Victor Rogerio Sousa Ferreira

Geração de tomografia cardíaca com contraste sintético a partir de imagens de tomografia cardíacas sem contraste utilizando modelos de difusão

São Luís - MA

Abril de 2024

Victor Rogerio Sousa Ferreira

Universidade Federal do Maranhão – UFMA

Centro de Ciências Exatas e Tecnologias

Departamento de Engenharia Elétrica

Programa de Pós-Graduação em Engenharia Elétrica

Geração de tomografia cardíaca com contraste sintético a partir de imagens de tomografia cardíacas sem contraste utilizando modelos de difusão

Dissertação de mestrado apresentada ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal do Maranhão na área de Ciência da Computação, como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

São Luís - MA

Abril de 2024

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Sousa ferreira, Victor rogerio.
Geração de tomografia cardíaca com contraste sintético a partir de imagens de tomografia cardíacas sem contraste utilizando modelos de difusão / Victor rogerio Sousa ferreira. - 2024.
64 f.

Coorientador(a) 1: Francesco Renna.
Orientador(a): Anselmo Cardoso de paiva.
Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, Universidade federal do maranhão - ufma, 2024.

1. Modelos de difusão. 2. Tradução de imagens. 3. Redes adversárias. 4. U-net. 5. . I. Cardoso de paiva, Anselmo. II. Renna, Francesco. III. Título.

Victor Rogerio Sousa Ferreira

Geração de tomografia cardíaca com contraste sintético a partir de imagens de tomografia cardíacas sem contraste utilizando modelos de difusão

Dissertação de mestrado apresentada ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal do Maranhão na área de Ciência da Computação, como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

Prof. Dr. Anselmo Cardoso de Paiva
Orientador

Prof. Dr. Francesco Renna
Coorientador

Prof. Dr. João Manuel Pedrosa
Universidade do Porto
Banca Examinadora

Prof. Dr. Aristófanes Correa Silva
Universidade Federal do Maranhão
Banca Examinadora

São Luís - MA
Abril de 2024

Este trabalho é dedicado à minha família.

Agradecimentos

Primeiramente, agradeço a Deus pela minha vida, saúde, oportunidades e principalmente pelos pais que Ele me deu. Gostaria de agradecer aos meus pais, José Rogério e Lindalva Sousa Ferreira, pela educação e por tudo que fizeram por mim. Agradeço-lhes também pelos valores e princípios que me ensinaram e por serem exemplos de como eu gostaria de ser na vida. Quero agradecer à minha irmã Victoria, que sempre me ajudou e esteve presente no que ela pôde durante esta jornada. Quero agradecer também aos meus tios, tias, primos e primas, em especial a minha tia Lúcia e minha tia Maria de Lourdes, que sempre estiverem comigo e foram, em muitos momentos, uma segunda mãe pra mim. E, por fim, quero agradecer à minha namorada, Hervyla Arouche, que tem sido minha companheira nestes últimos 6 anos, e, junto com minha família, é um porto seguro e foi de suma importância para minha saúde física e mental.

Agradeço também aos professores da Universidade Federal do Maranhão (UFMA) pelos conhecimentos transmitidos por eles. Um agradecimento especial ao Professor Anselmo pela oportunidade e confiança concedidas a mim, além da paciência que teve para que eu pudesse participar de seu laboratório e adquirir experiências de valor inestimável. Bem como ao Professor Francesco Renna da Universidade do Porto.

Agradeço à FAPEMA, CAPES e CNPQ pelos recursos cedidos que foram utilizados para o desenvolvimento desta dissertação. E também agradeço ao Núcleo de Computação Aplicada pelo espaço e materiais que foram disponibilizados para me auxiliar.

No laboratório, conheci e agradeço pessoas que foram essenciais para o meu desenvolvimento como aluno. São eles: Alan de Carvalho, Marcus Vinícius, Gabriel Costa, Alison Mendes, Daniel Lima, Saulo Rodrigues, Mário Freitas, Pedro Bernhard, João Pedro, Estephane Mendes, Ítalo Ferreira, Wesley Kelson e Alexandre Pessoa. Foram responsáveis por um conhecimento que não poderia adquirir de outra forma.

Não posso esquecer de agradecer aos meus amigos próximos. Alguns compartilharam experiências comigo desde os primórdios do mestrado e graduação, e outros foram se agregando no caminho, mas todos me ajudaram durante este período. Luiz Eduardo, Andrews Corrêa, Mauro de Oliveira, Kalynne Cardoso, Lyêssa Viana, João Vitor Gaspar, Camila de Carvalho, Rafaela Vitória, Thaylana Coimbra, Beneilton Leite, Ester Ramos e Tomaz Antonio.

Se você não for feliz agora
Alguém vai acabar sendo feliz
Com a sua felicidade por aí
(Daniel Furlan)

Resumo

Esta dissertação propõe o uso de um modelo de difusão adversária baseado em aprendizagem profunda para abordar a tradução de imagens de tomografia computadorizada (TC) sem contraste do coração em imagens sintéticas compatíveis com imagens adquiridas com injeção de contraste. O trabalho investiga desafios na tradução de imagens médicas combinando conceitos de redes adversárias generativas (GANs) e modelos de difusão. Os resultados foram avaliados por meio de algumas métricas como: Relação sinal-ruído de pico (PSNR), índice de similaridade estrutural (SSIM), Frechet Inception Distance (FID) e Raiz do Erro Médio Quadrado (RMSE) para demonstrar o desempenho do modelo na geração de imagens com contraste sintético, preservando a qualidade e a similaridade visual. O modelo proposto obteve os seguintes melhores resultados, sendo que para PSNR = 32,85, SSIM = 0,766 e FID = 42,348. O CyTran obteve o melhor RMSE, de 0,14, enquanto o modelo teve o pior de 0,2. Também é apresentada uma comparação dos resultados obtidos com as redes CyTran, Cycle-GAN e Pix2Pix. Embora os resultados obtidos sejam promissores, a análise do RMSE indica que ainda existem desafios a serem superados, destacando a necessidade de melhorias contínuas. A intersecção de GANs e modelos de difusão promete avanços futuros, contribuindo significativamente para a prática clínica.

Palavras-chave: Modelos de Difusão. Tradução de Imagens. Redes Adversárias. U-Net.

Abstract

This dissertation proposes the use of a deep learning-based adversarial diffusion model to address the translation of contrast-free computed tomography (CT) images of the heart into synthetic images compatible with images acquired with contrast injection. The work investigates challenges in medical image translation by combining concepts from generative adversarial networks (GANs) and diffusion models. The results were evaluated through several statistics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Frechet Inception Distance (FID), and Root Mean Square Error (RMSE) to demonstrate the model's performance in generating images with synthetic contrast while preserving quality and visual similarity. The proposed model achieved the following best results: PSNR = 32.85, SSIM = 0.766, and FID = 42.348. CyTran obtained the best RMSE of 0.14, while the model had the worst of 0.2. A comparison of the results obtained with CyTran, Cycle-GAN, and Pix2Pix networks is also presented. Although the obtained results are promising, the analysis of RMSE indicates that there are still challenges to be overcome, highlighting the need for continuous improvements. The intersection of GANs and diffusion models promises future advancements, significantly contributing to clinical practice.

Keywords: Diffusion Models. Image Translation. Adversarial Networks. U-Net.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ANN	Redes Neurais Artificais (do inglês, *Artificial Neural Network*)
CACS	Escore de Calcificação Arterial Aoronariana (do inglês, *Coronary Arterie Calcification Score*)
CCTA	Angiografia Coronariana (do inglês, *Coronary Computed Tomography Angiography*)
CCTA	Rede Neural Convolucional (do inglês, *Convolutional Neural Network*)
CECT	Tomografia Computadorizada (do inglês, *Contrast-Enchanced Cardiac Computed tomography*)
NECT	Tomografia Computadorizada (do inglês, *Non-Enchanced Cardiac Computed tomography*)
CT	Tomografia Computadorizada (do inglês, *Computed Tomography*)
DAC	Doenças Arterial Coronariana
DIC	Doenças Isquêmica do Coração
HU	Unidades Hounsfield (do inglês, *Hounsfield Units*)
IA	Inteligência Artificial
ICC	Coeficiente de Correlação Intraclasse (do inglês, *Intraclass Correlation coefficient*)
ML	Aprendizado de Máquina (do inglês, *Machine Learning*)
MSE	Erro Médio Quadrado (do inglês, *Mean Square Error*)
RMSE	Raiz do Erro Médio Quadrado (do inglês, *Root Mean Square Error*)
RELU	Unidade Linear Retificadora (do inglês, *Rectified Linear Unit*)
SSIM	índice de Medida da Similaridade Estrutural (do inglês, *Structural Similarity Index Measure*)
PSNR	Relação Sinal-Ruído de Pico (do inglês, *Peak Signal-to-Noise Ratio*)
FID	Distância de Frechet-Inception (do inglês, *Frechet Inception Distance*)
MAE	Erro Médio Absoluto (do inglês, *Mean Absolute Error*)

Sumário

1 Introdução

As doenças não transmissíveis (DNT) são condições médicas que não podem ser transmitidas de uma pessoa para outra e frequentemente resultam de uma combinação de fatores comportamentais, fisiológicos, ambientais e genéticos. Exemplos comuns de DNT incluem doenças cardiovasculares, como a doença arterial coronariana, insuficiência cardíaca, e cardiomiopatias, bem como doenças pulmonares, como a doença pulmonar obstrutiva crônica (DPOC) e a asma (PAHO, 2023). Estas condições têm um impacto significativo no sistema respiratório e cardiovascular, levando a alta morbidade e mortalidade quando não diagnosticadas e tratadas adequadamente.

Segundo a Organização Mundial da Saúde (OMS) (WHO, 2023), as DNT são responsáveis por 41 milhões de mortes por ano, representando 74% de todas as mortes globais. As doenças cardiovasculares são a principal causa dessas mortes, com 17,9 milhões de óbitos anuais, ressaltando a necessidade urgente de estratégias eficazes de prevenção, diagnóstico e tratamento (DONDI et al., 2021). Além de resultarem em mortalidade significativa, as doenças cardiovasculares causam incapacidade prolongada, impactando a qualidade de vida dos pacientes e gerando elevados custos econômicos e sociais.

O rastreamento de doenças através de técnicas de diagnóstico por imagem, como a tomografia computadorizada (TC) (CORBALLIS et al., 2023) (COUNSELLER; ABOEL-KASSEM, 2023), tem se mostrado importante na detecção precoce de doenças cardiovasculares e pulmonares. A TC, especialmente quando utilizada com contraste, permite a visualização detalhada das estruturas anatômicas, identificando precocemente patologias cardíacas e vasculares. Esta modalidade de imagem fornece informações valiosas sobre a anatomia e a função do coração, possibilitando intervenções terapêuticas oportunas.

No entanto, o uso de contraste em exames de TC não é isento de riscos, podendo causar reações adversas em alguns pacientes, que variam de leves a potencialmente fatais (COMPUTADORIZADA, 2007). A injeção de contraste iodado pode resultar em reações alérgicas, nefropatia induzida por contraste, entre outras complicações (BECKETT; MORIARITY; LANGER, 2015). Por isso, pacientes com alergia ao contraste iodado ou com insuficiência renal geralmente não podem fazer uma TC com contraste. Para aqueles com alergias, a TC com contraste apresenta risco de reações adversas. Da mesma forma, em pacientes com insuficiência renal, o contraste pode agravar a função renal (STACUL et al., 2011).

Por outro lado, uma TC com contraste é indicada para pacientes que não têm histórico de alergia ao contraste iodado e cuja função renal está adequada. O uso de contraste é especialmente útil na avaliação de tumores e infecções, pois melhora a vi-

sualização e a delineação da extensão das lesões. Estudos de vasos sanguíneos, como angiotomografia, dependem do contraste para identificar condições como aneurismas e tromboses. O contraste também é essencial para o detalhamento de órgãos e tecidos moles, permitindo uma distinção mais clara entre diferentes estruturas anatômicas e patologias que não seriam tão evidentes sem ele (JÚNIOR; YAMASHITA, 2001).

Dadas essas qualidades do exame de TC com contraste, destaca-se a necessidade de desenvolver alternativas que mantenham a qualidade diagnóstica das imagens enquanto minimizam os riscos associados ao uso de contraste. Neste contexto, a inteligência artificial (IA) tem se mostrado uma ferramenta promissora. Inovações na criação e interpretação de imagens médicas, impulsionadas pela IA, estão aprimorando a qualidade das imagens de contraste, aumentando a precisão diagnóstica e reduzindo o tempo necessário para análise.

Por exemplo, as Redes Generativas Adversárias (GANs) têm demonstrado capacidade de criar imagens em domínios específicos, como o aprimoramento de contraste em imagens médicas (SKANDARANI; JODOIN; LALANDE, 2023). Estas redes empregam um mecanismo adversário no qual um discriminador orienta um gerador para geração da imagem alvo. Além disso, os modelos de difusão, que caracterizam explicitamente pela geração de imagem através de ruído, têm mostrado sucesso na geração de imagens de alta qualidade (ROMBACH et al., 2022).

A aplicação da inteligência artificial (IA) na TC cardíaca oferece um grande potencial para melhorar o diagnóstico e o prognóstico. Esse exame de imagem possui características únicas que a tornam atraente para a aplicação de IA, como imagens detalhadas do coração e dos vasos sanguíneos, permitindo uma avaliação abrangente da anatomia cardíaca e da presença de quaisquer anomalias ou doenças.

No entanto, a complexidade da aplicação de IA na TC cardíaca está em ascensão devido a diversos fatores. A crescente quantidade de dados provenientes dos exames de TC cardíaca pode sobrecarregar os sistemas tradicionais de análise. Além disso, a interpretação das imagens requer um entendimento profundo da anatomia cardíaca, apresentando desafios para os algoritmos de IA. Garantir a precisão e confiabilidade dos resultados gerados pela IA, principalmente em decisões clínicas cruciais como diagnósticos de doenças cardíacas, é outro desafio importante.

Este trabalho justifica-se pela necessidade de desenvolver métodos que utilizem aprendizado profundo e facilitem a geração de imagens de TC sem contraste (NECT, do inglês *Non-Enchanced Computed Tomography*) para a TC cardíaca com contraste (CECT, do inglês *Contrast Enchanced Computed Tomography*), reduzindo os riscos associados ao uso de contraste e melhorando a precisão diagnóstica. O objetivo é explorar e implementar técnicas de IA para aprimorar a detecção precoce de doenças cardiovasculares, contribuindo para a redução da morbimortalidade associada a essas condições.

A integração de técnicas de redes adversárias com modelos de difusão podem trazer uma nova abordagem para o diagnóstico e o tratamento das doenças cardiovasculares, proporcionando uma ferramenta poderosa para a melhoria da saúde pública.

1.1 Justificativa

A tomografia computadorizada (TC) é uma técnica amplamente utilizada na identificação e avaliação de doenças cardiovasculares devido à sua capacidade de fornecer imagens detalhadas em alta resolução. Com a TC, é possível visualizar com precisão as estruturas cardíacas e vasculares, identificar obstruções, calcular o cálcio nas artérias coronárias e avaliar a extensão de danos após um evento cardiovascular, como um infarto.

Os exames com contraste são particularmente importantes na avaliação de doenças cardiovasculares, pois permitem uma visualização mais nítida e detalhada das estruturas vasculares e cardíacas. No entanto, é importante reconhecer que esses contrastes não estão isentos de riscos, sendo os efeitos adversos um deles. Por exemplo, o contraste iodado, comumente usado em exames como a angiografia, pode causar reações alérgicas e, em casos extremos, até mesmo danos renais.

Com o avanço da inteligência artificial (IA), temos testemunhado uma revolução na criação de imagens médicas. A IA pode ser empregada para aprimorar a qualidade das imagens de contraste, aumentando a precisão e reduzindo o tempo necessário para análise.

Além disso, modelos de difusão (ROMBACH et al., 2022) e Redes Generativa Adversárias(GANs, do inglês *Generative Adversarial Networks*) (KAZEMINIA et al., 2020) têm se destacado na criação de imagens de qualidade em domínios diferentes. Os modelos de difusão, por exemplo, são eficazes na geração de imagens realistas e detalhadas, enquanto as GANs são capazes de criar imagens em domínios específicos, como aprimoramento de contraste em imagens médicas. Essas tecnologias têm o potencial de aprimorar ainda mais a capacidade diagnóstica dos exames de imagem, permitindo uma detecção mais precoce e precisa de doenças cardiovasculares, o que é crucial para salvar vidas e melhorar a qualidade de vida dos pacientes.

Neste contexto, o presente trabalho tem por justificativa e consequente objetivo o desenvolvimento de um método para geração de imagens CECT utilizando aprendizado profundo.

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo desta dissertação é contribuir para o diagnóstico precoce de doenças não transmissíveis (DNT), particularmente as detectadas e diagnosticadas através de tomografia computadorizada cardíaca, propondo um método baseado em modelos de difusão para a geração de imagens sintéticas similares às tomografia computadorizada cardíaca com contraste (CECT), a partir de imagens de tomografia computadorizada cardíaca sem contraste (NECT).

1.2.2 Objetivos Específicos

Para atingir o objetivo principal deste trabalho, foi necessário atingir os seguintes objetivos específicos:

- Realizar análises comparativas entre arquiteturas de rede utilizadas e outras utilizadas no âmbito da geração de imagens em imagens médicas;
- Efetuar uma análise comparativa entre o método proposto e métodos utilizados na literatura, além de pontuar pontos fortes e limitações.
- Realizar a avaliação com medidas intrínsecas do método proposto.

1.3 Organização do trabalho

A organização desse trabalho está dividida em seis capítulos, sendo:

Capítulo 2 - Trabalhos Relacionados - capítulo dedicado a apresentar os trabalhos relacionados ao tema investigado, descrevendo seus métodos e resultados obtidos.

