Universidade Federal de Alagoas
Instituto de Computação

Dissertação de Mestrado

MMI-GAN: Multi Medical Imaging Translation using Generative Adversarial Network

Eduardo Felipe de Souza
eduardofelipe@ic.ufal.com

Orientador:
Prof. Dr. Marcelo Costa Oliveira

Maceió, 2020

Eduardo Felipe de Souza

MMI-GAN: Multi Medical Imaging Translation using Generative Adversarial Network

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Curso de Mestrado em Informática do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. Marcelo Costa Oliveira

Maceió, 2020

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

S729m Souza, Eduardo Felipe de.
MMI-GAN : Multi Medical Imaging Translation using Generative Adversarial Network / Eduardo Felipe de Souza. – 2020.
57 f. : il.

Orientador: Marcelo Costa Oliveira.
Dissertação (mestrado em Informática) - Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2020.

Bibliografia: f. 51-56.

1. *Generative Adversarial Network*. 2. Tradução de imagens. 3. Imagem por ressonância magnética. 4. Tomografia computadorizada. I. Título.

CDU: 004.932:543.429.2

UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Informática – PPGI
Instituto de Computação/UFAL
Campus A. C. Simões BR 104-Norte Km 14 BL 12 Tabuleiro do Martins
Maceió/AL - Brasil CEP: 57.072-970 | Telefone: (082) 3214-1401

Folha de Aprovação

EDUARDO FELIPE DE SOUZA

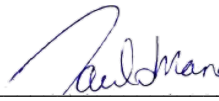
MMI-GAN: MULTI MEDICAL IMAGING TRANSLATION USING GENERATIVEADVERSARIAL NETWORK

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Informática da Universidade Federal de Alagoas e aprovada em 27 de NOVEMBRO de 2020.

Banca Examinadora:

Prof. Dr. MARCELO COSTA OLIVEIRA
UFAL – Instituto de Computação
Orientador

Prof. Dr. TIAGO FIGUEIREDO VIEIRA
UFAL – Instituto de Computação
Examinador Interno



Prof. Dr. PAULO MAZZONCINI DE AZEVEDO MARQUES
USP – Universidade de São Paulo
Examinador Externo

Agradecimentos

À Deus por estar sempre comigo.

À minha família por serem responsáveis por todo apoio nos meus estudos.

À minha namorada, Amanda Feitosa, pelo seu amor, carinho, paciência e apoio nos meus estudos.

Aos meus amigos e companheiros de curso: Lucas Lins, André Moabson e Bruno dos Anjos que me ajudou nas correções.

Ao meu orientador Marcelo pela oportunidade dada em seu laboratório, pela paciência nos ensinamentos e dedicação ao meu progresso nos estudos. Aos professores que aceitaram o convite de participarem da minha banca, Tiago e Paulo. Ao Anderson da secretaria do IC por todo auxílio e paciência.

A todos que me ajudaram direta ou indiretamente e torceram por mim.

Resumo

A tradução de imagens médicas é considerada uma nova fronteira no campo da análise de imagens médicas, com grande potencial de aplicação. No entanto, as abordagens existentes têm escalabilidade e robustez limitadas no manuseio de mais de dois domínios de imagens, uma vez que diferentes modelos devem ser criados independentemente para cada par de domínios. Para resolver essas limitações, desenvolvemos a MMI-GAN, uma nova abordagem para tradução entre múltiplos domínios de imagem, capaz de traduzir imagens intermodais (TC e RM) e intramodais (PD, T1 e T2) usando apenas um único gerador e um discriminador, treinados com dados de imagens de todos os domínios. Propomos uma arquitetura GAN que pode ser facilmente estendida a outras tarefas de tradução para o benefício da comunidade de imagens médicas. A MMI-GAN baseia-se nos avanços recentes na área das GANs (Generative Adversarial Network), utilizando uma estrutura adversária com uma nova combinação de perdas não adversárias, que permite o treino simultâneo de vários conjuntos de dados com diferentes domínios numa mesma rede, bem como a capacidade inovadora de traduzir com flexibilidade entre e intra/inter modalidades. As imagens traduzidas pelo MMI-GAN conseguiram obter MAE de 5.792, PSNR de 27.398, MI de 1.430 e SSIM de 0.900. Os seus resultados se mostraram, por muitas vezes estaticamente equiparáveis ou superiores a Pix2pix e em quase todas as traduções foi superior a Cyclegan.

Palavras-chaves: Generative Adversarial Networks, Tradução de Imagens, multi domínio, Ressonância Magnética, Tomografia Computadorizada.

Abstract

Medical image translation is considered a new frontier in the field of medical image analysis, with great potential for application. However, existing approaches have limited scalability and robustness in handling more than two image domains, since different models must be created independently for each pair of domains. To address these limitations, we developed MMI-GAN, a new approach for translation between multiple image domains, capable of translating intermodal (CT and RM) and intramodal (PD, T1 and T2) images using only a single generator and a discriminator, trained with image data from all domains. We propose a GAN architecture that can be easily extended to other translation tasks for the benefit of the medical imaging community. MMI-GAN is based on recent advances in the area of GANs (Generative Adversarial Network), using an adversary structure with a new combination of non-adversarial losses, which allows the simultaneous training of several data sets with different domains in the same network, as well as the innovative capacity to translate with flexibility between and inter/intra modalities. The images translated by MMI-GAN managed to obtain MAE of 5.792, PSNR of 27.398, MI of 1.430 and SSIM of 0.900. Its results were shown, often statically comparable or superior to Pix2pix and in almost all translations it was superior to Cyclegan.

Keywords: Generative Adversarial Networks, image translation, multi-domain, magnetic resonance, computed tomography.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

TC	Tomografia Computadorizada
RM	Ressonância Magnética
PET	do inglês *Positron Emission Tomography*
GAN	do inglês *Generative Adversarial Network*
PSNR	do inglês *Peak Signal-to-Noise Ratio*
SSIM	do inglês *Structural Similarity Index Measure*
PD	do inglês *Proton Density*
DL	do inglês *Deep Learning*
FCN	do inglês *Fully Connected Network*
IA	Inteligência artificial
CPU	do inglês *Central Process Unit*
GPU	do inglês *Graphics Processing Unit*
ML	do inglês *Machine Learning*
FDG	Fluoro-D-glicose
PIB	do inglês *Pittsburg compound B*
RF	Radio Frequência
CNN	do inglês *Convolutional Neural Network*
MRA	do inglês *Magnetic Resonance Angiography*

Sumário

1 Introdução

imagiologia médica é a técnica capaz de representar visualmente o corpo humano para fins de diagnóstico e de tratamento. É considerado um campo em constante mudança, com avanços significativos em seus métodos e tecnologias ao longo dos anos (IGLEHART, 2006). Vários métodos de imagem são utilizados para capturar informações sobre órgãos e tecidos, como radiografia, tomografia computadorizada, ressonância magnética, ultrassom, tomografia por emissão de pósitrons, entre outros. Os princípios físicos aplicados na obtenção das diferentes modalidades de imagem possuem características distintas que resultam em exames de imagem anatomicamente específicos, com diferentes dimensões (2D e 3D), resolução de contraste e resolução espacial (ARMANIOUS et al., 2020).

Dados os aspectos inerentes a cada modalidade, em muitos casos, o especialista define o diagnóstico ou tratamento do paciente a partir de informações contidas em duas ou mais modalidades de imagem combinadas. Um exemplo é a utilização de RM e TC na radioterapia, em que o especialista utiliza imagens de RM na análise de tecidos moles, devido à sua alta resolução anatômica, em conjunto com imagens de TC, que detalham as estruturas ósseas. Portanto, as técnicas fornecem ao especialista uma melhor discriminação dos tecidos e ajudam a avaliar os limites da infiltração tumoral (TANAKA et al., 2011) (ĐAN et al., 2013). Também há casos em que diferentes contrastes de uma mesma modalidade de imagem são usados para aumentar as informações diagnósticas, por exemplo, diferentes ponderações de RM (T1 e T2). As ponderação geram contrastes diferentes de uma mesma região anatômica, por exemplo, ao diagnosticar um tumor cerebral, o especialista usa imagens T1 para analisar a substância cinzenta e branca, enquanto T2 é usado para localizar tumores (JIN et al., 2019).

Portanto, em determinados casos a precisão do diagnóstico e a qualidade do tratamento do paciente estão diretamente relacionados a análise de diferentes modalidades de imagens médicas. Contudo, devido ao alto custo dos equipamentos de imagem (e.g., RM), o acesso a determinados exames é uma realidade distante para grande parte da população mundial que vive em países em desenvolvimento. Além disso, aspectos clínicos também inviabilizam a realização de exames de diferentes modalidades de imagem. Por exemplo, RM não é adequada para paciente com claustrofobia, pacientes portadores de marca-passo, próteses ortopédicas, etc; enquanto na TC, o especialista deve solicitar os exames respeitando o risco associado à exposição à radiação ionizante (KRUPA; BEKIESIŃSKA-FIGATOWSKA, 2015; LAM et al., 2015; LIMA; JUNIOR; OLIVEIRA, 2020). Neste contexto é clinicamente desejável gerar

imagens ausentes (e.g., TC) a partir de imagens existentes (e.g., RM), técnica essa conhecida como tradução de imagens (Fig. 1.1).

A técnica de tradução de imagens captura imagens de um domínio e as transforma para que tenham o estilo e as características das imagens do domínio de destino. Em analogia à tradução automática de linguagem, definimos a tradução image-to-image como a tarefa de traduzir uma representação possível de uma imagem em outra (ISOLA et al., 2017).

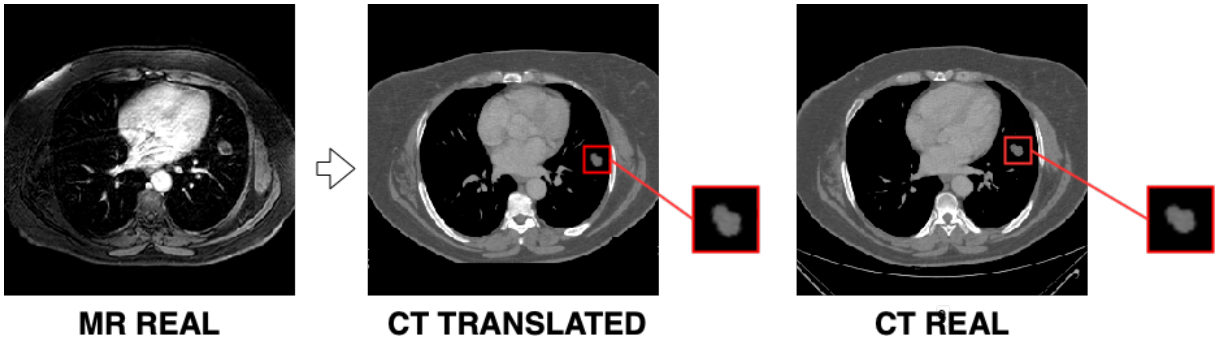


Figura 1.1: Exemplo de como a técnica de tradução pode prever os pixels de um nódulo em uma imagem de TC do tórax a partir de pixels de uma imagem de RM correspondente. Fonte: elaborado pelo autor.

Proposta por Goodfellow et al. (GOODFELLOW et al., 2014), as GANs são compostas por duas redes neurais, onde uma rede discriminadora tenta distinguir imagens reais de falsas, enquanto uma rede geradora aprende a produzir imagens sintéticas tão reais que impedem a rede discriminadora de julgá-las como falsas. Diversos modelos de GANs (LIU; BREUEL; KAUTZ, 2017; YI et al., 2017; LIU, 2019; LEI et al., 2019) foram propostos para a tradução de imagens, mas os modelos Pix2pix (ISOLA et al., 2017) e CycleGAN (ZHU et al., 2017) são os modelos mais utilizados (YI; WALIA; BABYN, 2019). A Pix2pix realiza a tradução do domínio de imagem A para B de forma supervisionada usando pares de imagens alinhadas, enquanto a CycleGAN realiza a tradução de forma não supervisionada, não utilizando dados pareados e alinhados no seu treinamento. Outra vantagem da CycleGAN é a sua capacidade de traduzir do domínio de imagem A para B e vice versa, pois possui dois geradores na sua arquitetura, diferentemente da Pix2pix que contém apenas um.

A tradução de imagens médicas foi investigada tanto em estudos intramodais (por exemplo, T1 e T2 RM) e em estudos intermodais (por exemplo, TC e RM). Em Dar et al (DAR et al., 2019), os autores usaram GANs com funções de perda baseadas em Pix2pix e Cyclegan para traduzir RM ponderada em T1 e T2 da região anatômica da cabeça. Além disso, os autores compararam os modelos de GANs com os métodos Multimodal (CHARTSIAS et al., 2017b) e Replica (JOG et al., 2017). Os experimentos foram avaliados em três bancos de dados diferentes: MIDAS (BULLITT et al., 2005), conjunto de dados IXI (BRAIN-IXI, 2020) e conjunto de dados BRATS (Menze et al., 2015). Usando Cyclegan, os autores obtiveram os melhores resultados para a relação sinal-ruído de pico (PSNR) de 27,37 (T1 para T2) e 27,33 (T2 para T1);

e usando Pix2pix obtiveram 27,51 (T1 a T2) e 27,37 (T2 a T1). Já com o Replica obtiveram 25,30 (T1 a T2) e 24,43 (T2 a T1); e usando Multimodal obtiveram 24,61 (T1 a T2) e 24,69 (T2 a T1). Os resultados de PSNR do método proposto usando o GAN superaram o método Multimodal e o método de réplica na tradução de ressonância magnética ponderada em T1 e T2. Em Yang et al. (YANG et al., 2020), os autores usaram uma arquitetura Pix2pix com uma rede U-Net (RONNEBERGER; FISCHER; BROX, 2015) como um gerador para tradução entre diferentes pesos RM (T1, T2, PD e T2-FLAIR) da cabeça. Os autores usaram cinco bancos de dados públicos (BRATS, Iseg2017 (WANG et al., 2019), MRBrain (MENDRIK et al., 2015), ADNI (INITIATIVE, 2020), RIRE (RIREDATASET, 2020)). Os melhores resultados obtidos foram PSNR de 28,99 (T1 a T2) e 24,04 (T2 a T1).

Em Lee et al. (LEE; MOON; YE, 2020), os autores utilizaram apenas um gerador e um discriminador baseado na rede StarGAN (CHOI et al., 2018) para a tradução multi-domain entre imagens de RM de T1-FLAIR, T2, T2-FLAIR e MAGiC T2-FLAIR (RM sintética) da cabeça utilizando um dataset privado. Contudo, diferente da StarGAN original que utiliza uma imagem por vez como entrada, os autores utilizaram três entradas pareadas de diferentes domínios de um total de quatro domínios, com isso a imagem do domínio ausente foi traduzida a partir das três imagens passadas como entrada. Aprendendo as características comuns aos pares de imagens, a rede foi treinada para estimar a imagem ausente combinando as informações das várias entradas. Os autores obtiveram o melhor resultado para tradução de MAGiC T2-FLAIR com 0.94 de SSIM (1.00 é o valor máximo para SSIM).

