>	<b>EXPERIMENTOS E RESULTADOS</b>	<b>45</b>
6.1	PROJEÇÕES ESPARSAS (QURE.AI)	45
6.2	RUÍDO DE <i>POISSON</i> (MAYO-CHALLENGE)	50
<b>7</b>	<b>CONCLUSÕES</b>	<b>54</b>
	<b>REFERÊNCIAS</b>	<b>56</b>

## 1 INTRODUÇÃO

A Tomografia Computadorizada (CT, do inglês *Computed Tomography*) é amplamente utilizada no diagnóstico por imagens tais como: detecção de embolismo pulmonar, lesões na cabeça e pescoço, lesões renais, detecção de ferro (LIGUORI et al., 2015); além de servirem como ferramenta de estudo em áreas como arqueologia (HUGHES, 2011), e também funcionarem como passo essencial na modelagem 3D de protótipos para cirurgias como a rinoplastia (GRAY et al., 2016), excisão tumoral (MATSUMOTO; MORRIS; ROSE, 2016) e a separação de gêmeos siameses (HEUER et al., 2019). Nesse sentido, as vantagens dos sistemas de CT sobre outros métodos de imagens são: os exames CT são mais rápidos que os de ressonância magnética e radiografia, além disso, são bastante eficientes na visualização de problemas em ossos e juntas, coágulos sanguíneos e ferimentos em órgãos internos (POWER et al., 2016).

O processo para geração de imagens tomográficas inicia-se com o escaneamento do paciente por um equipamento que registra centenas de radiografias durante revoluções em torno do corpo humano. Em seguida, algoritmos de reconstrução são utilizados para combinar as informações de todas as projeções radiográficas em um único modelo 3D que pode ser visualizado através de múltiplas fatias 2D (LIGUORI et al., 2015; POWER et al., 2016).

A obtenção das tomografias computadorizadas resulta na aplicação de grandes doses de radiação em pacientes (BRENNER; HALL, 2007), o que não necessariamente implicaria na existência de riscos mais sérios a pacientes sujeitos a exames esporádicos de CT. No entanto, no estudo feito por González *et al.* (GONZÁLEZ et al., 2009) chegou-se a conclusão de que mais de 29.000 casos de cânceres podem estar relacionado aos exames tomográficos, isso apenas nos Estados Unidos em 2007. Além disso, estimativas feitas por Pearce *et al.* (PEARCE et al., 2012) em 2011/2012 mostraram que a radiação proveniente de múltiplos exames tomográficos feitos em crianças podem aumentar em até três vezes o risco de câncer cerebral e leucemia. Em um estudo ainda mais recente (2018/2019), Shao *et al.* (SHAO et al., 2019) apontaram que a radiação das Tomografias Computadorizadas está ligada a um maior risco de câncer de tireoide e leucemia em adultos e em pessoas diagnosticadas com Linfoma não Hodgkin, quando mais novos. Logo, é consenso que a diminuição da dosagem de radiação emitida durante os exames tomográficos pode reduzir o risco de câncer.

Existem duas estratégias principais para amenizar a quantidade de radiação necessária em CTs. A primeira é a redução da quantidade de projeções de raios-X necessárias (amostragem

esparsa) para um escaneamento completo do paciente (MA; ZHU; ARCE, 2015) e a segunda é a diminuição da dosagem de radiação liberada durante a obtenção das projeções, o que pode ser feito através da redução no tempo de exposição aos raios-X ou redução da corrente de tubo de raios-X, que resulta em uma menor intensidade de radiação (SINGH et al., 2011). Como consequência, ambas as abordagens implicam no aumento de ruídos e distorções nas imagens reconstruídas. Nesse contexto, torna-se essencial a aplicação de técnicas de reconstrução que busquem minimizar o ruído e distorções nos resultados. Por consequência, recuperar imagens geradas a partir de aquisições com baixo uso de radiação é um dos grandes desafios da radiologia atualmente.

Redes neurais convolucionais (CNNs, do inglês *Convolutional Neural Networks*) vêm trazendo bons resultados na recuperação de imagens obtidas em Tomografias Computadorizadas de baixa dosagem (LDCT, do inglês, *Low-Dose Computed Tomography*). Abordando o problema de amostras esparsas, Humphries *et al.* (HUMPHRIES et al., 2019) aplicou uma CNN cuja arquitetura se baseia na DnCNN (ZHANG et al., 2017) e obteve reconstruções usando apenas 180, 90, e 60 projeções de raios-X. Cong *et al.* (CONG et al., 2019) aplicou uma ResNet em reconstruções com apenas 75 e 100 projeções de imagem. Han *et al.* (HAN; YE, 2018) propôs uma arquitetura baseada na U-net (RONNEBERGER; FISCHER; BROX, 2015) e aplicou em reconstruções com 60, 90, 120, 180, 240, e 360 projeções raio-X.

CNNs 2D são os métodos mais utilizados, visto que começaram a ser aplicadas no problema de recuperação de LDCTs, antes das arquiteturas 3D, como aponta Singh *et al.* (SINGH et al., 2020). Redes 2D possuem como principal característica, o processamento de fatias do volume (*slices*) de forma individual, enquanto as abordagens 3D processam volumes em 3 dimensões, geralmente separando um volume inteiro em pedaços menores (*patches*). Quando investigamos as vantagens e desvantagens de cada abordagem, as redes 2D usualmente demandam menos recursos computacionais, são mais rápidas no processamento e permitem maior facilidade na análise dos resultados de forma individual, já que não é necessário processar um volume inteiro para analisar somente um *slice*. Por outro lado, CNNs 3D possuem maior acesso a informações entre fatias de um volume ("entre fatias", pode ser entendido como relações entre os pixels de diferentes *slices* em uma mesma tomografia), visto que processa múltiplos *slices* de uma vez só, o que por sua vez, geralmente resulta em reconstruções mais fiéis e confiáveis. Além da já citada demanda computacional, um outro problema das arquiteturas 3D é que essa é uma abordagem mais suscetível ao problema de desbalanceamento de classes, que é o enviesamento do modelo em favor de uma determinada classe de pixels, por exemplo,

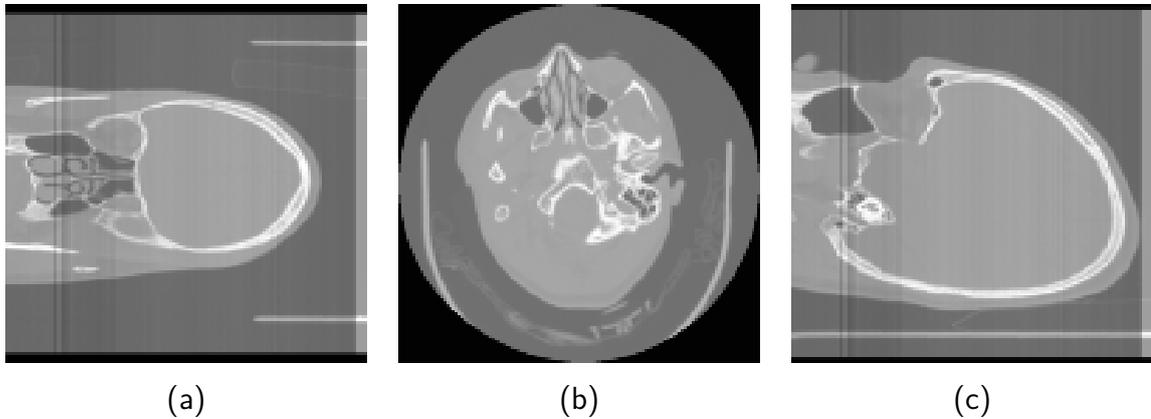
no contexto das redes 3D, são processados volumes completos, o que inclui informações de múltiplas fatias adjacentes. Isso pode levar a um desequilíbrio na distribuição de classes, pois certas classes ou regiões podem ser dominantes em um volume específico. Por exemplo, em uma tomografia, pode haver uma quantidade significativa de tecido ou órgão em uma região específica do volume, enquanto outras regiões podem conter principalmente fundo ou áreas menos relevantes (SINGH et al., 2020). Esse problema não é tão proeminente em redes 2D, visto que, geralmente, o modelo processa fatias inteiras separadamente o que, em tese, garante um balanceamento mais adequado entre as classes de pixels (ZHOU et al., 2018).

Dito isso, essa dissertação investiga a união entre as abordagens 2D e 3D, a qual visa amplificar as vantagens e minimizar os problemas de ambas, culminando em um método que apresenta resultados superiores aos obtidos com o uso de apenas uma solução 2D ou 3D individualmente. Tal abordagem, aproveita-se do processamento *intra-slice* das redes 3D sem a penalidade dos desbalanceamento de classes e altos custos computacionais. Neste cenário, parte-se da hipótese de que o processamento da combinação do resultado de três CNNs 2D, equivalentes aos três planos da tomografia (axial, coronal e sagital), pode gerar uma reconstrução melhor do que o uso de apenas uma CNN 2D ou 3D. Detalhadamente, as três redes primárias são treinadas de forma a aperfeiçoar diferentes visões de um mesmo volume LDCT, e posteriormente, os resultados das redes referentes as visões coronal e sagital são normalizados para o plano axial, em seguida, aproveitando-se da existência dos 3 canais de cores, a saída das 3 redes é unida em uma só imagem. Essa imagem é então processada por uma quarta rede 2D, que é funciona como método de fusão das 3 saídas iniciais, gozando de todas as informações *intra-slice* obtidas no processamento anterior. A Figura 1 mostra uma mesma tomografia vista de forma axial, coronal e sagital.

Cabe destacar algumas hipóteses similares que foram anteriormente aplicadas com sucesso em outros problemas envolvendo volumes CT e CNNs. Lessmann *et al.* (LESSMANN et al., 2018) apresentou uma aplicação de CNN que envolvia a segmentação de CTs obtidas utilizando uma dosagem menor de radiação na região torácica, onde buscava-se identificar a calcificação da artéria coronária. Noothout *et al.* (NOOTHOUT et al., 2018), por sua vez, utilizou-se de uma abordagem que levava em consideração três CNNs, uma para cada plano da tomografia (Axial, coronal e sagital), na segmentação de tomografias da região torácicas obtidas com baixa dosagem de radiação, cujo intuito era realçar a artéria aorta. Por fim, não foi encontrado nenhum trabalho que utilize a hipótese na reconstrução de volumes LDCT. Dessa forma, resultados oriundos da aplicação do método proposto nessa dissertação suportam tal hipótese

ao fim do estudo.

Figura 1 – Diferentes visões de uma mesma tomografia. (a) representa a visão/eixo sagital, (b) Axial e (c) Coronal.



Fonte: O Autor (2023).

O resto da Seção de introdução está dividida em duas Subseções, a Subseção 1.1 apresenta os objetivos do trabalho (o objetivo principal e os secundários) e a Subseção 1.2 apresenta a organização da dissertação no geral.

## 1.1 OBJETIVOS

Esta dissertação tem como objetivo investigar se o uso da combinação de redes neurais convolucionais pode resultar em reconstruções de volumes tomográficos de melhor qualidade do que as obtidas por apenas uma rede 2D ou 3D. Em suma, é esperado que os resultados obtidos pela combinação sejam superiores aos gerados com o uso de apenas uma CNN 2D ou 3D.

Para avaliar esta hipótese, os seguintes objetivos específicos foram definidos:

- Simulação da aquisição de LDCTs com projeções esparsas, a partir de uma base de dados adquirida com dosagem regular;
- Avaliação de implementações 2D e 3D de CNNs através da comparação com o método proposto;
- Investigação da aplicação do método proposto em uma base de dados LDCT não-simulados (redução direta de radiação).

## 1.2 ORGANIZAÇÃO DO TRABALHO

Este trabalho está dividido em 7 capítulos principais: Capítulo 1: Introdução; Capítulo 2: Fundamentação Teórica, onde apresentam-se as bases da dissertação; Capítulo 3: Trabalhos Relacionados, onde apresentam-se os métodos mais relevantes encontrados na literatura, com relação ao método proposto nessa dissertação; Capítulo 4: Método Proposto, onde apresenta-se formalmente, em duas fases, o método proposto; Capítulo 5: Ambiente Experimental, o qual dispõe da apresentação das bases de dados, arquiteturas e detalhes do processo de obtenção das LDCTs utilizadas nos experimentos; Capítulo 6: Experimentos e Resultados, onde estão contidos os experimentos feitos durante a confecção da presente dissertação, bem como a discussão dos resultados obtidos nos mesmos; e por último, Capítulo 7: Conclusões, onde expõe-se as contribuições e próximos passos do método aqui apresentado.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 TOMOGRAFIAS COMPUTADORIZADAS

Tomografias Computadorizadas (CT) são amplamente utilizadas na medicina para auxiliar em diagnósticos e procedimentos cirúrgicos. A aquisição dessas imagens ocorre em duas etapas: primeiro, os tomógrafos realizam rotações ao redor do paciente, enquanto capturam projeções tomográficas. Em média, são necessárias centenas dessas projeções para criar um volume tomográfico completo em alta resolução (REFAAT; SALEM, 2018). Cada projeção libera radiação ionizante no local da tomografia. Por fim, as projeções são utilizadas para reconstruir a imagem tomográfica, com o uso de algoritmos analíticos ou iterativos (LIGUORI et al., 2015; POWER et al., 2016).

O método analítico mais comum é a Retro-Projeção Filtrada (FBP), que usa um filtro unidimensional nos dados de projeção antes de retroprojetá-los no espaço da imagem (KAK; SLANEY, 2001). Por outro lado, os métodos iterativos otimizam um modelo matemático com base nas projeções adquiridas, atualizando continuamente uma estimativa inicial da imagem com base na comparação entre as projeções calculadas e as projeções reais. Um exemplo de método iterativo é a Técnica Iterativa de Reconstrução Simultânea (SIRT), que produz reconstruções visualmente mais agradáveis, mas em um tempo maior (ANDERSEN; KAK, 1984). Hoje em dia, existem várias técnicas de reconstrução tomográficas que podem ser simuladas com a ajuda da ferramenta/biblioteca Astra-toolbox (AARLE et al., 2015).

O resultado da reconstrução é um volume CT 3D, que pode ser visualizado por fatias em diferentes eixos: Axial (geralmente, o eixo horizontal, visto de cima); Coronal, visto de lado; Sagital, visto por trás. Essas diferentes visualizações permitem uma análise mais detalhada do órgão ou tecido em questão, conforme ilustrado na Figura 1.

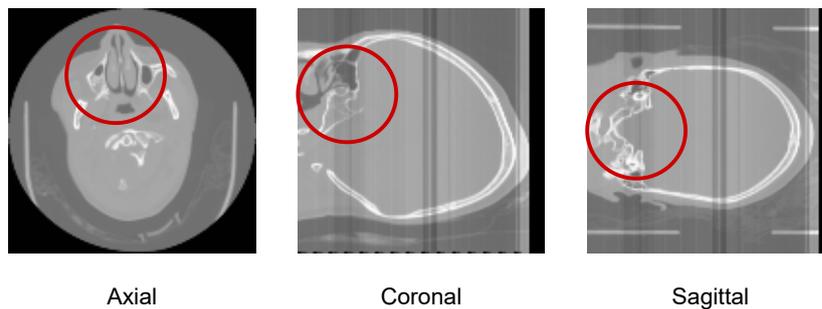
#### 2.1.1 Dosagem Normal de Radiação

As Tomografias Computadorizadas obtidas com dosagem normal de radiação (do inglês *Normal Dose Computed Tomography*, NDCT) é, como o próprio nome já diz, um volume CT obtido sem nenhuma redução nos níveis de radiação durante a aquisição das projeções tomográficas.

Esse tipo de aquisição tomográfica é amplamente utilizada devido a sua vantagem de

reproduzir com máxima fidelidade as estruturas internas do corpo humano (BOONE; STRAUSS, 2014). A Figura 2 mostra três fatias de um mesmo volume NDCT, demonstrando como as diferentes visões complementam-se e contribuem para uma melhor compreensão da anatomia do paciente. A visualização em diferentes planos (axial, coronal e sagital) permite aos médicos avaliar as estruturas internas de diferentes ângulos e identificar possíveis anomalias com maior precisão.

Figura 2 – Diferentes visões de um mesmo volume NDCT (dosagem normal). O círculo vermelho aponta como um conjunto de pixels se complementa nas diferentes visões.



Fonte: O Autor (2023).

### 2.1.2 Dosagem Reduzida de Radiação

No caso das Tomografias Computadorizadas obtidas com dosagem reduzida de radiação (anteriormente apresentadas como *Low-Dose Computed Tomography*, LDCT), trata-se de um volume CT cuja a aquisição foi feita com algum tipo de redução de radiação (MCCOLLOUGH et al., 2015). Essa redução, geralmente está ligada a:

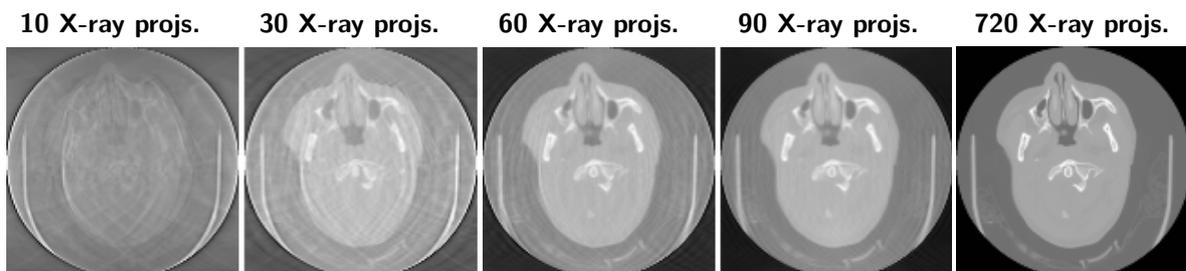
- Redução de tensão elétrica: a redução da tensão do tubo de raios-X durante a aquisição de tomografias computadorizadas pode reduzir a dose de radiação recebida pelo paciente (YANAGA et al., 2018). No entanto, a redução da tensão pode levar a uma redução da qualidade da imagem, o que pode ser compensado por outras técnicas, como a reconstrução iterativa de imagem;
- Uso de filtros: filtros podem ser utilizados para reduzir a dose de radiação recebida pelo paciente durante a aquisição de tomografias computadorizadas (MCCOLLOUGH; LENG; YU, 2015). Esses filtros são colocados entre o paciente e o detector, e podem ser utilizados para filtrar radiação desnecessária;

- Redução no número de projeções: sabendo que a radiação é aplicada no momento da captura das projeções, diminuir a quantidade de projeções necessárias para um *scan* completo resulta também na redução total da dosagem (MA; ZHU; ARCE, 2015).