Capítulo 3 - Fundamentação Teórica - neste capítulo, é explicado e abordado trabalhos relacionados ao tema, além dos conceitos necessários para a compreensão dos resultados e da metodologia. Nele é explicado os conceito, exames de tomografia computadorizada, inteligência artificial, aprendizado de máquina, aprendizado profundo, redes adversárias, redes generativas e modelos de difusão.

Capítulo 4 - Materiais e método - neste capítulo, são descritos os passos necessários para desenvolver e implementar o método de geração de imagem de contraste, além de explicar o funcionamento e o desenvolvimento de cada passo.

Capítulo 5 - Resultados e Discussão - neste capítulo, são apresentados os resultados obtidos mediante a implementação do método, além de comparações com outras arquiteturas e outros métodos

Capítulo 6 - Conclusão - capítulo dedicado a apresentar as considerações finais sobre o método proposto, além de descrever possíveis trabalhos futuros.

2 Trabalhos Relacionados

Este capítulo apresenta os trabalhos relacionados utilizados como referência para o desenvolvimento deste estudo. Os trabalhos aqui mencionados dizem respeito ao processo de geração de imagens de tomografia computadorizadas utilizando redes adversárias e modelos de difusão.

A introdução das Redes Adversárias Generativas (GANs), conforme proposto por Goodfellow et al. (2014), trouxe uma arquitetura inovadora para modelos generativos, baseada no princípio da rivalidade entre duas redes - o gerador e o discriminador. Nessa arquitetura, o gerador tem como objetivo produzir dados sintéticos indistinguíveis dos dados reais, enquanto o discriminador se esforça para diferenciar entre os dois. Através do treinamento adversário, as GANs alcançam o equilíbrio de Nash, convergindo a distribuição do gerador para os dados de treinamento, ou seja, as amostras geradas pelo gerador serão indistinguíveis das amostras reais presentes nos dados de treinamento, conforme avaliado pelo discriminador durante o treinamento adversário.

Na tradução de imagens, especialmente na análise de imagens médicas, as GANs são amplamente utilizadas por sua capacidade de aprender automaticamente padrões nos dados de entrada, de modo que o modelo possa gerar novos exemplos (saída) que poderiam existir no conjunto de dados original. Ao realizar a geração de imagens, o modelo mais simples mapeia de origem para destino através de um gerador treinado usando perda adversária (GOODFELLOW et al., 2014). Consequentemente, a abordagem de tradução baseada em GANs tem sido extensivamente adotada em várias aplicações.

As GANs condicionais se destacam na mapeamento de uma única fonte para um destino, melhorando a sensibilidade aos detalhes de alta frequência na estrutura tecidual em comparação com perdas tradicionais pixel a pixel. A integração de termos de perda adversaria tem se mostrado eficaz em aprimorar a precisão espacial e o realismo em imagens alvo geradas com GANs, superando modelos convolucionais convencionais.

Outros estudos, especificamente usando GANs, cujo foco principal está na síntese de imagens de tomografia computadorizada com contraste (CECT) a partir de varreduras de TC sem contraste (NECT), são Chun et al. (2022) e Seo et al. (2021). Eles empregam um *framework* de duas etapas baseado no processo padrão das GANs, que inclui a geração de imagens e a discriminação dessas imagens.

Neste contexto, as arquiteturas de rede sofisticadas utilizadas como geradores para os trabalhos citados anteriormente são a DenseNet de Huang et al. (2017), empregada por Chun et al. (2022) e a SPADE de Park et al. (2019), utilizada por Seo et al. (2021). Na primeira etapa, o gerador cria imagens de CECT sintéticas a partir das varreduras de

NECT, enquanto na segunda etapa, o discriminador avalia a autenticidade dessas imagens sintéticas, fornecendo *feedback* ao gerador para melhorar sua capacidade de gerar imagens indistinguíveis das reais.

No estudo de Chun et al. (2022), foram preparados pares de varreduras cardíacas NCT-CECT de 59 pacientes. Enquanto os valores médios (± desvio padrão) do erro absoluto médio (MAE), pico de relação sinal-ruído (PSNR) e índice de similaridade estrutural (SSIM) entre SCECT e CECT foram 20,66 ± 5,29, 21,57 ± 1,85, e 0,77 ± 0,06, respectivamente, esses valores foram 23,95 ± 6,98, 20,67 ± 2,34, e 0,76 ± 0,07 entre NCT e CECT.

Por outro lado, Seo et al. (2021) utilizaram a arquitetura SPADE para gerar imagens de CECT sintéticas. Eles treinaram sua rede com o conjunto de dados *IXI brain MRI*, que é um conjunto de dados de imagens de ressonância magnética do crânio, e um conjunto de dados para imagem do TC do abdômen. Para avaliar a as imagens geradas, esse estudo utiliza métricas como erro quadrático médio (MSE) e índice de similaridade estrutural de multi-escala (MS-SSIM) de, respectivamente, 0.271 e 0.949 para o *IXI brain MRI*, enquanto o conjunto de TC do abdômen obteve 0,097 e 0,559 nas mesmas métricas.

O modelo Pix2Pix em Zhu et al. (2020) é uma das abordagens projetadas para tarefas de tradução de imagem para imagem. Ele consiste em um gerador para criar imagens sintéticas e um discriminador para distinguir entre imagens reais e geradas. O treinamento envolve um processo adversário, onde o gerador tenta enganar o discriminador, e o discriminador busca identificar imagens falsas. Aplicações da arquitetura Pix2Pix, como a de Choi et al. (2021), utilizam a estrutura fundamental do modelo original pix2pix para gerar imagens sintéticas com contraste a partir de tomografias computadorizadas torácicas sem contraste, com a distinção de que as camadas convolucionais 2D são substituídas por suas equivalentes 3D. Este modelo compreende redes generativas e discriminadoras semelhantes a uma GAN convencional. A rede generativa é uma rede neural convolucional U-Net codificador-decodificador com conexões de salto. A rede discriminadora é um PatchGAN que classifica cada *patch* de pixel como real ou falso, e seu módulo convolucional é idêntico ao bloco de codificador do gerador.

A dissertação de Domingues (2022) compara o desempenho de dois modelos GANs, Pix2Pix-GAN e Cycle-GAN, na geração de imagens com contraste a partir de varreduras de TC sem contraste. O estudo explora as compensações de usar entradas 2D, 2.5D e 3D, empregando diferentes tipos de geradores e conjuntos de dados. O modelo 2D opera em fatias individuais, e um método foi desenvolvido para a criação das entradas 2.5D e 3D com apenas três imagens 2D sendo utilizadas para cada imagem 3D devido a limitações de recursos .A criação de imagens 3D envolveu a combinação da imagem atual com as imagens anteriores e posteriores. As métricas da dissertação do autor inclue índice de Medida da Similaridade Estrutural (SSIM), Razão Sinal-Ruído Máxima (PSNR), Erro

Quadrático Médio (MSE) e métrica de Dice para fidelidade em regiões de alto contraste.

O CyTran de Ristea et al. (2023) é um modelo baseado em GAN projetado para trabalhar com imagens de TC. Esta abordagem inovadora concentra-se na tradução bidirecional de tomografias computadorizadas (TC) com e sem contraste, mesmo quando as imagens não possuem emparelhamento direto. O CyTran visa abordar o desafio de gerar varreduras com contraste para pacientes que não podem receber contraste e para melhorar o alinhamento entre as varreduras de TC com e sem contraste. O método emprega uma arquitetura consistente em ciclo baseada em transformadores generativos adversárias projetados para transferir varreduras de TC entre diferentes fases de contraste. Inspirado no *framework* Cycle-GAN, o CyTran compreende dois discriminadores e dois geradores, permitindo o treinamento em imagens não pareadas por meio de uma perda de consistência de ciclo de vários níveis. Além de garantir consistência em nível de imagem, o CyTran utiliza perdas adicionais entre representações de características intermediárias para aprimorar ainda mais o desempenho do modelo. Esta estratégia abrangente contribui para a eficácia do modelo na tradução bidirecional de imagens de TC, oferecendo aplicações valiosas em cenários de imagiologia médica.

No entanto, as GANs apresentam seus desafios. Questões como menor confiabilidade na mapeamento de uma única amostra, convergência prematura do discriminador e pobre diversidade representacional levando ao colapso do modo podem comprometer a qualidade e a diversidade de amostras geradas (DHARIWAL; NICHOL, 2021). Apesar desses desafios, as GANs atualmente lideram em tarefas de geração de imagens, superando outros modelos com base em métricas como Índice de *Inception* e Precisão.

Recentemente, os modelos de difusão profunda se tornaram uma alternativa às GANs em tarefas de modelagem generativa em visão computacional (JING et al., 2019). Esses modelos são inspirados na termodinâmica não equilibrada, definindo uma cadeia de Markov de etapas de difusão para adicionar ruído aleatório aos dados lentamente e depois aprender a reverter o processo de difusão para construir amostras de dados desejadas a partir do ruído. A remoção de ruído é realizada por uma arquitetura de rede neural treinada para maximizar a correlação entre pixels adjacentes.

A técnica de difusão proporciona maior confiabilidade no mapeamento de rede e aprimora a qualidade e a diversidade das amostras geradas. Estruturado em dois pontos, o modelo de difusão opera por meio de um processo direto e reverso. No processo direto, os dados de entrada são gradualmente perturbados ao longo de várias etapas, adicionando ruído gaussiano. Já na etapa reversa, o modelo é treinado para recuperar os dados originais, revertendo o processo de difusão passo a passo. Essa abordagem inovadora oferece uma solução robusta e eficaz para gerar dados realistas em diversos contextos de visão computacional (HO; JAIN; ABBEEL, 2020).

Além disso, os modelos de difusão são facilmente adaptáveis, capazes de usar

diferentes arquiteturas como *Transformers* em Peebles e Xie (2023) e redes adversárias em Wang et al. (2022), alcançando resultados que superam a qualidade de modelos de difusão anteriores em métricas como razão pico sinal-ruído (PSNR, do inglês Peak Signal-to-Noise Ratio), comumente usada para avaliar a qualidade de uma imagem comprimida em relação à sua versão original, e SSIM para medir a similaridade entre duas imagens.

No trabalho de Azarfar et al. (2023), os autores apresentam vários trabalhos propondo arquiteturas de aprendizado profundo para reduzir ou eliminar o uso de meios de contraste administrados para adquirir tomografias computadorizadas clinicamente úteis.

Na Tabela 2.1, pode-se ver um resumo dos trabalhos relacionados mencionados neste capítulo, apontando os autores e as arquiteturas utilizadas.

Tabela 2.1 – Comparação entre arquiteturas de GANs

Artigos	Vantagens
(CHOI et al., 2021)	Bom em imagens torácicas
(SEO et al., 2021)	Bom em vários *datasets*
(CHUN et al., 2022)	Alta precisão gerando CECT
(DOMINGUES, 2022)	Análise detalhada do desempenho de diferentes arquiteturas para a geração de imagens
(RISTEA et al., 2023)	Tradução bidirecional de imagens de TC

Fonte: Autoral.

3 Fundamentação Teórica

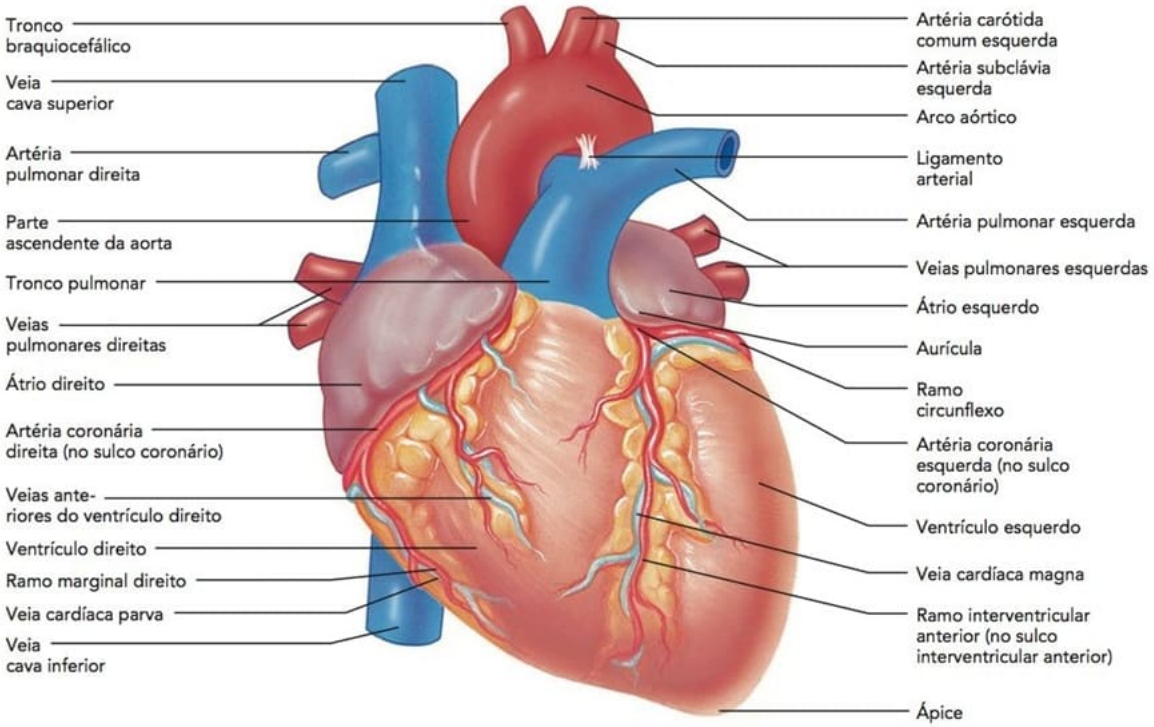
Este capítulo apresenta os conceitos relacionados a doença arterial coronariana, uso de aprendizado profundo e modelos de difusão para criação de imagens medicas, necessários para o bom entendimento do trabalho desenvolvido.

3.1 Doença Arterial Coronariana

As doenças cardiovasculares são a principal causa de morte no mundo, segundo a Organização Mundial da Saúde (OMS). Entre elas está a doença arterial coronariana (DAC), que é a causa de 120 mil mortes no Brasil por ano, segundo estimativa da Sociedade de Cardiologia do Estado de São Paulo (SOCESP, 2021). Além da alta morbidade, a DAC custa alto para os sistemas de saúde. Para um adulto de 40 anos de idade, o risco de desenvolver essa condição durante a vida é de 49% para homens e 32% para mulheres, de acordo com dados do estudo de Framinhgam (2021).

Figura 3.1 do coração com suas estratificações

Figura 3.1 – Artérias coronárias.



Fonte: Adaptado de(MATTOS, 2024).

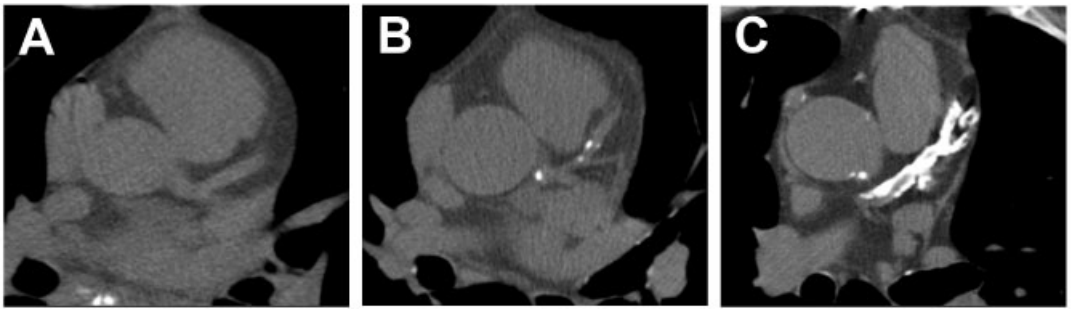
A Doença Arterial Coronariana (DAC) é uma condição resultante da obstrução gradual ou súbita das artérias coronárias por placas de gordura, processo conhecido como

aterosclerose. Essa obstrução reduz o fluxo sanguíneo nas artérias responsáveis por fornecer nutrientes e oxigênio ao músculo cardíaco (miocárdio), podendo levar a sintomas como angina (dor no peito) e, em casos mais graves, infarto do miocárdio (ataque cardíaco) e parada cardíaca (PINHO et al., 2010).

Um método não invasivo utilizado para avaliar a presença e extensão de calcificações nas artérias coronárias é o escore de cálcio coronário, realizado por tomografia computadorizada sem contraste. A presença de calcificações é frequentemente associada à existência de aterosclerose coronária (FERNANDES; BITTENCOURT, 2017).

A Figura 3.2 ilustra o escore de cálcio coronário de três pacientes com diferentes graus de calcificação na artéria descendente anterior: ausência de calcificação (A), calcificação leve (B) e calcificação acentuada (C). Esta representação visual auxilia na compreensão da progressão da doença e na interpretação dos resultados do escore de cálcio coronário.

Figura 3.2 – Exemplo de escore de cálcio coronariano



Fonte: (AZEVEDO; ROCHITTE; LIMA, 2012)

A tomografia computadorizada (TC) é uma técnica amplamente utilizada para diagnosticar uma variedade de condições médicas, incluindo a DAC. Desenvolvida para aplicações clínicas no início da década de 70, essa técnica oferece uma visão não invasiva das estruturas internas do corpo (JÚNIOR; YAMASHITA, 2001). Seu princípio é enviar raios-X ou radiação ionizante através de uma área de interesse, como as artérias coronárias, e medir a atenuação dos raios-X à medida que atravessam os tecidos. A imagem resultante aparece mais brilhante quando a atenuação é alta e mais escura quando é baixa (DOMINGUES, 2022).