No contexto intermodal, Nie et al. (NIE et al., 2017) traduziu imagens de RM em imagens de TC utilizando uma pilha/cascata de redes GANs 3D para melhorar o realismo das imagens tomográficas traduzidas. A ideia de utilizar uma pilha de redes de geradores tem origem no chamado modelo de Auto-Contexto (TU, 2008), no qual uma rede fornece sua saída como entrada adicional para uma rede sucessora, com o objetivo de fornecer informações de contexto e permitir refinamentos. Utilizando esta abordagem os autores obtiveram PSNR de 34.10. Porém, o trabalho de Nie et al. requer pares correspondentes de imagens de TC e RM para treinamento, contudo, isso é um problema na área de imagens médicas, pois na maioria das vezes os exames não são pareados. Em Wolterick et al. (WOLTERINK et al., 2017), os autores utilizaram uma CycleGAN para traduzir imagens 2D de RM em imagens TC da cabeça vice-versa, sem a necessidade de utilizar imagens de treinamento pareadas e registrados. A arquitetura da Cyclegan proposta obteve PSNR de 32.30, segundo os autores, a adoção de dados não pareados resultou numa melhora dos resultados de 1.70 (PSNR) em comparação ao uso de dados pareados.

Entretanto, as redes Pix2pix e CycleGAN foram projetadas para uma aplicação específica ou com uma capacidade limitada de modelagem, ou seja, para aprender a traduzir entre os k domínios de imagens médicas envolvidas, é necessário treinar k*(k - 1) geradores, por exemplo: se quisermos traduzir entre imagens de TC, T1, T2, proton density (PD) e T2-FLAIR, seriam necessários 20 geradores diferentes, quanto mais domínios envolvidos mais geradores são necessários. Outra limitação das redes Pix2pix e CycleGAN é o fato dos geradores não utili-

zarem completamente as informações globais comuns que podem ser aprendidas com imagens de todos os domínios (e.g bordas). Nas redes Pix2pix e Cyclegan os geradores não utilizam completamente todos os dados de treinamento, pois os geradores apenas aprendem usando pares de domínios utilizados pelo gerador, limitando a qualidade das imagens traduzidas (CHOI et al., 2018).

1.1 Objetivo Geral e Específico

Esta dissertação de mestrado tem como objetivo geral desenvolver um modelo computacional para a tradução de imagens médicas entre diferentes domínios utilizando redes GANs.

Como objetivos específicos pretendemos responder as seguintes perguntas de pesquisa:

1. O processo de tradução usando GANs é invariante ao pareamento intra/inter modalidade de exames de imagens médicas?
2. O processo de tradução usando GANs é invariante ao processo de registro de imagens médicas?
3. A abordagem multidomínio é capaz de obter resultados tão bons quanto as GANs mais relevantes?

1.2 Estrutura do Trabalho

A estrutura da dissertação está da seguinte forma:

- **Capítulo 2 - Trabalhos Relacionados:** Este capítulo apresenta os trabalhos relacionados na literatura que trabalharam tradução de imagens médica;
- **Capítulo 3 - Fundamentação Teórica:** Este capítulo apresente os principais conceitos utilizados neste trabalho como a preprocessamento das imagens médicas, bem como as técnicas utilizadas na tradução de imagens médicas;
- **Capítulo 4 - Materiais e Métodos:** Este capítulo apresenta como foi desenvolvido o modelo proposto para a classificação dos nódulos pulmonares;
- **Capítulo 5 - Resultados e Discussão:** Este capítulo apresenta os resultados obtidos e a discussão em torno deles e perante aos resultados obtidos pelos trabalhos relacionados;
- **Capítulo 6 - Conclusão:** Este capítulo finaliza este trabalho apresentando as conclusões e os planejamentos futuros.

2 Trabalhos Relacionados

A tradução de imagens médicas pode ser utilizada para diversas aplicações com diferentes tipos de imagens médicas, essas aplicações estão listadas na tabela 2.2.

2.1 Traduções de imagens de RM para TC

Em muitos cenários clínicos é de extrema importância a aquisição de imagens de TC. Isso no entanto coloca o paciente em risco de dano celular e câncer devido a exposição à radiação ionizante, o que motiva a tradução de imagens de TC a partir de RM. Em (NIE et al., 2017) foram traduzidas imagens de TC de imagens de RM correspondentes com a ajuda de uma pilha/cascata de redes GANs 3D totalmente convolucionais que foram treinadas com uma perda de reconstrução normal, perda de gradiente de imagem e adicionalmente com uma rede adversária para melhorar o realismo das imagens tomográficas sintéticas. A ideia de utilizar uma pilha de redes de geradores tem origem no chamado modelo de Auto-Contexto, no qual uma rede fornece sua saída como entrada adicional para uma rede sucessora, a fim de fornecer informações de contexto e permitir refinamentos.

Diferentemente de (NIE et al., 2017) que requer pares correspondentes de imagens de TC e RM para treinamento, (WOLTERINK et al., 2017) utiliza uma CycleGAN para transformar imagens de RM 2D em imagens de TC e vice-versa sem a necessidade de dados de treinamento emparelhados e co-registrados, porém desalinhamento entre imagens emparelhadas pode levar a erros nas imagens de TC sintetizadas. Em (ZHAO et al., 2018), os autores utilizaram GANs

ID	Tipo	Requerimentos	Descrição
1	Adversaria	Não possui requerimentos	Perda adversária introduzida pelo discriminador,assumindo a forma de perda de entropia cruzada, perda de dobradiça, perda de mínimos quadrados
2	Cíclica	Não possui requerimentos	Perda usada para garantir a similaridade durante a transformação em ciclo quando o par de treinamento não alinhado é fornecido
3	Imagem	Imagens de treinamento pareadas	Perda para garantir a similaridade da estrutura com o alvo
4	Estilo	Imagens de treinamento pareadas	Perda para garantir a similaridade do estilo e do conteúdo da imagem
5	Perceptual	Imagens de treinamento pareadas	Perda baseado em um domínio de recurso computado a partir de uma rede pré-treinada que espera estar em conformidade com a percepção visual
6	Borda	Imagens de treinamento pareadas	Semelhante a perda de gradiente, mas usa o mapa de recurso de gradiente como um peso para os pixels da imagem.
7	Recurso tumoral	Não possui requerimentos	Força os recursos de alto nível do tumor real e do traduzido a serem compartilhados
8	Tumoral	Não possui requerimentos	Limita a TC e a Unet à base de MRI sintética a produzir segmentações tumorais semelhantes, preservando assim os tumores
9	Gradiente	Dados pareados	Tenta manter as zonas com gradientes fortes (por exemplo, arestas)

Tabela 2.1: Resumo das perdas utilizadas nos artigos revisados

Artigo	Tradução	GAN	Perda	Métrica(s)	Dados
(NIE et al., 2017)	MRI - TC	Stack GAN	1, 3, 9	PSNR, MAE	Pareados
(WOLTERINK et al., 2017)	MRI - TC e TC - MRI	CycleGAN	1, 2	PSNR, MAE	Não pareados
(ZHAO et al., 2018)	MRI - TC	Stack GAN	5	PSNR, MAE	Pareados
(JIANG et al., 2018)	MRI - TC E TC - MRI	CycleGAN	1, 2, 7, 8	Segmentação	Não pareados
(JIN et al., 2018)	TC - MR	CycleGAN	1, 2, 3	PSNR, MAE	Pareados e não pareados
(CHARTSIAS et al., 2017a)	MR - TC e TC - MR	CycleGAN	1,2	Segmentação	Não pareados
(HIASA et al., 2018)	MR - TC	CycleGAN*	1, 2, 9	Segmentação, Informação mútua	Não pareados
(BI et al., 2017)	TC - PET	FCN + cGAN	1,3	PSNR, MAE, Detecção	Pareados
(BEN-COHEN et al., 2019)	TC - PET	cGAN	1,3	PSNR, MAE	Pareados
(WEI et al., 2018)	MR - PET	CascadeGAN	1,3	Voxels desmielinizados	Pareados
(DAR et al., 2019)	T1 - T2 e T2 - T1	CycleGAN	1,2	PSNR, SSIM	Não pareados
(YANG et al., 2018)	T1 - T2 e T2 - T1	cGAN	1,3	PSNR, MAE, informação mútua, segmentação, Registro de modalidade cruzada	Não pareados
(NIE et al., 2018)	3T - 7T MR	Cascade GAN	1, 3, 9	PSNR, MAE	Pareados
(OLUT et al., 2018)	T1, T2 - MRA	Pix2Pix*	1,3	PSNR, segmentação	Pareados
(YU et al., 2018)	T1 - FLAIR RM	cGan	1,3	PSNR, MAE, Segmentação	Pareados

Tabela 2.2: Resumo dos artigos suas traduções, métricas, GANs utilizadas e suas perdas.

condicionais para mapear dados de RM 3D da cabeça para sua TC equivalente, utilizando a mesma ideia de pilha de geradores utilizada em (NIE et al., 2017). Para validar a tradução realizada, as imagens traduzidas foram utilizadas na segmentação segmentação de estruturas ósseas do crânio a partir de dados anatômicos utilizando uma Unet 3D, sem expor o paciente à radiação. Para obter resultados viáveis de tradução de imagem para imagem, (NIE et al., 2017) propuseram a chamada discriminação de supervisão profunda, que similarmente à perda perceptual (ZHANG et al., 2018), utiliza as representações características de um modelo VGG16 (*transfer Learning*) para distinguir imagens de TC reais e sintéticas.

Assim como em (WOLTERINK et al., 2017) onde não existem dados pareados, (CHARTSIAS et al., 2017a) demonstrou o potencial para a síntese de imagens médicas usando uma arquitetura CycleGAN. Como não existe uma forma de avaliação direta das imagens sintéticas, já que não existem imagens verdadeiras, sua qualidade foi testada aproveitando os dados sintéticos gerados para obter melhores resultados na segmentação, sendo demonstrado que o treinamento com dados reais e sintéticos aumentou a precisão em 15% em comparação com dados reais.

Uma limitação importante das GANs é a falta de garantia de que os tumores ou lesões presentes em uma imagem sejam preservados durante a tradução de uma modalidade para a outra. Para lidar com esse problema, em (JIANG et al., 2018) foi proposto uma CycleGAN utilizando uma função de perda com conhecimento do tumor, além do objetivo de consistência do ciclo para traduzir imagens de RM a partir de imagens de TC com tumores. Depois de treinada, a GAN é utilizada para aumentar o conjunto de treinamento das imagens de RM originalmente muito pequeno para o treinamento de um modelo de segmentação de tumor baseado em Unet, demonstrando melhorias significativas nos resultados das segmentações.

Na radioterapia o uso de RM pode ser limitado devido ao aumento do uso de implantes

metálicos como marca-passos cardíacos e articulações artificiais, assim sendo, a TC é mais recomendada para esses casos, além disso, possui maior resolução de imagem, e menor artefato de movimento devido a sua alta velocidade de aquisição. Para melhorar a precisão do planejamento radioterápico baseado em TC, em (JIN et al., 2018) foi proposto uma abordagem sintética para produzir imagens de RM traduzidas a partir de imagens de TC do cérebro. Utilizando dados pareados e não pareados para resolver o problema de desalinhamento do contexto de treinamento não pareado, e aliviar a tarefa de registro rígido e os resultados confusos do treinamento emparelhado. Sendo capaz de traduzir de forma eficiente estruturas dentro de fatias cerebrais 2D complexas, como vasos cerebrais e ossos.

A RM é usada para diagnosticar a osteonecrose devido ao seu contraste superior nos tecidos moles, no entanto a ressonância magnética tem contraste pobre para estruturas ósseas, para delinear melhor as estruturas ósseas uma TC para auxiliar o diagnóstico é necessária. Para suprir essa necessidade de um TC complementar, (HIASA et al., 2018) propôs uma GAN que estende a Cyclegan e é capaz de traduzir MR para TC, adicionando a perda de consistência de gradiente para melhorar a precisão nos limites (bordas) e avaliando a tradução das imagens investigando a dependência da acurácia da síntese de imagens baseado no número de dados de treinamento e na incorporação da perda de consistência do gradiente na segmentação em imagens traduzidas.

2.2 Traduções de imagens de TC para PET

O uso de PET em conjunto com TC tornou-se um componente padrão de diagnóstico e tratamento em oncologia. O acúmulo de Fluoro-D-glicose (FDG) em PET em relação ao tecido normal é um marcador útil para muitos cânceres e pode ajudar na detecção e localização de lesões malignas. Além disso, a imagem PET/TC está se tornando uma importante ferramenta de avaliação para novas terapias medicamentosas (CIERNIK et al., 2003). (BI et al., 2017) propôs um novo método para traduzir dados de PET através de GAN de multicanais (M-GAN) para lidar com a tradução de TC para PET. A abordagem M-GAN tem a capacidade de capturar representações de recursos com um alto nível de informações semânticas, além de ser capaz de captar a entrada anotada (rótulo) para sintetizar regiões de alta captação, por exemplo, tumores em TC para restringir a consistência da aparência baseada nas informações anatômicas derivadas de TC. Os dados experimentais de 50 estudos de PET-TC com câncer de pulmão mostraram que a M-GAN fornece imagens PET mais realistas em comparação com os métodos convencionais de GAN. Além disso, o modelo de detecção de tumores de PET, treinado com os dados de PET sintético, apresentaram desempenho competitivo quando comparado ao modelo de detecção treinado com dados reais de PET (2,79% menor em termos de revocação).

Em (BEN-COHEN et al., 2019) foi apresentado um novo modelo para geração de imagens PET traduzidas a partir de TC utilizando redes FCN e GAN para gerar dados PET sintéticos a

partir de dados de entrada de TC, onde a GAN refina a saída sintetizada extraída da FCN. O PET sintetizado foi utilizado para redução de falso positivo em soluções de detecção de lesões. A avaliação quantitativa foi realizada utilizando um software de detecção de lesões existente. Os resultados mostraram uma redução de 28% na média de falsos positivos.

2.3 Traduções de imagens de RM para PET

A esclerose múltipla é uma doença desmielinizante do sistema nervoso central . Uma medida confiável do conteúdo de mielina tecidual é, portanto, essencial para entender a fisiopatologia da EM, acompanhar a progressão e avaliar a eficácia do tratamento. A tomografia por emissão de pósitrons com PIB foi proposta como um biomarcador promissor para medir as alterações no conteúdo de mielina (MCDONALD et al., 2001). No entanto a imagem de PET é invasiva devido à injeção de um traçador radioativo, já a MR é uma técnica não invasiva e amplamente disponível, mas as sequências de RM existentes não fornecem, um marcador confiável, específico ou direto de desmielinização ou remielinização. Em (WEI et al., 2018) foi proposto uma cascata de duas GANs condicionais chamada desenhista-refinadora, onde essas GANs possuem funções de perda adversarial projetada especificamente para prever o mapa de conteúdo de mielina traduzindo PET a partir de um conjunto de diferentes modalidades de RM utilizando uma Unet 3D para redes geradoras e redes discriminadoras com convoluções 3D. O problema de tradução é resolvido por um processo de refinamento de esboço no qual a GAN desenhista gera a informação anatômica e fisiológica preliminar e a GAN refinador refina e gera imagens refletindo o conteúdo de mielina tecidual no cérebro humano.