Como dito anteriormente, a aquisição de tomografias computadorizadas com baixas doses de radiação pode resultar em volumes ruidosos ou com problemas de contraste, o que leva a necessidade de aplicação de algoritmos de recuperação e limpeza que lidam com a diminuição da dosagem, como por exemplo: CNNs.

A Figura 3 apresenta um exemplo prático da importância da dosagem de radiação em exames de Tomografia Computadorizada. Ao comparar a mesma fatia de um volume reconstruída com diferentes quantidades de projeções, fica claro como a dosagem afeta a qualidade da imagem e, por consequência, a precisão do diagnóstico médico.

Figura 3 – Uma mesma fatia do volume reconstruída com diferentes quantidade de projeções (10, 30, 60, 90 e quantidade padrão do Qure.ai (720)).



Fonte: O Autor (2023).

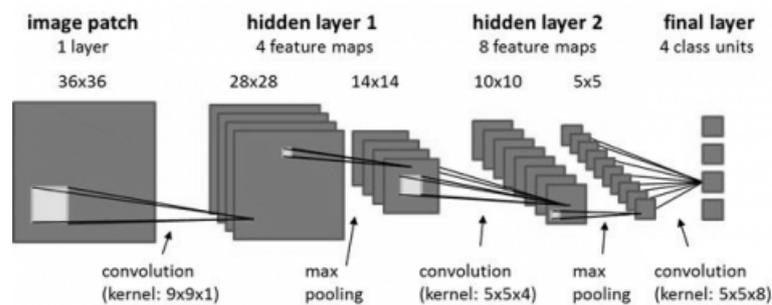
## 2.2 REDES NEURAI CONVOLUCIONAIS

Dados de imagens são representados como matrizes numéricas, onde cada valor refere-se a intensidade do pixel representado. Imagens monocromáticas são representadas por matrizes 2D (altura e largura) enquanto imagens coloridas são representadas por matrizes 3D (altura, largura e profundidade). Redes Neurais Convolucionais são as ferramentas de aprendizagem de máquina para lidar com esse tipo de dados (LECUN; BENGIO, 1995).

As chamadas Redes Neurais Convolucionais (CNN, do inglês *Convolutional Neural Network*), são formadas por camadas convolucionais, nas quais os pesos são filtros convolucionais (em inglês *Kernel*). Como definido em (ZHANG et al., 2023): Uma camada convolucional correlaciona os filtros convolucionais e as imagens de entrada e adiciona um viés (*bias*) escalar. Os

dois parâmetros da camada convolucional são os filtros convolucionais e o viés escalar. Uma outra operação característica das CNNs é o *MaxPooling*, operação não linear que reduz a dimensionalidade dos dados de entrada da CNN, através da aplicação de uma janela deslizante sobre a entrada. Por fim, em cada posição da janela é selecionado o valor máximo encontrado. Essa técnica ajuda a capturar características importantes, como bordas ou texturas, enquanto reduz a quantidade de parâmetros e a computação necessária nas camadas subsequentes da rede (GOODFELLOW; BENGIO; COURVILLE, 2016). A Figura 4 exemplifica uma CNN com duas camadas escondidas, duas operações de *MaxPooling* além de fornecer uma visão geral da arquitetura de uma CNN.

Figura 4 – Exemplo de CNN com 2 camadas escondidas. A CNN conta ainda com 2 operações de *max pooling*



Fonte: Trimble Inc. 2019, Acesso em Março de 2022. Disponível em: <https://images.app.goo.gl/z3uzXaKnmoBL8APEA>

Redes neurais "clássicas", como a Multilayer Perceptron (rede neural que usa camadas de neurônios para aprender padrões em dados), têm dificuldade de receber imagens como entrada, por estas serem dados não estruturados onde cada pixel é equivalente à dimensão do vetor de entrada, e então, devido ao fato de as imagens possuírem grandes volumes de pixels, seriam necessárias diversas entradas de dados ligadas aos neurônios, tornando o custo computacional muito alto (ZHANG et al., 2023). Por exemplo, uma imagem com 32 pixels de altura e 32 de largura, precisaria de 1024 (32x32 pixels) entradas para cada neurônio da camada de entrada, os quais possuiriam, no mínimo 1048576 (1024 x 1024 pixels) ligações com a primeira camada escondida. É notável que o custo computacional para somente uma imagem pequena já é muito alto. Além disso, existe também a necessidade de transformar a imagem em um vetor unidimensional, acarretando na perda de informações entre pixels de vizinhança próxima (ZHANG et al., 2023).

As CNNs são utilizadas em bases de dados de imagens por trabalhar melhor com dados não estruturados (LECUN; BENGIO; HINTON, 2015), uma vez que não é necessário nenhum

pré-processamento dos dados antes de serem fornecidos para o método, pois as CNNs detectam as features importantes de forma automática na camada convolucional. Além disso, por utilizarem-se da convolução matemática, que é uma operação matricial, a complexidade computacional diminui bastante em comparação às já citadas Redes Neurais clássicas, visto que cada filtro/kernel (matrizes com as quais a convolução opera na entrada) abrange uma quantidade maior de pixels por operação. Analogamente as redes clássicas, os filtros das camadas convolucionais representam os pesos. Os valores nas matrizes dos filtros são atualizados ao longo do treino da CNN, exatamente como nas MLPs.

Uma convolução é definida pela função:

$$f_{(j,k)} = \sigma\left(b + \sum_{l=-1}^1 \sum_{m=-1}^1 \omega_{l,m} \alpha_{j+l,k+m}\right)$$

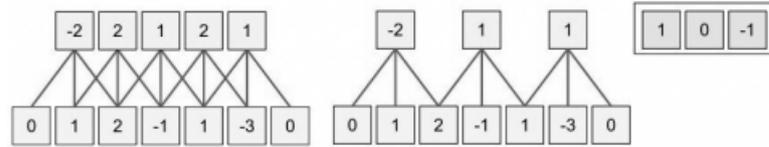
Em que,  $l$  e  $m$  são os índices de linha e coluna do filtro;  $\sigma$  é a função de ativação neural e  $b$  é o valor compartilhado do viés escalar;  $\omega$  é o filtro ou kernel;  $\omega_{l,m}$  é o valor do filtro na posição  $l,m$ ;  $\alpha_{j+l,k+m}$  é a ativação de entrada na posição  $\alpha_{x,y}$ , onde  $x = j + l$  e  $y = k + m$ .

As convoluções acontecem em toda a imagem. Durante o *forward-pass*, os filtros deslizam ao longo da matriz, realizando convoluções em cada pixel da entrada. O resultado das convoluções é um produto entre os elementos da entrada e o filtro que o sobrepôs, o qual será usado para indicar a correlação entre o filtro e a imagem em diversos pontos.

As camadas convolucionais possuem 4 hiper-parâmetros importantes: tamanho do filtro (*kernel size*), número de filtros (*depth*), *stride* e *padding*. O tamanho do filtro diz respeito ao tamanho (altura e largura) dos filtros convolucionais daquela camada. Por sua vez, número de filtros é o hiper-parâmetro responsável por regular quantos filtros serão usados na camada, em que cada filtro será responsável por detectar uma *feature*. O *stride* é o valor que determina o espaçamento do deslocamento dos filtros. a Figura 5 mostra como diferentes valores de *stride* afetam o resultado, de forma que quanto maior o valor do mesmo, menor é o tamanho do resultado. O *padding* nada mais é que o tamanho das bordas de preenchimento das matrizes de entrada (usualmente todos os elementos das bordas possuem o valor 0). Esse parâmetro geralmente é utilizado em conjunto com o *stride*, para poder normalizar o tamanho da entrada ao *stride*.

A grande vantagem das CNNs é poder utilizar o mesmo filtro em todo um mesmo canal (profundidade) da entrada, o que implica que poucos valores são utilizados para representar *features*, pois para cada canal os pesos serão compartilhados entre todos os valores, indepen-

Figura 5 – Exemplo de aplicação do *kernel*  $(1, 0, -1)$  com *strides* de tamanho 1 e 2. Nota-se que com *Stride* = 2 o resultado é duas vezes menor que *Stride* = 1.



Fonte: Github cs231n, Acesso em Março de 2023. Disponível em: <https://cs231n.github.io/convolutional-networks>.

dente do tamanho da entrada.

### 3 TRABALHOS RELACIONADOS

Apesar da escassez de trabalhos que promovem o uso de múltiplas CNNs no processo de reconstrução de LDCTs, é possível encontrar alguns estudos que idealizam diferentes abordagens que podem ser comparadas com o método proposto nessa dissertação. É importante, em um primeiro momento, separar métodos que fazem o uso dos diferentes eixos de visões (axial, coronal e sagital) das tomografias computadorizadas ((IBRAGIMOV et al., 2017; NOOTHOUT et al., 2018; MORTAZI et al., 2017; LIU; KANG, 2017), de trabalhos que apenas aplicam múltiplas CNNs na obtenção de seus resultados ((NANNI et al., 2020; KUSHIBAR et al., 2017; HUANG et al., 2020; KANG et al., 2019; GU; YE, 2021). Abaixo estão listadas e discutidas algumas das referências mais interessantes e relevantes encontradas na literatura.

#### 3.1 CNNS EM LOW-DOSE

Uma referência fundamental sobre o uso de CNNs na reconstrução de LDCTs é o artigo de (CHEN et al., 2017) de 2017. Nesse trabalho, os autores descrevem a reconstrução de LDCTs, realizada através de patches com o uso de uma CNN. Eles detalham o processo de seleção e funcionamento de cada camada da CNN durante a recuperação da LDCT. O artigo apresenta de maneira objetiva o uso da CNN na resolução do problema em questão, tornando-se uma referência valiosa na área.

Outro trabalho parecido é o proposto em (Chen; Zhang et al., 2017) em 2017, onde foi proposta a arquitetura de uma rede do tipo codificadora-decodificadora residual para *denoising* de LDCTs (REDCNN). Uma rede codificadora-decodificadora é uma arquitetura de rede neural que consiste em duas partes principais: um codificador e um decodificador. O codificador é responsável por mapear a entrada para um espaço latente, enquanto o decodificador mapeia o espaço latente de volta para a saída. O trabalho, novamente, descreveu em detalhes a arquitetura da rede proposta, a qual tornou-se um dos *baselines* da literatura no que se refere a *denoising* de LDCTs. A Figura 6 mostra a arquitetura da REDCNN, como vista no artigo original.

Em 2018 (YANG et al., 2018) propôs um método que utiliza uma rede Generativa-Adversarial em junção com uma *perceptual loss* e distância de Wasserstein na recuperação das LDCTs. (GOODFELLOW et al., 2014) define as redes Generativas-Adversarial (*Generative Adversarial*

Figura 6 – A Figura abaixo mostra a arquitetura da REDCNN, proposta em (Chen; Zhang et al., 2017). A parte de cima mostra o processo de codificação da arquitetura, enquanto a parte de baixo mostra o processo inverso, de decodificação.

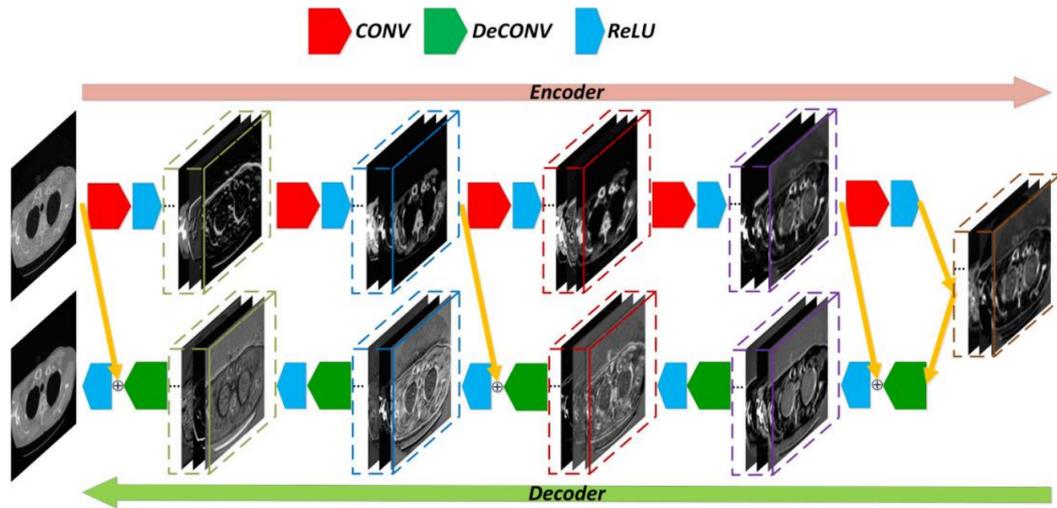


Imagem disponível em (Chen; Zhang et al., 2017).

*Networks* (GANs)) como um tipo de modelo de aprendizagem profunda que consiste em duas redes neurais, o gerador e o discriminador, que competem entre si em um jogo minimax. O gerador tenta criar amostras que imitam um conjunto de dados de treinamento, enquanto o discriminador tenta distinguir as amostras geradas do conjunto de dados real. A Perda perceptual é uma técnica de otimização em redes neurais que usa uma representação de alta dimensionalidade da imagem, geralmente obtida a partir de uma rede pré-treinada de classificação de imagens, para medir a diferença entre a imagem gerada e a imagem original. Em vez de simplesmente comparar os valores de pixel das imagens, a perda perceptual considera aspectos mais abstratos, como textura, cor e forma, permitindo que a rede gere imagens mais realistas e convincentes. Essa abordagem é comumente usada em tarefas de geração de imagens, como super-resolução, colorização e síntese de imagens, e tem se mostrado eficaz na melhoria da qualidade das imagens geradas (JOHNSON; ALAHI; FEI-FEI, 2016). A distância de Wasserstein (também conhecida como distância de transporte ótimo) é uma métrica utilizada para medir a diferença entre duas distribuições de probabilidade. Ela é definida como a quantidade mínima de trabalho necessário para transformar uma distribuição na outra, onde o trabalho é medido como o produto entre a quantidade de massa movida e a distância percorrida, como explicado por (GURULINGAPPA; JAKKULA; HU, 2017). A ideia principal do artigo, é buscar minimizar a distância de Wasserstein entre a distribuição real dos dados e a distribuição gerada pela rede geradora, ao mesmo tempo que busca-se minimizar a diferença perceptual entre as *features*

da imagem gerada pela rede geradora e da imagem NDCT. A Figura 7 mostra a visão geral do método proposto.

Figura 7 – A Figura abaixo mostra a visão geral do método proposto em (YANG et al., 2018). Na "Part 1" observa-se a arquitetura da rede geradora, enquanto nas "Part 2 e 3" lista-se o funcionamento da *perceptual loss* e da rede discriminadora respectivamente.