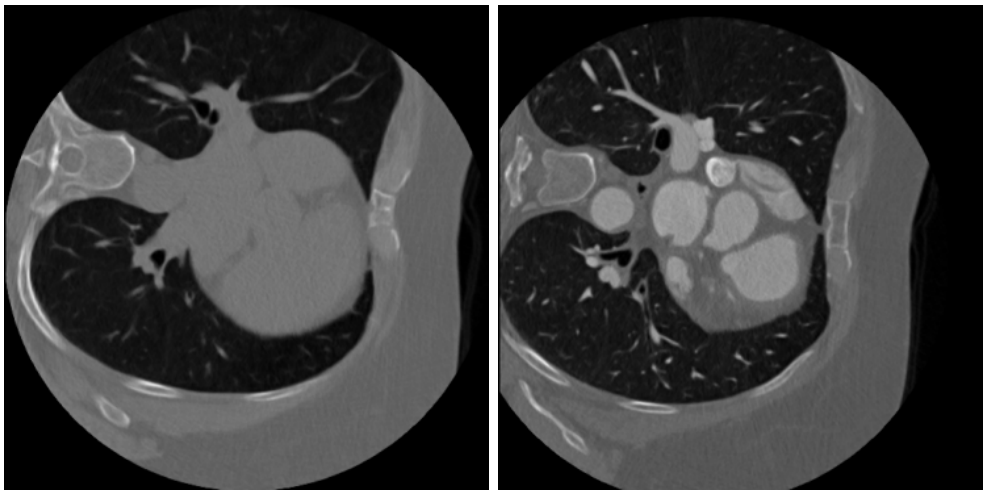
Um importante uso da TC no contexto da DAC é a quantificação do cálcio nas artérias coronárias. Os exames de cálcio, obtidos por meio da TC, são amplamente empregados para detectar a presença de calcificação nessas artérias. A presença de cálcio está associada a um risco aumentado de eventos cardíacos, como ataques cardíacos, indicando um bloqueio significativo das artérias (NABEL; BRAUNWALD, 2012).

A evolução tecnológica na TC, incluindo a técnica de varredura espiral (ou helicoidal), trouxe avanços significativos na rapidez e na capacidade de adquirir imagens detalhadas das artérias coronárias. Essa técnica permite realizar exames em um curto

período de tempo, o que é especialmente útil para pacientes que podem estar agitados ou intolerantes a procedimentos mais demorados (JÚNIOR; YAMASHITA, 2001).

A angiografia coronariana por cateter, considerada o padrão-ouro para avaliação da DAC, é outro método que pode ser empregado em conjunto com a TC. Nessa técnica, um cateter é inserido nas partes próximas das artérias coronárias e o fluido de contraste radiopaco é então injetado. Isso facilita a visualização das artérias coronárias e permite observar o fluxo sanguíneo para detectar obstruções com mais facilidade (BARSNESS; HOLMES, 2011). No entanto, é importante notar que o uso do contraste pode desencadear efeitos colaterais como calor no corpo, leve aceleração dos batimentos cardíacos, náuseas, vômitos e, em casos raros, alergias graves (COMPUTADORIZADA, 2007).

Figura 3.3 – Exemplo de imagem adquirida por TC, sem(a) e com(b) contraste do mesmo paciente



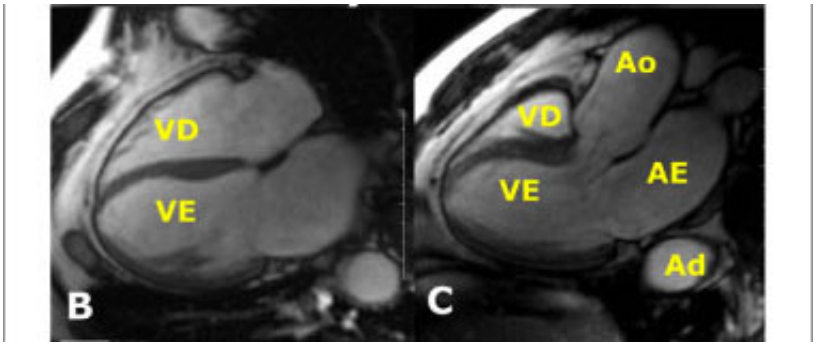
(a) (b)

Fonte: Adaptado de (WOLTERINK, 2016).

Outro método de diagnóstico por imagem que não invasivo é a ressonância magnética (RM), muito utilizado na avaliação da condição de órgãos e estruturas moles, como seios, vasos sanguíneos e encéfalo (HAGE; IWASAKI, 2009). Esse método utiliza ondas de radiofrequência para obter informações dos íons de hidrogênio, sem radiação ionizante. Os núcleos dos átomos de hidrogênio, quando expostos a um campo magnético, alinham-se e são temporariamente estimulados por ondas de radiofrequência, refletindo a energia recebida. Estes sinais são captados e transformados em imagens. Essas imagens são usadas para avaliar aspectos cardiovasculares como anatomia, função ventricular, detecção de infarto, viabilidade, isquemia e fluxo sanguíneo (SARA et al., 2014).

A RM é valiosa na análises de fluxo e angiografias, fornecendo informações detalhadas sobre o fluxo sanguíneo nas artérias coronárias e outras estruturas cardíacas, incluindo a velocidade, direção e possíveis obstruções (SARA et al., 2014).

Figura 3.4 – Exemplo de imagem de ressonância magnética



Fonte: (HAGE; IWASAKI, 2009)

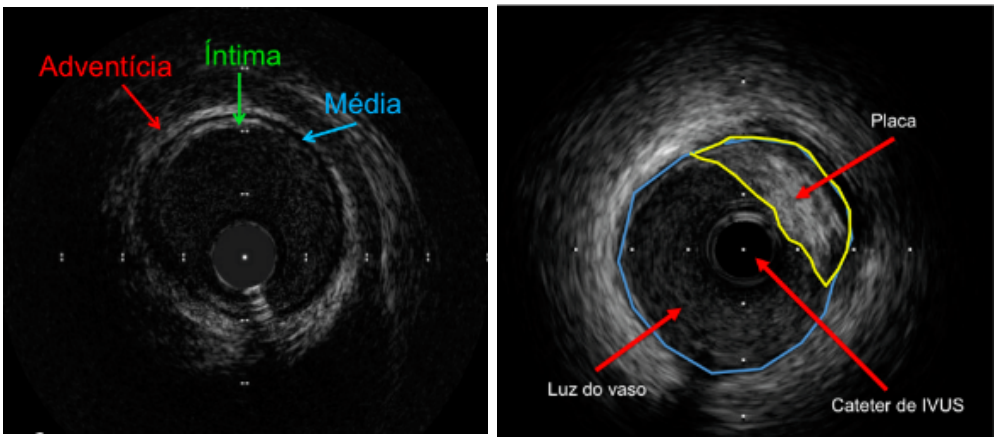
Outra técnica comumente usada para visualização de tecidos moles e estruturas do corpo é a ultrassonografia (USG). Assim como outros exames mencionados, a ultrassonografia é não invasiva e envolve o uso de um transdutor colocado sobre a região do corpo a ser examinada. Esse transdutor emite ondas sonoras de alta frequência que penetram nos tecidos do corpo e, ao encontrar diferentes densidades, geram ecos. Esses ecos são então capturados pelo transdutor e convertidos eletronicamente em uma imagem em tempo real (LEITE et al., 2014).

Se for identificado um padrão irregular nos batimentos cardíacos, pode indicar que o coração não está recebendo uma quantidade adequada de sangue devido à DAC (BARSNESS; HOLMES, 2011). Na Figura 3.5 há exemplos de uma artéria sem calcificação e outra com calcificação durante o exame de ultrassom. A artéria é composta por três camadas: adventícia, íntima e média. Já na Figura 3.5b, é visível a presença de uma placa (acúmulo de cálcio), um cateter de Ultrassom Intra-Coronário (IVUS) e o vaso sanguíneo, representado nesse caso pela artéria coronária.

3.2 Aprendizado Profundo

O aprendizado profundo é uma vertente do aprendizado de máquina que envolve redes neurais artificiais (ANN, do inglês *Artificial neural network*) compostas por múltiplas camadas interconectadas de nós. Essas redes, construídas sobre camadas anteriores, são fundamentais para refinamento e otimização de previsões ou categorizações. Nesta seção, são explorados os princípios fundamentais das ANN, começando com os blocos de construção essenciais e avançando para uma discussão das redes neurais convolucionais aplicadas especificamente a imagens.

Figura 3.5 – Exemplo de exame de ultrassom de artéria : (a) sem calcificação e (b) com calcificação.



(a) (b)

Fonte: Adaptado de (MELO, 2017).

3.2.1 Redes neurais artificiais

ANNs são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano. Elas consistem em unidades interconectadas, conhecidas como neurônios, organizadas em camadas. As informações fluem através dessas camadas, passando por operações matemáticas que atribuem pesos às conexões entre os neurônios. Durante o treinamento, a rede ajusta esses pesos para aprender padrões nos dados de entrada e fazer previsões ou classificações. Essas redes são amplamente utilizadas em uma variedade de aplicações, como reconhecimento de padrões, processamento de linguagem natural e visão computacional(PATTANAYAK, 2023).

Na Figura 3.6 é possível observar um exemplo de *perceptron*, a unidade básica de processamento em ANN, e uma rede *feed-forward*, que consiste em camadas de neurônios dispostas sequencialmente, onde a saída de uma camada serve como entrada para a próxima camada. Essa estrutura é chamada de ”feed-forward” porque as informações fluem em uma única direção, da entrada para a saída, sem ciclos ou *feedbacks*. A ANN pode ter qualquer número de camadas ocultas e cada camada pode ter qualquer número de neurônios.

Figura 3.6 – Exemplo de Perceptron e Feed-Forward.

Fonte: Autoral.

3.2.2 Neurônios

A entrada para um neurônio na camada l é a soma das saídas y_{l-1} da camada anterior multiplicada pelos pesos w_l^i, onde i é o índice do neurônio. Para obter a saída um *bias* w_l^i é adicionado e uma função de ativação não linear é usada no resultado para produzir a saída w_l^i. Isso pode ser representado pela Equação 2.1, onde N_l é o número de neurônios na camada l.

$$y_l^i = \sigma(\sum_{j=1}^{N_{l-1}} y_{l-1}^j w_l^{j,i} + b_l^i) \tag{3.1}$$

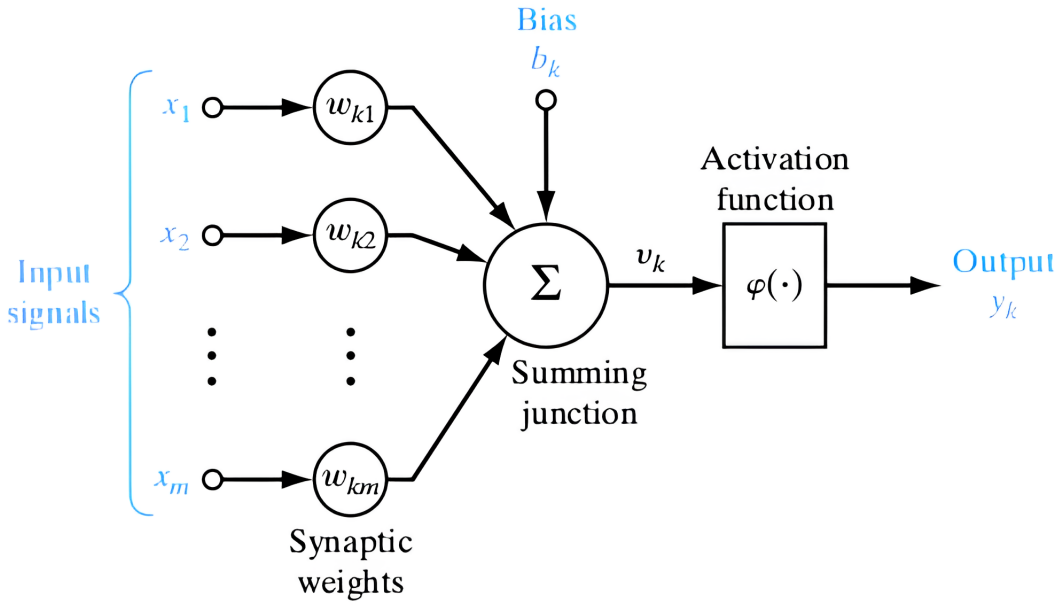
A soma na Equação 3.1 também pode ser representada como o produto escalar do vetor $y_l - 1$ e w_l^i, resultando na Equação 3.2.

$$y_l^i = \sigma((y_l - 1)^T w_l^i + b_l^i) \tag{3.2}$$

O vetor de saída da camada l pode ser representado pela Equação 3.3, onde $W_l - 1$ é uma matriz N_{l-1} por N_l.

$$y_l = \sigma((y_{l-1})^T w_l + b_l) \tag{3.3}$$

Figura 3.7 – Exemplo de neurônio de uma ANN.



Fonte: (COELHO, 2017).

O processo pode ser representado pela Figura 3.7, onde x representam as entradas, w representam os pesos, b é o bias, e φ é a função de ativação.

3.2.3 Função de ativação

As operações descritas anteriormente são lineares, de forma que, por mais complexa que seja a rede neural, ela só será capaz de captar relações lineares entre as variáveis de entrada e saída. De forma a tornar o modelo capaz de estabelecer relações não-lineares, os resultados de cada camada são processados pelas funções de ativação (APICELLA et al., 2021).

A principal razão pelas quais as funções de ativação são utilizadas, portanto, é para conferir essa capacidade não-linear ao processamento realizado pelas redes neurais. Isso é especialmente importante nas camadas escondidas da rede neural. Nas camadas de saída, as funções de ativação podem ter finalidades específicas, dependendo do objetivo da rede neural (APICELLA et al., 2021).

As escolhas possíveis para a função de ativação σ incluem sigmoide, Tangente Hiperbólica(tanh), unidade linear retificadora(ReLU do inglês, *Rectified Linear Unit*) entre outras. A maioria das arquiteturas modernas usa alguma variação da função de ReLU. A ReLU é definida pela Equação 3.4 (GLOROT; BORDES; BENGIO, 2011).

$$\begin{cases} 0 : x < 0 \\ \quad x : x \geq 0 \end{cases} \tag{3.4}$$

Um dos primeiros estudos que mostraram melhorias de desempenho de redes equipadas com funções de ativação baseadas em retificadores foi (GLOROT; BORDES; BENGIO, 2011), onde modelos com funções de ativação ReLU melhoraram o desempenho se comparados a redes com unidades sigmoides. O principal benefício do uso de funções de ativação retificadas é evitar o problema do *vanishing gradient*, que tem sido um dos principais problemas para redes profundas por muitos anos (APICELLA et al., 2021).

3.2.4 Função de perda

As funções de perda medem quão bem um modelo pode aproximar o resultado desejado e são usadas durante o treinamento para otimizar os parâmetros do modelo. Elas medem a diferença entre os resultados previstos e esperados do modelo, e o objetivo do treinamento é minimizar essa diferença (TERVEN et al., 2023).

A *Binary Cross Entropy* (BCE), também conhecida como perda logarítmica, é uma função de perda comumente utilizada para problemas de classificação binária. Ela mede a dissimilaridade entre a probabilidade prevista de uma classe e o rótulo de classe verdadeiro (TERVEN et al., 2023). Essa função de perda é usada para medir a diferença entre as probabilidades previstas pelo modelo e as classes verdadeiras dos dados. Ela é calculada para cada par de entrada e saída, onde y é a classe verdadeira (que pode ser 0

ou 1) e p é a probabilidade prevista pelo modelo de que a entrada pertence à classe 1, é definida pela Equação 3.5

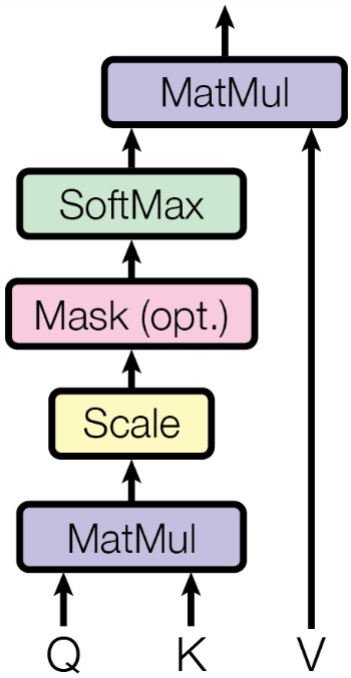
$$L(y, p) = -(y \log(p) + (1 - y) \log(1 - p)) \tag{3.5}$$

Quando a classe verdadeira é 1 ($y = 1$), a função de perda mede o quão próximo a probabilidade prevista p está de 1. Quando a classe verdadeira é 0 ($y = 0$), a função de perda mede o quão próximo a probabilidade prevista p está de 0.

3.2.5 Atenção (Attention)

A *Attention* é uma técnica utilizada em redes neurais para permitir que o modelo atribua pesos diferentes a diferentes partes da entrada durante o processamento. Isso significa que o modelo pode se concentrar em partes específicas dos dados de entrada, atribuindo mais importância a certos elementos e menos a outros (VASWANI et al., 2017). A *Attention* é comumente usada em tarefas de processamento de sequências, onde o modelo precisa entender quais partes da sequência de entrada são relevantes para gerar cada parte da sequência de saída.

Figura 3.8 – Exemplo de um tipo de arquitetura de *Attention*.



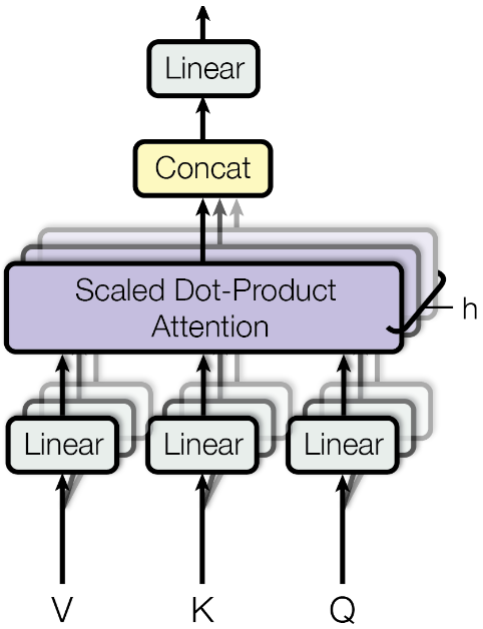
Fonte: (VASWANI et al., 2017).