2.4 Traduções de imagens de T1 para T2

Dadas as dificuldades encontradas em exames prolongados devido a repetidas aquisições, apenas um subconjunto de contrastes pode ser coletado com qualidade adequada, particularmente em pacientes pediátricos e idosos. A síntese multi-contraste de ressonância magnética pode ser útil nas piores situações, onde as imagens foram corrompidas ou mesmo estão indisponíveis. Em (DAR et al., 2019) dois métodos de tradução de ressonância magnética multi-contraste baseado em cGANs foram propostos, sendo uma para o caso das imagens de T1 e T2 registradas e outro para não registradas. O uso de funções de perda contraditória juntamente com as perdas perceptuais e de pixels no caso de imagens registradas, e uma perda de consistência de ciclo para imagens não registradas melhoraram ainda mais a síntese. Baseadas nas CycleGANs, o método proposto aproveita as informações das seções transversais vizinhas em cada volume para aumentar a precisão da síntese.

Em (YANG et al., 2018) a estrutura explora conjuntamente informações em pixels e as representações de alto nível como tumores cerebrais, estrutura cerebral e etc. Primeiro foi pro-

posto um método para o registro de modalidade cruzada, fundindo os campos de deformação para adotar as informações de modalidade cruzada de modalidades traduzidas. Em segundo lugar, foi proposto uma abordagem para a segmentação por ressonância magnética, a segmentação multicanal traduzida, onde as modalidades dadas juntamente com as modalidades traduzidas, são segmentadas por redes totalmente convolucionais (FCN) em uma maneira multicanal. Ambos métodos adotam com sucesso as informações de modalidade cruzada para melhorar o desempenho sem adicionar dados extras.

2.5 Traduções de imagens de T1 para FLAIR RM

Diferentes modalidades de RM podem indicar alterações teciduais induzidas por tumores a partir de diferentes perspectivas, beneficiando a segmentação do tumor cerebral quando consideradas em conjunto. Clinicamente, a RM ponderada em T1 é a modalidade de imagem de RM mais comumente utilizada, embora possa não ser a melhor opção para o contorno do tumor cerebral (HAACKE et al., 1999). (YU et al., 2018) utiliza cGAN 3D para a síntese de imagens FLAIR e um método de fusão adaptativa local para melhor representar os detalhes das imagens FLAIR sintetizadas. O método proposto pode efetivamente lidar com a tarefa de segmentação de tumores cerebrais que variam em aparência, tamanho e localização entre as amostras. As imagens finais sintetizadas, juntamente com as imagens T1, são processadas por um modelo CNN 3D para segmentação do tumor cerebral.

2.6 Traduções de imagens de T1, T2 para MRA

A angiografia por ressonância magnética tornou-se um contraste essencial da RM para imagens e avaliação da anatomia vascular e de doenças relacionadas. As aquisições de MRA são tipicamente solicitadas para intervenções vasculares (SAILER et al., 2013), mas em cenários típicos, as sequências de MRA podem estar ausentes nas varreduras de pacientes. Em (OLUT et al., 2018) é proposta a chamada GAN dirigível para sintetizar imagens de MRA de exames de RM ponderados em T1 e T2, onde a GAN condicional e dirigível combina um gerador tipo ResNet (HE et al., 2016) com um discriminador PatchGAN, uma perda do tipo l1 entre imagens reais e sintetizadas, bem como uma perda de filtro direcionável para promover reconstruções fiéis de estruturas vasculares.

De todos os tipos de imagens médicas contidas na revisão a MR foi a modalidade de imagem mais comum utilizada no processo de tradução utilizando GANs. Isso se deve em parte pelo fato de imagens de RM possuírem múltiplas sequências que são rotineiramente adquiridas para fornecer informações complementares, devido ao tempo de aquisição significativo para adquirir cada sequência, as GANs mantêm o potencial de reduzir o tempo de aquisição da RM se o número de sequências adquiridas puder ser reduzido.

Juntamente com todas as utilidades positivas das GANs apresentadas durante toda a revisão, a literatura existente também destaca algumas deficiências e desafios na utilização de GANs em imagens médicas.

(FLORKOW et al., 2019) demonstra o impacto negativo causado pela presença de erros de registro na fase de treinamento de uma GAN, treinada em uma tarefa de tradução de RM para TC. O desempenho do modelo foi degradado como resultado de uma crescente proporção de erros de registro no conjunto de treinamento, onde um único par mal registrado no conjunto de treinamento foi capaz de alterar drasticamente o desempenho da tradução das imagens quantitativa e qualitativamente.

(COHEN; LUCK; HONARI, 2018) chama atenção contra o uso de imagens geradas para interpretação médica. Foi observado que as redes CycleGANs podem estar sujeitas a viés devido à correspondência da distribuição de dados do domínio de destino, que é adquirida a partir dos dados de treinamento, e que pode ser bem diferente da distribuição de dados de teste. Além dos dados não pareados também foi observado o viés com GAN condicional (para dados pareados) quando os dados fornecidos no domínio de destino têm uma representação excessiva ou insuficiente de algumas classes.

3 Fundamentação Teórica

3.1 Aquisição de imagens médicas

3.1.1 TC

A TC (Fig. 3.1) começou a ser utilizada como método diagnóstico no início da década de 1970, teve sua difusão nos anos 1980 e, atualmente, é cada vez mais utilizada na prática médica. Desde seu aparecimento até os dias atuais vem sofrendo muitas evoluções, com a criação de aparatos (*scanners*) cada vez mais complexos, contribuindo para a redução no tamanho dos aparelhos, diminuição do tempo de aquisição, melhoria na qualidade da imagem, novas aplicações e maior flexibilidade no tratamento dos dados (MOURÃO; OLIVEIRA, 2018).

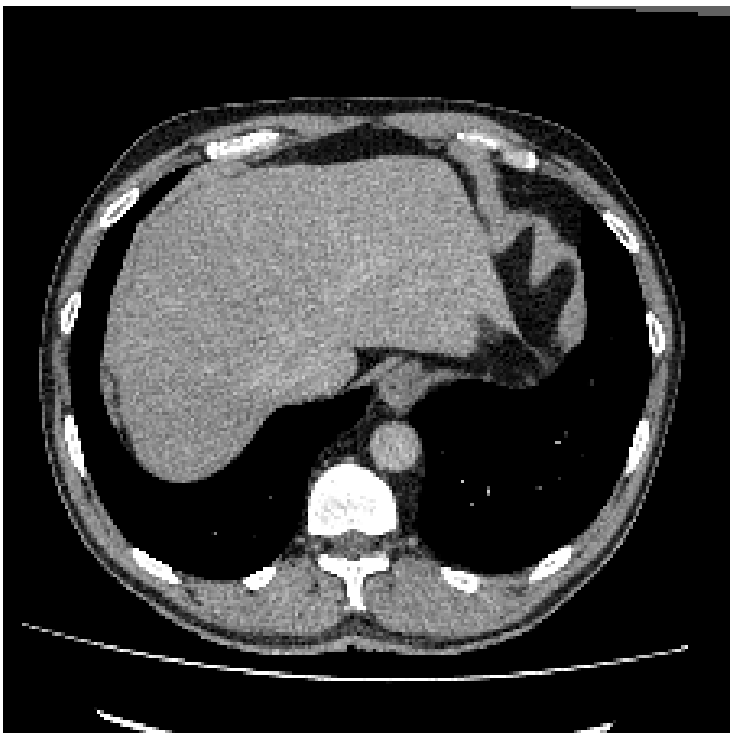


Figura 3.1: Exemplo de uma imagem de TC. Fonte: elaborado pelo autor.

No aparelho de tomografia um feixe de raios X é gerado em um dos lados do paciente (Fig. 3.2). O feixe de raios X é atenuado por absorção e dispersado à medida que passa pelo paciente. Detectores sensíveis no lado oposto do paciente medem a transmissão de raios X através do

corte. Essas medições são repetidas sistematicamente de diferentes direções, enquanto o tubo de raios X é pulsado à medida que gira 360° em torno do paciente. Valores de TC são atribuídos para cada pixel da imagem por meio de um algoritmo computacional, que usa como dados essas medições dos raios X transmitidos. Os valores de pixel são proporcionais à diferença na média entre a atenuação dos raios X do tecido no voxel e a da água. Quanto maior a absorção do feixe pelo tecido, mais branco esse tecido aparece na imagem, uma vez que há grande absorção e pouca radiação ultrapassa o objeto, e quanto menor a absorção do feixe pelo tecido, mais preto ele se apresenta na imagem. A unidade empregada é o Hounsfield (H), em homenagem a Godfrey Hounsfield, inventor da TC. A água tem valor de 0 H na escala Hounsfield, que vai de –1.024 H para o ar até +3.000 a 4.000 H para osso muito denso. Em geral, o tecido ósseo varia entre +400 H e +1.000 H; os tecidos moles entre +40 H e +80 H; a gordura entre –60 H e –100 H; o tecido pulmonar entre –400 H e –600 H, e o ar, –1.000 H (BRANT; HELMS, 2008).

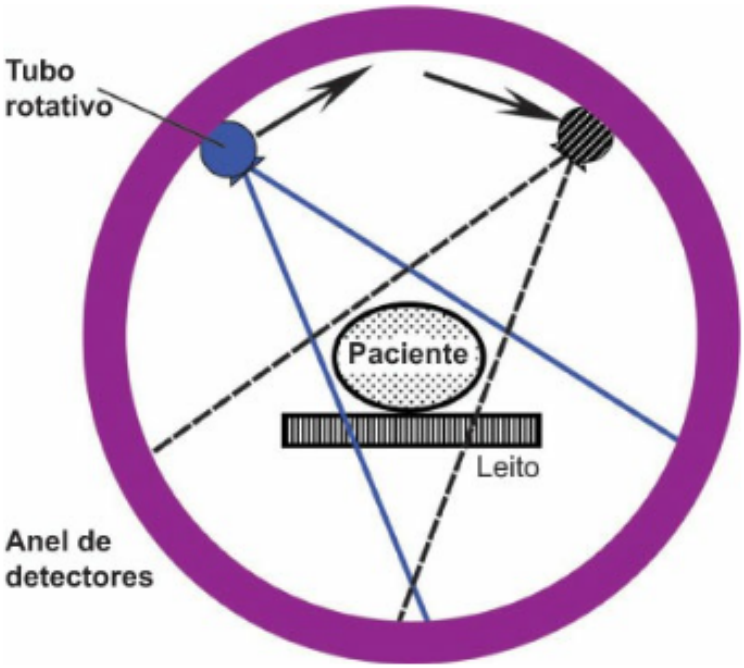


Figura 3.2: Diagrama de um scanner de TC. Fonte: (BRANT; HELMS, 2008)

As dimensões dos voxels são determinadas por algoritmos computacionais selecionados para a reconstrução e pela espessura do corte escaneado. A maioria dos *scanners* de TC possibilita especificações para espessura do corte entre 0.5 mm e 10 mm. Os dados para um corte individual, com rotação de 360° do tubo de raios X, normalmente são adquiridos em 1 segundo ou menos (MOURÃO; OLIVEIRA, 2018). As vantagens da TC em relação à RM incluem rapidez na varredura, superioridade nos detalhes do tecido ósseo e apresentação de calcificações. A aquisição do exame de TC geralmente, limita-se ao plano axial; entretanto, as imagens podem ser reformatadas nos planos sagital, coronal ou oblíquo ou como imagens tridimensionais. A TC com multidetectores possibilita a aquisição de voxels isotrópicos cuboides de tamanho

igual nos três lados. Voxels isotrópicos tornam possível a reconstrução da imagem em diferentes planos com a mesma qualidade.

3.1.2 RM

RM é uma técnica que produz imagens internas do paciente por meio de campos magnéticos e ondas de rádio. Enquanto a TC analisa apenas um parâmetro do tecido do paciente, a atenuação de raios X, a ressonância magnética consegue analisar diversas características, entre elas a densidade de hidrogênio (prótons), tempos de relaxação T1 e T2 dos tecidos e o fluxo sanguíneo nos tecidos. O contraste para tecidos moles fornecido pela RM é muito melhor do que o de qualquer outra modalidade de imagem. São as diferentes densidades protônicas disponíveis nos tecidos que contribuem para que o sinal de RM consiga fazer a distinção entre um tecido e outro. A maioria dos tecidos pode ser distinguida por meio de diferenças significativas em seus tempos de relaxação T1 e T2 específicos. T1 e T2 são características do ambiente molecular tridimensional que circunda cada próton no tecido que está sendo examinado. T1 mede a capacidade do próton de trocar energia com a matriz química adjacente. É uma medida da rapidez com que um tecido se torna magnetizado. T2 representa a rapidez com que determinado tecido perde sua magnetização. O fluxo sanguíneo tem um efeito complexo sobre o sinal de RM e pode aumentar ou diminuir a intensidade desse sinal nos vasos sanguíneos (MOURÃO; OLIVEIRA, 2018).

Em termos simplificados, a RM se baseia na capacidade que um pequeno número de prótons do corpo tem de absorver e emitir ondas de rádio quando o corpo é colocado sob a influência de um forte campo magnético. Tecidos diferentes absorvem e emitem a energia das ondas de rádio a taxas específicas, detectáveis e características. As RM são obtidas expondo o paciente a campos magnéticos de potências que variam entre 0,02T e 7T, dependendo do equipamento que está sendo utilizado. A escolha do equipamento é feita de acordo com a preferência e a disponibilidade local. Um pequeno número de prótons nos tecidos do paciente se alinha de forma resultante ao eixo do campo magnético principal e, subsequentemente, é desalinhado pela aplicação de gradientes de RF. Quando o gradiente de RF é desligado, os prótons desalinhados tornam a alinhar-se com o campo magnético principal, liberando um pequeno pulso de energia, o qual é detectado, localizado e depois processado por um algoritmo computacional semelhante ao empregado na TC a fim de produzir uma imagem anatômica. A localização do corte é determinada pela aplicação de um gradiente de seleção de corte, que aumenta gradualmente de intensidade ao longo do eixo Z. Os pulsos de baixa energia liberados pelos tecidos são posteriormente localizados por meio de codificação de frequência em uma direção (eixo X) e codificação de fase, na outra direção (eixo Y). As imagens podem ser obtidas em qualquer plano anatômico, ajustando-se os gradientes do campo magnético nos eixos X, Y ou Z. Como o sinal de RM é muito fraco, normalmente é necessário um tempo prolongado de imagem para que seja obtida uma imagem de boa qualidade. (MOURÃO; OLIVEIRA, 2018).

As principais vantagens da RM, incluem sua incrível resolução de contraste para tecidos moles, a capacidade de fornecer imagens em qualquer plano anatômico e a ausência de radiação ionizante. A RM é limitada em sua capacidade de mostrar detalhes de ossos densos ou calcificações; além disso, tem tempos de aquisição longos se comparada com a TC, disponibilidade limitada em certas regiões geográficas e alto custo (BRANT; HELMS, 2008).