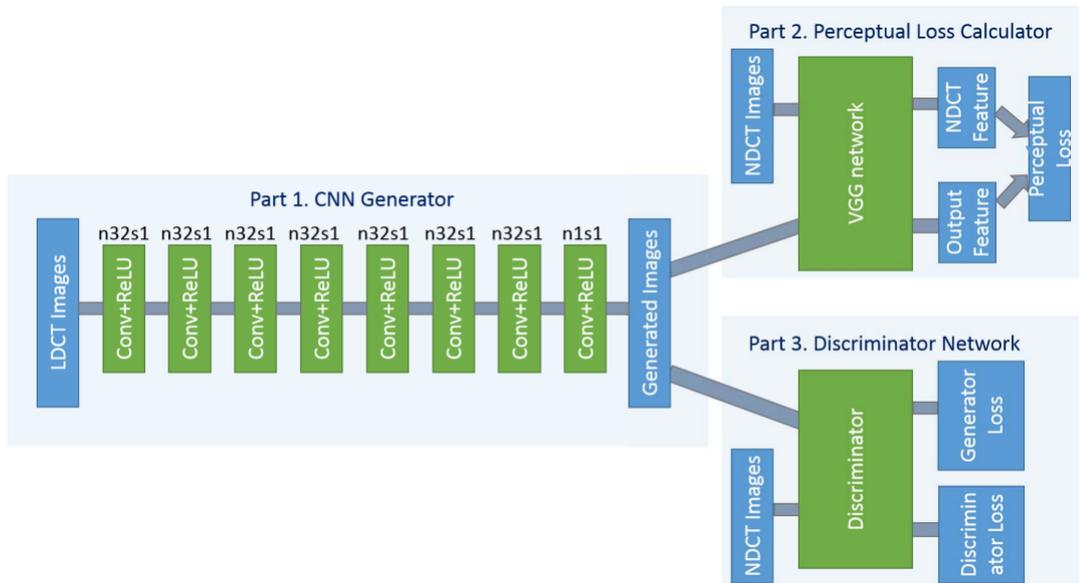
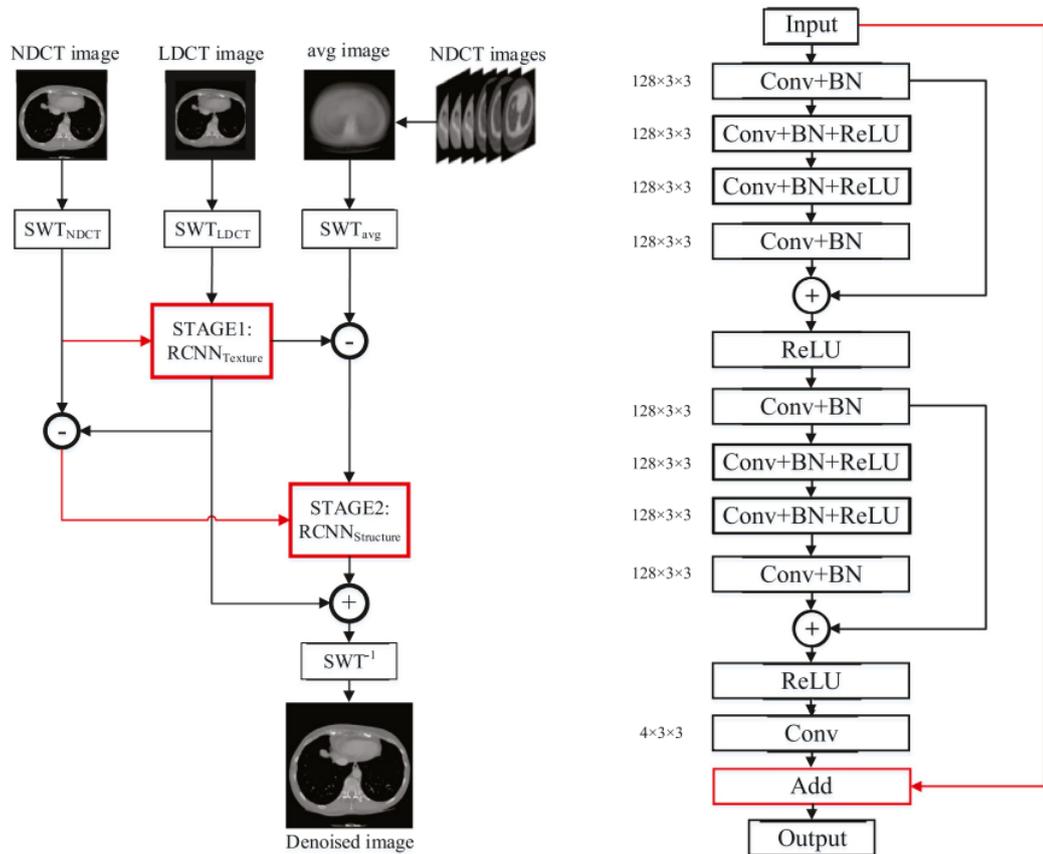


Imagem disponível em (YANG et al., 2018).

### 3.2 MÚLTIPLAS CNNs EM LOW-DOSE

Em 2019, (HUANG et al., 2020) propôs o uso de duas CNNs residuais no *denoising* de LDCTs. O processo geral pode ser melhor entendido com a explicação separada do papel de cada CNN: A primeira tem como objetivo fazer o *denoising* inicial do LDCT, e suas particularidades são o uso de uma transformada wavelet estacionária (SWT, *stationary wavelet transform*) e *perceptual loss* como *loss function*. Já a segunda, busca melhorar ainda mais os resultados obtidos na primeira etapa. Uma *wavelet* é uma função que decompõe e descreve ou representa outra função em suas componentes de frequência, esse processo é conhecido como "transformada wavelet (*wavelet transform*)" (HUBBARD, 1996), uma transformada wavelet estacionária, ao decompor uma função, não diminuirá a resolução da mesma (JAMALUDDIN et al., 2015). A Figura 8 apresenta o diagrama geral do método e a arquitetura das redes. O treinamento e teste foram feitos em LDCTs simulados e os resultados foram majoritariamente positivos. O ponto a se notar nesse trabalho é o uso de "etapas" (podemos dizer que cada CNN corresponde a uma etapa, nesse sentido).

Figura 8 – A Figura abaixo mostra a visão geral do método proposto em (HUANG et al., 2020). Em (a), observe o fluxo de entrada e processamento das imagens, enquanto (b) apresenta a arquitetura da CNN primária. É importante notar que a primeira CNN é chamada de RCNN(texture), pois essa tem como objetivo fazer a limpeza inicial de forma a extrair a textura principal do CT.



(a) The overall process of TS-RCNN.

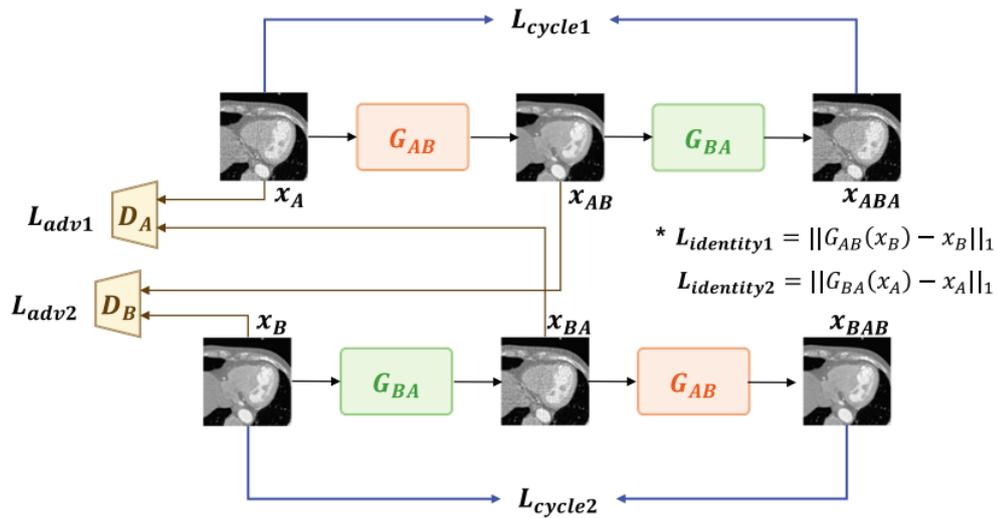
(b) The network structure of RCNN<sub>Texture</sub> in (a).

Imagem disponível em (HUANG et al., 2020).

Apesar da maioria dos métodos de recuperação de LDCTs serem baseados em aprendizagem supervisionada, em 2019, (KANG et al., 2019) propôs uma abordagem não-supervisionada utilizando o CycleGAN. Esse método é composto por duas redes geradoras, onde uma é responsável por mapear imagens do domínio LDCT para o domínio NDCT, enquanto a outra faz o inverso, mapeando do domínio NDCT para o LDCT. Para isso, são utilizados dois discriminadores adversariais, um para cada gerador. A Figura 9 ilustra o método. É importante ressaltar que essa abordagem não resultou na criação de novas estruturas ou adição de ruídos, que são problemas típicos de aprendizagem não supervisionada. O uso dessa abordagem mostra que a utilização de múltiplas CNNs pode levar a resultados melhores, mesmo em problemas não-supervisionados.

Recentemente, (GU; YE, 2021) adaptou o método não-supervisionado de *denoising* Cycle-

Figura 9 – O método proposto em (KANG et al., 2019) consiste em utilizar duas redes geradoras,  $G_{AB}$  e  $G_{BA}$ , em conjunto com dois discriminadores adversariais,  $D_A$  e  $D_B$ , para realizar a recuperação de imagens de tomografia computadorizada de baixa dose (LDCT) por meio de imagens de tomografia computadorizada de dose normal (NDCT). A figura abaixo apresenta uma visão geral do método, onde  $A$  e  $B$  denotam os domínios LDCT e NDCT, respectivamente. Após o processamento de cada imagem pelas redes geradoras, a qualidade dos resultados é avaliada pelos discriminadores correspondentes. Uma das características interessantes do método é a utilização de três funções de perda diferentes: adversarial, cíclica e identidade.



A Figura está disponível em (KANG et al., 2019)

GAN, utilizando apenas uma rede geradora e aplicando camadas de normalização adaptativa (*adaptive instance normalization*, AdaIN) para investigar melhor o processo de transformação de imagens NDCT em LDCT. As mudanças no método estão ilustradas nas Figuras 10 e 11. Os resultados da nova versão superaram o método original, mesmo com apenas metade dos parâmetros, graças às camadas AdaIN, que possibilitaram a investigação das etapas de processamento das imagens em níveis intermediários. O trabalho é importante, mesmo utilizando apenas uma rede, devido à otimização em relação ao método anterior que usava múltiplas redes.

### 3.2.1 Métodos que utilizam as múltiplas visões

A técnica de utilizar múltiplas visões já foi explorada com sucesso em segmentação de imagens médicas, como no caso do estudo realizado por (IBRAGIMOV et al., 2017). Neste trabalho, as múltiplas visões foram utilizadas para a segmentação da veia porta em imagens de tomografia computadorizada (CT). Foram empregadas três redes neurais convolucionais (CNNs), uma para cada visão, e também cadeias de Markov para aprimorar os resultados

Figura 10 – A Figura abaixo mostra a mudança feita por (GU; YE, 2021) no método proposto em (KANG et al., 2019). A mudança mais óbvia é o uso de apenas uma rede geradora  $G$  ao invés de duas ( $G_{AB}$  e  $G_{BA}$ ). Além disso, é possível ver as camadas AdaIN  $F$ .

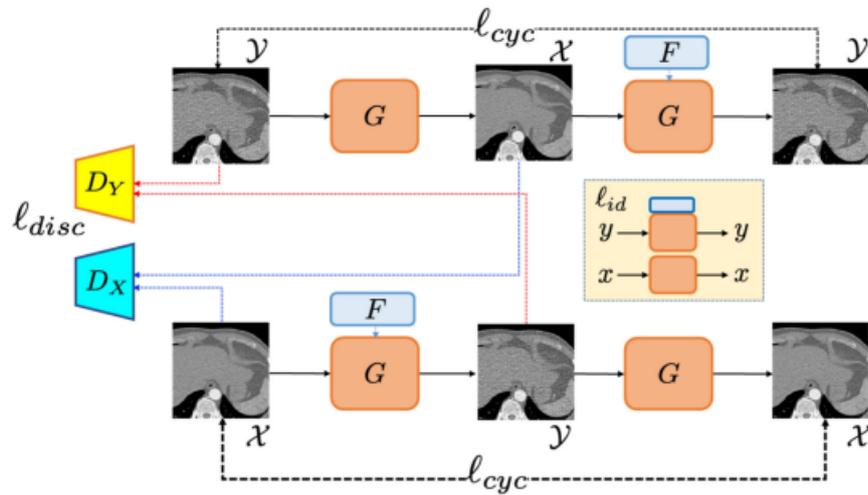


Imagem disponível em (GU; YE, 2021).

Figura 11 – Arquitetura da rede geradora com as camadas AdaIN. A arquitetura original era uma U-net padrão onde, as camadas de normalização de *batch* foram substituídas pelas camadas AdaIN.

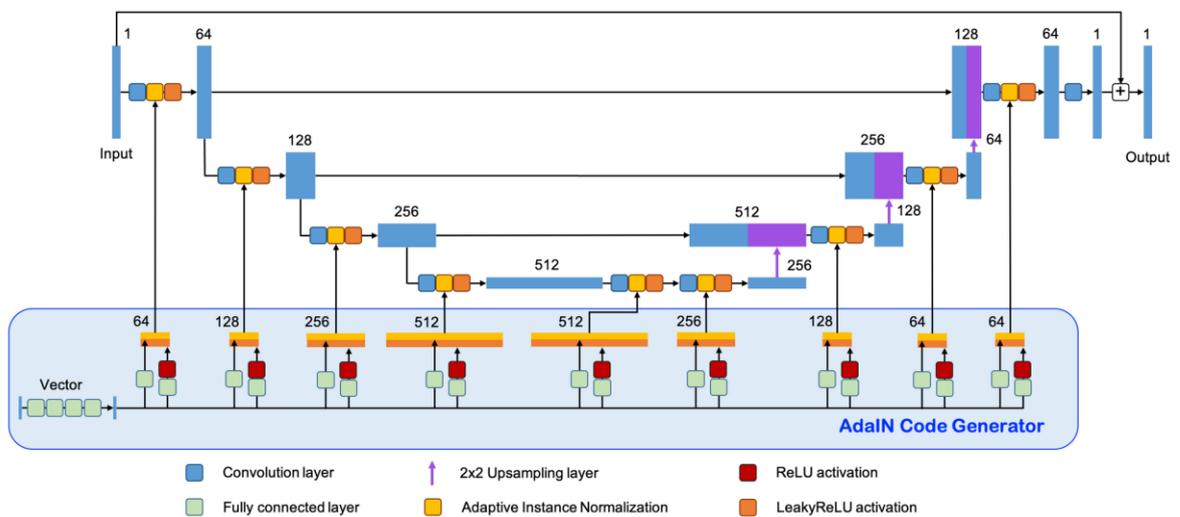


Imagem disponível em (GU; YE, 2021).

finais. A Figura 12 apresenta uma ilustração esquemática que demonstra o funcionamento das redes e fornece uma visão geral do processo.

Uma aplicação de segmentação de imagens médicas foi abordada em (MORTAZI et al., 2017), onde se optou pelo uso de múltiplas visões para reduzir os custos computacionais de redes 3D. Segundo os autores, a modelagem do problema em 2D reduziria o custo computacional e o uso de múltiplas visões poderia gerar resultados equivalentes ou até mesmo melhores do

Figura 12 – Na figura é possível observar o esquema geral do método proposto por (IBRAGIMOV et al., 2017), que utilizou múltiplas visões no processo de segmentação da veia porta em imagens CT. O método consistiu na aplicação de três redes neurais convolucionais, cada uma processando uma visão separadamente e criando três visualizações melhoradas da veia porta. Para criar a segmentação final, as saídas das três redes foram combinadas por meio de uma média. Essa abordagem permitiu uma melhoria significativa nos resultados de segmentação da veia porta em comparação com métodos que utilizam apenas uma única visão.

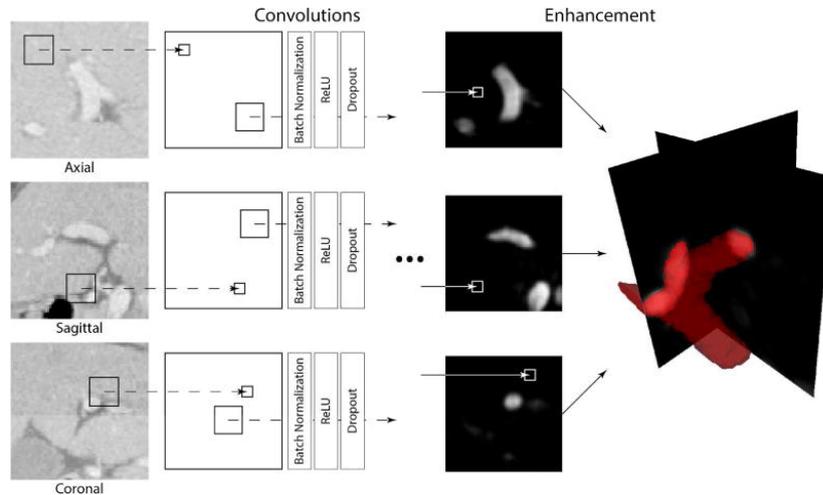


Imagem disponível em (IBRAGIMOV et al., 2017).

que os obtidos com redes 3D. O fluxo do método é apresentado na Figura 13 e é semelhante ao processo descrito em (IBRAGIMOV et al., 2017), com exceção do fato de que não se utiliza cadeias de Markov e a área de segmentação é diferente.

Figura 13 – Esquema geral do uso das múltiplas visões por (MORTAZI et al., 2017). Novamente, múltiplas CNNs são utilizadas no processamento de cada eixo e, para o resultado final, é utilizado uma estratégia de fusão que, medindo a confiabilidade de cada CNN, busca maximizar a fidelidade da segmentação.

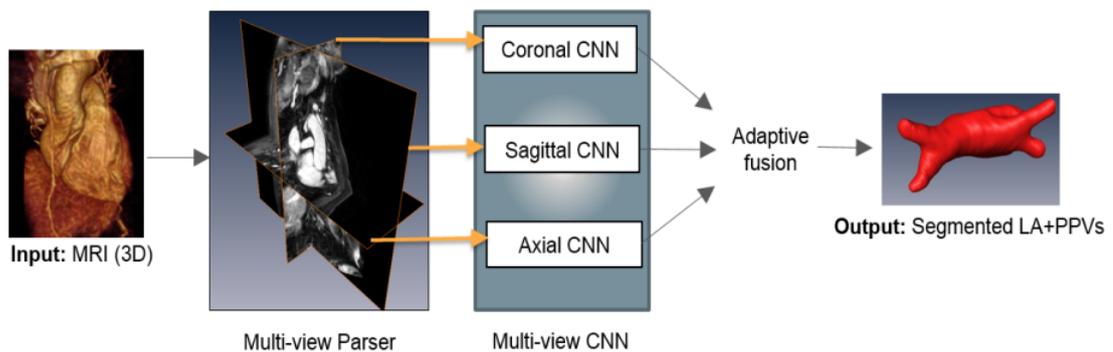


Imagem disponível em (MORTAZI et al., 2017).