A atenção própria, também conhecida como auto-atenção ou *self-attention*, é uma forma específica de atenção em que o modelo calcula as interações entre diferentes elementos dentro da mesma sequência. Em uma camada de *self-attention*, cada elemento da sequência é comparado com todos os outros elementos para determinar a importância relativa.

Isso permite que o modelo capture dependências de longo alcance entre os elementos da sequência e é especialmente útil em tarefas envolvendo sequências de comprimento variável.

Multi-head Attention: A *Multi-head Attention* é uma extensão da auto-atenção, em que várias ”cabeças” de atenção são usadas para capturar diferentes representações da relação entre os elementos da sequência. Em cada cabeça, os pesos de atenção são calculados separadamente, geralmente projetando os dados de entrada em diferentes subespaços de representação. As saídas dessas cabeças de atenção são então combinadas linearmente e projetadas de volta para o espaço de saída (VASWANI et al., 2017). A ideia principal por trás do mecanismo é que cada elemento na sequência aprenda a coletar de outras cabeças na sequência, como mostra na Figura 3.9.

Figura 3.9 – Exemplo de um *Multi-head Attention*.



Fonte: (VASWANI et al., 2017).

Multi-head Attention é comumente usada em modelos de linguagem como o *Transformer*, onde ajuda a capturar diferentes aspectos da relação entre palavras em uma frase.

3.2.6 Rede Neural Convolucional

As redes neurais convolucionais (CNN do inglês, *Convolutional Neural Network*) incluem mais de uma camada convolucional. Essas redes são compostas por neurônios, pesos que podem ser ajustados durante o treinamento, utilizam funções de perda e podem ser treinadas usando os mesmos princípios (KJERLAND, 2017). No entanto, elas se distinguem na maneira como os pesos e resultados são gerados, sendo que as camadas convolucionais são capazes de capturar informações espaciais presentes nas imagens. Uma

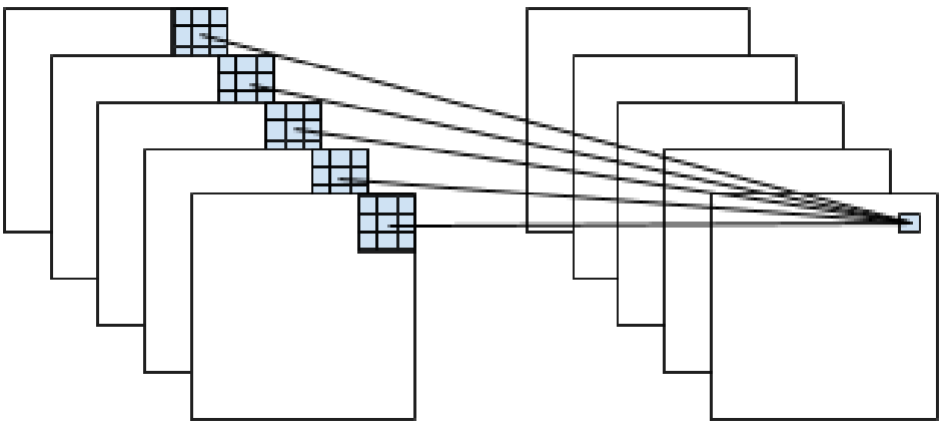
CNN típica é formada por diversas camadas em sequência, conectadas por pesos que funcionam como *kernels* de filtros convolucionais (KAMNITSAS et al., 2016).

Cada camada $l \in [1, L]$ possui C_l mapas de características. Nas CNN 2D, os mapas de características são grades 2D de neurônios que identificam características do mapa de características da camada anterior. Denotando por $K_l^{m,n}$ o *kernel*, uma matriz de pesos aprendidos $W_l^{m,n}$ do n-ésimo mapa de características na camada anterior ao m-ésimo mapa de características na camada l. As ativações do m-ésimo mapa de características na l-ésima camada podem ser representadas pela Equação 3.6, onde $*$ indica a operação de convolução (KAMNITSAS et al., 2016).

$$y_l^m = \sigma\left(\sum_{n=1}^{C_{l-1}} K_l^{m,n} * y_{l-1}^n + b_l^m\right) \tag{3.6}$$

Essa equação representa a soma dos resultados da convolução de cada mapa de características da camada anterior pelo *kernel* 2D. Para obter todas as ativações, os *kernels* são deslizados através dos mapas de características, o que significa que todos os neurônios no m-ésimo mapa de características da l-ésima camada compartilharão os mesmos pesos $W_l^{m,n}$. Todos os mapas de características de uma mesma camada têm as mesmas dimensões, e todos os *kernels* têm os mesmos tamanhos. Na Figura 3.10, podemos observar um exemplo de convolução entre os *kernels* e os mapas de características.

Figura 3.10 – Exemplo de camada de convolução 2D de um *kernel* 3x3 e 5 mapas de características.



Fonte: (KJERLAND, 2017).

Além disso, as CNNs incluem camadas de *Pooling*. Geralmente, as camadas de *Pooling* vêm após as camadas de convolução. O *pooling* é uma etapa crucial em sistemas baseados em convolução que reduz a dimensionalidade dos mapas de características. Ele combina um conjunto de valores em um número menor de valores, ou seja, realiza a redução na dimensionalidade do mapa de características. Essa operação transforma a representação

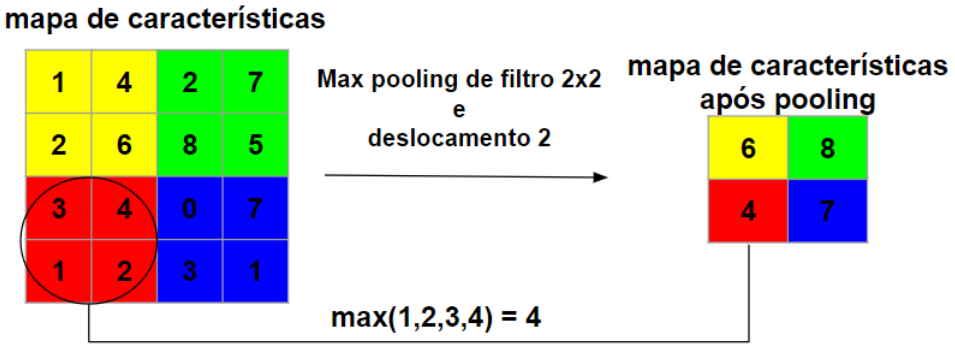
conjunta de recursos em informações valiosas, preservando informações úteis e descartando informações irrelevantes. Os operadores de *pooling* fornecem uma forma de invariância de transformação espacial, ao mesmo tempo em que reduzem a complexidade computacional para camadas superiores, eliminando algumas conexões entre camadas convolucionais (SUN et al., 2017).

O *Max-Pooling* é um operador que pode ser aplicado para reduzir a amostragem das saídas convolucionais, reduzindo assim a variabilidade. O operador *max-pooling* transmite o valor máximo dentro de um grupo de R ativações. A m-ésima camada *max-pooled* é composta por j filtros relacionados da seguinte forma (SUN et al., 2017):

$$P_{j,m} = max(h_{j,(m-1)N+r}) \tag{3.7}$$

Onde N é um deslocamento que permite a sobreposição entre regiões de agrupamento quando $N < R$. A camada de *pooling* diminui a dimensionalidade de saída de K convolucional para $M = \frac{K-r}{N+1}$ e resultando em uma camada de $p = [p_1, ..., p_m]$. Um exemplo de operação *max-pooling* pode ser visto na Figura 3.11.

Figura 3.11 – Exemplo de operação de *max pooling*.



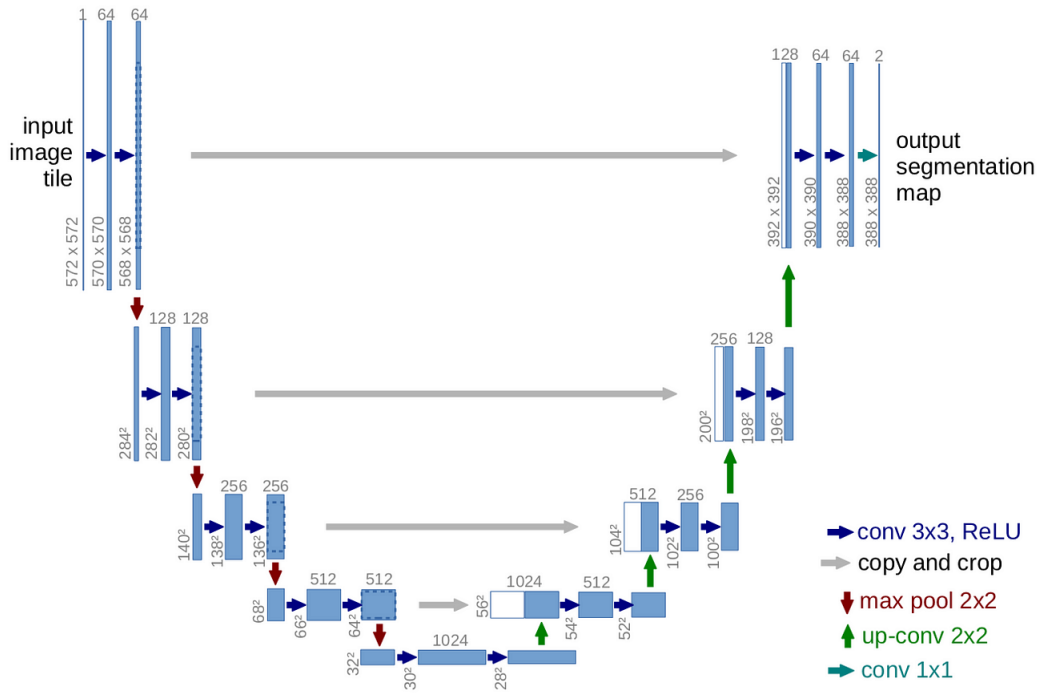
Fonte: Adaptada de (SUN et al., 2017).

3.2.7 U-Net

A arquitetura U-net foi concebida principalmente para segmentação de imagens (RONNEBERGER; FISCHER; BROX, 2015). Seu design fundamental consiste em dois caminhos distintos. O primeiro caminho, conhecido como caminho de contração ou codificador, assemelha-se a uma rede convolucional tradicional e é responsável pela extração de características. Enquanto isso, o segundo caminho, denominado caminho de expansão ou decodificador, é composto por operações de convolução ascendente e concatenações com características provenientes do caminho de contração. Essa abordagem de expansão capacita a rede a aprender informações detalhadas e localizadas. A U-net resultante

apresenta uma estrutura quase simétrica, formando uma configuração em formato de ”U”, como ilustrado na Figura 3.12.

Figura 3.12 – Arquitetura U-Net.



Fonte: (RONNEBERGER; FISCHER; BROX, 2015).

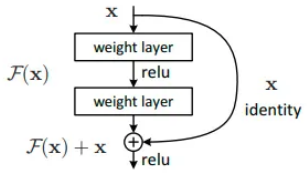
O caminho de contração utiliza uma arquitetura típica da CNN. Cada bloco no caminho de contração consiste em duas convoluções 3×3 sucessivas seguidas por uma função de ativação ReLU e uma camada de *max pooling*. Este arranjo é repetido várias vezes. A novidade do U-net vem na segunda parte, chamada de caminho expansivo, em que cada estágio *up-sampling* do mapa de características usando convolução ascendente 2×2. Em seguida, o mapa de características da camada correspondente no caminho de contração é concatenado no mapa de características com *up-sampling*. Isto é seguido por duas convoluções 3×3 sucessivas e ativação ReLU. No estágio final, uma convolução 1×1 adicional é aplicada para reduzir o mapa de características ao número necessário de canais e produzir a imagem segmentada. O recorte é necessário porque os pixels nas bordas possuem a menor quantidade de informações contextuais e, portanto, precisam ser descartadas. Isso resulta em uma rede semelhante a um formato de U e, mais importante, propaga informações contextuais ao longo da rede, o que permite segmentar objetos em uma área usando o contexto de uma área sobreposta maior (RONNEBERGER; FISCHER; BROX, 2015).

3.2.8 ResNet

A rede residual (ResNet) é uma arquitetura de CNN desenvolvida por He et al. (2015). Ela foi concebida para suportar centenas ou milhares de camadas convolucionais. As arquiteturas CNN anteriores não eram capazes de se adaptar a um grande número de camadas, o que resultava num desempenho limitado. No entanto, quando se acrescentavam mais camadas, os pesquisadores deparavam-se com o problema do *vanishing gradient*.

A arquitetura ResNet consiste em múltiplos blocos residuais, cada um dos quais inclui várias camadas de convolucionais e ligações de atalho (*Skip Connections*) que contornam as camadas convolucionais, adicionando a entrada original à saída do bloco. As camadas totalmente conectadas utilizam uma ativação *softmax* para produzir uma distribuição de probabilidades (HE et al., 2015).

Figura 3.13 – Exemplo de Bloco Residual.



Fonte: (HE et al., 2015).

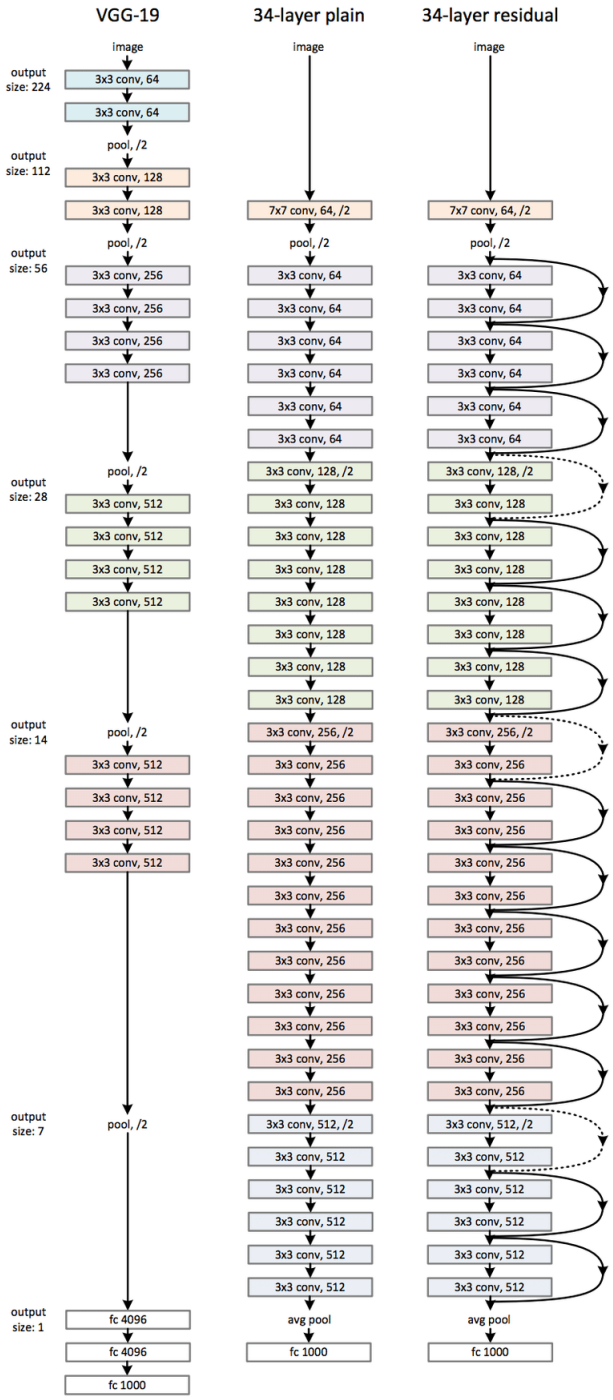
Assim, a ResNet oferece uma solução inovadora para mitigar o problema do *vanishing gradient* ao empregar conexões de atalho. A ResNet empilha múltiplos mapeamentos de identidade, onde camadas convolucionais são inseridas, e utiliza essas conexões de atalho para saltar camadas específicas, reutilizando as ativações da camada anterior. Essa abordagem de conexões de atalho permite acelerar o treinamento inicial, pois comprime a rede em menos camadas, além de facilitar o fluxo do gradiente durante o treinamento.

Depois, quando a rede é treinada novamente, todas as camadas são expandidas e as partes restantes da rede - conhecidas como partes residuais - podem explorar mais o espaço de características da imagem de entrada. Essa abordagem permite que a ResNet se adapte a um grande número de camadas convolucionais, superando o problema do *vanishing gradient* e alcançando desempenho superior em tarefas de visão computacional.

3.2.9 MobileNet

MobileNet é arquiteturas de CNNs desenvolvida por Howard et al. (2017), especialmente para realizar tarefas de visão computacional em dispositivos móveis e embarcados com recursos computacionais limitados. A característica principal do MobileNet é sua eficiência computacional durante a inferência, tornando-o adequado para dispositivos como *smartphones*, *tablets* e dispositivos de IoT.

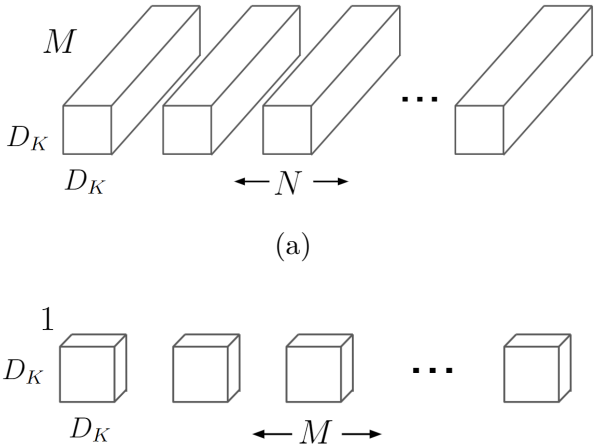
Figura 3.14 – Exemplo de ResNets.



Fonte: (HE et al., 2015).