3.2 Registro

O registro de imagem é o processo de transformar imagens em um sistema de coordenadas comum, de modo que os pixels correspondentes representam pontos biológicos homólogos, ou seja, alinhar imagens para que os recursos correspondentes possam ser facilmente relacionados, como mostrado na Fig. 3.3. Por exemplo, o registro pode ser usado para obter um referencial anatomicamente normalizado, no qual regiões do cérebro de diferentes modalidade ou ponderações de exames podem ser comparadas. No nosso trabalho usaremos o registro rígido para alinhar as imagens. Nesse tipo de registro a imagem é tratada como um corpo rígido, que pode ser transladado e girado, mas não pode ser deformado.

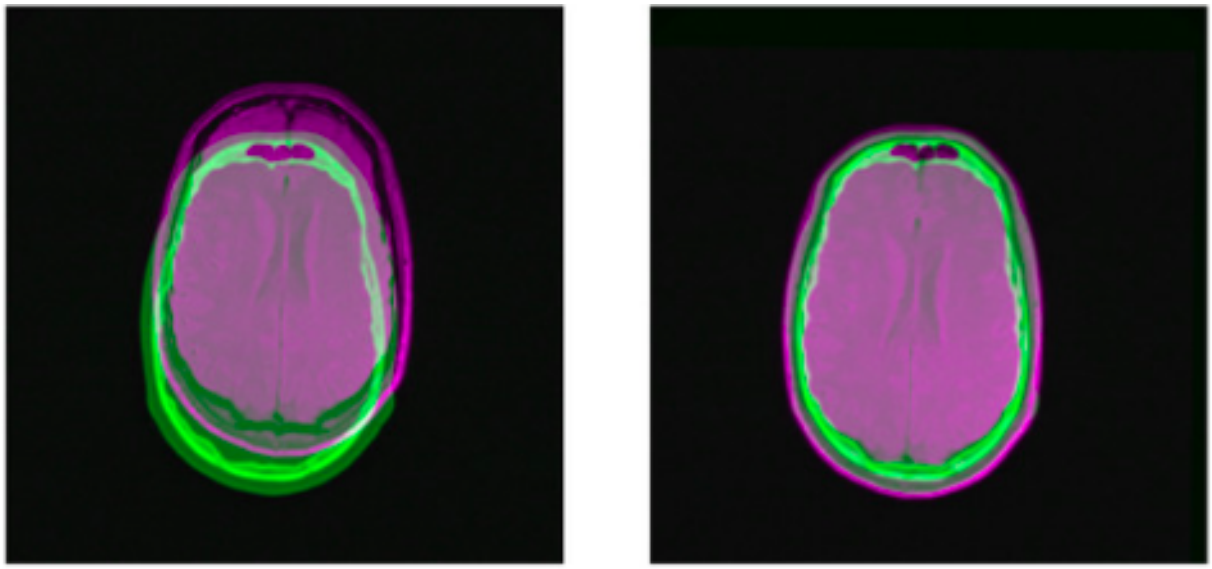


Figura 3.3: A imagem estática representada em verde, a imagem movida representada em rosa. A imagem da esquerda mostra as imagens antes do registro e a imagem da direita mostra as imagens depois do registro. Fonte: (KLEIN et al., 2010)

O registro é representado matematicamente como uma transformação $T(x)$ que representa o mapeamento espacial de pontos da imagem fixa p para pontos na imagem em movimento q. Isso estabelece uma correspondência para cada pixel na imagem fixa para uma posição na imagem em movimento, Figura 3.4.

Uma métrica de similaridade fornece uma medida de quão bem a imagem fixa corresponde à imagem em movimento. Essa medida forma um critério quantitativo a ser otimizado por um

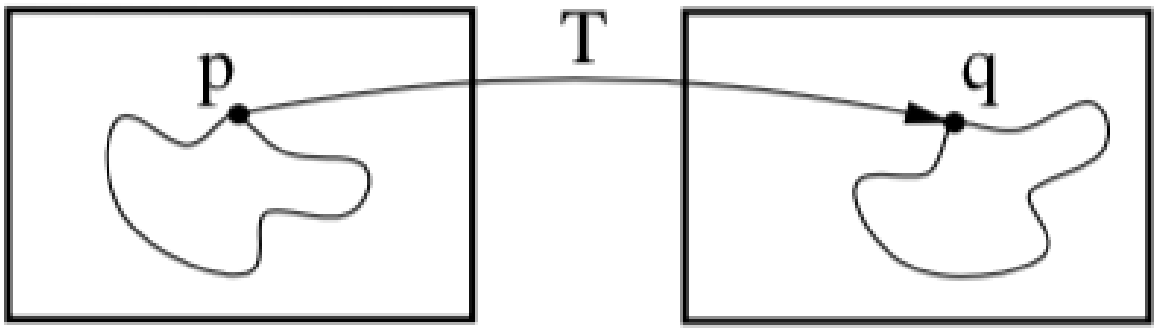


Figura 3.4: O registro de imagens é a tarefa de localizar uma transformação espacial mapeando uma imagem para outra. Esquerda é a imagem fixa e direita é a imagem em movimento. Fonte: (KLEIN et al., 2010)

otimizador sobre o espaço de busca definido pelos parâmetros da transformação. Em geral, o procedimento de registro é formulado como um problema de otimização no qual uma função de custo C é minimizada em relação a T. Matematicamente, $IM(x)$ é ajustado para corresponder a $IF(x)$ encontrando uma transformação de coordenadas $T(x)$ que faz $IM(T(x))$ alinhado espacialmente com $IF(x)$. Isso significa simplesmente que o otimizador ajusta os parâmetros da transformação de maneira a minimizar a diferença entre as duas imagens (KLEIN et al., 2010).

A métrica é um componente chave no processo de registro. Ele usa informações da imagem fixa e móvel para calcular um valor de similaridade. A derivada desse valor nos diz em qual direção devemos mover a imagem em movimento para melhor alinhamento. A imagem em movimento é movida em pequenos passos e o processo é repetido até que um critério de convergência seja atingido. A métrica pode usar intensidades dos pixels, posições de ponto, recursos de imagem pré-computados ou qualquer coisa que possamos otimizar. É necessário apenas definir uma métrica para isso (KLEIN et al., 2010).

3.3 Redes Neurais Convolucionais

Uma CNN é um algoritmo de Aprendizado Profundo que pode captar uma imagem de entrada, para alimentar a rede e atribuir importância (pesos e vieses que podem ser aprendidos) a vários aspectos e características objetos da imagem e ser capaz de diferenciar um do outro. O pré-processamento exigido em uma CNN é muito menor em comparação com outros algoritmos de classificação. Enquanto nos métodos primitivos os filtros são feitos à mão, com treinamento suficiente, as CNNs têm a capacidade de aprender esses filtros.

Modelos seguindo a arquitetura do tipo CNN tem tido um impacto maior na área de informática médica, sendo considerados a abordagem mais efetiva atualmente para tradução de imagens (ISOLA et al., 2017).

CNNs são uma arquitetura específica entre as várias existentes (U-Net, Resnet etc) que seguem o conceito de DL, com o foco específico em traduções de imagens. As CNNs combinam três ideias arquitetônicas para garantir algum grau de invariância de deslocamento, escala e distorção, que são: campos receptivos locais, compartilhamento de pesos e subamostragem espacial ou temporal (LECUN et al., 1998).

As CNNs possuem a capacidade de aprender características invariantes a rotação, translação e transformações afins de forma hierárquica, ou seja, são capazes de aprender dados abstratos independente de fatores externos (FUKUSHIMA, 1980; LECUN; KAVUKCUOGLU; FARABET, 2010). Para uma rede neural ser considerada uma CNN basta conter pelo menos uma camada de convolução entre as suas camadas, sendo que as CNNs são normalmente compostas de camadas de convolução, subamostragem e totalmente conectadas, sendo utilizadas e concatenadas de acordo com o objetivo da rede. As camadas de convolução e subamostragem são responsáveis por extrair características das imagens e a camada totalmente conectada é responsável pelo processo de classificação dos padrões da imagem de entrada (Figura 3.5).

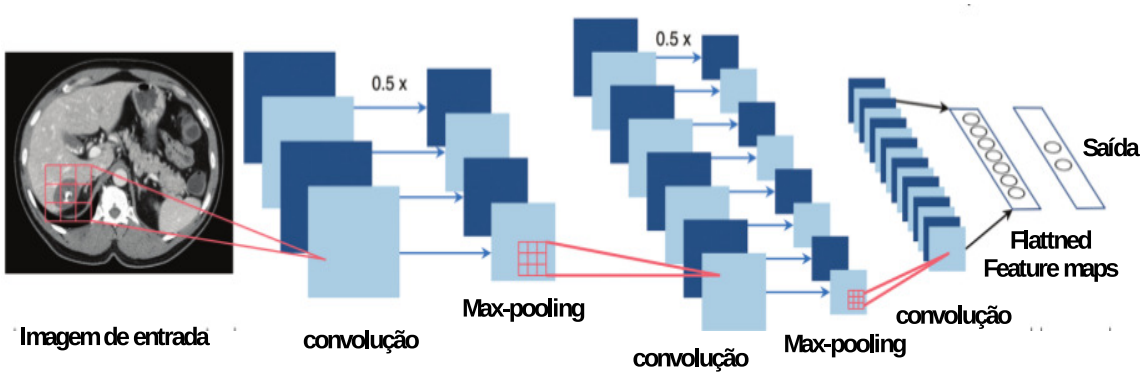


Figura 3.5: Exemplo de uma rede neural. Fonte: (CHARTRAND et al., 2017).

A camada de convolução tem como objetivo aplicar operações de convolução na imagem de entrada para extrair características para o treinamento da rede (Fig. 3.6). Os filtros são ajustados durante a fase de treinamento para que a rede possa ser capaz de encontrar características relevantes para a construção e atualização da rede e o resultado obtido da aplicação dos filtros é conhecida como *feature map*. A profundidade da saída de uma convolução é igual a quantidade de filtros aplicados. Quanto mais profundas são as camadas das convoluções, mais detalhados são os traços identificados com o *activation map*. Assim, a camada de convolução recebe como entrada um conjunto de características e aplica diversos filtros de convolução onde cada um dos filtros tem o função de enaltecer uma característica diferente, como na primeira camada, por exemplo, em que as linhas e gradientes (características mais gerais) em diferentes orientações são destacadas (LECUN; KAVUKCUOGLU; FARABET, 2010).

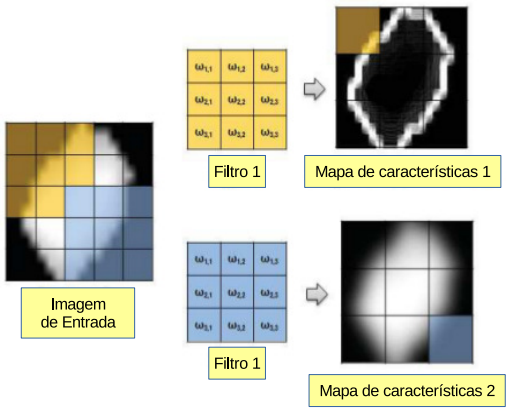


Figura 3.6: exemplo de um filtro de convolução. Fonte: (SILVA; PAIVA; SILVA, 2017).

As camadas de subamostragem são utilizadas logo após as camadas de convolução, possuem a mesma quantidade de planos, mas com menos células complexas de campos receptivos menores que computam a média (*mean-pooling*) do mapa de características através da aplicação de uma função não-linear, que fazem uma subamostragem do mapa de características, como a sigmóide-logística ou tangente hiperbólica, ponderadas por um coeficiente de treino e somadas com um bias treinável, reduzindo, assim, a resolução do mapa de características e a sensibilidade à modificações topológicas e distorções.

As camadas de subamostragem quando inseridas em uma CNN são utilizadas logo após uma camada de convolução, para reduzem progressivamente a resolução espacial ou dimensão dos mapas de características, propagando (selecionando) as características mais importantes, reduzindo assim a quantidade de parâmetros, o tempo de treinamento na rede, reduzindo a chance de *overfitting* (quando o modelo se adapta ao dados de treinamento e é incapaz de generalizar). Este tipo de camada permite a aprendizagem ideal de *features* invariantes ao deslocamento e à distorções em cada *feature map* melhorando a capacidade de generalização do conhecimento da rede (LECUN et al., 1998; CRUZ-ROA et al., 2014; PEIXOTO; CáMARA-CHáVEZ; MENOTTI, 2015).

Despois da extração dos atributos da imagem de entrada usando as camadas de convolução e subamostragem, os atributos são fornecidos como entrada para a camada totalmente conectada, que pode conter uma ou mais camadas dependendo da arquitetura da rede. O seu objetivo desta é classificar a imagem de entrada.

3.4 GAN

Em 2014, quando Ian Goodfellow criou as GANs (GOODFELLOW et al., 2014) permitiu que os computadores fossem capazes de gerar dados muitos próximos da realidade usando duas redes neurais separadas, sendo que o treinamento de uma depende dos resultados da outra. As GANs não foram a primeira técnica para geração de dados realísticos, mas devido aos seus

resultados, inovação e versatilidade elas foram capazes de se diferenciar de todo o resto. As GANs alcançaram resultados notáveis, que eram considerados impossíveis para um sistema computacional, como a capacidade de gerar imagens falsas com qualidade semelhante à do mundo real, transformar um rabisco em uma imagem semelhante a uma fotografia (ISOLA et al., 2017) ou transformar um vídeo de um cavalo em uma zebra em execução (ZHU et al., 2017), tudo isso sem um gigantesco volume de dados para treinamento e sem a necessidade desses dados serem cuidadosamente mapeados.

Uma GAN é uma técnica de aprendizado de máquina que utilizada a ideia de uma competição, um jogo de soma zero entre duas redes neurais, para aprender a gerar exemplos falsos indistinguíveis de dados reais, a partir de um dado conjunto treinamento, tendo como ideal o equilíbrio de Nash (ou seja, ponto de sela), onde em um jogo envolvendo dois jogadores, nenhum jogador tem a ganhar mudando sua estratégia unilateralmente.

As duas redes são chamadas de Gerador e Discriminador. O objetivo do gerador é produzir dados indistinguíveis do conjunto de dados de treinamento. O objetivo do Discriminador é determinar corretamente se um exemplo específico é real (ou seja, proveniente do conjunto de dados de treinamento) ou falso (criado pelo Gerador).

Geralmente partindo de um vetor de ruídos aleatórios, o Gerador aprende a produzir exemplos realistas. Ele faz isso indiretamente, através do feedback que recebe das predições do Discriminador, onde cada vez que o Discriminador classifica uma imagem sintética como real, o Gerador recebe um feedback positivo, e com isso ele sabe que fez algo bom. Para cada vez que o Discriminador detecta corretamente uma imagem sintética criada pelo Gerador, o Gerador recebe a informação de que errou e que precisa melhorar. O fluxo da GAN é mostrado na figura 3.7.