Em 2017, (KUSHIBAR et al., 2017) examinou uma abordagem que emprega o uso de diferentes visões como entrada para uma CNN 2.5D (ROTH et al., 2014) no processo de segmentação de estruturas subcorticais de tomografias do cérebro. A Figura 14 apresenta o diagrama com-

pleto do funcionamento do método que, além das múltiplas visões, utiliza também *features* espaciais extraídas através do uso do que eles chamam de "Atlas probabilístico estrutural", que são informações estatísticas sobre estruturas corporais extraídas de diversos indivíduos através de uma média global ou local da estrutura de interesse (GOOD et al., 2001). Apesar de ser, também, um método de segmentação como os citados anteriormente, esse aqui possui algumas particularidades que ajudam a entender e expandir o poder do uso das múltiplas visões, além disso, o fato de apenas uma rede fazer a análise da informação axial, coronal e sagital, dá um indicativo do poder de processamento *multi-view* de uma CNN, validando-a como método de fusão das três visões (que é o papel da quarta CNN no método proposto nesta dissertação).

Figura 14 – A Figura abaixo mostra a visão geral do método proposto em (KUSHIBAR et al., 2017). Em (A) é possível ver a arquitetura da CNN 2.5D, que possui diferentes entradas para cada visão e para as informações espaciais do Atlas. As visões são, inicialmente, processadas pela mesma organização estrutural (apresentada em (B)) e posteriormente, são utilizadas em conjunto com a informação espacial do Atlas na obtenção da segmentação final. A camada final (*softmax*) possui a probabilidade de cada estrutura subcortical (ao todo, são 15 possíveis estruturas), e a segmentação é feita com base na estrutura de maior probabilidade.

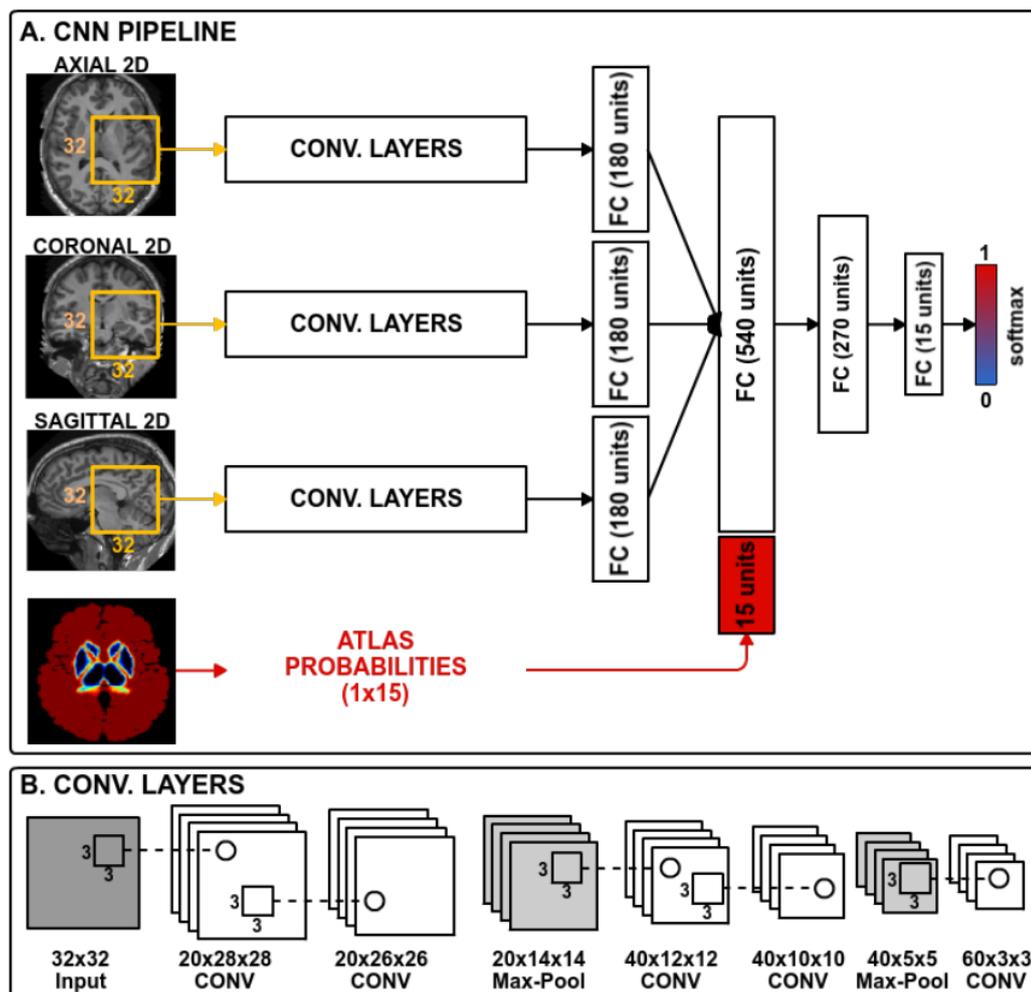


Imagem disponível em (KUSHIBAR et al., 2017).

---

O artigo de (ZHENG et al., 2021) apresentou uma abordagem eficaz para a detecção de nódulos pulmonares, com foco em nódulos menores e mais difíceis de detectar. Para isso, eles utilizaram uma CNN do tipo *encoder-decoder* que processou fatias de cada uma das três visões, com a marcação de possíveis nódulos adicionada na saída. Esses resultados foram então fornecidos como entrada para uma CNN 3D, que teve como objetivo reduzir a quantidade de falsos positivos. Este é um exemplo promissor do uso de CNNs para a fusão de informações, em que uma CNN 3D é usada como decisora de diferentes saídas.

O trabalho de maior correlação com o proposto nesta dissertação é (LIU; KANG, 2017), o qual utilizou uma CNN 2D como entrada de múltiplas fatias de uma mesma tomografia, com o objetivo de a rede extrair relações entre os pixels das diferentes imagens. Os canais de cores foram utilizados como entrada das três visões e a rede utilizou essa informação espacial extra para classificar a possível existência de um tumor nos pulmões. Essa aplicação é interessante pois mostra que utilizar uma CNN como método de intermediar e pesquisar essas informações entre as fatias leva a bons resultados.

Até o momento, não foram encontrados métodos que fazem uso do processamento das múltiplas visões (axial, coronal e sagital) no processo de recuperação/reconstrução de volumes LDCT. Além disso, até o momento da escrita desta dissertação, o uso de uma CNN 2D no processo de fusão das múltiplas visões é uma novidade. Logo, o método proposto é inovador de duas formas: uso de múltiplas visões para processamento LDCT e aplicação de CNN 2D como método de fusão.

## 4 MÉTODO PROPOSTO

A apresentação do método pode ser dividida em duas fases para maior clareza e compreensão: processamento *multi-view* e fusão (JUNIOR et al., 2022). Durante a fase de processamento *multi-view*, cada visão (axial, coronal e sagital) é processada separadamente por uma respectiva CNN, todas com a mesma arquitetura. Cada rede trabalha com as mesmas tomografias, mas em visões diferentes, usando fatias em tons de cinza (um canal de cor) como entrada. Isso significa que cada fatia do volume é processada iterativamente por cada rede.

O uso das três visões (axial, coronal e sagital) permite uma exploração mais completa das relações entre as fatias do volume LDCT, sem a necessidade do uso de uma rede 3D, que poderia acarretar em maior complexidade computacional. O objetivo principal é obter diversidade em cada uma das visões, ou seja, extrair detalhes que podem ser decisivos para o resultado final da reconstrução.

Como explicado anteriormente, três redes neurais convolucionais (CNNs) são utilizadas para processar separadamente cada visão da mesma tomografia, explorando informações dispostas em diferentes ângulos. É importante destacar que, apesar de acessarem as mesmas informações de cada volume, as redes analisam as informações de diferentes ângulos, o que pode resultar em informações complementares e exclusivas em cada visão. Por exemplo, um pixel analisado na visão axial pode ter uma vizinhança completamente diferente ao ser investigado na visão coronal ou sagital, o que pode enriquecer a reconstrução final com informações que não estão presentes em outras visões. O resultado é a obtenção de três reconstruções diferentes do mesmo volume LDCT que, idealmente, contêm informações complementares e diversificadas.

Para ilustrar melhor, seja  $V$  um volume de entrada tridimensional com dimensões  $n_1 \times n_2 \times n_3$ , onde  $n_1$ ,  $n_2$  e  $n_3$  são o número de voxels em cada dimensão. O objetivo é reconstruir uma imagem LDCT  $I$ , representada por uma matriz bidimensional de dimensões  $m_1 \times m_2$ , onde  $m_1$  e  $m_2$  representam o número de pixels em cada dimensão.

Cada visão do volume de entrada é processada separadamente por uma CNN, usando fatias em tons de cinza (um canal de cor) como entrada. A saída de cada CNN é uma matriz bidimensional de dimensões  $m_1 \times m_2$ . Sejam  $f_A$ ,  $f_C$  e  $f_S$  as funções de mapeamento aprendidas pelas CNNs para a visão axial, coronal e sagital, respectivamente. Dessa forma, a saída do processamento *Multi-View* para a visão axial pode ser representada como a imagem reconstruída  $I_A = f_A(V_A)$ , onde  $V_A$  é o volume de entrada visto na visão axial. Analogamente,

a saída para as visões coronal e sagital pode ser representada como  $I_C = f_C(V_C)$  e  $I_S = f_S(V_S)$ , respectivamente.

É importante informar que, para a segunda fase (fusão), é necessário que o volume de entrada seja completamente processado nas diferentes visões. Em suma, todas as fatias de um mesmo volume devem ser recuperadas nas três visões, resultando em três reconstruções tomográficas intermediárias.

Na fase de fusão, os resultados de saída das redes do processamento *multi-view* são combinados para a reconstrução final, que é realizada por uma quarta CNN com uma arquitetura semelhante às outras três, exceto por uma pequena adaptação (mudança na quantidade de canais/filtros de entrada) para permitir a entrada de imagens com três canais. A ideia aqui é combinar as saídas da primeira etapa em um único tensor através dos canais de cores, e é devido essa combinação que justifica-se a necessidade da reconstrução intermediária completa dos volumes tomográficos na fase *multi-view*, visto que as três fatias serão extraídas de uma mesma visão nos volumes intermediários. Dessa forma, garante-se que a fusão será feita apenas com a mesma fatia reconstruída por diferentes visões.

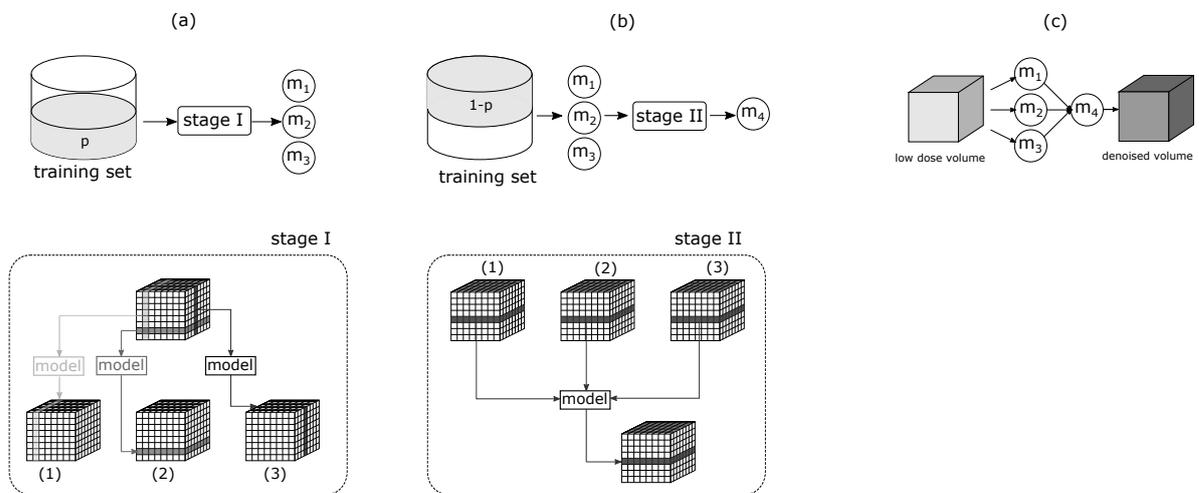
A quarta rede é responsável por combinar as informações obtidas pelas três redes do processamento *multi-view*. A combinação dessas informações é feita através do uso dos canais de cores como mecanismo de entrada, unindo as três fatias processadas em uma única imagem. Cada canal de cor representa um eixo de visão e a combinação das informações permite obter um resultado final mais completo e com maior riqueza de detalhes.

No entanto, existem alguns desafios a serem superados na utilização desta rede de fusão. Em primeiro lugar, como dito anteriormente, é necessário padronizar as três reconstruções de entrada para que estejam na mesma visão (geralmente a axial), surgindo então a obrigação de obter reconstruções intermediárias completas do volume de entrada em cada visão. Em segundo lugar, a arquitetura original da rede precisa ser adaptada para receber imagens com três canais de cores, o que é feito através da alteração manual no código da CNN. Essas alterações são necessárias para permitir que a rede possa processar adequadamente as informações obtidas pelas três redes de processamento *multi-view* e gerar um resultado final mais preciso e completo.

Novamente, para melhor visualização, ilustramos o método de fusão, seja  $F$  a função de mapeamento aprendida pela CNN de fusão. A saída da segunda fase do método é a imagem LDCT final  $I = F(I_A, I_C, I_S)$ , que é a reconstrução final obtida pela fusão das imagens reconstruídas para cada uma das três visões.

Em resumo, o método pode ser representado matematicamente como  $I = F(I_A, I_C, I_S)$ , onde  $I_A = f_A(V_A)$ ,  $I_C = f_C(V_C)$  e  $I_S = f_S(V_S)$  são as imagens reconstruídas para cada visão do volume de entrada  $V$ , e  $F$  é a função de mapeamento aprendida pela CNN de fusão. A Figura 15 ilustra, de forma geral, o método proposto.

Figura 15 – O esquema para treinamento do método em duas etapas: (a) e (b). Na etapa (a), Uma partição  $p_1$  do conjunto de treino  $T$  é usado pra treinar a primeira fase de combinação sendo as redes referentes a sagital (1), axial (2), e coronal (3). Como resultado, um conjunto de modelos  $\{m_1, m_2, m_3\}$  é gerado. Em (b), uma partição  $p_2$  de  $T$  é processado pelos modelos treinados anteriormente, gerando tuplas  $((\hat{v}_k^{m_1}, \hat{v}_k^{m_2}, \hat{v}_k^{m_3}), v_k)$ ; usando conjuntos de tuplas, o modelo final (fusão)  $m_4$  é treinado. Em (c), os modelos treinados anteriormente são organizados de forma a computar  $\hat{v}^{m_4} = m_4(\hat{v}^{m_1}, \hat{v}^{m_2}, \hat{v}^{m_3})$ .



Fonte: O Autor (2023).

## 5 AMBIENTE EXPERIMENTAL

### 5.1 DATASETS

Ao todo, foram utilizados duas bases de dados para os experimentos. A primeira, Qure.Ai (CHILAMKURTHY et al., 2018), foi utilizada nos experimentos que dizem respeito ao problema das projeções esparsas, enquanto a segunda, Mayo-Challenge (MAYO..., 2017), protagonizou a segunda rodada de experimentos, cuja simulação das LDCTs foi feita antes da reconstrução (através da adição de ruído de *Poisson*). Essa separação de experimentos foi feita por consequência do intervalo entre um experimento e outro. Inicialmente foram feitos os experimentos apenas com os dados simulados e, posteriormente, foi decidido fazer novos experimentos com dados não-simulados, de forma a conduzir novos testes com o método proposto e investigar qual o melhor cenário para aplicação do mesmo.

#### 5.1.1 Qure.ai

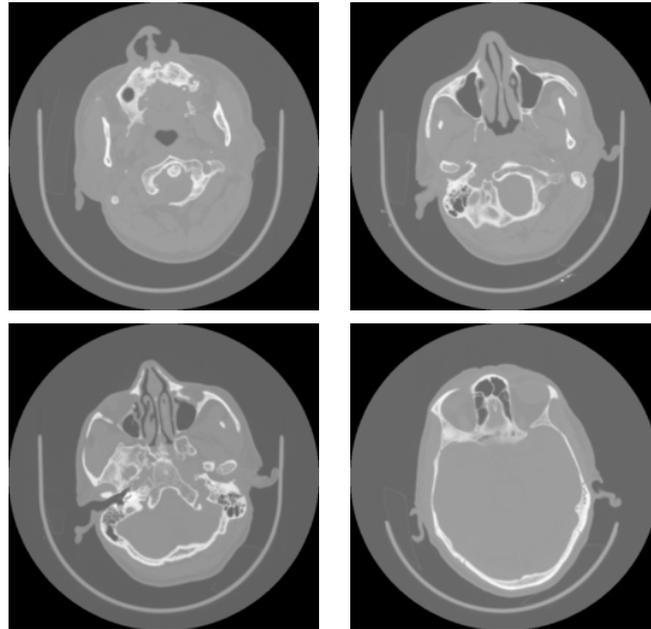
Qure.ai CQ500 *dataset* é uma base de dados que possui 258 tomografias computadorizadas (cada uma com, em média, 256 fatias) disponíveis. A base é composta por tomografias computadorizadas da cabeça e pescoço, coletadas de 20 diferentes centros radiológicos na Índia. Como resultado, a base de dados contém CTs adquiridas com doses normais de radiação.