O MobileNet utiliza uma técnica chamada convolução separável em profundidade (*Depthwise Separable Convolution*, que divide a convolução padrão em duas etapas: uma convolução por canal (convolução em profundidade) e uma convolução ponto a ponto (convolução ponto a ponto). Isso reduz significativamente a quantidade de cálculos necessários, resultando em um modelo mais leve e rápido em comparação com arquiteturas convencionais de CNN (HOWARD et al., 2017).

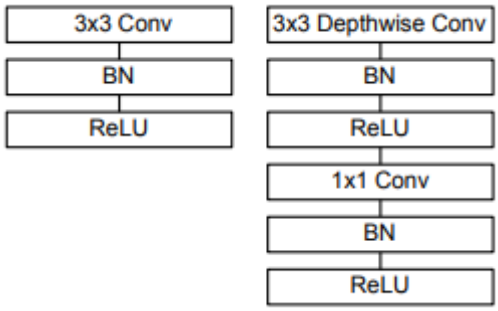
Figura 3.15 – Exemplos de convolução convencional (a) e convolução proposta pela MobileNet (b)



Fonte: (HOWARD et al., 2017).

Essa técnica tem como objetivo minimizar a complexidade computacional e o tamanho do modelo. A Figura 3.16 ilustra como a convolução padrão é decomposta em convolução em profundidade e convolução ponto a ponto. O desempenho do MobileNet pode variar de acordo com a versão e configuração específica, mas geralmente oferece um equilíbrio entre precisão e eficiência computacional, sendo amplamente utilizado em diversas aplicações de visão computacional em dispositivos móveis.

Figura 3.16 – Exemplo de CNN a direita e MobileNet a esquerda.



Fonte: (HOWARD et al., 2017).

3.2.10 Redes Generativas Adversárias

O conceito de Redes Generativas Adversárias (GANs) representa uma abordagem inovadora na geração de novas amostras com base em uma distribuição existente. Essa técnica, proposta em 2014 por Goodfellow et al. (2014), revoluciona tanto o aprendizado

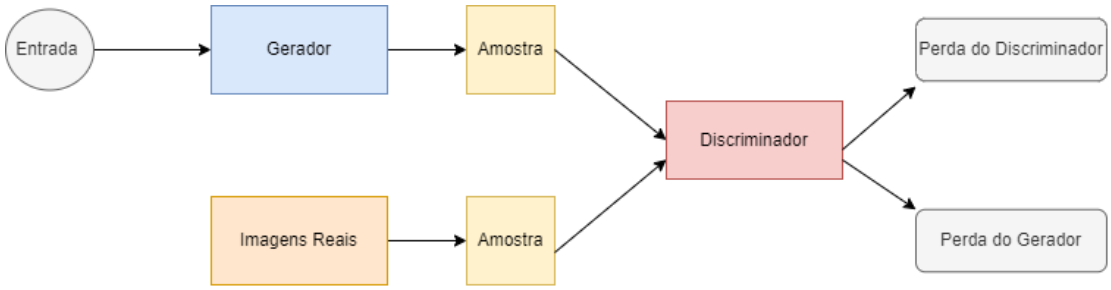
supervisionado quanto o não supervisionado, capacitando-se para modelar distribuições de dados complexas e de alta dimensionalidade. Ao treinar o gerador e o discriminador de forma adversaria, os GANs promovem um refinamento constante na capacidade de geração de imagens realistas, impulsionando avanços em diversas áreas, incluindo a arte digital, reconhecimento de padrões, geração de imagens e muito mais.

Tipicamente, os GANs são compostos por dois modelos principais: o gerador e o discriminador. O gerador é responsável por produzir exemplos sintéticos, enquanto o discriminador avalia esses exemplos, distinguindo entre os verdadeiros e os falsos.

No processo de treinamento, o gerador busca enganar o discriminador, gerando exemplos que sejam classificados como reais. Por sua vez, o discriminador é constantemente aprimorado para identificar melhor as amostras falsas. Esta dinâmica competitiva entre os dois modelos leva a uma melhoria contínua na capacidade de geração do gerador.

As GANs podem operar com diferentes funções de perda para o gerador e o discriminador, essenciais para orientar o processo de aprendizado. Essas perdas são usadas quando o gerador está tentando enganar o discriminador para melhorar a si mesmo. O gerador mistura imagens que ele produziu com imagens originais e as envia para o discriminador. O discriminador então as classificará como reais ou falsas. Após isso, o gerador atualizará seus pesos com a perda medida obtida do discriminador. O gerador gerará novas amostras, e a perda que ele produziu será usada pelo discriminador para ajustar seus pesos também.

Figura 3.17 – Arquitetura GAN.



Fonte: Autoral.

No contexto da tradução de imagem para imagem, um foco específico desta dissertação, destacam-se dois tipos de GANs: o Pix2Pix-GAN e o Cycle-GAN. Ambos os modelos têm aplicações únicas na transformação de imagens entre domínios distintos.

3.2.11 PatchGAN

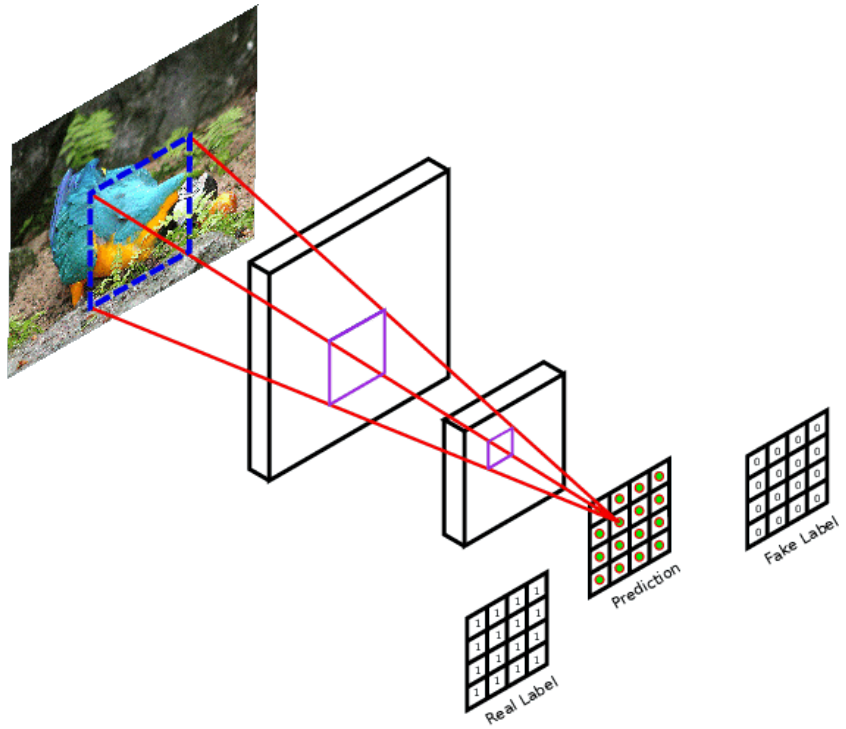
O PatchGAN é uma abordagem inovadora para a classificação de imagens, que visa lidar com as limitações das perdas L1 e L2 em problemas de geração de imagens. Embora essas perdas possam capturar com precisão as baixas frequências, frequentemente

resultam em imagens borradas, especialmente nas altas frequências. Para contornar essa questão, o PatchGAN restringe o discriminador GAN para modelar apenas a estrutura de alta frequência, enquanto utiliza um termo L1 para forçar a correção de baixa frequência (ISOLA et al., 2017).

Essa abordagem é implementada através de um discriminador que se concentra na classificação de *patches* N × N em uma imagem, em vez de avaliar a imagem como um todo. Ao classificar *patches* menores, o PatchGAN permite uma melhor captura da estrutura em manchas de imagens locais, contribuindo para a nitidez e coerência das imagens geradas (ISOLA et al., 2017). Uma das configurações mais comuns é a classificação de *patches* 70 × 70, devido à sua eficiência e desempenho (DOMINGUES, 2022).

Ao realizar essa classificação de *patches*, o discriminador procura discernir se cada *patch* é real ou falso, e essa informação é agregada por meio de uma média para produzir a resposta final do discriminador (DOMINGUES, 2022).

Figura 3.18 – Exemplo de aplicação da PatchGAN.



Fonte: (ISOLA et al., 2017).

3.2.12 Cycle-GANs

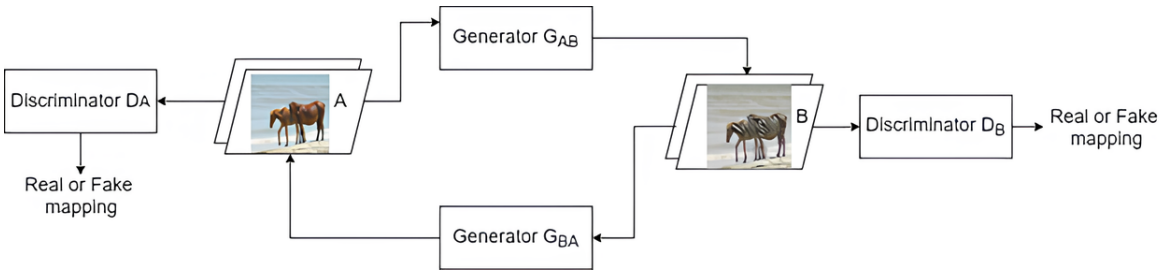
A Rede Adversaria Generativa em Ciclo, conhecida como Cycle-GAN, é uma arquitetura de aprendizado profundo amplamente empregada na tradução de imagem para imagem, pois enfrenta esse desafio possibilitando a tradução de imagem de forma não supervisionada. Ao aprender a mapear entre dois domínios de imagem a partir de

conjuntos de dados não emparelhados, o Cycle-GAN elimina a necessidade de exemplos em pares, permitindo gerar imagens do conjunto de dados B a partir de A e também reconstruir imagens originais a partir das geradas. Isso é conhecido como consistência de ciclo (ZHU et al., 2020).

A estrutura do Cycle-GAN é composta por dois elementos principais: geradores e discriminadores. Cada domínio possui um gerador encarregado de converter imagens de um domínio para outro, enquanto os discriminadores têm a tarefa de discernir entre imagens reais e geradas. Geralmente baseados em CNNs, os geradores aprendem a modificar imagens capturando as características fundamentais do domínio de destino. Enquanto isso, os discriminadores são treinados para distinguir entre imagens reais e aquelas geradas pelos geradores. Esse processo de treinamento adversaria motiva os geradores a produzir traduções realistas capazes de enganar os discriminadores (ZHU et al., 2020).

A arquitetura desse método se resume ao uso de um par de gerador e discriminador para gerar novos conjuntos de imagens e outro par para reconstruir a imagem original. Durante o treinamento, os discriminadores são expostos a dois tipos de imagens: algumas são reais, o que lhes permite distinguir posteriormente as imagens falsas fornecidas para enganá-los

Figura 3.19 – Exemplo de arquitetura da Cycle-GAN.



Fonte: (HELWAN, 2024).

Tanto os discriminadores quanto os geradores serão treinados no processo adversário de soma zero. Isso significa que, para um modelo melhorar, o outro precisa piorar. Neste caso, o objetivo de ambos os modelos é minimizar sua perda para se aprimorarem mutuamente.

3.2.13 Pix2pix

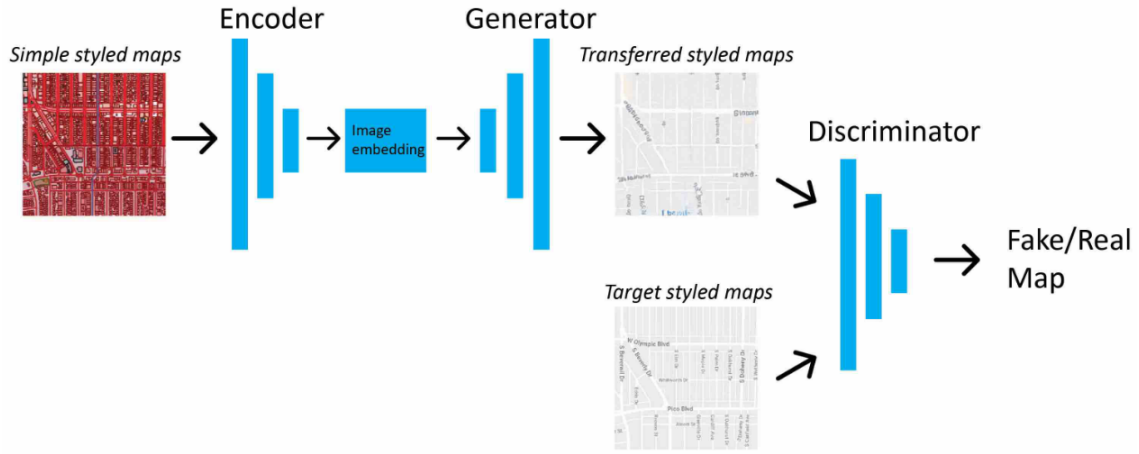
O Pix2Pix é uma arquitetura de tradução condicional de imagem para imagem que usa um objetivo de GAN condicional combinado com uma perda de reconstrução (ISOLA et al., 2017).

A contribuição chave do Pix2Pix é a introdução do conceito de PatchGAN para a função de perda adversária. Em vez de classificar a autenticidade da imagem como um todo, o discriminador do PatchGAN opera a nível de *patch*, ou seja, avalia pequenas

regiões das imagens de entrada e de saída. Isso permite uma avaliação mais localizada e detalhada da autenticidade das imagens geradas (ISOLA et al., 2017).

Além disso, o Pix2Pix utiliza uma arquitetura baseada em U-Net para o gerador, que é composta por um codificador que reduz a dimensionalidade da entrada e um decodificador que aumenta a dimensionalidade para gerar a saída. Essa arquitetura é eficaz para preservar detalhes finos da imagem durante o processo de tradução.

Figura 3.20 – Exemplo de arquitetura da Pix2pix.



Fonte:(KANG; GAO; ROTH, 2019).

3.2.14 Modelo de Difusão

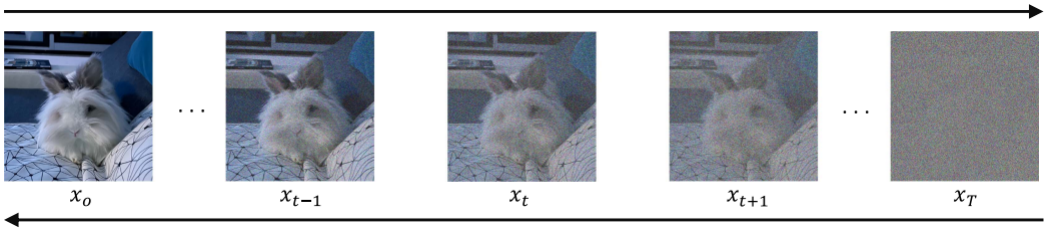
O modelo probabilístico de difusão (HO; JAIN; ABBEEL, 2020), também conhecido como ”modelo de difusão”, é uma cadeia de Markov parametrizada que é treinada usando inferência variacional para produzir amostras que correspondem aos dados após um tempo finito. Eles pertencem a uma categoria de modelos gerativos que operam ao adicionar gradualmente ruído (ruído Gaussiano) a um sinal de entrada (como uma imagem, texto ou áudio) e, em seguida, aprendem a remover esse ruído do sinal para produzir novas amostras. O modelo aprende gradualmente a remover o ruído. Este processo de remoção de ruído aprendido gera novas imagens de alta qualidade a partir de sementes aleatórias (imagens com ruído aleatório).

Assim, o procedimento de treinamento envolve duas fases: o processo de difusão direta (*Forward process*), que adiciona gradualmente ruído aos dados na direção oposta à amostragem até que o dado seja completamente ruidoso, e o processo de difusão reverso (*Reverse process*), que vai lentamente removendo o ruido do dados, até que ele a mesma ou uma nova amostra.

Na primeira fase, ocorrem múltiplas etapas em que ruído de baixo nível é adicionado a cada imagem de entrada, variando a escala do ruído a cada etapa. Os dados de treinamento

são gradualmente corrompidos até se tornarem ruído gaussiano puro.

Figura 3.21 – Exemplo do processo de difusão



Fonte: (CROITORU et al., 2023).

A última fase é representada pela reversão do processo de difusão direta. Nesse procedimento iterativo, o ruído é removido sequencialmente, resultando na recriação da imagem original, como pode ser visto na Figura 3.21. Portanto, durante a inferência, as imagens são geradas reconstruindo-as gradualmente a partir de ruído branco aleatório. O ruído subtraído em cada passo de tempo é estimado por meio de uma rede neural, frequentemente baseada na arquitetura U-Net utilizando as adaptações propostas por Ho, Jain e Abbeel (2020), garantindo a preservação das dimensões e permitindo o processo de difusão.

3.2.15 Forward Process

Durante a *Forward Process*, o processo começa com uma amostra de dados não corrompidos retirados da distribuição $p(x_o)$. Em cada passo t, o ruído gaussiano é adicionado à imagem anterior x_{t-1} para criar uma versão corrompida x_t, definida através do processo Markoviano definido pela Equação 3.8. Isso é modelado como uma sequência de distribuições condicionais gaussianas, onde β_t controla a escala do ruído e o I é a matriz identidade (CROITORU et al., 2023), tendo as mesmas dimensões que a imagem de entrada x_0. A distribuição condicional é dada por 3.8

$$q(x_t|x_{t-1}) := \mathcal{N}(x_t; p_1 - \beta_t x_{t-1}, \beta_t I) \tag{3.8}$$

Essencialmente, cada x_t é uma versão ruidosa da imagem anterior, e o processo continua até T, sendo que esse é número de etapas de difusão e $\beta_1, \ldots, \beta_T \in [0, 1]$ são hiper parâmetros que representam a variação do ruído através das etapas de difusão (HO; JAIN; ABBEEL, 2020).