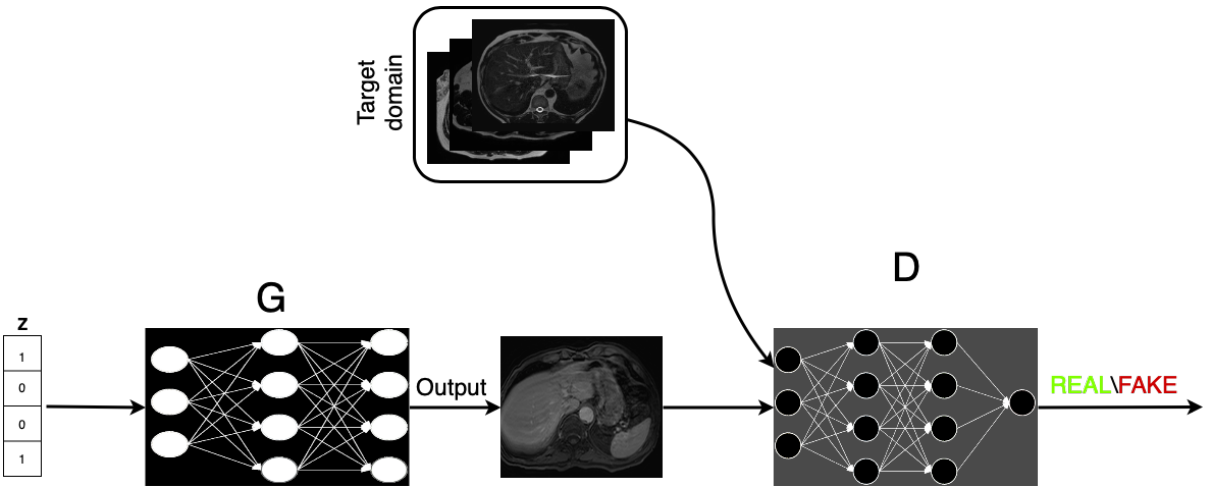


Figura 3.7: A GAN consiste em duas redes neurais: o gerador G e o discriminador D. A entrada para G é z é um vetor de ruído aleatório e sua saída é uma imagem próxima das imagens do domínio alvo. O discriminador irá julgar se a imagem gerada faz parte ou não do domínio dos dados reais. Fonte: elaborado pelo autor.

O Discriminador também melhora a cada classificação, quando é dado um feedback se ele

conseguiu detectar a falsificação. Assim, a medida que o Gerador melhora a produção de dados de aparência realista, o Discriminador melhora ao distinguir dados falsos dos reais. Ambas as redes continuam a melhorar simultaneamente através deste jogo de gato e rato.

Para aprender a distribuição do gerador P_g sobre os dados **x**, definimos antes as variáveis de ruído de entrada $P_z(z)$, então representamos um mapeamento para o espaço de dados como G(z; θ_g), onde G é uma função diferenciável representada por um perceptron multicamada com parâmetros θ_g. Também definimos um segundo perceptron multicamada D(x; θ_d) que gera um único escalar. D(x) que representa a probabilidade de que x foi proveniente dos dados em vez de P_g. Nós treinamos D para maximizar a probabilidade de atribuir o rótulo correto a ambos os exemplos de treinamento e amostras de G. Treinamos simultaneamente G para minimizar o $\log(1 - D(G(z)))$. Em outras palavras, D e G jogam o minmax de dois jogadores com função de valor $V(G, D)$ como na equação 3.1 .

$$\min_G \max_D \mathcal{L}_{GAN}(G, D) = E_{x \sim P_{data}(x)}[\log D(x) + E_{z \sim P_z(z)} \log(1 - D(G(z)))] \tag{3.1}$$

Saindo da abstração, as GANs têm como analogia um falsificador de dinheiro e o banco, sendo o Gerador representado pelo falsificador e o discriminador representado pelo banco. O falsificador decide falsificar dinheiro, cada vez que ele produz um novo lote de notas falsas, ele envia suas notas para um banco local e tenta depositar o dinheiro como real. Se as notas forem detectadas como falsas, o falsificador sabe que precisa melhorar na falsificação. Assim como um gerador cujo exemplo foi rejeitado como falso, o falsificador não sabe exatamente onde errou, tudo o que ele sabe é que precisa melhorar alguma coisa.

Da mesma forma, o banco precisa continuar a melhorar. Aceitar dinheiro falso pode causar prejuízo, portanto, se o banco descobrir que recentemente aceitou notas falsas como um depósito real, poderá investir em novas técnicas para detecção de notas falsas.

Como resultado, o banco agora pode reconhecer as notas falsas de alta qualidade que antes passavam pelas falhas do sistema. O que por sua vez leva o falsificador a melhorar, e o ciclo de feedback continua até que ambos não tenham o que melhorar.

3.5 CGAN

Redes adversárias geradoras podem ser estendidas a um modelo condicional se tanto o gerador quanto o discriminador estiverem condicionados a alguma informação adicional y, onde y pode ser qualquer tipo de informação auxiliar, como rótulos de classe ou dados de outras modalidades. Podemos realizar o condicionamento alimentando y tanto no discriminador quanto no gerador como camada de entrada adicional(MIRZA; OSINDERO, 2014).

No gerador, o ruído de entrada anterior $p_z(z)$ e y são combinados em representação oculta conjunta, e a estrutura de treinamento adversária permite considerável flexibilidade em como essa representação oculta é composta. No discriminador y e x (conjunto real) são apresentados

como entradas para uma função discriminativa. A função objetivo de um jogo minimax de dois jogadores seria como na equação 3.2:

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) = E_{x \sim P_{data}(x)}[\log D(x|y) + E_{z \sim P_z(z)} \log (1 - D(G(z|y)))] \quad (3.2)$$

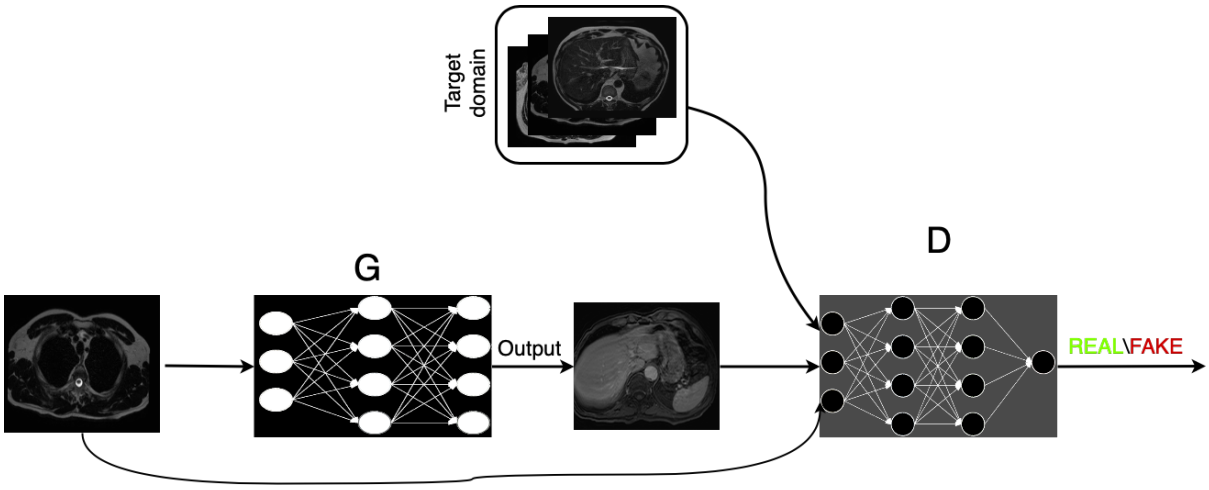


Figura 3.8: Na CGAN ocorre o mesmo processo da GAN, mas agora a entrada passa a ter uma imagem condicionante no lugar de um vetor de ruído aleatório. Fonte: elaborado pelo autor.

3.6 PIX2PIX

Assim como as GANs aprendem um modelo gerativo de dados, GANs condicionais (cGANs) aprendem um modelo generativo condicional. Isto torna as cGANs adequados para tarefas de tradução de imagem para imagem, onde nós condicionamos em uma imagem de entrada e geramos uma imagem de saída correspondente, alcançando bons resultados em uma ampla variedade de problemas (ISOLA et al., 2017).

Abordagens anteriores acharam benéfico misturar o objetivo do GAN com uma perda mais tradicional, como a distância L2 (PATHAK et al., 2016). O trabalho do discriminador permanece inalterado, mas o gerador é encarregado de não apenas enganar o discriminador, mas também de estar perto da saída real em um sentido L2. Também exploramos essa opção, usando a distância L1 em vez de L2, já que L1 incentiva menos desfoque:

$$\mathcal{L}_{L1}(G) = E_{x,y,z}[||y - G(x, z)||_1] \quad (3.3)$$

O objetivo final é:

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (3.4)$$

Sem z, a rede ainda poderia aprender um mapeamento de x para y, mas produziria saídas

determinísticas e, portanto, não corresponderia a nenhuma distribuição diferente de uma função delta. GANs condicionais anteriores reconheceram isso e forneceram ruído gaussiano z como uma entrada para o gerador além de x (WANG; GUPTA, 2016), além disso as arquiteturas do gerador e do discriminador foram modificados daqueles em (RADFORD; METZ; CHINTALA, 2015). O gerador e o discriminador usam módulos da forma convolution-BatchNorm-ReLu (IOFFE; SZEGEDY, 2015).

Muitas soluções anteriores (WANG; GUPTA, 2016; JOHNSON; ALAHI; FEI-FEI, 2016) usaram uma rede decodificadora-codificadora (HINTON; SALAKHUTDINOV, 2006). Em tal rede, a entrada é passada através de uma série de camadas que progressivamente reduzem a amostragem, até uma camada de gargalo, no ponto em que o processo é invertido. Essa rede exige que todo o fluxo de informações passe por todas as camadas, incluindo o gargalo. Para muitos problemas de tradução de imagens, há uma grande quantidade de informações de baixo nível compartilhadas entre a entrada e a saída, e seria desejável transferir essas informações diretamente pela rede. Por exemplo, no caso de tradução RM para TC, a entrada e a saída compartilham a localização das bordas.

Para dar ao gerador um meio de contornar o gargalo para informações como essa, adicionamos conexões ignoradas, seguindo a forma geral de uma "U-Net" (RONNEBERGER; FISCHER; BROX, 2015). Especificamente, adicionamos ligações ignoradas entre cada camada i e camada n - i, onde n é o número total de camadas. Cada conexão ignorada simplesmente concatena todos os canais na camada i com aqueles na camada n - i.

A perda de L2 e L1 produzem resultados borrados em problemas de geração de imagens (LARSEN et al., 2015). Embora essas perdas não estimulem a nitidez de alta frequência, em muitos casos elas capturam com precisão as baixas frequências. Para problemas onde este é o caso, não precisamos de uma estrutura inteiramente nova para reforçar a correção nas baixas frequências. L1 já será suficiente.

Isso motiva a restrição do discriminador da GAN para modelar apenas a estrutura de alta frequência, contando com um termo L1 para forçar a correção de baixa frequência. Para modelar frequências altas é suficiente restringir nossa atenção à estrutura em patches de imagens locais. Portanto, foi projetada uma arquitetura discriminadora chamada de PatchGAN, que apenas penaliza a estrutura na escala dos patches. Este discriminador tenta classificar se cada patch N X N em uma imagem é real ou falsa. Passamos este discriminador de forma convolucional através da imagem, calculando a média de todas as respostas para fornecer o resultado de D.

Tal discriminador modela efetivamente a imagem como um campo aleatório de Markov, assumindo a independência entre os pixels separados por mais de um diâmetro de patch. Esta conexão foi previamente explorada em (LI; WAND, 2016), e é também uma suposição comum em modelos de textura (GATYS; ECKER; BETHGE, 2015) e estilo (GATYS; ECKER; BETHGE, 2016). Portanto, o PatchGAN pode ser entendido como uma forma de perda de textura/estilo.

3.7 CycleGAN

Obter dados de treinamento emparelhados pode ser difícil e caro. Por exemplo, apenas alguns conjuntos pequenos de dados existem para tarefas como a segmentação semântica (CORDTS et al., 2016). A obtenção de pares de entrada-saída para tarefas gráficas, como a estilização artística, pode ser ainda mais difícil, uma vez que a saída desejada é altamente complexa, exigindo normalmente a autoria artística. Para muitas tarefas, como a transfiguração de objetos, a saída desejada não é nem mesmo bem definida.

Portanto, buscamos um algoritmo que aprenda a traduzir entre domínios sem exemplos de entrada-saída pareados. Assumimos que existe alguma relação subjacente entre os domínios e procuramos aprender essa relação. Embora não tenhamos supervisão na forma de exemplos emparelhados, podemos explorar a supervisão no nível de conjuntos: recebemos um conjunto de imagens no domínio X e um conjunto diferente no domínio Y.

Um mapeamento $G : XY$ tal que a saída $\hat{y} = G(x)$, $x \in X$, é indistinguível das imagens $y \in Y$ por um discriminador treinado para classificar $\hat{y}$ além de y. Teoricamente, esse objetivo pode induzir uma distribuição de saída sobre $\hat{y}$ que corresponda à distribuição empírica $p_{data(y)}$ (em geral, isso requer que G seja estocástico) (GOODFELLOW et al., 2014). O G ótimo então traduz o domínio X para um domínio $\hat{y}$ distribuído identicamente a Y. No entanto, tal tradução não garante que uma entrada individual x e a saída y sejam emparelhadas de forma significativa, existem infinitos mapeamentos G que irão induzir a mesma distribuição ao longo de $\hat{y}$. Além disso, na prática é difícil otimizar o objetivo contraditório isoladamente: procedimentos padrão geralmente levam ao conhecido problema de mínimo local para um métrica de erro, onde todas as imagens de entrada são mapeadas para a mesma imagem de saída e a otimização não consegue progresso (GOODFELLOW, 2016).

Essas questões exigem mais estrutura para o objetivo. Portanto, exploramos a propriedade de que a tradução deve ser consistente em ciclo, no sentido de que se traduzirmos, por exemplo, uma frase do inglês para o francês e traduzi-la do francês para o inglês, devemos voltar na sentença original (BRISLIN, 1970). Matematicamente, se tivermos um tradutor $G : XY$ e outro tradutor $F : YX$, então G e F devem ser inversos um do outro, e ambos os mapeamentos devem ser bijeções. Aplicando esse pressuposto estrutural treinando simultaneamente o mapeamento G e F e adicionando uma perda de consistência do ciclo (ZHOU et al., 2016) que encoraja $F(G(x)) \approx x$ e $G(F(y)) \approx y$. Combinando essa perda com perdas contraditórias nos domínios X e Y, obtém-se um novo objetivo total de tradução imagem-imagem não pareada.