Uma vantagem importante na utilização dessa base de dados, em específico, é a grande variedade de scans. Os exames foram realizados em diferentes tomógrafos, em diferentes locais, por diferentes profissionais. Graças a isso, diminui-se a possibilidade de *overfitting* do método, visto que as diferentes configurações na aquisição, simulam melhor a diversidade real do mundo. A Figura 16 contém 4 exemplos de fatias de diferentes volumes CT da base de dados utilizada.

Cada um dos volumes CT presentes na Qure.ai originalmente possuem diferentes dimensões geométricas, tendo em média tamanhos de 256x256x256 pixels. Sabendo disso, cada tomografia foi dividida de forma que suas dimensões de altura e largura se iguallassem á profundidade do volume. Tal divisão foi repetida para todos os três eixos, anteriormente apresentados na Figura 1.

Como dito, os volumes CT possuem dimensões de tamanho diferente, um scan com 245 fatias no eixo axial, resulta em um volume no eixo coronal, cujas fatias possuem dimensões

Figura 16 – Alguns exemplos de tomografias encontradas na base de dados Qure.ai CQ500.



Fonte: O Autor (2023).

245x256 pixels (por exemplo). Isso é um problema devido ao tamanho da entrada das três redes que compõem a primeira parte do método proposto, onde, por vezes, a fatia resultante do corte necessitava de um complemento em uma de suas dimensões, como um *padding* de pixels pretos. Além disso, inicialmente presou-se pela igualdade na quantidade de exemplos utilizados entre todas as redes e, devido às diferenças nas dimensões, a quantidade de fatias também variava entre os eixos, o que geraria o problema de tamanhos diferentes de treino, validação e teste entre as CNNs responsáveis pelo processamento de cada eixo (um eixo poderia ter mais fatias que outros, o que mudaria a separação dos dados entre os eixos).

Dado tal problema, a solução foi a normalização da base de dados, na qual, para os scans que possuíam dimensões com valores superiores a 256, as fatias excessivas foram retiradas. Analogamente, scans que não possuíam as dimensões com valores iguais a 256 tiveram fatias completamente escuras adicionadas até que todos os eixos se iguallassem dimensionalmente.

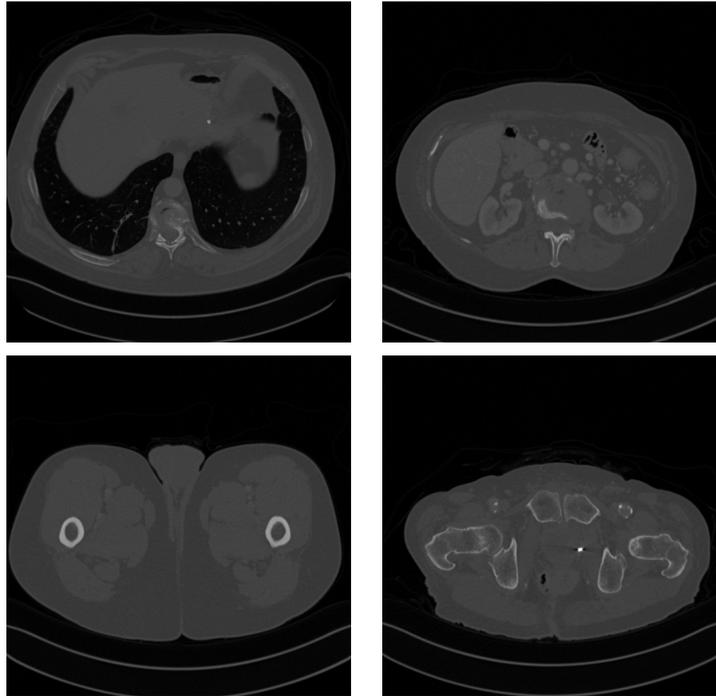
### 5.1.2 Mayo-Challenge

O *dataset* Mayo-Challenge possui *scans* da área torácica de 10 diferentes pacientes, os quais foram obtidos sem redução de radiação. Para a obtenção dos volumes LDCT, foi adicionado antes da reconstrução, ruído de *Poisson*, resultando em reconstruções LDCT (reconstruções

com dosagem normal foram geradas sem o ruído de *Poisson*).

Esse dataset foi utilizado durante o "*Low Dose CT Grand Challenge*" que ocorreu ao longo do ano de 2016 e tinha como objetivo testar o desempenho de diferentes algoritmos no processo de *denoising* dos volumes LDCT. A Figura 17 contém 4 exemplos de fatias de diferentes volumes do Mayo-Challenge.

Figura 17 – Alguns exemplos de tomografias encontradas na base de dados Mayo-Challenge.



Fonte: O Autor (2023).

Assim como o Qure.ai, o Mayo-Challenge possui volumes com quantidades variadas de fatias (apesar de todas as fatias possuírem dimensões 512x512 pixels) então, para que não houvesse problemas no processamento das visões coronal e sagital, foi feito o preenchimento com pixels pretos nas imagens (coronal e sagital) cujas dimensões não possuíam  $mod32 = 0$ .

Com essa base de dados, foi possível testar o método em imagens relativamente maiores que as da Qure.ai, além da diferença na área do corpo (a primeira base de dados com foco na cabeça e pescoço, enquanto a Mayo-Challenge com foco no tórax). Além disso, foi possível também verificar a efetividade do método proposto no problema da redução direta de radiação (que foi simulada na Mayo-challenge).

Para os experimentos desta dissertação, foi utilizada apenas a versão *1mm B30* da Mayo-Challenge, visto que é a versão com maior nível de ruído (MAYO..., 2017).

### 5.1.3 Simulação da aquisição em Low-Dose

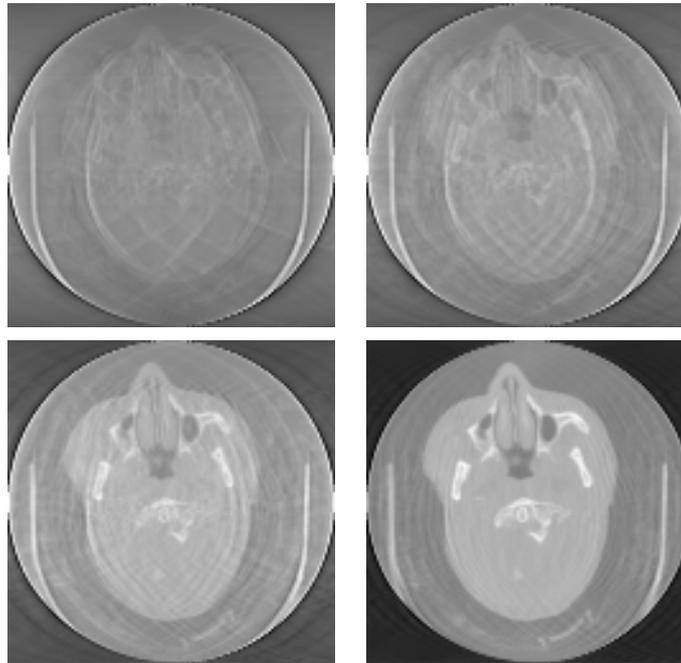
Utilizando as CTs disponíveis na bases de dados Qure.ai como *phantoms*, o ambiente de obtenção de tomografias computadorizadas de feixe cônico (CBCT, do inglês *Cone Beam Computed Tomography*) obtidas com projeções esparsas (*sparse-view*) foi simulado utilizando o ASTRA-toolbox (AARLE et al., 2015). As simulações foram realizadas para a aquisição de volumes com 10, 30, 60 e 90 projeções de raio-X para cada CT, o que representa uma redução de 32, 24, 12 e 8 vezes o número de projeções usados em um scan CBCT feito durante uma radioterapia guiada por imagem (IGRT, do inglês *Image Guided Radiation Therapy*) (HU et al., 2010).

As projeções foram reconstruídas utilizando a técnica SIRT (GORDON; BENDER; HERMAN, 1970) para gerar  $\mathbf{v}_{in}$ . Através do uso da ASTRA-toolbox, as simulações foram feitas em cada eixo (Axial, Coronal e sagital) separadamente. O método mais popular de Retroprojeção filtrada (o FPB) (KAK; SLANEY, 2001) não foi utilizado pois, como dito anteriormente, o SIRT fornece resultados melhores em situações que incluem ruído e projeções esparsas (MU; PARK, 2020).

O banco de dados de 258 LDCT de projeções esparsas  $\mathbf{v}_{in}$  e seus respectivos CTs *phantoms*  $\mathbf{v}_{sd}$  da Qure.ai:  $\{(\mathbf{v}_{in}^1, \mathbf{v}_{sd}^1), (\mathbf{v}_{in}^2, \mathbf{v}_{sd}^2), \dots, (\mathbf{v}_{in}^{258}, \mathbf{v}_{sd}^{258})\}$ , assim como as 10 LDCT da Mayo-Challenge, foram divididos (de forma não estratificada) em três grupos de dados de forma randômica, compostos respectivamente, 40%, 40%, e 20% de toda a base de dados. Os dois primeiros grupos de dados foram utilizados para treinamento da fase *multi-view* e fusão, respectivamente do método proposto, e o terceiro grupo foi usado para teste. A divisão foi feita de forma a igualar a quantidade de dados utilizados no treino de ambas as fases do método. A separação dos dados foi a mesma para todos os experimentos, por exemplo: No caso da Mayo-Challenge, 3 tomos foram utilizados para treinar as 3 redes iniciais, da mesma forma que outros 3 foram utilizados para treinar a rede de fusão (sobrando 2 tomos para teste e 1 para validação em cada etapa).

A Figura 18 mostra exemplos LDCT da Qure.ai, simuladas através do Astra-Toolbox.

Figura 18 – Simulações LDCT da base de dados Qure.ai CQ500. Cada imagem representa (respectivamente, da esquerda pra direita, de cima para baixo) reconstruções com 10, 30, 60 e 90 projeções.

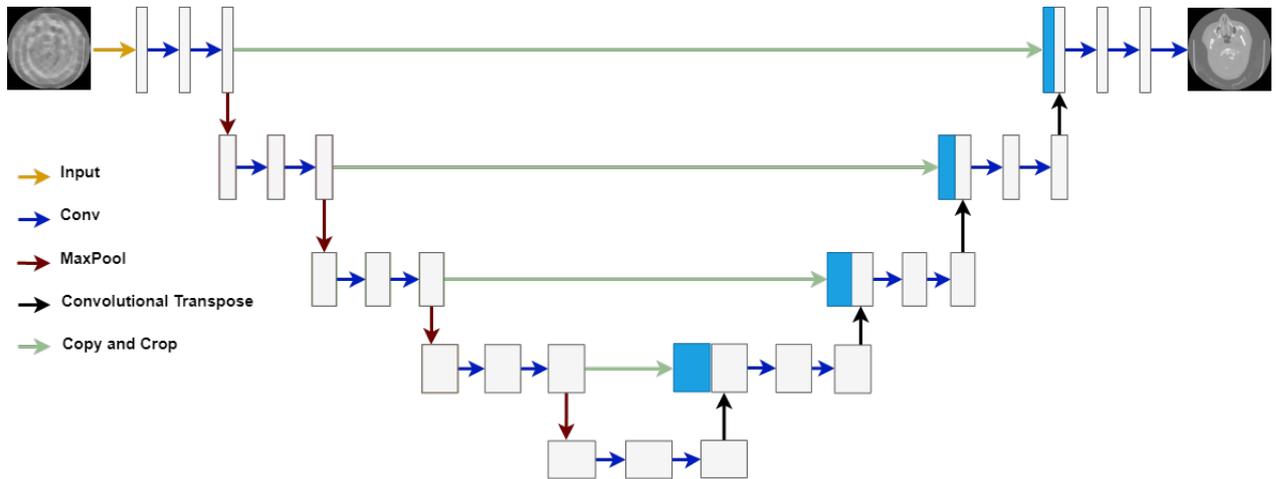


Fonte: O Autor (2023).

## 5.2 ARQUITETURAS

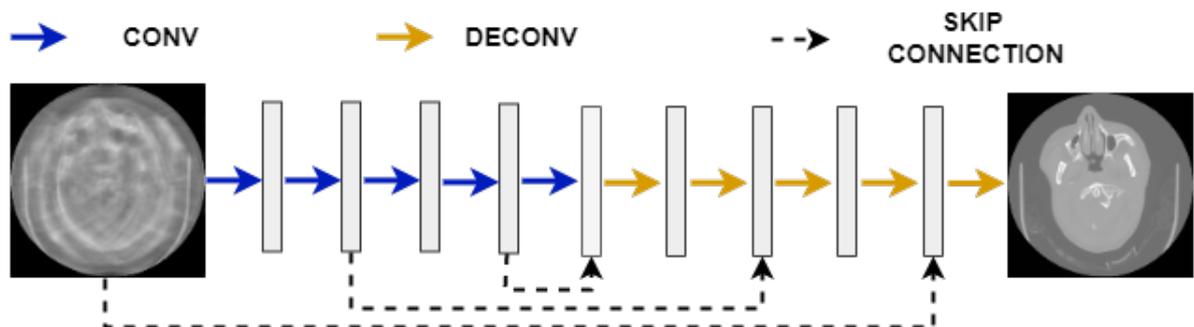
Ao todo, foram utilizadas 3 arquiteturas diferentes, tanto a versão 2D quanto a 3D das CNNs foram utilizadas. Para os experimentos com a base de dados Qure.ai, foram utilizadas as arquiteturas U-net (RONNEBERGER; FISCHER; BROX, 2015) e RED-CNN (Chen; Zhang et al., 2017). Já no caso da Mayo-Challenge, foram utilizadas a RED-CNN e a GRC-GAN (ALMEIDA et al., 2021). A GRC-GAN foi escolhida como forma de testar uma arquitetura mais próxima ao estado da arte (a GRC-GAN foi escolhida após o fim dos experimentos na base Qure.ai, por isso não foi utilizada na primeira rodada de experimentos), ao mesmo tempo que a RED-CNN e U-net foram escolhidas como *baselines* bem estabelecidos da literatura, mas é importante ressaltar que o método possui a premissa de funcionar para a maior parte das arquiteturas que sejam adequadas para a recuperação de LDCTs. As Figuras 19, 20 e 21 mostram a arquitetura geral das redes utilizadas.

Figura 19 – Descrição da arquitetura da U-net utilizada neste trabalho. U-net 2D e 3D diferem-se pelo uso de convoluções 2D e 3D.



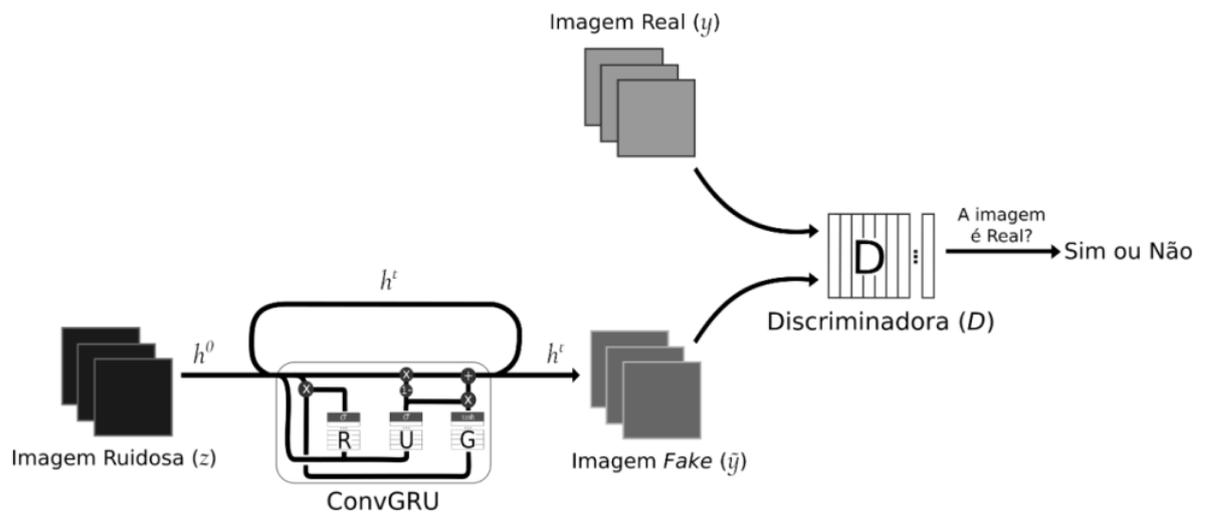
Fonte: O Autor (2023).

Figura 20 – Descrição da arquitetura da RED-CNN utilizada neste trabalho. RED-CNN 2D e 3D diferem-se pelo uso de convoluções 2D e 3D.



Fonte: O Autor (2023).

Figura 21 – Descrição da arquitetura da GRC-GAN utilizada neste trabalho. GRC-GAN 2D e 3D diferem-se pelo uso de convoluções 2D e 3D. Fonte (ALMEIDA et al., 2021) (Imagem cedida pelo autor).



Fonte: (ALMEIDA et al., 2021) (Imagem cedida pelo autor).