Uma propriedade importante desta formulação recursiva é que ela também permite a amostragem direta de x_t, quando t é retirado de uma distribuição uniforme, i.e., $\forall t \sim U(\{1, \ldots, T\})$:

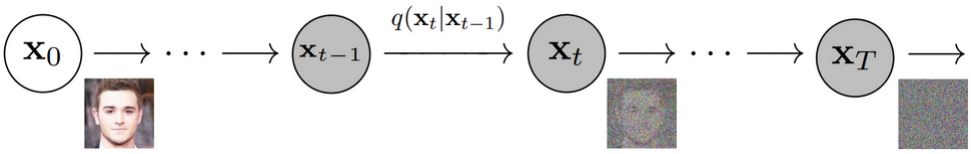
$$p(x_t|x_0) = \mathcal{N}\left(x_t; \widehat{\beta}_t x_0, (1 - \widehat{\beta}_t) I\right) \tag{3.9}$$

Onde $\widehat{\beta}_t = \prod_{i=1}^{t} \alpha_i$ e $\alpha_t = 1 - \beta_t$. Essencialmente, a Equação 3.9 mostra que podemos amostrar qualquer versão ruidosa x_t via um único passo, se tivermos a imagem original x_0 e fixarmos uma programação de variância β_t (CROITORU et al., 2023).

Se traduzir processo para as novas variáveis, então x_t é amostrado de $p(x_t|x_0)$ da seguinte forma:

$$x_t = \sqrt{\widehat{\beta}_t \cdot x_0} + \sqrt{(1 - \widehat{\beta}_t) \cdot z_t} \tag{3.10}$$

Figura 3.22 – Exemplo de processo direto



Fonte: Adaptado de (HO; JAIN; ABBEEL, 2020).

A variação $(\beta_t)_{t=1}^{T}$ é escolhida de forma que $\widehat{\beta}_T \to 0$, então, de acordo com a equação anterior. Além disso, essa equação também é usada os no processo reverso $p(x_{t-1}|x_t)$ têm a mesma forma funcional que o processo direto $p(x_t|x_{t-1})$. Intuitivamente, esta última afirmação é verdadeira quando x_t é criado com um passo muito pequeno, pois torna-se mais provável que x_{t-1} venha de uma região próxima à onde x_t é observado, o que nos permite modelar essa região com uma distribuição gaussiana (CROITORU et al., 2023). Para conformar-se às propriedades mencionadas, Ho, Jain e Abbeel (2020) escolhe $(\beta_t)_{t=1}^{T}$ para serem constantes linearmente crescentes entre $\beta_1 = 10^{-4}$ e $\beta_T = 2 \times 10^{-2}$, onde $T = 1000$.

3.2.16 Reverse Process

Após o processo de difusão ter corrompido gradualmente os dados da imagem, o processo reverso, conhecido como *denoising*, é realizado para recuperar a imagem x_t para a original x_0. Isso é feito usando uma rede neural que recebe a imagem ruidosa x_t e aprende a prever os parâmetros da distribuição gaussiana $p(x_{t-1}|x_t)$ que geraria a imagem original sem ruído (HO; JAIN; ABBEEL, 2020).

$$p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu(x_t, t), \Sigma(x_t, t)) \tag{3.11}$$

A distribuição condicional é dada pela média $\mu(x_t, t)$ é aprendida pela rede neural, enquanto a covariância dada por $\Sigma(x_t, t)$ se torna uma constante (HO; JAIN; ABBEEL, 2020). Onde $z_\theta(x_t, t)$ é a estimativa do ruído no momento t o $\mu(x_t, t)$ é definido pela Equação 3.12.

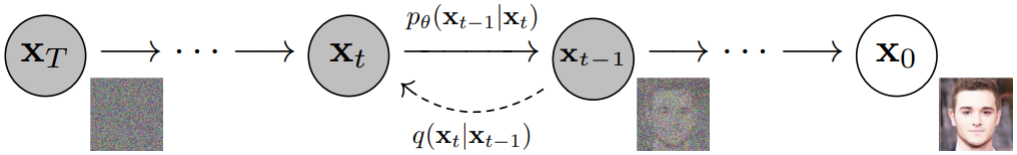
$$\mu_\theta = \frac{1}{\sqrt{\alpha_t}} \cdot \left(x_t - \frac{\sqrt{1-\alpha_t}}{\sqrt{1-\hat{\beta}_t}} \cdot z_\theta(x_t, t) \right) \tag{3.12}$$

Essas simplificações desbloquearam uma nova formulação do objetivo L_{vlb}, que mede, para um passo de tempo aleatório t do processo direto, a distância entre o ruído real z_t e a estimativa de ruído $z_\theta(x_t, t)$ do modelo:

$$L_{\text{simple}} = \mathbb{E}_{t\sim[1,T]}\mathbb{E}_{x_0\sim p(x_0)}\mathbb{E}_{z_t\sim\mathcal{N}(0,I)}\|z_t - z_\theta(x_t, t)\|_2^2, \tag{3.13}$$

onde $\mathbb{E}$ é o valor esperado, e $z_\theta(x_t, t)$ é a rede que prevê o ruído em x_t. O x_t é amostrado através da Equação 3.10, onde uma imagem aleatória x_0 do conjunto de treinamento. O processo generativo ainda é definido por $p_\theta(x_{t-1}|x_t)$, mas a rede neural não prevê a média e a covariância diretamente. Em vez disso, ela é treinada para prever o ruído da imagem, e a média é determinada de pela Equação 3.12 enquanto a covariância é fixada em um valor constante.

Figura 3.23 – Exemplo de processo reverso



Fonte: (HO; JAIN; ABBEEL, 2020) Modificada.

3.3 Métricas de avaliação e Qualidade de Imagens

Para cada imagem sem contraste de entrada, o modelo gera uma imagem de contraste. Isso é então comparado com a verdadeira imagem de contraste, que é do mesmo tamanho da gerada, e o modelo é avaliado dessa maneira. Para avaliar quão bem o método realiza a geração de imagem existem algumas métricas especializadas para lidar com GANs (BORJI, 2018).

O Erro Médio Quadrado (MSE do inglês, *Mean Square Error*), Raiz do Erro Médio Quadrado (RMSE do inglês, *Root Mean Square Error*) e Erro Médio Absoluto (MAE do inglês, *Mean Absolute Error*) são métricas simples que calculam a diferença entre os valores dos pixels das imagens reais e geradas (BORJI, 2018). O MSE é a média dos quadrados das diferenças, enquanto o RMSE é a raiz quadrada do MSE, fornecendo uma medida da dispersão dos erros. MAE, calcula a média das diferenças absolutas entre os valores reais e os valores previstos por um modelo. Todos são úteis para avaliar a discrepância entre as imagens. As Equações de MAE, MSE E RSME são definidas em, respectivamente, 3.14, 3.15 e 3.16.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.14}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.15}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.16}$$

Onde n é o número total de observações. , y_i são os valores reais. e $\hat{y}_i$ são os valores previstos pelo modelo.

A Relação Sinal-Ruído de Pico (PSNR do inglês, Peak Signal-to-Noise Ratio) é uma métrica comumente usada para medir a qualidade de uma imagem, especialmente em tarefas de reconstrução de imagem e compressão de imagem. Ele quantifica a relação entre o sinal máximo possível (geralmente representado pela faixa dinâmica da imagem) e o ruído introduzido durante o processo de reconstrução ou compressão (BORJI, 2018). O PSNR é expresso em decibéis (dB) e quanto maior o valor do PSNR, melhor é a qualidade da imagem.

Comparando a imagem reconstruída (ou comprimida) com a imagem original, calcula-se a diferença entre os valores de pixel das duas imagens. Essa métrica é sensível a pequenas diferenças de pixel e fornece uma medida numérica direta da fidelidade da reconstrução. Além disso, tal métrica não tem valor absoluto, servindo apenas para comparar métodos de compressão de perda geralmente ou, como é o caso deste trabalho, métodos de reconstrução (NASCIMENTO, 2015). O PSNR pode não corresponder perfeitamente à percepção humana de qualidade visual, especialmente em imagens altamente comprimidas, onde artefatos visuais podem ser perceptíveis mesmo com altos valores de PSNR (BORJI, 2018), a equação que define o PSNR é 3.17.

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right) \tag{3.17}$$

Além do PSNR, o índice de Medida da Similaridade Estrutural (SSIM do inglês, Structural Similarity Index Measure) é uma métrica amplamente utilizada para avaliar a similaridade entre duas imagens. Ele leva em consideração três aspectos principais: a similaridade de luminância, a similaridade de contraste e a similaridade estrutural (BORJI, 2018). Em essência, o SSIM compara as estruturas de pixels de duas imagens e calcula a semelhança entre elas, considerando tanto informações de intensidade quanto de estrutura, definida pela equação: 3.18.

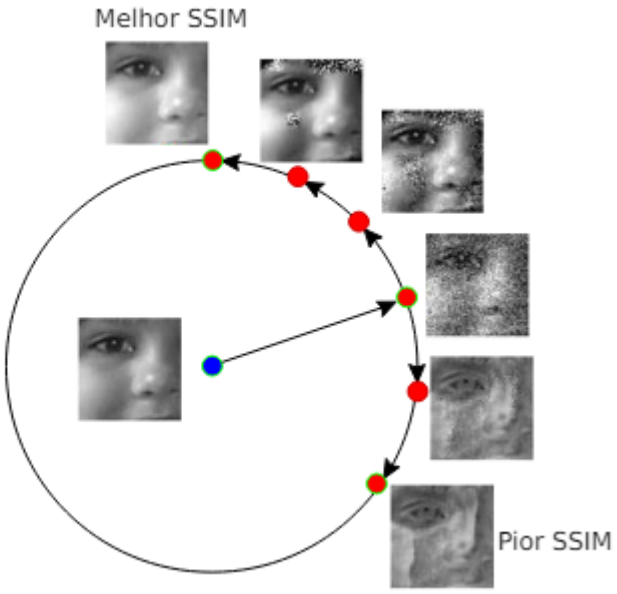
$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3.18}$$

Onde:

- x e y são as duas imagens que estão sendo comparadas.
- μ_x e μ_y são as médias de x e y, respectivamente.
- σ_x^2 e σ_y^2 são as variâncias de x e y, respectivamente.
- σ_{xy} é a covariância entre x e y.
- C_1 e C_2 são constantes para estabilizar a divisão com números pequenos no caso de baixos valores de variância.

O valor do SSIM varia entre -1 e 1 mas, geralmente, se encontra entre 0 e 1. Assume-se 1 o valor resultante quando duas imagens idênticas são comparadas, e 0 quando duas imagens completamente diferentes são dadas como entrada (NASCIMENTO, 2015).

Figura 3.24 – Comparação ilustrativa entre o PSNR e o SSIM. Observe que todas as imagens na circunferência do círculo possuem a mesma PSNR em relação à original



Fonte: (NASCIMENTO, 2015).

O *Fréchet Inception Distance* (FID) é uma métrica avançada de avaliação de modelos generativos que mede a semelhança entre as distribuições de características de imagens reais e geradas. Ao contrário de métricas tradicionais como PSNR e SSIM, que se concentram na fidelidade visual das imagens, o FID vai além, considerando a diversidade e a complexidade das características das imagens (YU; ZHANG; DENG, 2021). O FID utiliza uma rede neural pré-treinada para extrair características de alto nível das imagens

e, em seguida, calcula a distância de Fréchet entre as distribuições dessas características, fornecendo uma medida quantitativa da qualidade das imagens geradas (BORJI, 2018).

Em essência, um FID mais baixo indica uma melhor correspondência entre as características das imagens reais e geradas, o que geralmente é desejável em modelos generativos. Uma vez que o FID leva em consideração tanto a fidelidade visual quanto a diversidade das imagens geradas, ele oferece uma avaliação mais abrangente da qualidade do modelo generativo, ajudando os pesquisadores a entender e aprimorar a capacidade do modelo de capturar e reproduzir as características essenciais das imagens reais (BORJI, 2018).

$$\text{FID}(p, q) = ||\mu_p - \mu_q||^2 + \text{tr}(\Sigma_p + \Sigma_q - 2(\Sigma_p\Sigma_q)^{1/2}) \tag{3.19}$$

Onde:

- p e q representam as distribuições das características (por exemplo, características extraídas de imagens reais e geradas).
- μ_p e μ_q são os vetores de média das distribuições p e q, respectivamente.
- Σ_p e Σ_q são as matrizes de covariância das distribuições p e q, respectivamente.
- $||\mu_p - \mu_q||^2$ é a distância euclidiana ao quadrado entre os vetores de média.
- $\text{tr}(\Sigma_p + \Sigma_q - 2(\Sigma_p\Sigma_q)^{1/2})$ é a soma dos elementos da matriz resultante da soma das covariâncias menos duas vezes a raiz quadrada da matriz de covariância resultante do produto das covariâncias.

A NCC é uma técnica que avalia a correlação entre duas imagens. Ela é comumente usada para determinar o quanto uma imagem se assemelha a outra em termos de padrões e características gerais. A NCC calcula a similaridade entre duas imagens encontrando a correlação entre os valores dos pixels das imagens, normalizando-os para levar em conta variações na intensidade, definida pela Equação: 3.20.

$$\text{NCC}(x, y) = \frac{\sum_{p,q}(x_{p,q} - \mu_x)(y_{p,q} - \mu_y)}{\sqrt{\sum_{p,q}(x_{p,q} - \mu_x)^2} \cdot \sqrt{\sum_{p,q}(y_{p,q} - \mu_y)^2}} \tag{3.20}$$

Onde:

- x e y são as duas imagens que estão sendo comparadas.
- $x_{p,q}$ e $y_{p,q}$ são os pixels das imagens x e y nas posições (p, q), respectivamente.
- μ_x e μ_y são as médias de intensidade de x e y, respectivamente.

4 Materiais e Métodos

Neste capítulo, será apresentado a arquitetura proposta para geração de imagens de tomografia computadorizada sintéticas e os materiais utilizados nela.

4.1 Base de Imagens

O conjunto de dados utilizado foi obtido do Orca Score na plataforma Grand Challenge (WOLTERINK, 2016). Ele inclui imagens de tomografia computadorizada cardíaca de pacientes escaneados em quatro tomógrafos diferentes, de quatro fornecedores e em quatro hospitais distintos. O conjunto de dados consiste em imagens de pacientes escaneados em quatro tomógrafos diferentes, fornecidos por quatro fornecedores distintos e em quatro hospitais diferentes. Cada um dos 72 pacientes possui uma imagem de tomografia sem injeção de contraste (NECT) e uma tomografia computadorizada com injeção de contraste (CECT). O tamanho original das imagens 356x356xZ onde $Z \in [42, 64]$, e o total de imagens do conjunto de dados 6209 e imagens, com 2812 para teste e 3397 para treino e validação.

4.2 Pré-processamento

A realização de etapas específicas de pré-processamento nas imagens adquiridas foi essencial para melhorar a qualidade geral do modelo. Uma das etapas importantes foi a conversão das imagens originais da escala de *Hounsfield* para uma escala de cinza com valores de pixel entre 0 e 255. Essa conversão foi realizada para simplificar o processamento e permitir a aplicação adequada de técnicas de filtragem de ruído nas imagens.

O procedimento de conversão envolveu o mapeamento e a normalização dos valores dos pixels das imagens. Isso foi feito para garantir que os objetos presentes nas imagens não fossem perdidos durante o processo e para reduzir os valores associados a cada pixel, facilitando o processamento subsequente.

Para garantir uma normalização consistente em todas as imagens, o mapeamento foi realizado utilizando os valores máximos e mínimos presentes na escala de *Hounsfield* entre todas as imagens. Esse processo assegurou que os pixels não fossem ajustados de maneiras distintas em diferentes imagens, garantindo assim a uniformidade no pré-processamento e evitando distorções na análise posterior.

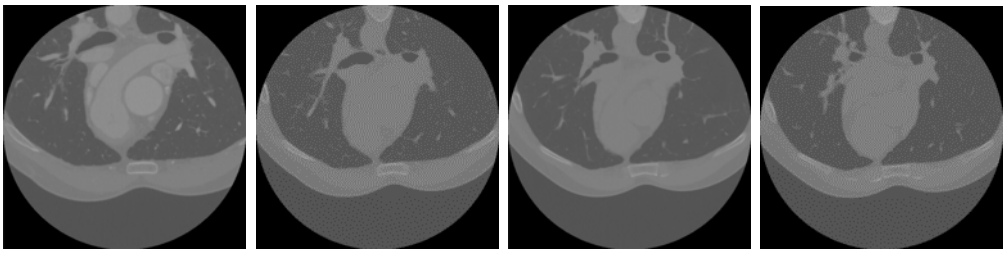
Há alguns problemas menores de formatação e clareza no texto. Aqui está uma versão revisada:

Após essa etapa, ocorreu o emparelhamento das tomografias computadorizadas com contraste (CECT) e sem contraste (NECT) de cada paciente, selecionando aquelas na mesma posição com alto índice de similaridade. Esta abordagem garantiu que apenas as imagens mais relevantes e correspondentes fossem utilizadas para refinar a análise do modelo.

Para este processo, foram utilizadas duas métricas de avaliação de similaridade entre imagens: SSIM e NCC, conforme descrito na Seção 3.3.

Assim, as imagens foram correlacionadas utilizando SSIM e NCC. Em geral, imagens com índices de similaridade acima de 82%, dentro do mesmo paciente, foram consideradas equivalentes e separadas para os experimentos. Esse limiar foi escolhido devido à presença de muitas imagens com índices de 78-80% de SSIM no conjunto de imagens, que poderiam estar relacionadas a várias imagens dentro do paciente, mas não representavam cortes significativos e acabavam por criar falsas correlações.