O treinamento adversário pode em teoria aprender mapeamentos G e F que produzem saídas distribuídas identicamente como domínios alvo Y e X, respectivamente (estritamente falando, isso requer que G e F sejam funções estocásticas) (GOODFELLOW, 2016). No entanto, com capacidade suficiente, uma rede pode mapear o mesmo conjunto de imagens de entrada para qualquer permutação aleatória de imagens no domínio de destino, onde qualquer um dos mapeamentos aprendidos pode induzir uma distribuição de saída que corresponda à distribuição de

destino. Assim, perdas adversárias sozinhas não podem garantir que a função aprendida possa mapear uma entrada individual x_i para uma saída desejada y_i. Para reduzir ainda mais o espaço de possíveis funções de mapeamento, eles devem ser consistentes em ciclos: para cada imagem x do domínio X, o ciclo de tradução da imagem deve ser capaz de trazer x de volta à imagem original, ou seja, $xG(x)F(G(x)) \approx x$. Isso é chamado de consistência de ciclo para frente. Da mesma forma, para cada imagem y do domínio Y, G e F também devem satisfazer a consistência do ciclo de retrocesso: $yF(y)G(F(y)) \approx y$.

$$\mathcal{L}_{cyc}(G, F) = E_{x \sim P_{data}(x)}[||F(G(x) - x||_1] + E_{y \sim P_{data}(y)}[||F(G(y) - y||_1]) \tag{3.5}$$

O objetivo geral é ter o jogo minmax na equação 3.6:

$$\mathcal{L}_{cyc}(G, F, D_x, D_y) = \mathcal{L}_{GAN}(G, D_y, X, Y) + \mathcal{L}_{GAN}(G, D_x, X, Y) + \lambda \mathcal{L}_{cyc}(G, F) \tag{3.6}$$

3.8 StarGAN

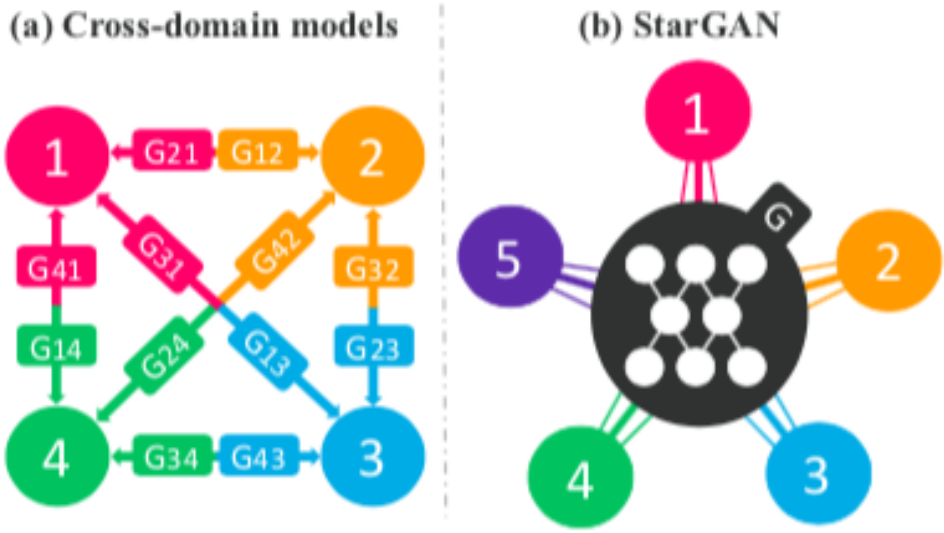


Figura 3.9: Comparação entre modelos de tradução entre múltiplos domínios. (a) Para lidar com vários domínios, os modelos entre domínios devem ser criados para cada par de domínio de imagens. (b) A StarGAN é capaz de aprender mapeamentos entre múltiplos domínios usando um único gerador. Fonte: (CHOI et al., 2018)

A StarGAN (CHOI et al., 2018) propõe uma abordagem GAN para tratar a tradução de imagem-para-imagem de vários domínios em um único conjunto de dados, sendo capaz de incorporar vários conjuntos de dados contendo diferentes conjuntos de rótulos para realizar com flexibilidade as traduções usando qualquer um desses rótulos.

O objetivo da Stargan é treinar um único gerador G que aprende mapeamentos entre múltiplos domínios. Para conseguir isso, treina G para traduzir uma imagem de entrada x em uma

imagem de saída y condicionada no rótulo de domínio de destino c, $G(x, c)y$, gerando aleatoriamente o rótulo de domínio de destino c para que G aprenda a traduzir com a imagem de entrada. Para alcançar essa condição, foi adicionado um classificador auxiliar no discriminador D e foi imposto uma perda de classificação de domínio a D e G. Ou seja, decompôs o objetivo em dois termos: uma perda de classificação de domínio de imagens reais usadas para otimizar D, e uma perda de classificação de domínio de imagens falsas usadas para otimizar G (CHOI et al., 2018).

Perda Adversária é incorporada para tornar as imagens geradas indistinguíveis das imagens reais. Perda de Reconstrução é utilizada para minimizar as perdas adversárias e de classificação, G é treinado para gerar imagens que sejam realistas e classificadas em seu domínio de destino correto. No entanto, minimizar essas perdas não garante que as imagens traduzidas preservem o conteúdo de suas imagens de entrada, alterando apenas a parte relacionada ao domínio das entradas. Para aliviar este problema, aplicamos uma perda de consistência cíclica ao gerador. onde G toma na imagem traduzida $G(x, c)$ e o rótulo de domínio original $c0$ como entrada e tenta reconstruir a imagem original x. Adotando a norma L1 como a perda de reconstrução.

Uma vantagem importante da StarGAN é que ela incorpora simultaneamente vários conjuntos de dados contendo diferentes tipos de rótulos, para que a StarGAN possa controlar todos os rótulos na fase de teste, mas esse tipo de abordagem é um problema devido a variação dos rótulos e pelo fato de que as informações dos rótulos são conhecidas apenas parcialmente por cada conjunto de dados. Isso é problemático porque a informação completa sobre o vetor c0 é necessária ao reconstruir a imagem de entrada x da imagem traduzida $G(x, c)$. Para aliviar esse problema, foi introduzido um vetor de máscara m que permite que a StarGAN ignore rótulos não especificados e se concentre no rótulo explicitamente conhecido fornecido por um conjunto de dados específico. Na StarGAN, usamos um vetor unidimensional n-dimensional para representar m, sendo n o número de conjuntos de dados. Além disso, definimos uma versão unificada do rótulo como um vetor (CHOI et al., 2018).

4 Materiais e Métodos

A seção de Materiais e Métodos foi dividida em duas seções principais. A primeira (seção 4.1) apresenta a arquitetura da Multi Medical Imaging Translation using Generative Adversarial Network (MMI-GAN) e o seu pipeline para a tradução multi-domain de imagens médicas; enquanto a segunda (seção 4.2) apresenta os experimentos realizados usando a MMI-GAN e o seu baseline de comparação, além do treinamento e a avaliação dos experimentos.

4.1 Arquitetura da MMI-GAN

A Fig. 4.1 apresenta a visão geral da MMI-GAN, em (A) é realizado o pré-processamento da base de imagem, onde as imagens usadas no treinamento da MMI-GAN devem ser pareadas e alinhadas usando registro de imagem. Após o pré-processamento, é realizada a etapa de treinamento dos dois componentes principais da rede: o **Generator** (seção 4.1.1) e o **Discriminator** (seção 4.1.2). O **Generator** (Fig. 4.1-B) recebe uma imagem da *training data* em conjunto com a target label, indicando o destino do domínio da tradução. O Gerador então cria uma nova imagem que chamaremos de imagem traduzida. O **Discriminator** (Fig. 4.1-(C)) recebe um par de imagens, podendo ser um par real, contendo a imagem de origem e sua correspondente no domínio indicado pela target label (*real pair* 4.1-D), ou um par contendo a imagem de origem e a imagem traduzida para o domínio de destino (Fig. 4.1-E). A partir desse par de imagens, o **Discriminator** tentará prever se é um real pair ou se é um par que contém uma imagem traduzida (fake). O **Discriminator** também é responsável em avaliar se a entrada foi classificada no domínio de imagem correto.

4.1.1 Gerador

Um dos problemas da tradução de imagens médicas é o mapeamento de uma imagem de entrada de alta resolução para uma imagem de saída de alta resolução, pois a entrada e a saída da rede são de domínios e aparências diferentes. As modalidades diferem quanto à representação de uma mesma região anatômica, implicando em uma grande quantidade de informação globalmente compartilhada entre a entrada e a saída. Logo, na tradução de imagens é desejável transferir a informação sobre a localização de determinadas características (e.g., borda de objetos) de maneira direta e rapidamente pela rede.

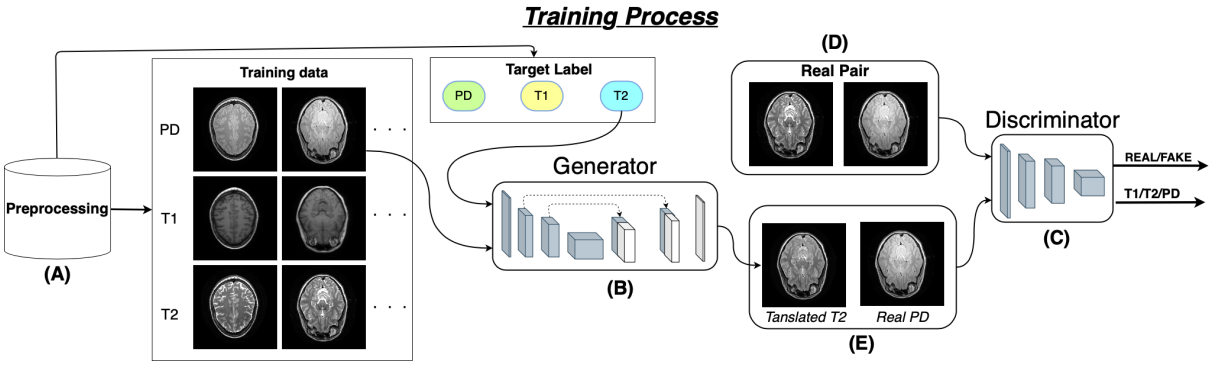


Figura 4.1: Visão geral do pipeline MMI-GAN. Fonte: elaborado pelo autor.

Devido a essa característica, utilizamos uma rede encoder-decoder do tipo U-Net (RONNEBERGER; FISCHER; BROX, 2015), devido a sua capacidade de transferir diretamente informações entre camadas através de skips na rede, onde no encoder (Fig. 4.2-A), a imagem de entrada é passada por N camadas de convolução com filtros de tamanho 4 e um número crescente de filtros para cada camada, iniciando com 64 e sempre dobrando para a próxima camada, e também strides de 2 e padding same.

A MMI-GAN possui a flexibilidade de ajustar a quantidade de camadas convolucionais dependendo da resolução da imagem de entrada. As camadas de convoluções são seguidas de uma ativação LeakyReLU (slope de 0.2) e BatchNormalization (exceto a primeira camada), chamaremos essa composição de bloco de encoder. Diferentemente da ReLU padrão, a LeakyRelu permite passar um pequeno gradiente para valores negativos permitindo a MMI-GAN estabilizar o treinamento do gerador, evitando a pouca variabilidade na tradução (mode collapse). Além disso, a MMI-GAN utiliza BatchNormalization para normalizar os recursos de saída das camadas e também aumentar a velocidade de treinamento. Essas técnicas são usadas com o objetivo de reduzir progressivamente o tamanho da imagem de entrada até a bottleneck (Fig. 4.2-D).

A bottleneck (Fig. 4.2-D) é a parte da rede que interliga o encoder ao decoder (Fig. 4.2-B). Na bottleneck a saída do encoder passa por mais um bloco de encoder com filtro de tamanho 4 e o dobro de filtros da última camada do encoder, para finalizar o processo de redução da amostra. Posteriormente, ainda na bottleneck, o processo de redução da amostra é revertido usando as camadas de upSampling, convolução (ativação ReLU) e BatchNormalization.

Após o bottleneck, em um processo contrário ao encoder, a MMI-GAN utiliza o decoder para expandir a amostra até obter a imagem traduzida. O decoder possui N camadas de upSampling, que dobram o tamanho da entrada, seguidas por convolução com stride de 1, padding same e ativação ReLU e BatchNormalization (exceto a última camada), chamaremos a junção entre upsampling, convolução e BatchNormalization de bloco de decoder. A quantidade de blocos de decoder é sempre igual a quantidade de blocos de encoder, onde para cada bloco de encoder existe um bloco de decoder interligado por skip, o último bloco de encoder é concatenado com o primero bloco de decoder, assim sucessivamente para todos os blocos de encoder

e decoder. As camadas de upsampling em conjunto com as camadas de convolução têm como objetivo dobrar as dimensões da entrada e reduzir a quantidade de filtros até chegarmos na resolução da imagem de entrada do encoder. O batchNormalization foi usado para normalizar os recursos de saída das camadas e também aumentar a velocidade de treinamento. O uso destas técnicas permitiu reduzir progressivamente o tamanho da imagem

A U-net com as skips entre blocos de encoder e decoder permitiu o compartilhamento das informações da imagem de entrada com a imagem de saída (bordas), porém, a sua arquitetura só pode ser utilizada para traduzir de um domínio de imagens para outro. Diante desta limitação, utilizamos o target label embedding (Fig. 4.2-C) na bottleneck da U-net para que a tradução possa ser guiada para o domínio definido, por exemplo: se tivermos como entrada uma imagens no domínio de T1 e TC como target label, teremos como saída uma imagem no domínio de TC; se trocarmos apenas a target label para T2, teremos como saída uma imagem no domínio de T2. A target label no gerador funciona como um *switch* para a tradução, guiando a tradução por múltiplos domínios, fazendo com que as informações globais de todos os domínios envolvidos sejam utilizadas.

A arquitetura da target label embedding possui parâmetros dependentes da resolução de saída da última camada de convolução do encoder, porque após a imagem ser processada pelo encoder, ela é concatenada com a saída da target label embedding. Dessa forma, são necessárias as seguintes camadas:

1. Embedding com dimensão de entrada igual a 3, referente a quantidade de domínios que estamos trabalhando e dimensão de saída igual a dimensão de saída da última camada de convolução do encoder dividida por 2*N, sendo N o número de camadas de convolução do encoder, para que ao final da target label embedding pudesse ser remodelada para o mesmo formato da saída do encoder;

2. Uma camada flatten, uma camada dense com a quantidade de unidades equivalente a dimensão de saída da camada anterior, BatchNormalization, Reshape, LeakyReLU, UpSampling para aumentarmos a resolução;

3. A camada de convolução com tamanho de kernel igual a 5 e a mesma quantidade de filtros usada pela última camada de convolução do encoder, para que a saída da target label embedding tenha o mesmo formato da saída do encoder, podendo assim, ser concatenada para entrada da bottleneck.