### 5.3 AMBIENTE DE TESTES

Para os experimentos, foi utilizada uma máquina com placa gráfica NVidia Geforce 1080 ti com 12GB de VRAM. Todos os experimentos foram feitos em Python, com o uso de bibliotecas como: Tensorflow, Pytorch, Astra-Toolbox, Numpy, Pandas etc. Os hiper-parâmetros foram: taxa de aprendizado de  $1e-5$ , tamanho do *batch* foi de 1 e para as redes 3D o tamanho do *patch* foi de  $32 \times 32 \times 32$  pixels enquanto nas redes 2D não houve o uso de *patches*, ou seja, a fatia inteira foi usada como entrada. A função de perda utilizada em todos os experimentos foi a função de erro quadrático médio (*MSE, Mean Squared Error*, uma vez que não seria possível utilizar funções de perda perceptual para todas as arquiteturas. O número de épocas para os experimentos com a Qure.ai foi 50 e, devido restrições de tempo, para a Mayo-Challenge o número de épocas foi 1. Para avaliar os resultados, foram utilizadas 3 métricas: *Structural similarity index measure* SSIM (WANG et al., 2004), que mede a similaridade entre a reconstrução e a imagem original (*Ground Truth*). O *Peak signal-to-noise ratio* PSNR (GONZALEZ; WOODS; EDDINS, 2008), que investiga a taxa de ruído entre a reconstrução e a imagem original (GT). E a *Normalized root mean squared error* NRMSE (GONZALEZ; WOODS; EDDINS, 2008), que diz a diferença normalizada entre a reconstrução e a imagem original. A seguir, dispõe-se dos experimentos e os resultados.

## 6 EXPERIMENTOS E RESULTADOS

### 6.1 PROJEÇÕES ESPARSAS (QURE.AI)

Para a Qure.AI foram feitas 4 rodadas de experimentos, uma para cada número de projeções de raios-X. A Figura 22 mostra exemplos de LDCTs recuperados com a RED-CNN 2D, RED-CNN 3D e o método proposto. Visualmente, é possível perceber que o método proposto apresenta melhores resultados e, o ganho mais expressivo é perceptível em maiores reduções de projeção. As métricas dispostas abaixo das imagens também confirmam que o método proposto obteve melhor desempenho em todos os cenários, alcançando os melhores valores de SSIM, PSNR e RMSE entre os modelos.

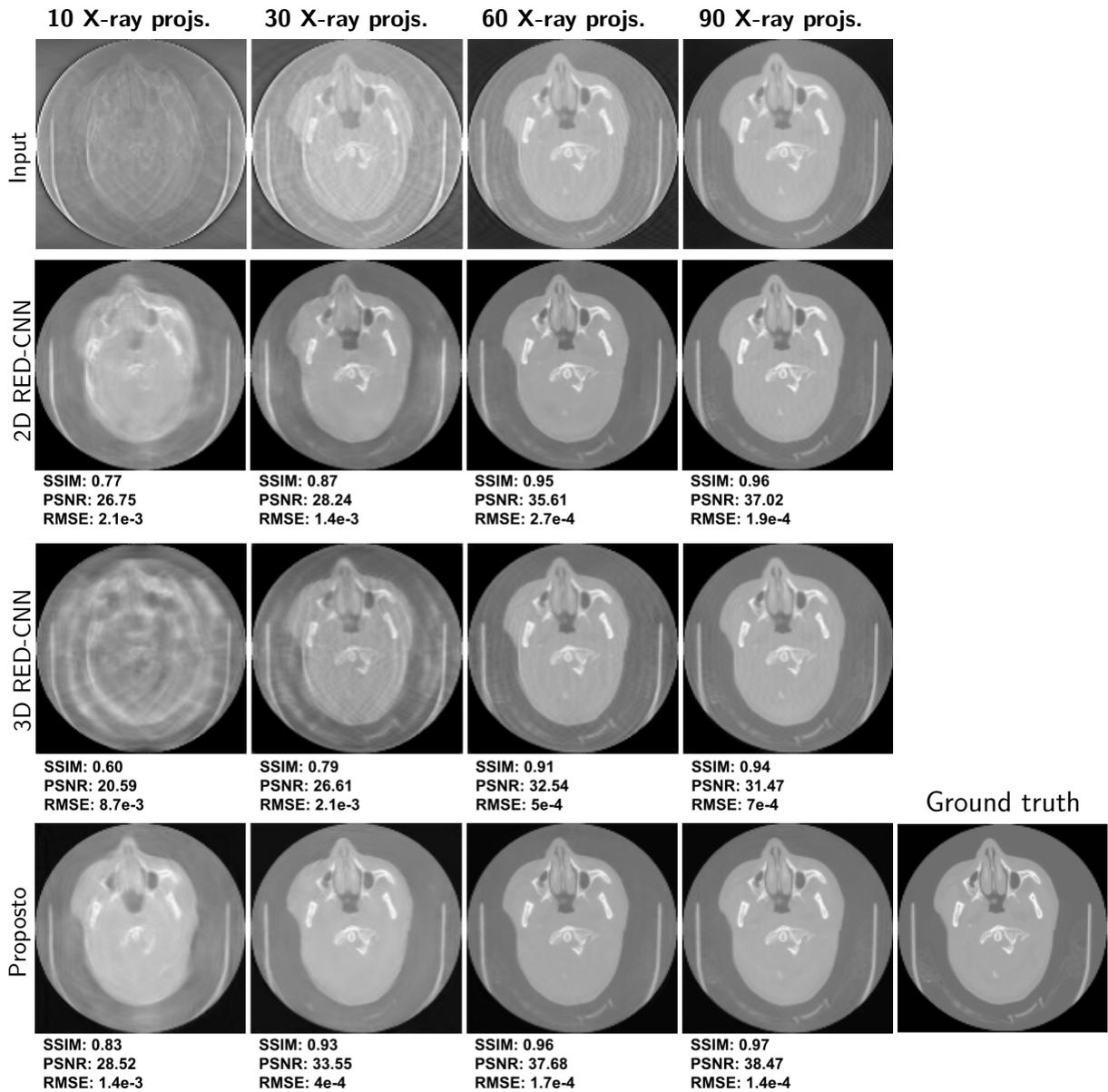
A Figura 23 dispõe dos mesmos exemplos de LDCTs usados no experimento anterior, dessa vez, recuperados com a U-net 2D, U-net 3D e o método proposto. De forma parecida com os anteriormente discutidos, os resultados obtidos nos experimentos com as U-nets também são visualmente bons, porém, não apresentam uma vantagem tão expressiva. Analisando-as quantitativamente, percebe-se que o método proposto obteve as melhores métricas de qualidade em todos os cenários, com exceção dos experimentos com 10 projeções.

As Figura 24 e 25 apresentam, para uma melhor análise qualitativa visual, os mapas de calor dos resultados apresentados anteriormente nas Figuras 22 e 23. Os mapas de calor indicam as zonas de maior e menor concentração de erro, as quais são indicadas com a intensidade da cor; quanto mais claro, maior o erro. Através desses mapas, nota-se que até mesmo em experimentos os quais os resultados dos três métodos não pareciam diferir muito, há uma distinção da concentração e intensidade dos erros, e é ainda mais notável a qualidade da reconstrução do método proposto nos experimentos mais complexos (menores quantidades de projeções).

Para uma análise quantitativa, as Tabelas 1 e 2 apresentam o desempenho dos métodos utilizados nos experimentos. São dispostas as médias e desvio padrão das métricas qualitativas (SSIM, PSNR e RMSE) utilizadas para avaliar a aplicação dos métodos no conjunto de teste (cerca de 53 volumes, ou 20% do total de volumes).

Observando os resultados dispostos nas tabelas 1 e 2, o método proposto obteve desempenho superior aos outros métodos em todos os cenários. Algumas outras conclusões notáveis: a média do SSIM do método proposto aplicado na U-Net para 30 projeções é melhor que a média do SSIM das U-net 2D e 3D referentes a 90 projeções. Já no PSNR, o método proposto

Figura 22 – Resultados gerados pela RED-CNN 2D, RED-CNN 3D, e o método proposto aplicado a RED-CNN para quatro fatias LDCT reconstruídas de 10, 30, 60, e 90 projeções X-ray.

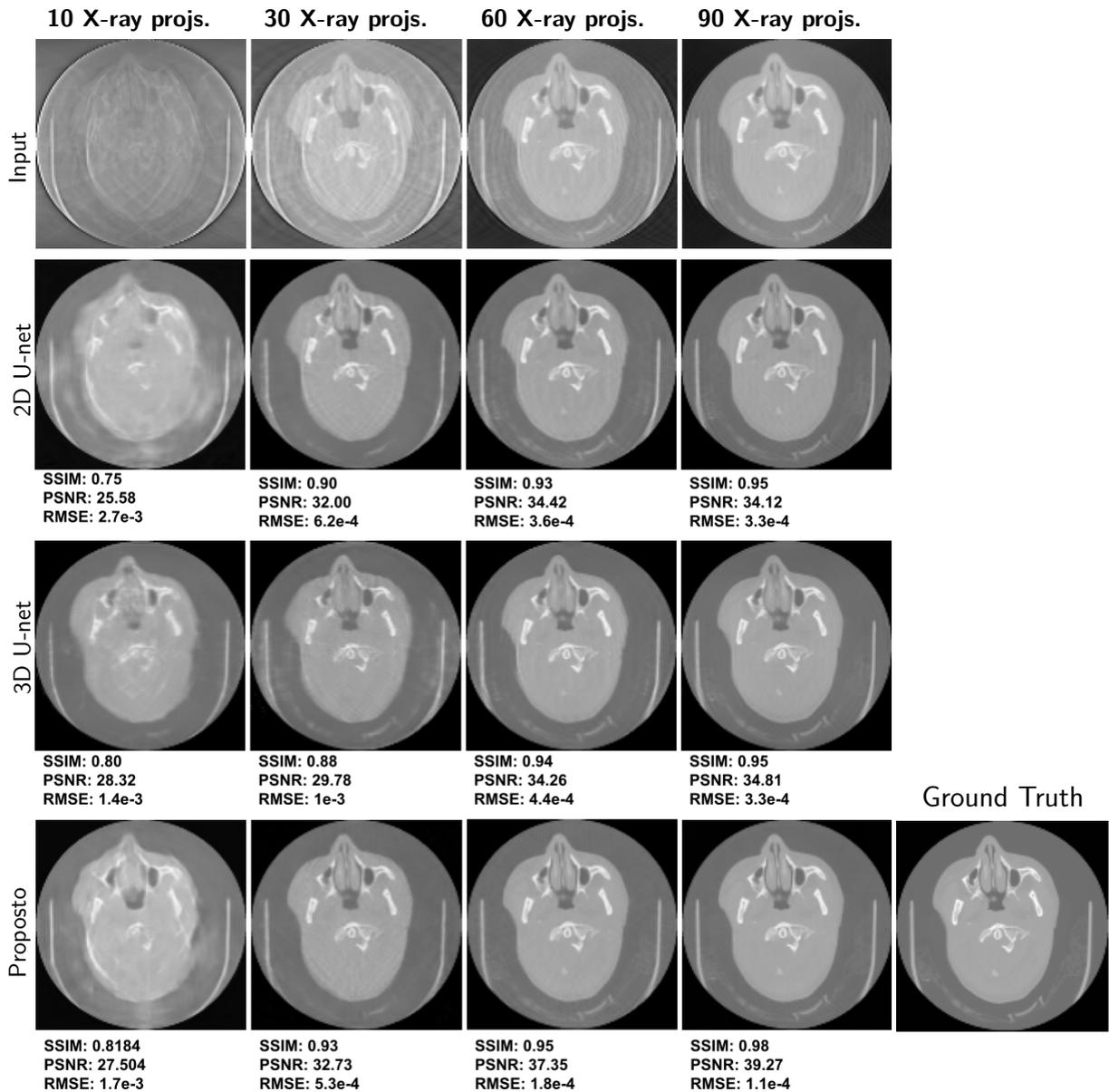


Fonte: O Autor (2023).

aplicado na RED-CNN para o processamento das LDCTs com 60 projeções, obteve média comparável ou maior que aquelas obtidas pelas RED-CNN 2D, 3D e U-Net 2D e 3D. Com relação ao RMSE, o método proposto aplicado a RED-CNN se saiu melhor em LDCTs com 30 projeções, que a RED-CNN 3D aplicada a LDCTs com 90 projeções.

Uma outra observação a ser feita é a superioridade da RED-CNN 2D sobre a RED-CNN 3D. Não foi feito um estudo completo para entender o porquê disso acontecer, porém, um palpite é a quantidade de dados que no caso da arquitetura 3D, é menor em quantidade (por exemplo: para a CNN 2D o volume tomográfico é dividido em fatias, o que acaba gerando

Figura 23 – Resultados gerados pela U-net 2D, U-net 3D, e o método proposto aplicado a U-net para quatro fatias LDCT reconstruídas de 10, 30, 60, and 90 X-ray projeções X-ray.

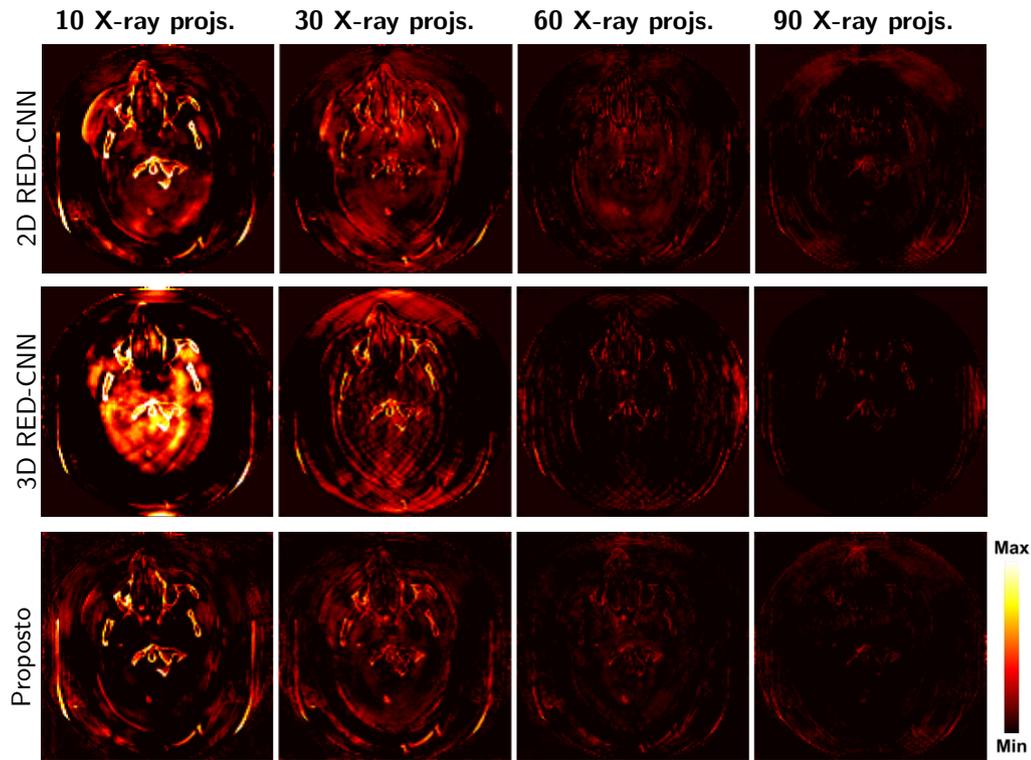


Fonte: O Autor (2023).

uma quantidade maior de entradas, enquanto a arquitetura 3D processa por volumes, que leva a um menor volume de entrada).

Por fim, é interessante notar que a variação das médias de SSIM, PSNR e RMSE, aumenta conforme a quantidade de projeções diminui. O método proposto, no geral, apresenta melhor desempenho quando comparado com os métodos bases, tal vantagem é ainda mais perceptível nos experimentos com menores quantidade de projeções (10 e 30).

Figura 24 – Mapas de calor indicando as zonas com maior erro nos resultados da RED-CNN.



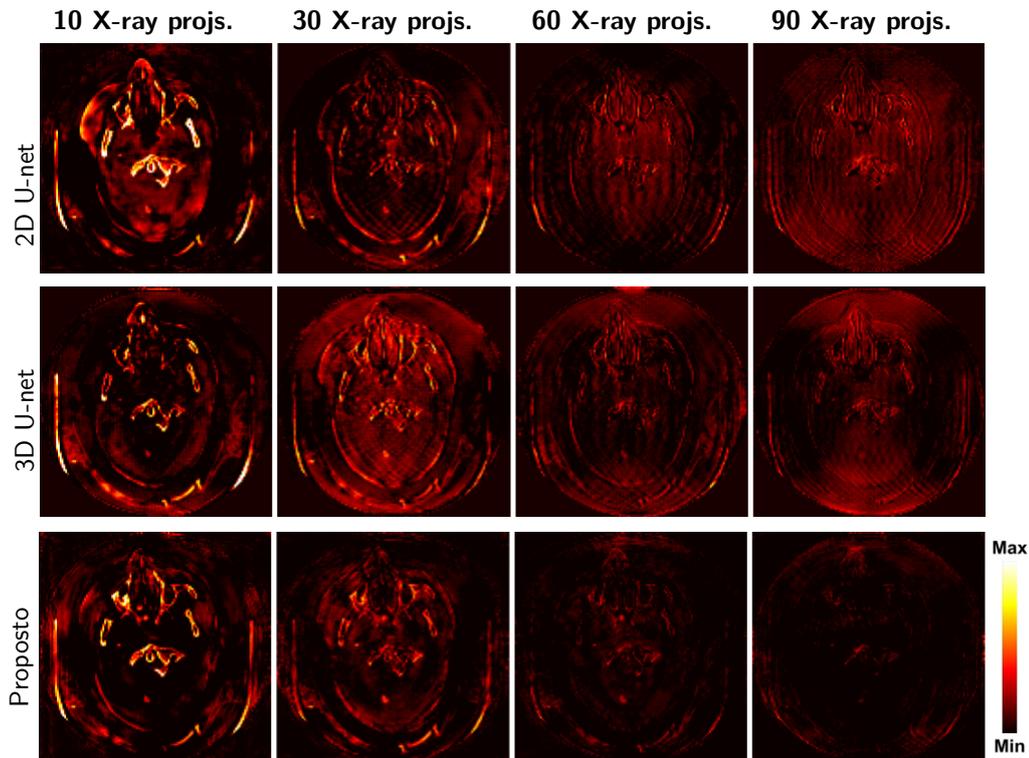
Fonte: O Autor (2023).