Figura 4.1 – Exemplo da correlação de similaridade adquirida após o pré-processamento

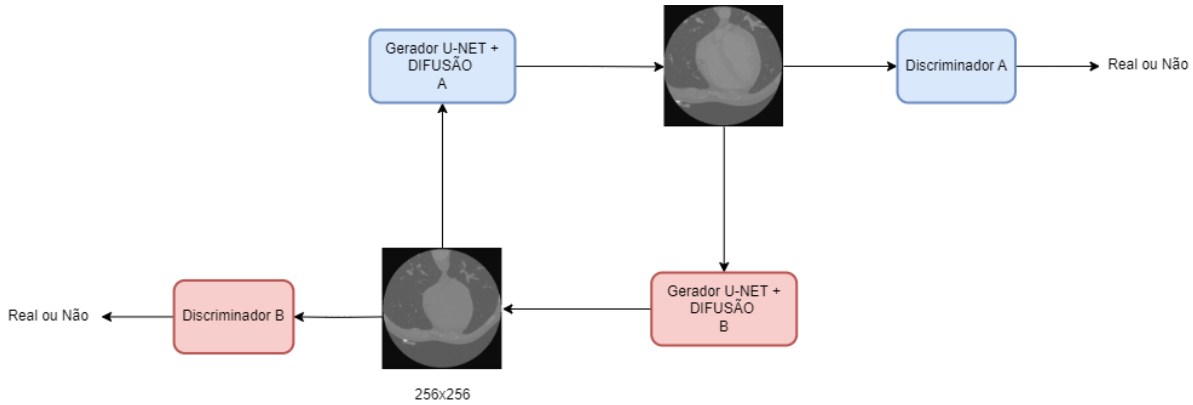


(Referência com Contraste) (Acima de 82%) (Entre 78% e 82%) (Abaixo de 78%)

Fonte: Autoral.

4.3 Arquitetura de geração de imagens sintéticas proposta

Figura 4.2 – Arquitetura do método proposto.

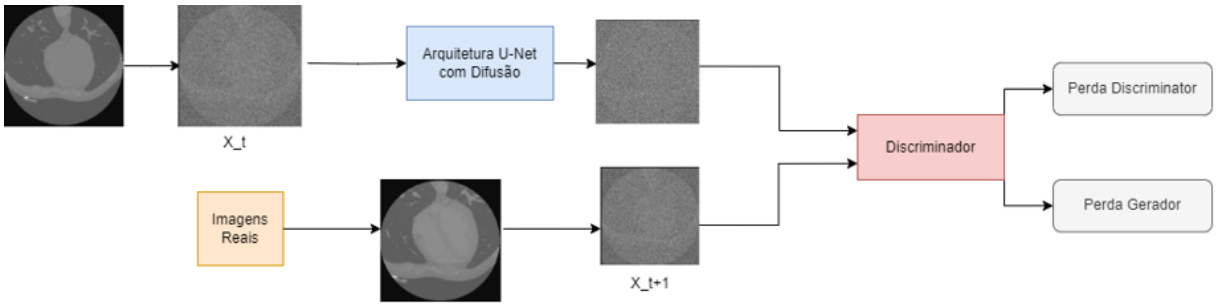


Fonte: Autoral.

Para a geração das imagens sintéticas, propõe-se um método baseado em uma abordagem iterativa e adversária, no qual duas redes são treinadas para competir entre si, usando como base a arquitetura Cycle-GAN proposta por Zhu et al. (2020). Este método consiste em duas GANs de difusão, cada uma responsável pela geração de imagens com e sem contrastes, a partir de suas contrapartes, e pela classificação das imagens como reais ou não, como mostrado na Figura 4.2.

Tudo isso ocorre dentro de um sistema iterativo, onde as redes generativas e discriminadoras se retroalimentam (ZHU et al., 2020). Na abordagem proposta nesta dissertação, os geradores e discriminadores usados são respectivamente, redes baseadas na arquitetura U-NET e redes CNNs, conforme ilustrado na Figura 4.3.

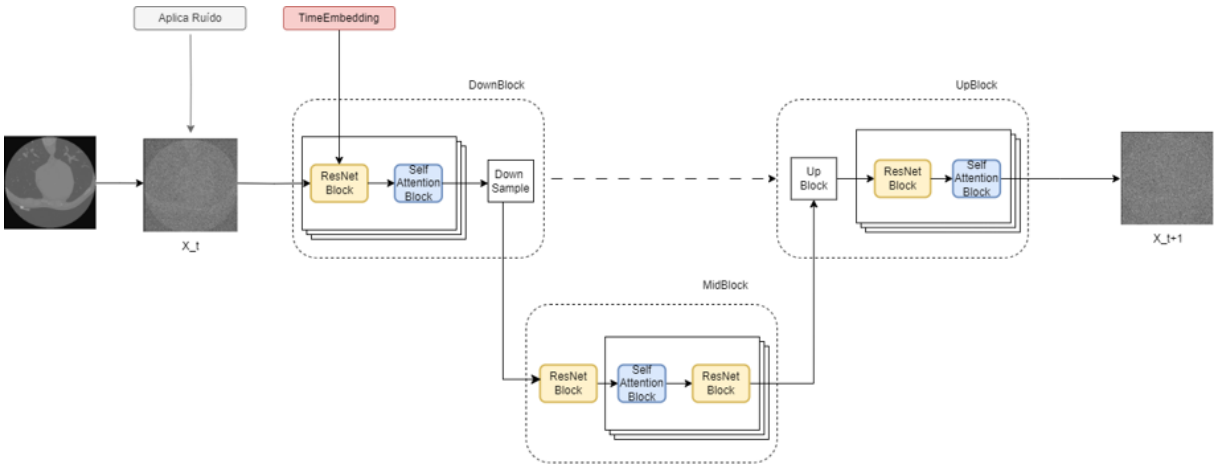
Figura 4.3 – Arquitetura da GAN de difusão



Fonte: Autoral.

Cada rede geradora é uma implementação que inclui uma série de modificações na arquitetura U-Net original, como blocos residuais e *multi-head attention*, além de adicionar *embeddings* de passos de tempo T (HO; JAIN; ABBEEL, 2020). Esta arquitetura da Figura 4.4 foi utilizada, pois apresenta bons resultados na tarefa de tradução de imagens.

Figura 4.4 – Arquitetura U-NET utilizada



Fonte: Autoral.

Na arquitetura U-NET utilizada, dentro dos blocos residuais na camada de *Down-Block*, foi integrado um bloco de *embedding* temporal que recebe um tensor unidimensional

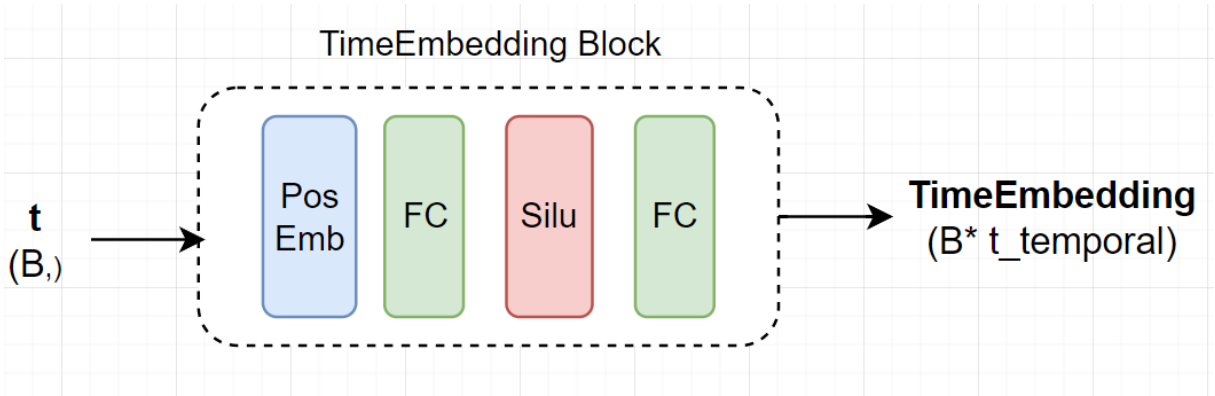
com passos temporais t e tamanho b, que é o tamanho de lote, e gera uma representação de tamanho t_{temporal} para cada um desses passos temporais no lote.

A informação de qual t o modelo se encontra é crucial para o ele determinar quanto de ruído existe na imagem (HO; JAIN; ABBEEL, 2020). Esse bloco converte os passos de tempo inteiros em representações vetoriais usando o espaço do *embedding*. Para isso, emprega-se a incorporação de posição senoidal e sua complementar. As transformações utilizadas são as seguintes:

- Para a incorporação de posição senoidal: $PE^1_{(t,i)} = \sin\left(\frac{t}{10000^{\frac{i}{d-1}}}\right)$
- Para a incorporação de posição senoidal complementar: $PE^2_{(t,i)} = \cos\left(\frac{t}{10000^{\frac{i}{d-1}}}\right)$

Posteriormente, essas representações são processadas por duas camadas lineares separadas por ativação, resultando na representação final do passo. Para o espaço de incorporação, os autores Ho, Jain e Abbeel (2020) optam por usar a incorporação de posição senoidal usando transformações. Quanto à ativação, unidades lineares sigmoide foram empregadas.

Figura 4.5 – Arquitetura do *Embedding* temporal



Fonte: Autoral.

5 Resultados e Discussão

Neste capítulo são apresentados e discutidos os resultados obtidos nos experimentos realizados para a validação da metodologia proposta. Além dos resultados numéricos, são apresentados resultados visuais, que permitem observar as semelhanças entre as imagens geradas pelo modelo e as originais.

Os experimentos foram todos usando o conjunto de dados OrcaScore (WOLTERINK, 2016) e os resultados apresentados foram obtidos de acordo com experimentos comparando diversas arquiteturas e modelos com os processamentos propostos.

Todos os experimentos foram realizados em um *hardware* com as seguintes especificações: CPU Intel(R) Core(TM) i9-7980XE @ 2.60GHz, placa gráfica NVIDIA GeForce 3090 com 24 GB de VRAM e sistema operacional Ubuntu 18.04.5 LTS.

Tabela 5.1 – Distribuição das imagens após pre-processamento

	Contraste	Sem Contraste
Treino	200	200
Teste	100	100
Validação	50	50

Fonte: Autoral.

5.1 Experimentos

Durante o treinamento dos modelos de difusão, foram realizadas 200 épocas utilizando uma taxa de aprendizado de 10^{-4}. Os parâmetros do modelo foram configurados da seguinte maneira: o tempo T foi definido como 1000, enquanto os passos de difusão k foram ajustados para 100, resultando em 10 etapas de difusão ($\frac{T}{k}$). Pesos λ foram atribuídos aos termos de consistência cíclica e perda adversaria, com $lambda_1 = 0,5$ e $\lambda_2 = 1$, respectivamente. Além disso, foram definidos limites inferior e superior para o agendamento da variância de ruído, com $\beta_1 = 10^{-4}$ e $\beta_T = 2 \times 10^{-2}$.

O padrão de referência para todas as imagens de TC sem contraste foi definido de forma independente por dois especialistas usando uma ferramenta personalizada. Seguindo o protocolo padrão fornecido pelos fornecedores comerciais, os voxels da imagem com intensidades superiores a 130 HU foram destacados.

Durante os experimentos, foram testadas três diferentes codificadores para a arquitetura U-Net utilizada no modelo de difusão: um utilizando MobileNet, outro utilizando

ResNet e o último utilizando um bloco de ResNet modificado com um bloco de atenção própria (*self-attention*).

Os experimentos realizados serão comparados através das métricas RMSE, SSIM, PSNR e FID. Todas as comparações foram feitas tendo a mesma divisão de treino e teste, garantindo uma comparação justa.

5.1.1 Qualidade da imagem

Os testes foram realizados correspondendo à comparação entre as redes Cycle-GAN (ISOLA et al., 2017), a Cycle-GAN com *SkipResidual* (DOMINGUES, 2022), Cytran (RISTEA et al., 2023) e o modelo proposto. Esta comparação é realizada com o objetivo de validar o uso da modificação proposta por Ho, Jain e Abbeel (2020) para gerar imagens com maior qualidade. Os resultados desta comparação podem ser vistos na Tabela 5.2.

Tabela 5.2 – Comparação da qualidade da imagem

Arquitetura	RMSE	PSNR
Cycle-GAN (ZHU et al., 2020)	0,15	18,22
Cycle-GAN (DOMINGUES, 2022)	0,17	15,57
Cytran (RISTEA et al., 2023)	**0,14**	29,66
Modelo com difusão MobileNet	0,22	18,30
Modelo com difusão ResNet	0,17	21,55
Modelo com difusão proposto	0,2	**32,85**

Fonte: Autoral.

Como observado na Tabela 5.2, a arquitetura da Cycle-GAN de Ristea et al. (2023) apresentou melhores resultados na diferença de valores de pixels, entretanto, quando se trata da qualidade da imagem, avaliado pelo PSNR, o modelo proposto foi o melhor demonstrado. Alguns exemplos de imagens geradas do usando conjunto OrCaScore e utilizando a U-Net com difusão a pode ser visto na Figura 5.1.

5.1.2 Similaridade da imagem

Os testes foram realizados correspondendo à comparação entre as redes Cycle-GAN de Zhu et al. (2020), a Pix2pix de Isola et al. (2017), Cytran de Ristea et al. (2023) e o modelo proposto acerca da de geração de imagem. Esta comparação é realizada de forma a validar o uso da modificação proposta por Ho, Jain e Abbeel (2020) na tarefa de geração de imagens médicas. Esta comparação pode ser vista na Tabela 5.3.

Como pode ser observado na Tabela 5.3, a arquitetura proposta da Cycle-GAN apresentou melhores resultados na diferença absoluta de pixels, junto com a Cytran

Tabela 5.3 – Comparação entre similaridade de imagens

Arquitetura	FID	SSIM
Cycle-GAN (ZHU et al., 2020)	129,423	0,635
Pix2pix (ISOLA et al., 2017)	210,232	0,545
Cytran (RISTEA et al., 2023)	58,787	0,745
Modelo com difusão MobileNet	180,683	0,444
Modelo com difusão Resnet	150,644	0,523
Modelo com difusão proposto	**42,348**	**0,766**

Fonte: Autoral.

(RISTEA et al., 2023). Quanto à similaridade e características semelhantes, avaliadas pelo SSIM e pelo FID, o modelo proposto foi o melhor demonstrado.

5.2 Comparação com trabalhos relacionados

Todos essas arquiteturas foram testadas utilizando a divisão de treino, teste e validação descrita na Seção 4.1. Após os experimentos, o método proposto apresenta resultados positivos próximos aos trabalhos anteriores, como pode ser visto na Tabela 5.4

Tabela 5.4 – Comparação entre redes

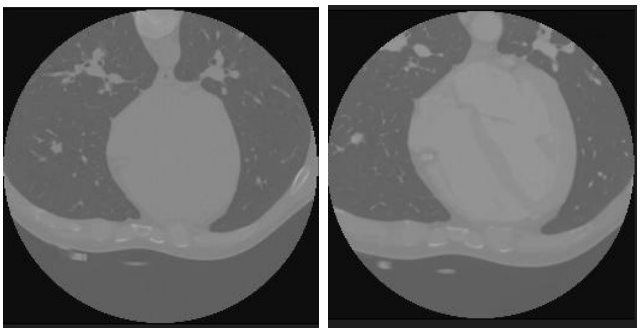
Arquitetura	PSNR	SSIM	RMSE	FID
Cycle-GAN (ZHU et al., 2020)	18,22	0,635	0,15	129,423
Cycle-GAN (DOMINGUES, 2022)	15,569	0,433	0,17	105,22
Pix2pix (ISOLA et al., 2017)	15,66	0,545	0,16	210,232
Pix2pix (DOMINGUES, 2022)	16,661	0,503	0,16	239,445
Cytran (RISTEA et al., 2023)	29,66	0,745	**0,14**	58,787
Modelo com difusão MobileNet	18,30	0,444	0,22	180,683
Modelo com difusão Resnet	21,55	0,523	0,17	150,644
Modelo com difusão proposto	**32,85**	**0,766**	0,20	**42,348**

Fonte: Autoral.

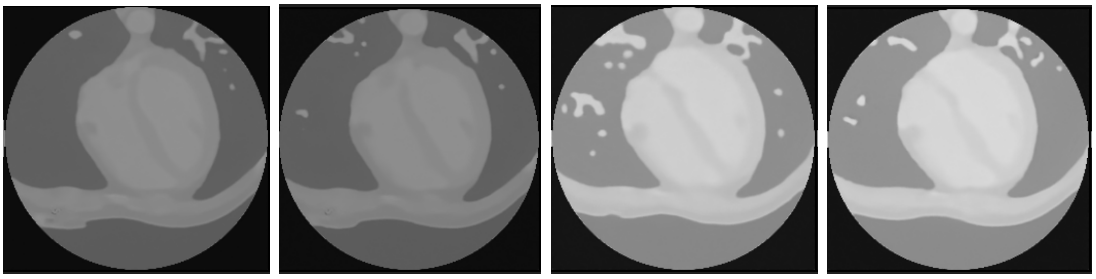
A Cycle-GAN de Zhu et al. (2020) apresenta um PSNR de 18,22 e um SSIM de 0,635, indicando uma qualidade de imagem razoável, mas não excepcional. O RMSE é 0,15 e o FID é 129,423, sugerindo que a fidelidade das imagens geradas não é alta e que há espaço para melhorias significativas na preservação das características das imagens originais.

A versão da Cycle-GAN discutida por (DOMINGUES, 2022) apresenta um PSNR de 15,569 e um SSIM de 0,433, mostrando uma qualidade de imagem inferior em comparação à versão original. O RMSE é 0,17 e o FID é 105,22, indicando melhorias em relação à fidelidade (menor FID), mas ainda com qualidade de imagem abaixo do esperado.