4.1.2 Discriminador

O discriminador (Fig. 4.3) tem como entrada a concatenação da imagem de origem (Fig. 4.3-A) e a imagem de destino (Fig. 4.3-B), sendo processado por quatro blocos de encoder com

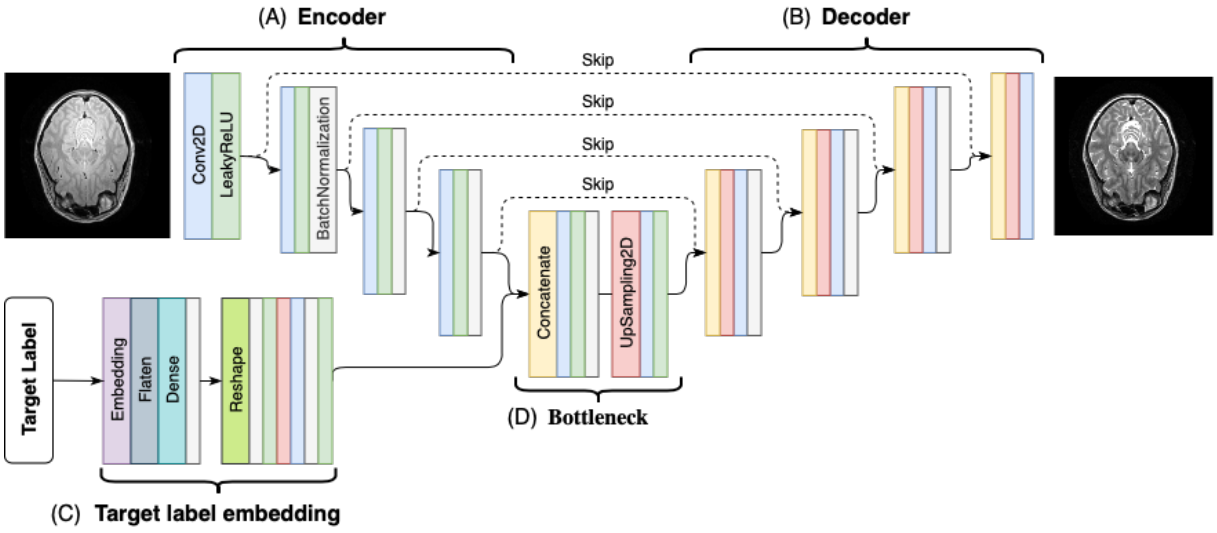


Figura 4.2: Estrutura geral do gerador. Fonte: elaborado pelo autor.

64, 128, 256 e 512 filtros, após isso o discriminador se divide e temos duas saídas: a saída da Patchgan (Fig. 4.3-C) e saída do classificador auxiliar (Fig. 4.3-D).

Neste trabalho utilizamos o discriminador PatchGAN proposto por Isola et al. (ISOLA et al., 2017). O PatchGAN determina se a imagem de destino é uma representação plausível da imagem de origem. A técnica aplica filtro convolucional igual a 1 e calcula a média de todos os patches da imagem para classificá-la em real ou falsa.

A segunda saída é a classificação da imagem segundo o domínio da imagem de destino, para alcançar essa condição, adicionamos um classificador auxiliar no discriminador (C_{aux}), impondo a perda de classificação de domínio para auxiliar o gerador e o discriminador, ou seja, utilizaremos a perda de classificação para avaliar imagens reais e otimizar o discriminador; e a perda de classificação de domínio de imagens falsas para otimizar o gerador. Para o C_{aux} utilizamos uma camada de flatten e uma camada densa com ativação softmax, pois o objetivo é avaliar a probabilidade da imagem pertencer a cada um dos 3 domínios de imagens usados neste trabalho.

Funções objetivo

Na MMI-GAN empregamos três funções objetivos diferentes para otimizar os parâmetros da nossa rede. A perda adversarial, baseada na estrutura GAN original (GOODFELLOW et al., 2014), definida conforme a equação (4.1).

$$\mathcal{L}_{adv}(G, D) = E_{x,y}[\log D(x, y)] + [E_{x,l} \log (1 - D(x, G(x, l)))] \tag{4.1}$$

Onde G é rede geradora e D é a rede discriminadora, x é a imagem de entrada e l a label

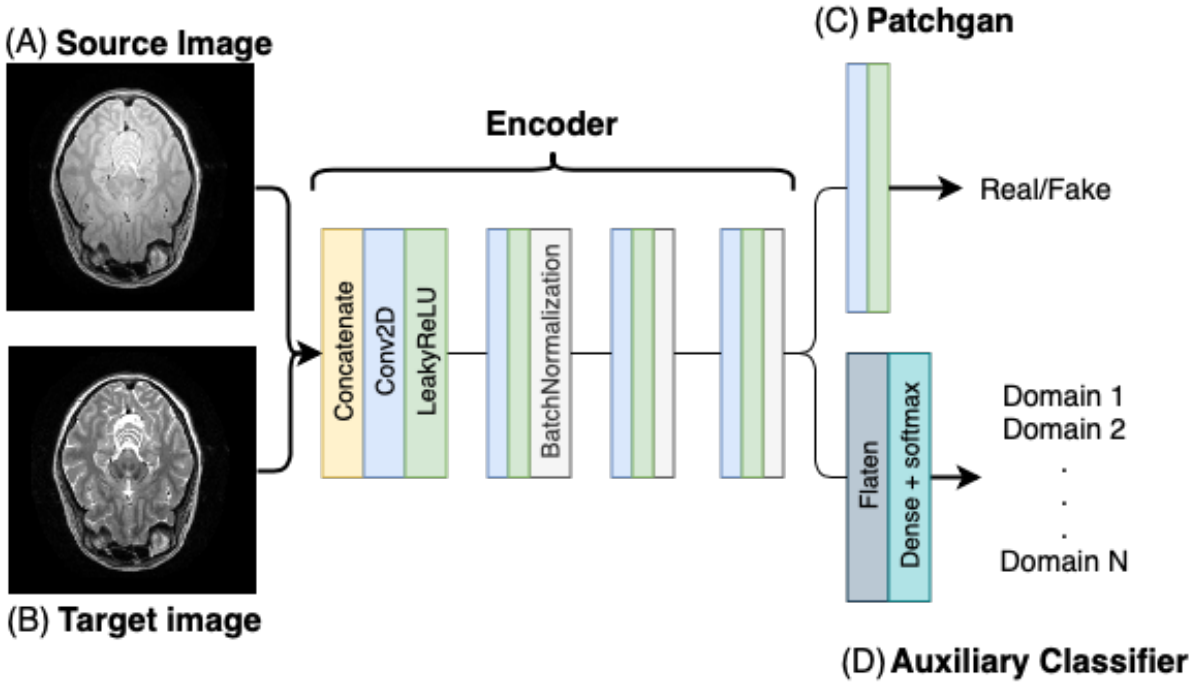


Figura 4.3: Estrutura geral do Discriminador. O par de imagens de entrada é concatenado e passa por um processo de codificação até o ponto em que é dividido em dois classificadores, (i)classifica se a imagem é real ou falsa, (ii)classifica a imagem em relação ao domínio a que pertence. Fonte: elaborado pelo autor.

indicando para qual domínio devemos traduzir, $G(x, l)$ é a imagem traduzida e y é a imagem de *ground truth* (alvo), respectivamente. Utilizamos a abordagem PatchGAN, onde similarmente, a perda adversária avalia se o patch de entrada é real ou gerado sinteticamente

As GANs de tradução de imagem para imagem que dependem exclusivamente da função de perda adversária não produzem resultados consistentes (ISOLA et al., 2017). Mais especificamente, as imagens de saída podem não compartilhar uma estrutura global semelhante à imagem alvo desejada. Para amenizar esse problema, uma perda de reconstrução de pixel, como a perda de $L1$ Eq. 4.2 foi incorporada, calculando o erro absoluto médio (MAE) entre a imagem de ground truth e a imagem gerada:

$$\mathcal{L}_{L1}(G) = \|y - G(x, l)\|_1 \tag{4.2}$$

Na Eq. 4.3 temos que o resultado do discriminador representa a probabilidade da imagem pertencer a classe, ao minimizar esse objetivo conseguimos fazer com que o discriminador aprenda a classificar a imagen no seu domínio de origem e o gerador gere outras que podem ser

classificadas no domínio indicado pela label de entrada l.

$$\mathcal{L}_{DomainsClassification}(G, D) = -E_{x,y,l}[\log C_{aux}(y, l)] - E_{x,y,l}[\log C_{aux}(G(x, l), l)] \quad (4.3)$$

Finalmente as funções de otimização podem ser descritas como na Eq. 4.4 para o gerador e Eq. 4.5 para o discriminador.

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{DomainsClassification} \quad (4.4)$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_2 \mathcal{L}_{DomainsClassification} + \lambda_3 \mathcal{L}_{L1} \quad (4.5)$$

Onde os λ representam os pesos dados a cada perda, que controlam as suas importâncias no processo de treinamento.

4.2 Experimentos

Os experimentos foram realizados em uma unidade de processamento gráfico (GPU) NVidia Titan X com 12 Gigabytes de RAM, 3584 cores, velocidade de 1.5 GHz e arquitetura Pascal. A biblioteca Keras (v.2.2.4) com Tensorflow (v.2.1.0) foi usada como backend. Os experimentos foram realizados em duas bases distintas com intuito de avaliar a MMI-GAN em diferentes situações: na tradução intermodalidade (RM-TC) e intramodalidades (T1, T2 e PD), em diferentes regiões anatômicas (tórax e cabeça) e com conjuntos de treinamentos com dimensões diferentes.

Bancos de dados e pré-processamento

Banco de dados de tórax. Este estudo prospectivo foi aprovado pelo nosso Comitê de Pesquisa institucional com o consentimento informado de todos os pacientes. As imagens foram cedidas pelo Hospital das Clínicas de Ribeirão Preto - Centro de Ciência das Imagens, processo hcrp 3733/2017. Todos os exames foram anônimos para garantir a privacidade dos pacientes. O banco de dados inclui exames de imagens de RM e TC do tórax de 37 pacientes, sendo 17 pacientes do sexo masculino e 20 do sexo feminino com uma média de idade de 60 anos. Entretanto, foram removidos oito pacientes que possuíam exames com artefatos de movimento.

O protocolo clínico de RM de tórax incluiu o auxílio do procedimento de retenção de respiração do paciente, imagens ponderadas em T1 pós-contraste (tempo de repetição 3.9539 ms, tempo de eco 1.874 ms e ângulo de inversão 10°), espessura do corte de 4 mm, resolução

de contraste de 8 bits e com resoluções espacias variando em 224x224, 256x256 e 288x288; imagens ponderadas em T2 (tempo de repetição 1175 ms, tempo de eco 160 ms e ângulo de inversão 90°), espessura do corte de 5mm, resolução de contraste de 8 bits e com resoluções espacias variando em 448x448, 512x512 e 560x560. Todos os exames foram obtidos por um scanner Philips Achieva 1.5T. O database é composto por 3286 cortes de T1 e 3286 cortes de T2. Devido a variação entre as resoluções espaciais, adotamos a menor resolução encontrada (224x224) para todos os exames. T1 possui o contraste e brilho de (778,1498); e T2 possui contraste e brilho de (921,1950).

As imagens de TC foram adquiridas helicoidalmente em um scanner Philips Brilliance Big Bore TC (120 kVp e 102 mAs), com espessura do corte de 1 mm, resolução espacial de 512x512, resolução de contraste de 12 bits, window level de 23 e width de 350. Assim como nos exames de RM, o database de TC é composto por 3286 cortes. A base foi escalonada para 224x224 para que RM e TC tivessem a mesma resolução

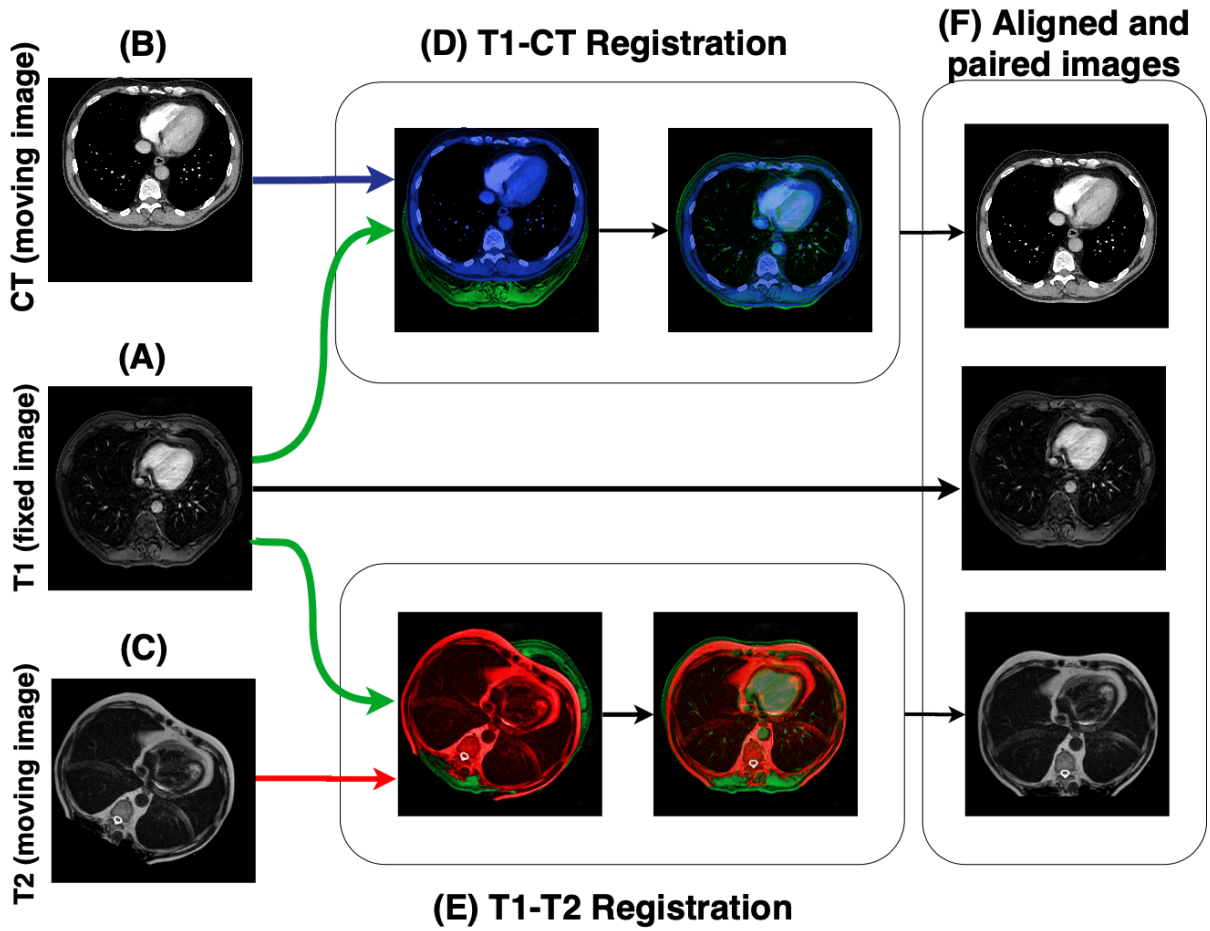


Figura 4.4: Registro de imagem T1-T2 e T1-TC. Fonte: elaborado pelo autor.