Tabela 1 – SSIM, PSNR and RMSE (Média  $\pm$  desvio padrão) obtidos usando os métodos baseados na RED-CNN para supressão de ruídos e artefatos em LDCTs reconstruídos com projeções-esparsas. Resultados referentes ao conjunto de teste. Os melhores resultados estão destacados em negrito.

X-ray Projs.	Método	SSIM	PSNR	RMSE
10	RED-CNN 2D	0,81 $\pm$ 0,07	27,3 $\pm$ 3,04	2,5e <sup>-3</sup> $\pm$ 4e <sup>-3</sup>
	RED-CNN 3D	0,64 $\pm$ 0,09	22,6 $\pm$ 3,2	8e <sup>-3</sup> $\pm$ 1,6e <sup>-2</sup>
	<b>Proposto</b>	<b>0,905 <math>\pm</math> 0,04</b>	<b>30,4 <math>\pm</math> 3,2</b>	<b>1,1e<sup>-3</sup> <math>\pm</math> 1,1e<sup>-3</sup></b>
30	RED-CNN 2D	0,902 $\pm$ 0,05	31,3 $\pm$ 4,6	1,2e <sup>-3</sup> $\pm$ 2e <sup>-3</sup>
	RED-CNN 3D	0,86 $\pm$ 0,06	28,6 $\pm$ 4,4	2,9e <sup>-3</sup> $\pm$ 7e <sup>-3</sup>
	<b>Proposto</b>	<b>0,95 <math>\pm</math> 0,02</b>	<b>34,9 <math>\pm</math> 3,7</b>	<b>4e<sup>-4</sup> <math>\pm</math> 7e<sup>-4</sup></b>
60	RED-CNN 2D	0,95 $\pm$ 0,02	36,7 $\pm$ 5,08	4e <sup>-4</sup> $\pm$ 1e <sup>-3</sup>
	RED-CNN 3D	0,92 $\pm$ 0,05	34,3 $\pm$ 4,5	9e <sup>-4</sup> $\pm$ 3e <sup>-3</sup>
	<b>Proposto</b>	<b>0,97 <math>\pm</math> 0,02</b>	<b>38,3 <math>\pm</math> 3,8</b>	<b>2e<sup>-4</sup> <math>\pm</math> 6e<sup>-4</sup></b>
90	RED-CNN 2D	0,96 $\pm$ 0,02	38,6 $\pm$ 4,7	3e <sup>-4</sup> $\pm$ 9e <sup>-4</sup>
	RED-CNN 3D	0,94 $\pm$ 0,04	36,2 $\pm$ 4,6	6e <sup>-4</sup> $\pm$ 3e <sup>-3</sup>
	<b>Proposto</b>	<b>0,97 <math>\pm</math> 0,01</b>	<b>39,6 <math>\pm</math> 3,7</b>	<b>1e<sup>-4</sup> <math>\pm</math> 4e<sup>-4</sup></b>

Fonte: O Autor (2023).

Figura 25 – Mapas de calor indicando as zonas com maior erro nos resultados da U-net.



Fonte: O Autor (2023).

Tabela 2 – SSIM, PSNR and RMSE (Média  $\pm$  desvio padrão) obtidos usando os métodos baseados na U-net para supressão de ruídos e artefatos em LDCTs reconstruídos com projeções-esparsas. Resultados referentes ao conjunto de teste. Os melhores resultados estão destacados em negrito.

X-ray Projs.	Método	SSIM	PSNR	RMSE
10	U-net 2D	0,83 $\pm$ 0,08	27,3 $\pm$ 3,08	2,6e <sup>-3</sup> $\pm$ 4e <sup>-3</sup>
	U-net 3D	0,90 $\pm$ 0,04	27,6 $\pm$ 3,40	<b>9e<sup>-4</sup> <math>\pm</math> 2e<sup>-3</sup></b>
	<b>Proposto</b>	<b>0,91 <math>\pm</math> 0,05</b>	<b>31,29 <math>\pm</math> 4,05</b>	1e <sup>-3</sup> $\pm$ 1,2e <sup>-3</sup>
30	U-net 2D	0,92 $\pm$ 0,03	32,7 $\pm$ 2,97	7e <sup>-4</sup> $\pm$ 2,8e <sup>-3</sup>
	U-net 3D	0,93 $\pm$ 0,04	31,6 $\pm$ 3,3	7e <sup>-4</sup> $\pm$ 2e <sup>-3</sup>
	<b>Proposto</b>	<b>0,953 <math>\pm</math> 0,022</b>	<b>35,9 <math>\pm</math> 4,02</b>	<b>3e<sup>-4</sup> <math>\pm</math> 2e<sup>-4</sup></b>
60	U-net 2D	0,94 $\pm$ 0,03	35,7 $\pm$ 3,5	4,1e <sup>-4</sup> $\pm$ 2,5e <sup>-3</sup>
	U-net 3D	0,951 $\pm$ 0,033	33,09 $\pm$ 1,44	4,3e <sup>-4</sup> $\pm$ 1,2e <sup>-3</sup>
	<b>Proposto</b>	<b>0,959 <math>\pm</math> 0,022</b>	<b>37,05 <math>\pm</math> 1,9</b>	<b>2,1e<sup>-4</sup> <math>\pm</math> 2e<sup>-4</sup></b>
90	U-net 2D	0,943 $\pm$ 0,02	36,5 $\pm$ 2,6	2,7e <sup>-4</sup> $\pm$ 2,5e <sup>-4</sup>
	U-net 3D	0,95 $\pm$ 0,02	38,7 $\pm$ 2,5	3e <sup>-4</sup> $\pm$ 8e <sup>-4</sup>
	<b>Proposto</b>	<b>0,98 <math>\pm</math> 0,009</b>	<b>39,8 <math>\pm</math> 2,7</b>	<b>1e<sup>-4</sup> <math>\pm</math> 1e<sup>-4</sup></b>

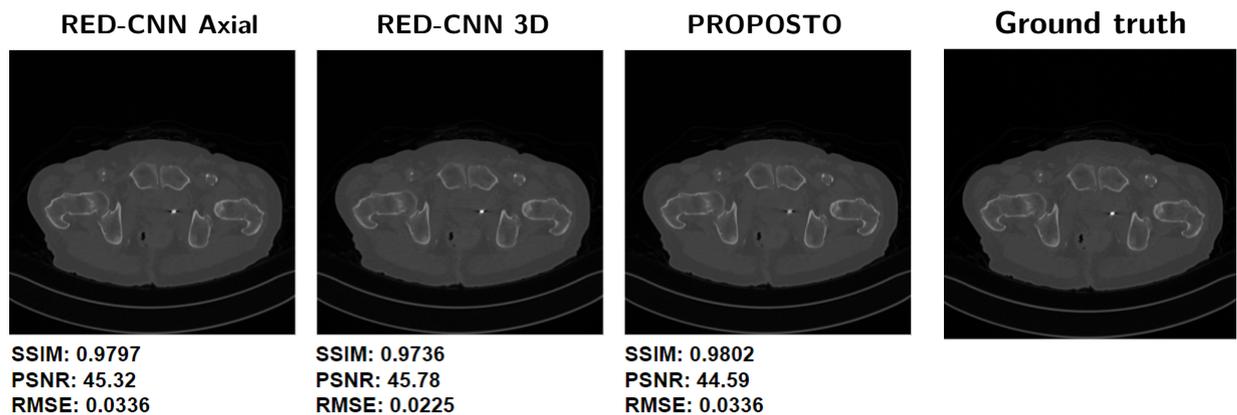
Fonte: O Autor (2023).

## 6.2 RUÍDO DE *POISSON* (MAYO-CHALLENGE)

Os experimentos com a base de dados Mayo-Challenge foram feitos de forma análoga aos experimentos com a Qure.ai, através de treinamento, validação e teste, tanto o método proposto quanto redes 2D e 3D passaram pelo mesmo processo. É importante lembrar que a base de dados Mayo-Challenge possui como diferencial a adição de ruído de *Poisson* que, semelhante ao caso de projeções esparsas, leva a simulação dos efeitos do menor número de fótons detectados durante a aquisição de uma imagem LDCT (através da diminuição da tensão ou dos tubos de aquisição, mas não é especificado em (MAYO..., 2017)).

A primeira arquitetura testada foi a RED-CNN. A Figura 26 mostra o resultado de uma fatia para os três métodos baseados em RED-CNN.

Figura 26 – Resultados gerados pela RED-CNN 2D, RED-CNN 3D, e o método proposto aplicado a RED-CNN para a recuperação de uma fatia obtida com redução de radiação.



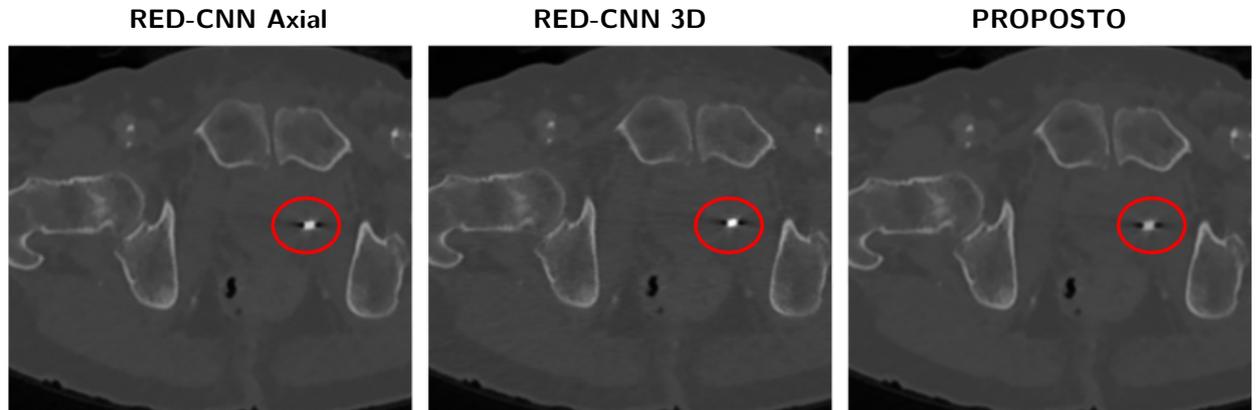
Fonte: O Autor (2023).

Baseado nesse exemplo, o método proposto obteve um SSIM maior que a RED-CNN 2D e a RED-CNN 3D, mas perdeu para a RED-CNN 3D em ambas as outras métricas (PSNR e NRMSE). É importante notar, porém, que todos os três resultados foram visualmente muito semelhantes, o que torna complicado encontrar, a olho nu, diferenças mais severas entre eles.

Através de um zoom, a Figura 27 nos permite analisar algumas sutilezas entre os 3 resultados. É possível perceber que a diferença mais expressiva entre as três saídas está no contraste da região destacada em vermelho. Mesmo com o zoom, ainda é difícil perceber outras regiões com grande distinção. No geral, o zoom nas imagens deixa ainda mais claro a diferença mínima entre a fatia recuperada pelos três métodos. Por fim, a saída resultante dos 3 métodos de recuperação, ficaram bastante fiéis, dificultando uma conclusão sobre qual foi o resultado

superior.

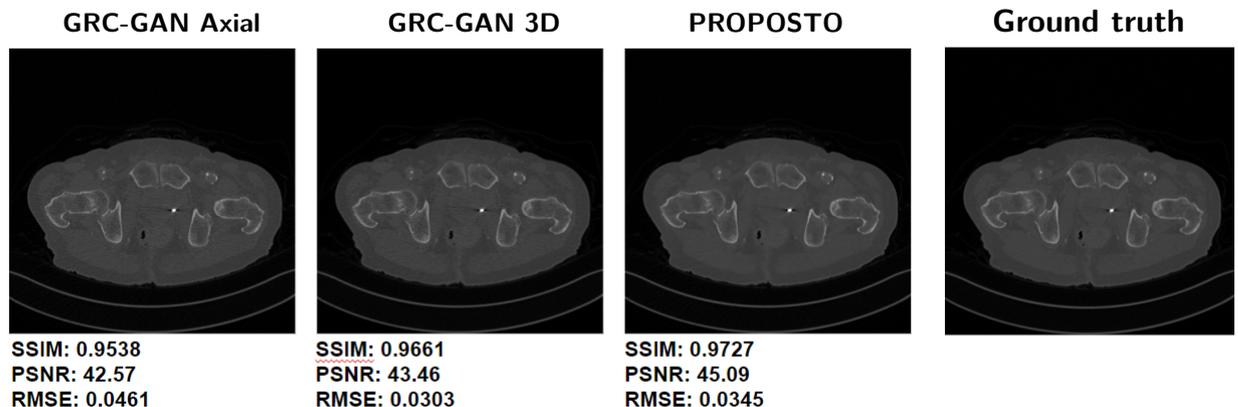
Figura 27 – Zoom das três reconstruções baseadas na RED-CNN. A área circulado em vermelho mostra a região com maior distinção entre as três figuras.



Fonte: O Autor (2023).

Para a segunda rodada de experimentos, foi utilizada a arquitetura GRC-GAN, a qual foi submetida aos mesmos conjuntos de treino e teste da RED-CNN. A Figura 28 dispõe do resultado de uma fatia para os três métodos baseados em GRC-GAN.

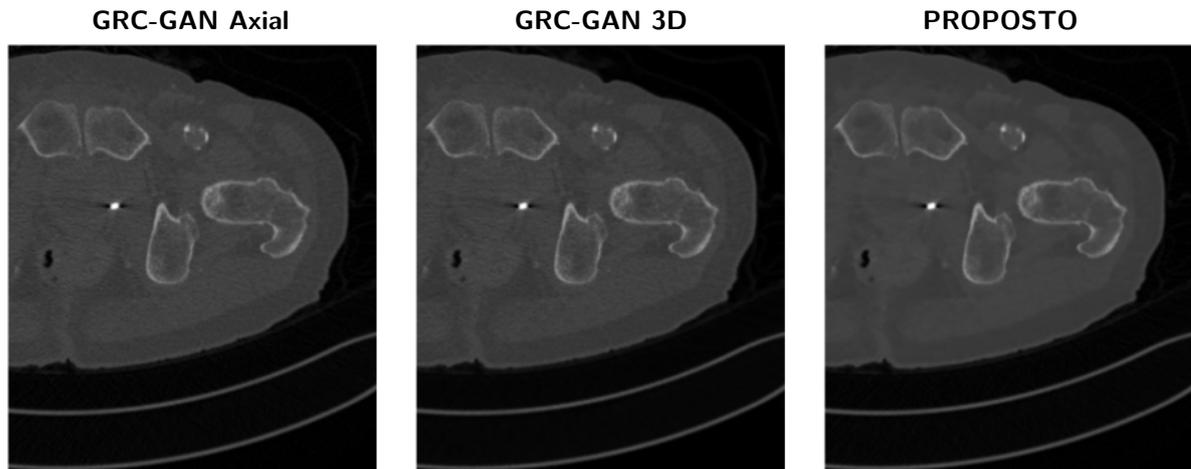
Figura 28 – Resultados gerados pela GRC-GAN 2D, GRC-GAN 3D, e o método proposto aplicado a GRC-GAN para a recuperação de uma fatia obtida com redução de radiação.



Fonte: O Autor (2023).

Através da análise das fatias recuperadas pelos métodos baseados na GRC-GAN, assim como o foi com a RED-CNN, visualmente as 3 saídas resultantes da aplicação dos métodos ficaram muito parecidos, dificultando dizer qual o melhor método apenas por uma visualização direta. Por outro lado, no SSIM e PSNR, o método proposto obteve o melhor resultado, perdendo pra GRC-GAN 3D apenas no NRMSE. Para ajudar na análise, a Figura 29 apresenta um zoom das reconstruções apresentadas anteriormente.

Figura 29 – Zoom das três fatias recuperadas utilizando métodos baseados na GRC-GAN.



Fonte: O Autor (2023).

Utilizando o zoom, é possível perceber que a fatia recuperada com o método proposto possui um aspecto mais "liso", enquanto naquelas resultantes da aplicação dos outros métodos (GRC-GAN 2D e 3D), faz-se presente uma espécie de granulação sutil. Ainda assim, é bem difícil fazer uma distinção elaborada entre as três.

Para uma análise quantitativa geral, a Tabela 3 mostra as médias e desvio padrão do SSIM, PSNR e RMSE para os experimentos baseados nas redes RED-CNN e GRC-GAN.

Tabela 3 – SSIM, PSNR e NRMSE (Média  $\pm$  desvio padrão) obtidos usando os métodos baseados na RED-CNN e GRCGAN para recuperação de LDCTs obtidos com redução direta de radiação. Resultados referentes ao conjunto de teste. Os melhores resultados estão destacados em negrito.