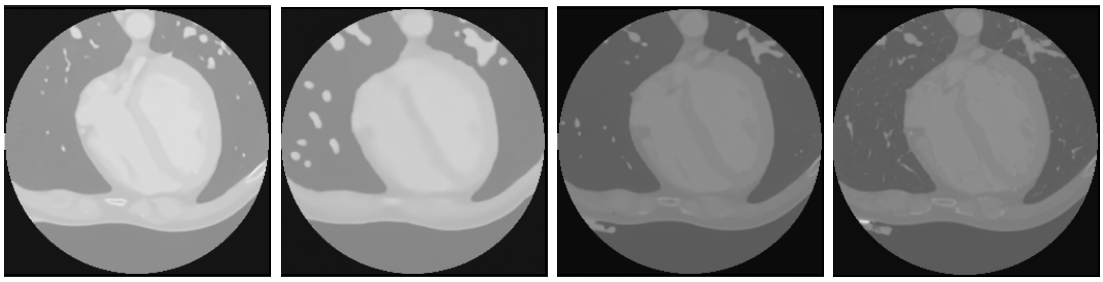
Figura 5.1 – Exemplos de imagens geradas pela base OrcaScore pelos modelos comparados no experimento. Onde a e b são as imagens originais, e de c até h são as imagens de contraste geradas pelo modelo



(a) NECT (b) CECT



(c) CECT CycleGan (d) CECT CycleGan (DOMINGUES, 2022) (e) CECT Pix2Pix (f) CECT Pix2Pix (DOMINGUES, 2022)



(g) CECT ResNet (h) CECT Mobile (i) CECT Cytran (j) CECT Proposto

Fonte: Autoral.

O Pix2pix proposto por Isola et al. (2017) apresenta um PSNR de 15,66 e um SSIM de 0,545, o que reflete uma precisão e qualidade de imagem inferiores. O RMSE é 0,16 e o FID é 210,232, o mais alto da tabela, indicando uma baixa fidelidade na geração de imagens.

A versão do Pix2pix discutida por (DOMINGUES, 2022) apresenta um PSNR de 16,661 e um SSIM de 0,503, mostrando uma leve melhora em comparação à versão original do Pix2pix. O RMSE é 0,16 e o FID é 239,445, sugerindo que, apesar de uma leve melhoria em qualidade, a fidelidade ainda é uma preocupação significativa.

O Cytran de (RISTEA et al., 2023) se destaca com um PSNR de 29,66 e um SSIM de 0,745, indicando uma boa qualidade de imagem. O RMSE é o menor da tabela, com 0,14, e o FID é 58,787, sugerindo alta fidelidade e precisão nas reconstruções das imagens.

O modelo com difusão MobileNet apresenta um PSNR de 18,30 e um SSIM de 0,444, indicando qualidade de imagem inferior. Além disso, o RMSE é 0,22, o mais alto da tabela, e o FID é 180,683, sugerindo baixa fidelidade. Em contraste, o modelo com difusão Resnet mostra melhorias em qualidade de imagem, com um PSNR de 21,55 e um SSIM de 0,523. O RMSE desse modelo é 0,17, e o FID é 150,644, indicando que, embora a fidelidade tenha melhorado, ainda precisaria ser aprimorada.

O modelo proposto com Cycle-Gan e difusão apresenta os melhores resultados no PSNR de 32,85 e o SSIM de 0,766, dentre modelos analisados, sinalizando boa qualidade e precisão nas imagens geradas. o FID é o menor da tabela, com 42,348, refletindo uma maior fidelidade na geração de imagens. Embora o O RMSE é 0,20 seja maior que o do Cytran, ele permanece dentro da margem dos modelos testados, destacando este modelo como o mais eficaz para gerar imagens de alta qualidade e fidelidade.

Um comportamento que pode ser observado nas imagens das 5.1 é que modelos que utilizam a Cycle-GAN tendem a gerar imagens mais escuras, mantendo-se mais próximas às versões originais. Em contraste, os modelos que utilizam apenas GANs, como o Pix2pix, produzem versões mais claras das imagens. Essa diferença sugere que a Cycle-GAN pode preservar melhor a intensidade luminosa das imagens originais, enquanto as GANs simples tendem a alterar esse aspecto, resultando em imagens mais claras.

6 Conclusão

A aplicação de contraste em exames de tomografia computadorizada (TC) para o coração é uma tarefa importante para o diagnóstico de algumas doenças. Métodos de geração de imagem e de sem contraste podem auxiliar os profissionais de saúde no diagnóstico e contribuir para salvar a vida do paciente.

Este trabalho aborda um desafio na área de geração de imagens médicas: a tradução de imagens de TC cardíaca com e sem contraste utilizando uma arquitetura CycleGAN baseada em modelos de difusão. Ao combinar geradores de difusão na arquitetura mencionada, foi possível superar algumas complexidades associadas à tradução de imagens médicas. A combinação de conceitos de GANs e modelos de difusão obteve resultados muito promissores. Avaliados por meio de métricas como PSNR, SSIM e FID, destaca-se a notável capacidade do modelo em gerar imagens cardíacas com contraste, mantendo qualidade e semelhança visual.

No entanto, a análise detalhada do RMSE revela desafios persistentes, sugerindo a presença de variações que demandam uma compreensão mais aprofundada para aprimorar a consistência e fidelidade das imagens geradas. Apesar desses desafios, o modelo desenvolvido entrega resultados significativos, reconhecendo-se, contudo, a necessidade de melhorias contínuas para lidar com as variações nas imagens produzidas. A interseção entre GANs e modelos de difusão mostra-se promissora, abrindo caminho para futuras pesquisas e desenvolvimentos na tradução de imagens médicas, contribuindo assim de forma significativa para o avanço desta área crucial na prática clínica.

Considerando os resultados dos experimentos e a comparação com outros trabalhos, pode-se afirmar que o método consegue realizar a tarefa de forma satisfatória. Compreende-se ainda que os resultados levantam indícios de que o método proposto, em um futuro próximo, pode de fato auxiliar os profissionais de saúde.

Para trabalhos futuros, sugere-se a aplicação das imagens geradas em tarefas de segmentação e classificação, a fim de verificar a validade das imagens para diagnósticos de doenças reais. Nesse sentido, seria interessante realizar uma avaliação das imagens segmentando apenas a região do coração. Ao estudar mais a fundo as imagens, percebeu-se que a maior parte da discrepância de valores vem da região ao redor do coração. Portanto, segmentar apenas a região cardíaca poderia proporcionar uma análise mais precisa e direcionada.

Além disso, outra sugestão para trabalhos futuros é explorar o uso de outros tipos de arquiteturas e formulações para o modelo de difusão. A abordagem utilizada neste trabalho baseou-se em DDPM. No entanto, existem outras arquiteturas e formulações

disponíveis que podem oferecer percepções diferentes ou melhorias adicionais na geração de imagens médicas. Explorar essas alternativas poderia enriquecer ainda mais os resultados e contribuir para avanços significativos na área.

Referências

APICELLA, A. et al. A survey on modern trainable activation functions. *Neural Networks*, Elsevier, v. 138, p. 14–32, 2021. Citado na página 30.

AZARFAR, G. et al. Applications of deep learning to reduce the need for iodinated contrast media for ct imaging: a systematic review. *International Journal of Computer Assisted Radiology and Surgery*, Springer, v. 18, n. 10, p. 1903–1914, 2023. Citado na página 23.

AZEVEDO, C. F.; ROCHITTE, C. E.; LIMA, J. A. Escore de cálcio e angiotomografia coronariana na estratificação do risco cardiovascular. *Arquivos brasileiros de cardiologia*, SciELO Brasil, v. 98, p. 559–568, 2012. Citado na página 25.

BARSNESS, G. W.; HOLMES, D. R. *Coronary Artery Disease: New Approaches Without Traditional Revascularization*. [S.l.]: Springer Science & Business Media, 2011. Citado 2 vezes nas páginas 26 e 27.

BECKETT, K. R.; MORIARITY, A. K.; LANGER, J. M. Safe use of contrast media: what the radiologist needs to know. *Radiographics*, Radiological Society of North America, v. 35, n. 6, p. 1738–1750, 2015. Citado na página 15.

BORJI, A. *Pros and Cons of GAN Evaluation Measures*. 2018. Citado 3 vezes nas páginas 45, 46 e 48.

CHOI, J. W. et al. Generating synthetic contrast enhancement from non-contrast chest computed tomography using a generative adversarial network. *Scientific reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 20403, 2021. Citado 2 vezes nas páginas 21 e 23.

CHUN, J. et al. Synthetic contrast-enhanced computed tomography generation using a deep convolutional neural network for cardiac substructure delineation in breast cancer radiation therapy: a feasibility study. *Radiation Oncology*, Springer, v. 17, n. 1, p. 83, 2022. Citado 3 vezes nas páginas 20, 21 e 23.

COELHO, M. *Fundamentos De Redes Neurais*. 2017. Disponível em: <https://www2.decom.ufop.br/imobilis/fundamentos-de-redes-neurais/#google_vignette>. Citado na página 29.

COMPUTADORIZADA, T. Reações adversas imediatas ao contraste iodado intravenoso em. *Rev Latino-am Enfermagem*, v. 15, n. 1, 2007. Citado 2 vezes nas páginas 15 e 26.

CORBALLIS, N. et al. Ct angiography compared to invasive angiography for stable coronary disease as predictors of major adverse cardiovascular events-a systematic review and meta-analysis. *Heart & Lung*, Elsevier, v. 57, p. 207–213, 2023. Citado na página 15.

COUNSELLER, Q.; ABOELKASSEM, Y. Recent technologies in cardiac imaging. *Frontiers in Medical Technology*, Frontiers, v. 4, p. 984492, 2023. Citado na página 15.

CROITORU, F.-A. et al. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 2023. Citado 2 vezes nas páginas 43 e 44.

DHARIWAL, P.; NICHOL, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, v. 34, p. 8780–8794, 2021. Citado na página 22.

DOMINGUES, R. A. D. Automatic contrast generation from contrastless cts. 2022. Citado 8 vezes nas páginas 21, 23, 25, 40, 54, 55, 56 e 57.

DONDI, M. et al. Integrated non-invasive cardiovascular imaging: a guide for the practitioner. 2021. Citado na página 15.

FERNANDES, J. L.; BITTENCOURT, M. S. Escore de cálcio coronariano: onde e quando faz a diferença na prática clínica. *Rev. Soc. Cardiol. Estado de Säo Paulo*, p. 88–95, 2017. Citado na página 25.

FRAMINHGAM. *Escore de risco global (ERG) de Framingham*. 2021. Disponível em: <https://linhasdecuidado.saude.gov.br/portal/obesidade-no-adulto/unidade-de-atencao-primaria/planejamento-terapeutico/escore-risco-global-framingham/>. Citado na página 24.

GLOROT, X.; BORDES, A.; BENGIO, Y. Deep sparse rectifier neural networks. In: JMLR WORKSHOP AND CONFERENCE PROCEEDINGS. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. [S.l.], 2011. p. 315–323. Citado na página 30.

GOODFELLOW, I. J. et al. *Generative Adversarial Networks*. 2014. Citado 2 vezes nas páginas 20 e 38.

HAGE, M. C. F. N. S.; IWASAKI, M. Imagem por ressonância magnética: princípios básicos. *Ciência Rural*, SciELO Brasil, v. 39, p. 1275–1283, 2009. Citado 2 vezes nas páginas 26 e 27.

HE, K. et al. *Deep Residual Learning for Image Recognition*. 2015. Citado 2 vezes nas páginas 36 e 37.

HELWAN, A. *Introduction to CycleGAN*. 2024. Disponível em: <https://abdulkaderhelwan.medium.com/introduction-to-cyclegan-db1978e8f924>. Citado na página 41.

HO, J.; JAIN, A.; ABBEEL, P. *Denoising Diffusion Probabilistic Models*. 2020. Citado 8 vezes nas páginas 22, 42, 43, 44, 45, 51, 52 e 54.

HOWARD, A. G. et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017. Citado 3 vezes nas páginas 36, 37 e 38.

HUANG, G. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4700–4708. Citado na página 20.

ISOLA, P. et al. *Image-to-Image Translation with Conditional Adversarial Networks*. 2017. 5967-5976 p. Citado 6 vezes nas páginas 40, 41, 42, 54, 55 e 57.

JING, Y. et al. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, IEEE, v. 26, n. 11, p. 3365–3385, 2019. Citado na página 22.

JÚNIOR, E. A.; YAMASHITA, H. Aspectos básicos de tomografia computadorizada e ressonância magnética. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 23, p. 2–3, 2001. Citado 3 vezes nas páginas 16, 25 e 26.

KAMNITSAS, K. et al. Deepmedic for brain tumor segmentation. In: SPRINGER. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 2*. [S.l.], 2016. p. 138–149. Citado na página 33.

KANG, Y.; GAO, S.; ROTH, R. E. Transferring multiscale map styles using generative adversarial networks. *International Journal of Cartography*, Taylor & Francis, v. 5, n. 2-3, p. 115–141, 2019. Citado na página 42.

KAZEMINIA, S. et al. Gans for medical image analysis. *Artificial Intelligence in Medicine*, Elsevier, v. 109, p. 101938, 2020. Citado na página 17.

KJERLAND, Ø. *Segmentation of Coronary Arteries from CT-scans of the heart using Deep Learning*. Dissertação (Mestrado) — NTNU, 2017. Citado 2 vezes nas páginas 32 e 33.

LEITE, K. K. d. A. et al. Ultrassonografia e deglutição: revisão crítica da literatura. *Audiology-Communication Research*, SciELO Brasil, v. 19, p. 412–420, 2014. Citado na página 27.

MATTOS, L. *Sistema Cardiovascular*. 2024. Disponível em: <https://anatomia-papel-e-caneta.com/sistema-cardiovascular/>. Citado na página 24.

MELO, E. P. *Ultrassom intracoronário ou OCT: qual o melhor para avaliar a anatomia coronariana?* 2017. Disponível em: <https://cardiopapers.com.br/ultrassom-intracoronario-ou-oct-qual-o-melhor-para-avaliar-a-anatomia-coronariana>. Acesso em: 02 Janeiro de 2024. Citado na página 28.

NABEL, E. G.; BRAUNWALD, E. A tale of coronary artery disease and myocardial infarction. *New England Journal of Medicine*, Mass Medical Soc, v. 366, n. 1, p. 54–63, 2012. Citado na página 25.

NASCIMENTO, T. P. do. *O Uso do SSIM Como Critério de Convergência em Abordagens Bayesianas Variacionais de Superresolução Multiframe*. 2015. Disponível em: <https://repositorio.ufes.br/server/api/core/bitstreams/ab623148-94f2-4327-814e-e9f7c9073186/content>. Citado 2 vezes nas páginas 46 e 47.

PAHO. *Doenças cardiovasculares*. 2023. Disponível em: <https://www.paho.org/pt/topicos/doencas-cardiovasculares>. Citado na página 15.

PARK, T. et al. Semantic image synthesis with spatially-adaptive normalization. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 2337–2346. Citado na página 20.

PATTANAYAK, S. *Pro Deep Learning with TensorFlow 2.0: A Mathematical Approach to Advanced Artificial Intelligence in Python*. [S.l.]: Springer, 2023. Citado na página 28.

PEEBLES, W.; XIE, S. Scalable diffusion models with transformers. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2023. p. 4195–4205. Citado na página 23.

PINHO, R. A. d. et al. Doença arterial coronariana, exercício físico e estresse oxidativo. *Arquivos Brasileiros de Cardiologia*, SciELO Brasil, v. 94, p. 549–555, 2010. Citado na página 25.

RISTEA, N.-C. et al. Cytran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast ct translation. *Neurocomputing*, Elsevier BV, v. 538, p. 126211, jun. 2023. ISSN 0925-2312. Disponível em: <http://dx.doi.org/10.1016/j.neucom.2023.03.072>. Citado 5 vezes nas páginas 22, 23, 54, 55 e 57.

ROMBACH, R. et al. High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2022. p. 10684–10695. Citado 2 vezes nas páginas 16 e 17.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. [S.l.], 2015. p. 234–241. Citado 2 vezes nas páginas 34 e 35.

SARA, L. et al. Ii diretriz de ressonância magnética e tomografia computadorizada cardiovascular da sociedade brasileira de cardiologia e do colégio brasileiro de radiologia. *Arquivos brasileiros de cardiologia*, SciELO Brasil, v. 103, p. 1–86, 2014. Citado na página 26.

SEO, M. et al. Neural contrast enhancement of ct image. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. [S.l.: s.n.], 2021. p. 3973–3982. Citado 3 vezes nas páginas 20, 21 e 23.

SKANDARANI, Y.; JODOIN, P.-M.; LALANDE, A. Gans for medical image synthesis: An empirical study. *Journal of Imaging*, MDPI, v. 9, n. 3, p. 69, 2023. Citado na página 16.

SOCESP. *A DOENÇA ARTERIAL CORONÁRIA (DAC) CORRESPONDE A UM CONJUNTO DE DOENÇAS*. 2021. Disponível em: <https://socesp.org.br/noticias/area-medica/a-doenca-arterial-coronaria-dac-corresponde-a-um-conjunto-de-doencas/>. Citado na página 24.

STACUL, F. et al. Contrast induced nephropathy: updated esur contrast media safety committee guidelines. *European radiology*, Springer, v. 21, p. 2527–2541, 2011. Citado na página 15.

SUN, M. et al. Learning pooling for convolutional neural network. *Neurocomputing*, Elsevier, v. 224, p. 96–104, 2017. Citado na página 34.

TERVEN, J. et al. Loss functions and metrics in deep learning. a review. *arXiv preprint arXiv:2307.02694*, 2023. Citado na página 30.

VASWANI, A. et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017. Citado 2 vezes nas páginas 31 e 32.

WANG, Z. et al. Diffusion-gan: Training gans with diffusion. *arXiv preprint arXiv:2206.02262*, 2022. Citado na página 23.

WHO. *Noncommunicable diseases*. 2023. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>. Citado na página 15.

WOLTERINK, J. M. *An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework*. 2016. 2361-2373 p. Citado 3 vezes nas páginas 26, 49 e 53.

YU, Y.; ZHANG, W.; DENG, Y. Frechet inception distance (fid) for evaluating gans. *China University of Mining Technology Beijing Graduate School*, 2021. Citado na página 47.

ZHU, J.-Y. et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. 2020. Citado 5 vezes nas páginas 21, 41, 51, 54 e 55.