Método	SSIM	PSNR	RMSE
RED-CNN 2D	0,991 $\pm$ 0,003	44,9 $\pm$ 1,64	3e <sup>-1</sup> $\pm$ 6e <sup>-2</sup>
RED-CNN 3D	0,987 $\pm$ 0,06	<b>46,2 <math>\pm</math> 3,27</b>	<b>8e<sup>-3</sup> <math>\pm</math> 2e<sup>-2</sup></b>
<b>Proposto</b>	<b>0,9945 <math>\pm</math> 0,002</b>	46,04 $\pm$ 1,2	2e <sup>-1</sup> $\pm$ 6e <sup>-2</sup>
GRC-GAN 2D	0,957 $\pm$ 0,01	40,2 $\pm$ 0,92	5e <sup>-1</sup> $\pm$ 1e <sup>-1</sup>
GRC-GAN 3D	0,98 $\pm$ 0,06	43,4 $\pm$ 3	<b>2e<sup>-3</sup> <math>\pm</math> 2e<sup>-2</sup></b>
<b>Proposto</b>	<b>0,995 <math>\pm</math> 0,003</b>	<b>45,6 <math>\pm</math> 1,3</b>	3e <sup>-1</sup> $\pm$ 5e <sup>-2</sup>

Fonte: O Autor (2023).

Através das médias e desvio padrão, quando isolamos o SSIM, percebe-se que o método proposto performou melhor que aplicação das redes de forma individual, em ambos os experimentos, obtendo uma pequena vantagem nos experimentos baseados na RED-CNN e uma vantagem mais expressiva naqueles baseados na GRC-GAN. Ambas as versões 3D das arquiteturas

turas foram as melhores no RMSE. Já para o PSNR, o método proposto obteve performance inferior nos experimentos baseados na RED-CNN, onde a versão 3D da rede alcançou melhores resultados porém, nos experimentos com métodos baseados na GRC-GAN, o método proposto obteve uma vantagem bastante expressiva quando comparado ao segundo melhor (novamente a rede 3D).

## 7 CONCLUSÕES

Esta dissertação apresentou uma Combinação de Rede Neurais Convolucionais 2D para reconstrução de Tomografias Computadorizadas com Baixa Dosagem. O método apresentado, buscou explorar a correlação entre pixels de um mesmo volume através do processamento conjunto de suas visões axial, sagital e coronal. A abordagem apresentada, foi desenvolvida com o conceito de divisão em duas fases: Processamento *multi-view*, onde cada visão da tomografia (axial, coronal e sagital) é processada separadamente por uma respectiva CNN, com o objetivo de obter informações complementares no resultado da recuperação de cada uma das visões; Fusão, que busca combinar as fatias recuperadas na fase anterior utilizando uma quarta CNN, onde espera-se que as relações entre as três visões sejam exploradas de forma a obter uma boa reconstrução final.

O método foi testado utilizando duas fases principais de experimentos: A primeira com dados referentes a redução do número de projeções de raio-X, e a segunda referente a aplicação de ruído de *Poisson* antes da reconstrução das tomografias.

Os resultados dos experimentos com a base de dados Qure.ai (recuperação de LDCTs simuladas com projeções esparsas) mostraram que, no geral, o método proposto performa melhor que o uso de apenas uma rede 2D ou 3D, suprimindo quantidades maiores de ruído e distorções referentes às reduções na quantidade de projeções raios-X. Não só isso, a performance do método proposto foi notavelmente maior que os outros métodos a conforme a redução no número de projeções aumentou.

Já na segunda rodada de experimentos, ocorreu a aplicação dos métodos na base de dados Mayo-Challenge (a qual recebeu o ruído de *Poisson*). O método proposto, apesar de obter resultados majoritariamente positivos (perdendo apenas em algumas métricas), não apresentou melhora muito significativa com relação aos métodos 2D e 3D. Na verdade, baseado-se nas médias, desvio padrão e exemplos, é difícil concluir que o método proposto obteve performance melhor que a aplicação individual de uma CNN 2D ou 3D. O motivo pode ser a baixa quantidade de ruído da base Mayo-Challenge, onde a simples aplicação de um método 2D ou 3D já resulta em uma saída bastante limpa, tornando a aplicação de uma segunda etapa (fusão) não tão necessária.

Para os experimentos, foram utilizadas as arquiteturas RED-CNN, U-net e GRC-GAN, o que mostra a flexibilidade do método com relação ao uso de diferentes modelos da literatura,

---

podendo-se estender a outros métodos de processamento de LDCTs.

Apesar dos resultados não tão expressivos dos experimentos referentes a Mayo-Challenge, o método proposto ainda sim obteve resposta majoritariamente positiva e animadora se focarmos principalmente nos resultados obtidos durante os experimentos de recuperação de LDCTs obtidas com projeções esparsas. Além disso, levando em consideração a análise quantitativa das métricas de qualidade (em ambas as bases de dados), pode-se dizer que o método proposto performou em nível igual ou superior a aplicação de apenas uma CNN 2D ou 3D, excluindo os experimentos com a RED-CNN no Mayo-Challenge.

É de suma importância destacar que a primeira fase de experimentos com o método proposto (experimentos com a base de dados Qure.Ai 6.1), resultou na confecção de um *paper* que foi posteriormente apresentado na conferencia IJCNN em 2022 (JUNIOR et al., 2022). A segunda fase de experimentos (feitos com a mayo-challenge 6.2), foi feita após a publicação e apresentação do *paper* original.

Após ambas as rodadas de experimentos, os próximos passos envolvem testar o método em uma base de dados de projeções-esparsas não simuladas, ou aplicar o método em dados que fazem uma combinação de estratégias de redução de radiação, como projeções-esparsas e redução da tensão elétrica nos tubos de raio-X ou a aplicação de filtros. Além disso, a exploração de experimentos utilizando redes recorrentes surge como uma possibilidade interessante. Por fim, é promissor aplicar o método em outros desafios, como a segmentação de LDCTs.

## REFERÊNCIAS

- AARLE, W. V. et al. The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, Elsevier, v. 157, p. 35–47, 2015.
- ALMEIDA, M. B. de; PEREIRA, L. F. A.; REN, T. I.; CAVALCANTI, G. D. C.; SIJBERS, J. The gated recurrent conditional generative adversarial network (grc-gan): application to denoising of low-dose ct images. In: *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.: s.n.], 2021. p. 129–135.
- ANDERSEN, A. H.; KAK, A. C. Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm. *Ultrasonic Imaging*, SAGE Publications Sage UK: London, England, v. 6, n. 1, p. 81–94, 1984.
- BOONE, J. M.; STRAUSS, K. J. *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. [S.l.]: Saunders, 2014.
- BRENNER, J.; HALL, 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.
- Chen, H.; Zhang et al. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, v. 36, n. 12, p. 2524–2535, 2017.
- CHEN, H.; ZHANG, Y.; ZHANG, W.; LIAO, P.; LI, K.; ZHOU, J.; WANG, G. Low-dose ct via convolutional neural network. *Biomed. Opt. Express*, Optica Publishing Group, v. 8, n. 2, p. 679–694, Feb 2017. Disponível em: <<https://opg.optica.org/boe/abstract.cfm?URI=boe-8-2-679>>.
- CHILAMKURTHY, S. et al. Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. *The Lancet*, Elsevier, v. 392, n. 10162, p. 2388–2396, 2018.
- CONG, W. et al. Deep-learning-based breast ct for radiation dose reduction. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Developments in X-Ray Tomography XII*. [S.l.], 2019. v. 11113, p. 111131L.
- GONZÁLEZ, A. D. et al. Projected cancer risks from computed tomographic scans performed in the united states in 2007. *Archives of internal medicine*, American Medical Association, v. 169, n. 22, p. 2071–2077, 2009.
- GONZALEZ, R.; WOODS, R.; EDDINS, S. *Digital Image Processing*. [S.l.]: Pearson Education, 2008.
- GOOD, C. D.; JOHNSRUDE, I. S.; ASHBURNER, J.; HENSON, R. N.; FRISTON, K. J.; FRACKOWIAK, R. S. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, v. 14, n. 1, p. 21–36, 2001. ISSN 1053-8119. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1053811901907864>>.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: The MIT Press, 2016.
- GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. *Advances in neural information processing systems*, v. 27, p. 2672–2680, 2014.

- GORDON, R.; BENDER, R.; HERMAN, T. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *Journal of theoretical Biology*, Academic Press, v. 29, n. 3, p. 471–481, 1970.
- GRAY, E. et al. Estimation of nasal tip support using computer-aided design and 3-dimensional printed models. *JAMA facial plastic surgery*, American Medical Association, v. 18, n. 4, p. 285–291, 2016.
- GU, J.; YE, J. C. Adain-based tunable cyclegan for efficient unsupervised low-dose ct denoising. *IEEE Transactions on Computational Imaging*, v. 7, p. 73–85, 2021.
- GURULINGAPPA, H.; JAKKULA, V.; HU, Y. A new wasserstein distance measure for bernoulli random variables. *Journal of Statistical Theory and Practice*, Taylor Francis, v. 11, n. 2, p. 351–362, 2017.
- HAN, Y.; YE, C. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1418–1429, 2018.
- HEUER, G. et al. Separation of craniopagus twins by a multidisciplinary team. *New England Journal of Medicine*, Mass Medical Soc, v. 380, n. 4, p. 358–364, 2019.
- HU, W. et al. Use of kilovoltage x-ray volume imaging in patient dose calculation for head-and-neck and partial brain radiation therapy. *Radiation Oncology*, Springer, v. 5, n. 1, p. 29, 2010.
- HUANG, L.; JIANG, H.; LI, S.; BAI, Z.; ZHANG, J. Two stage residual cnn for texture denoising and structure enhancement on low dose ct image. *Computer Methods and Programs in Biomedicine*, v. 184, p. 105115, 2020. ISSN 0169-2607. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169260719301038>>.
- HUBBARD, B. B. *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. USA: A. K. Peters, Ltd., 1996. ISBN 1568810474.
- HUGHES, S. Ct scanning in archaeology. In: \_\_\_\_\_. [S.l.: s.n.], 2011. ISBN 978-953-307-723-9.
- HUMPHRIES, T. et al. Comparison of deep learning approaches to low dose ct using low intensity and sparse view data. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2019: Physics of Medical Imaging*. [S.l.], 2019. v. 10948, p. 109484A.
- IBRAGIMOV, B. et al. Combining deep learning with anatomical analysis for segmentation of the portal vein for liver sbrrt planning. *Physics in medicine and biology*, ePublish, v. 23, p. 8943–8958, 2017.
- JAMALUDDIN, F. N.; AHMAD, S. A.; NOOR, S. B. M.; HASSAN, W. Z. W.; YAACOB, A.; ADAM, Y. Performance of dwt and swt in muscle fatigue detection. In: *2015 IEEE Student Symposium in Biomedical Engineering & Sciences (ISSBES)*. [S.l.: s.n.], 2015. p. 50–53.
- JOHNSON, J.; ALAHI, A.; FEI-FEI, L. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*. 2016.
- JUNIOR, C. A. A.; PEREIRA, L. F. A.; CAVALCANTI, G. D. C.; REN, T. I. Ensemble of convolutional neural networks for sparse-view cone-beam computed tomography. In: *2022 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2022. p. 1–7.

- KAK, A. C.; SLANEY, M. *Principles of Computerized Tomographic Imaging*. [S.l.]: Society of Industrial and Applied Mathematics (SIAM), 2001.
- KANG, E.; KOO, H. J.; YANG, D. H.; SEO, J. B.; YE, J. C. Cycle-consistent adversarial denoising network for multiphase coronary ct angiography. *Medical Physics*, v. 46, n. 2, p. 550–562, 2019. Disponível em: <<https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13284>>.
- KUSHIBAR, K.; VALVERDE, S.; GONZÁLEZ-VILLÀ, S.; BERNAL, J.; CABEZAS, M.; OLIVER, A.; LLADO, X. Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. 09 2017.
- LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time-series. In: \_\_\_\_\_. *The handbook of brain theory and neural networks*. [S.l.]: MIT Press, 1995.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LESSMANN, N. et al. Automatic calcium scoring in low-dose chest ct using deep neural networks with dilated convolutions. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 2, p. 615–625, 2018.
- LIGUORI, C. et al. Emerging clinical applications of computed tomography. *Medical devices (Auckland, N.Z.)*, US National Library of Medicine National Institutes of Health, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4467659/>>.
- LIU, K.; KANG, G. Multiview convolutional neural networks for lung nodule classification. *International Journal of Imaging Systems and Technology*, v. 27, p. 12–22, 03 2017.
- MA, J.; ZHU, S.; ARCE, G. R. Sparse-view computed tomography: The role of prior image-constrained compressed sensing in medical imaging. *IEEE Signal Processing Magazine*, 2015.
- MATSUMOTO, S.; MORRIS, M.; ROSE, S. 3-dimensional printed anatomic models as planning aids in complex oncology surgery. *JAMA oncology*, American Medical Association, v. 2, n. 9, p. 1121–1122, 2016.
- MAYO Clinic Low Dose CT Grand Challenge. 2017. <<https://www.aapm.org/GrandChallenge/LowDoseCT/>>. Acesso em: 08 de Junho de 2023.
- MCCOLLOUGH, C. H.; BUSHBERG, J. T.; FLETCHER, J. G.; ECKEL, L. J. Answers to common questions about the use and safety of ct scans. *Mayo Clinic Proceedings*, v. 90, n. 10, p. 1380–1392, 2015.
- MCCOLLOUGH, C. H.; LENG, S.; YU, L. Ct dose reduction and dose management tools: overview of available options. *Radiographics*, Radiological Society of North America, v. 35, n. 4, p. 1189–1204, 2015.
- MORTAZI, A. et al. Cardiacnet: Segmentation of left atrium and proximal pulmonary veins from mri using multi-view cnn. In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2017*. Cham: Springer International Publishing, 2017. p. 377–385. ISBN 978-3-319-66185-8.

- MU, C.; PARK, C. Sparse filtered sirt for electron tomography. *Pattern Recognition*, Elsevier, v. 102, p. 107253, 2020.
- NANNI, L.; COSTA, Y.; AGUIAR, R. et al. Ensemble of convolutional neural networks to improve animal audio classification. 2020.
- NOOTHOUT, M. et al. Automatic segmentation of thoracic aorta segments in low-dose chest ct. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2018: Image Processing*. [S.l.], 2018. v. 10574, p. 105741S.
- PEARCE, M. et al. Radiation exposure from ct scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *The Lancet*, 2012. Disponível em: <[https://doi.org/10.1016/S0140-6736\(12\)60815-0](https://doi.org/10.1016/S0140-6736(12)60815-0)>.
- POWER, S. et al. Computed tomography and patient risk: Facts, perceptions and uncertainties. *World Journal of Radiology*, US National Library of Medicine National Institutes of Health, 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5183924/>>.
- REFAAT, M. M.; SALEM, M. E. Radiation dose and image quality optimization in ct imaging. In: *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*. [S.l.]: IGI Global, 2018. p. 165–193.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015.
- ROTH, H. R.; LU, L.; SEFF, A.; CHERRY, K. M.; HOFFMAN, J.; WANG, S.; LIU, J.; TURKBEBY, E.; SUMMERS, R. M. A new 2.5d representation for lymph node detection using random sets of deep convolutional neural network observations. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2014*. Springer International Publishing, 2014. p. 520–527. Disponível em: <[https://doi.org/10.1007/978-3-319-10404-1\\_65](https://doi.org/10.1007/978-3-319-10404-1_65)>.
- SHAO, Y.-H. et al. Exposure to Tomographic Scans and Cancer Risks. *JNCI Cancer Spectrum*, v. 4, n. 1, 11 2019. Pkz072. Disponível em: <<https://doi.org/10.1093/jncics/pkz072>>.
- SINGH, S.; KALRA, M. K.; GILMAN, M. D.; HSIEH, J.; PIEN, H. H.; DIGUMARTHY, S. R.; SHEPARD, J.-A. O. Adaptive statistical iterative reconstruction technique for radiation dose reduction in chest ct: a pilot study. *Radiology*, Radiological Society of North America, v. 259, n. 2, p. 565–573, 2011.
- SINGH, S. P.; WANG, L.; GUPTA, S.; GOLI, H.; PADMANABHAN, P.; GULYÁS, B. 3d deep learning on medical images: A review. *Sensors*, v. 20, n. 18, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/18/5097>>.
- WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. *Image Quality Assessment: From Error Visibility to Structural Similarity*. [S.l.]: Cambridge University Press, 2004.
- YANAGA, Y.; AWAI, K.; NAKAURA, T.; SASAKI, Y. Reduced tube voltage and current techniques in ct. *Japanese Journal of Radiology*, Springer, v. 36, n. 3, p. 161–169, 2018.
- YANG, Q.; YAN, P.; ZHANG, Y.; YU, H.; SHI, Y.; MOU, X.; KALRA, M. K.; ZHANG, Y.; SUN, L.; WANG, G. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, v. 37, n. 6, p. 1348–1357, 2018.

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. *Dive into Deep Learning*. 2023.

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.

ZHENG, S.; CORNELISSEN, L. J.; CUI, X.; JING, X.; VELDHUIS, R. N. J.; OUDKERK, M.; OOIJEN, P. M. A. van. Deep convolutional neural networks for multiplanar lung nodule detection: Improvement in small nodule identification. *Medical Physics*, v. 48, n. 2, p. 733–744, 2021. Disponível em: <<https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14648>>.

ZHOU, Y.; ZENG, Z.; QI, S.; XIAO, J.; GUAN, S.; ZHANG, Y. Evaluation of cnn approaches for imbalanced medical image classification. In: PMLR. *International Conference on Artificial Intelligence and Statistics*. [S.l.], 2018.