Após o registro dos exames (TC, T1 e T2), foram criados os conjuntos de imagens pareadas (Fig. 6-4.5-A) e não pareadas (Fig. 6-4.5-B). As imagens registradas foram utilizadas em ambos conjuntos, com 3286 imagens de cada um dos três domínio para ambos conjuntos. A diferença entre os conjuntos é que o conjunto de imagens pareadas compreende o conjunto em que as três

imagens do mesmo paciente são relativas ao mesmo corte, porém, em diferentes domínios (TC, T1 e T2). Diferentemente do conjunto não pareados, onde as imagens não possuem qualquer relação, ou seja, não existe relação entre os domínios, região anatômica e paciente

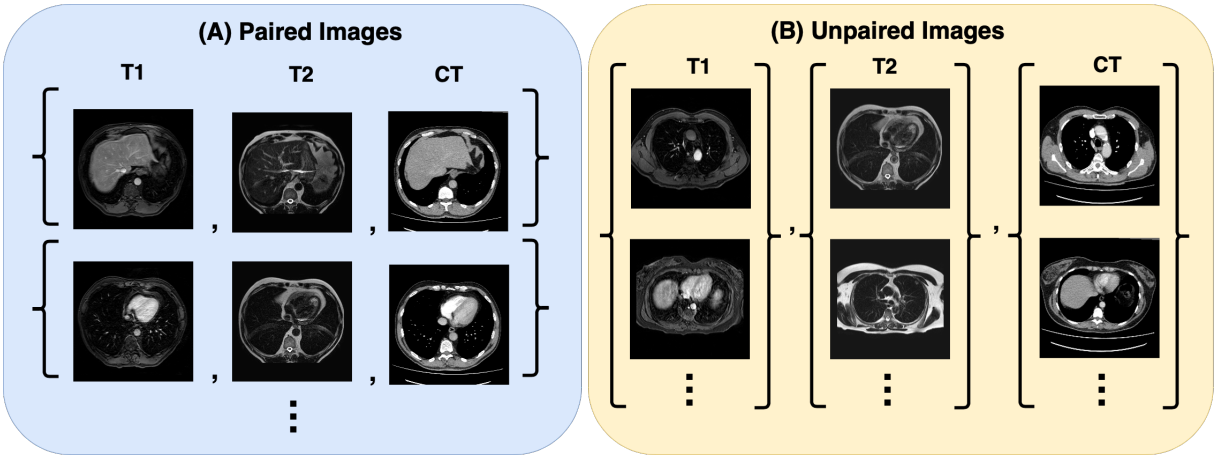


Figura 4.5: O conjunto emparelhado (à esquerda) consiste em exemplos de imagens TC, T1 e T2, nos quais há correspondência entre elas. No não pareado (direita), nenhuma informação é fornecida a respeito da correspondência entre as imagens. Fonte: elaborado pelo autor.

Banco de dados de cabeça. Com o objetivo de permitir a pesquisa reprodutível, avaliamos a MMI-GAN em um dataset público, BRAIN IXI (BRAIN-IXI, 2020). O database também permitiu avaliar a MMI-GAN em uma base maior e de outra região anatômica (cabeça). O BRAIN IXI possui 73 exames de cabeça de T1, T2 e PD obtidos por um scanner GE 1.5T system, totalizando 9928 cortes de T1 , 9928 de T2 e 9928 de PD, as imagens possuem 16 bits com resolução de 256x256, T1 possui brilho e contraste de (19, 1666), T2 possui respectivamente brilho e contraste de (5, 638) e PD possui respectivamente brilho e contraste de (6, 1416).

Definir T1 como imagem fixa no registro entre T1-PD e T1-T2 permitiu que os exames fossem transformados para o sistema de coordenadas comum de T1, possibilitando também o alinhamento entre os exames de PD e T2. Após o registro dos exames (PD, T1 e T2), foram criados os conjuntos de imagens pareadas e não pareadas. As imagens registradas foram utilizadas nos conjuntos pareado e não pareado, com 9928 imagens de cada um dos três domínio (T1, T2 e PD) para os conjuntos pareados e não-pareados.

A diferença entre os conjuntos é que o conjunto de imagens pareadas compreende o conjunto em que as três imagens do mesmo paciente são relativas ao mesmo corte, porém, em diferentes domínios (PD, T1 e T2). Diferentemente do conjunto não pareados, onde as imagens não possuem qualquer relação, ou seja, não existe relação entre os domínios, região anatômica e paciente.

Arquiteturas da linha de base

Neste trabalho comparamos a MMI-GAN com as duas principais GANs utilizadas para tradução de imagens médicas: Pix2pix (ISOLA et al., 2017) e Cyclegan (ZHU et al., 2017).

A Pix2pix (Fig. 4.6-A) tem como característica a tradução unidirecional, pois para aprender a traduzir de um domínio de imagem para outro a Pix2pix precisa de imagens correspondentes entre dois domínios diferentes. Por esse motivo, utilizamos seis Pix2pix para a tradução no banco de dados de tórax, uma para cada par de domínios diferentes: T1-T2, T1-TC, T2-T1, T2-TC, TC-T1 e TC-T2. Para a tradução no banco de dados de cabeça também utilizamos seis Pix2pix, uma para cada par de domínios diferente de T1-T2, T1-PD, T2-T1, T2-PD, PD-T1 e PD-T2.

A CycleGAN, por outro lado, não precisa de imagens correspondentes entre domínios para realizar o seu aprendizado. O seu aprendizado consiste em um mapeamento bidirecional entre os domínios de forma não pareada, utilizando dois geradores (Fig. 4.6-B G1 e G2) e dois discriminadores (Fig. 4.6-B D1 e D2), traduzindo a input image para o domínio de destino, mas como não tem uma imagem para comparar quantitativamente para ajustar seu loss, ela traduz de volta para o domínio da imagem de entrada, e com isso consegue uma imagem para gerar seu loss. Neste trabalho usamos três Cyclegans para fazer a tradução entre os três domínios diferentes no banco de dados de tórax (T1-T2, T1-TC e T2-TC). Para o banco de cabeça, também utilizamos três Cyclegans, para as traduçoes:T1-T2, T1-PD e T2-PD.

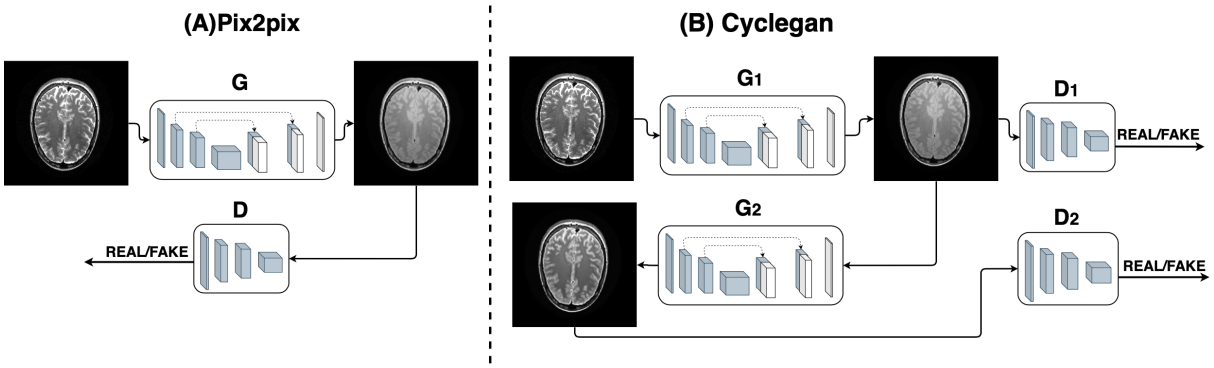


Figura 4.6: Arquiteturas da Pix2pix e da Cyclegan. Fonte: elaborado pelo autor. Fonte: elaborado pelo autor.

O gerador da Cyclegan e da Pix2pix utilizaram a U-Net com as mesmas camadas e parâmetros utilizados na MMI-GAN, com exceção da camada de concatenação da target label embedding, ausente neste modelo. O discriminador da Cyclegan e da Pix2pix utilizaram a Patchgan com as mesmas camadas e parâmetros da MMI-GAN, com exceção do classificador auxiliar, ausente neste modelo.

Treinamento

Para uma comparação justa, todas as arquiteturas (MMI-GANs, Pix2pixs e Cyclegans) foram treinadas usando o mesmo conjunto de treinamento, a mesma duração de épocas e o mesmo número de batchs do gerador para cada atualização do discriminador, além de configurações idênticas de parâmetros (Tab. 4.1). Para cada um dos domínios de imagens as arquiteturas

foram treinadas utilizando 8160 cortes para treinamento e 1768 cortes para teste do banco de cabeça; e 2973 cortes para treinamento e 313 cortes para teste do banco de tórax.

GAN	Epoch	Batch size	Optimizer	Learning rate	Decay	Type of training
MMI-GAN	30	10	Adam	0.0002	0.5	Paired
Pix2pix	30	10	Adam	0.0002	0.5	Paired
Cyclegan	30	10	Adam	0.0002	0.5	Unpaired

Tabela 4.1: Parâmetros utilizados no treinamento das GANs.

MMI-GAN. Para o treinamento da MMI-GAN, em cada época dividimos o conjunto de treinamento em batchs, onde para cada batch escolhemos aleatoriamente input images nos três domínios usados. Para cada input image escolhemos aleatoriamente uma target label para guiar a tradução, e definimos que se a target label indicar o mesmo domínio da input image, eliminamos a target label e escolhemos outra aleatoriamente. Em seguida, passamos o batch de input images em conjunto com suas respectivas target labels para o gerador realizar as traduções. Depois disso, calculamos o loss (Eq. 4.2) entre cada imagem traduzida e seu respectivo ground truth.

Para treinar o discriminador, ora usamos como entrada imagens reais e ora usamos imagens traduzidas pelo gerador para o discriminador avaliar se as imagens traduzidas estão próximas das reais. Para calcular o loss do discriminador (Eq. **??**) fizemos uma média entre o loss das imagens reais e o loss das imagens traduzidas, o mesmo procedimento foi realizado no calculo do loss do classificador auxiliar. Depois da avaliação do discriminador, os pesos do gerador e do discriminador são atualizados.

Pix2pix. Para o treinamento das seis Pix2pixs, em cada época dividimos o conjunto de treinamento em batchs, onde para cada batch escolhemos aleatoriamente input images no domínio de origem e suas respectivas imagens pareadas no domínio de destino (imagem alvo). Em seguida, da-se inicio ao processo de tradução, onde o gerador tenta traduzir o batch de input images para o domínio de destino.

Para cada input image temos uma imagem traduzida, esse conjunto (input image e imagem traduzida) é usado como entrada para o discriminador, que avalia se o par é real ou falso. Em seguida o par de imagem real, que é o par de imagem que realmente são obtido do mesmo paciente em diferentes domínios, passa pela mesma avaliação do par de imagem que contem imagem traduzida.

Depois da avaliação do discriminador, os seus pesos e do gerador são atualizados, que utiliza o loss do discriminador para melhorar a tradução. O loss é baseado na semelhança entre os pixels das imagens traduzidas e da ground truth, sendo responsável por calcular a diferença entre a imagem traduzida e a ground truth. Os dois losses são utilizados para atualizar a rede geradora, tanto o loss do discriminador, obtido pela avalição do realismo da imagem (real ou fake), quanto da diferença das imagens

Cyclegan. Para o treinamento das três Cyclegans, em cada época dividimos o conjunto de treinamento em batchs, onde para cada batch escolhemos aleatoriamente input images, mas diferentemente da Pix2pix e da MMI-GAN, que usam pares (input image, target image), a Cyclegan utiliza um treinamento não pareado.

Pré-processamento de cabeça. Assim como no banco de cabeça, utilizamos a ferramenta SimpleElastix para registrar os exames de PD, T1 e T2. As imagens foram normalizadas para o intervalo de 0 a 255 níveis de cinza. Para o registro entre as imagens de T1 e T2, definimos T1 como imagem fixa e T2 como imagem movida, e adotamos os seguintes parâmetros: Mutual Information como métrica de similaridade, Gradient Descent Line Search como otimizador e interpolação linear. Para o registro entre T1 e PD, T1 foi definida como imagem fixa e PD como imagem movida, e usamos os parâmetros: Mutual Information como similarity metric, optimizer Gradient Descent Line Search como otimizador e interpolação linear

Depois de traduzir o batch de imagens, o discriminador recebe como entrada as imagens traduzidas em conjunto com suas respectivas input images com o objetivo de avaliar se o par é real ou falso. Após este passo, o discriminador recebe as imagens reais com suas input images para avaliação (real ou falso), e gera o loss do discriminador que irá atualizar os seus pesos. Logo, utilizamos o loss cíclico com o loss do discriminador para atualizar o gerador

Métricas de Avaliação

Os resultados foram avaliados usando as principais métricas de similaridade entre imagens: MAE (SAMMUT; WEBB, 2010), PSNR (HORE; ZIOU, 2010), MI (KRASKOV; STÖGBAUER; GRASSBERGER, 2004) e SSIM (WANG et al., 2004). Neste trabalho as métricas foram usadas para comparar quão próximas estão as imagens traduzidas pelas diferentes arquiteturas das suas respectivas ground truths.

As imagens traduzidas e suas ground truth foram comparadas por meio do erro médio absoluto (MAE), sendo definido pela Eq. 4.6, onde m e n representam o total de linhas e colunas da imagem, N o total de pixels da imagem, $y(i,j)$ e $y_p(i,j)$ representam os valores dos pixels da ground truth e das imagens traduzidas.

$$MAE = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}|y(i,j) - y_p(i,j)| \tag{4.6}$$

Utilizamos o PSNR para avaliar a razão entre o valor máximo possível de uma imagem e o erro que afeta a sua qualidade de representação. Para calcular a relação sinal ruído de pico, precisamos primeiro calcular o erro quadrático médio cuja for fórmula é dada por:

$$MSE = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}(y(i,j) - y_p(i,j))^2 \tag{4.7}$$

Na equação 4.7, onde m e n representam o total de linhas e colunas da imagem, N o total

de pixels da imagem e $y(i,j)$ e $y_p(i,j)$ representam valores de pixel das imagens verdadeiras e traduzidas respectivamente. A fórmula PSNR é dada pela Equação 4.8, onde R indica o valor máximo que um pixel pode atingir, no nosso caso $R = 255$.

$$PSNR = 10 \log_{10} \frac{R^2}{MSE} \tag{4.8}$$

A informação mútua (MI) foi utilizada como medida de similaridade que não pressupõe uma relação funcional prévia entre as imagens comparadas. Em vez disso, pressupõe uma relação estatística que pode ser capturada pela análise da entropia conjunta das imagens. Na Fig. 4.7 temos uma representação de como funciona a informação Mutual, na qual X e Y são imagens. A área contida por ambos os círculos é a entropia conjunta H(X, Y). O círculo à esquerda é a entropia individual H(X), a entropia condicional sendo H(X | Y). O círculo à direita é H(Y), a entropia condicional sendo H(Y | X), e na interseção entre H(X) e H(Y) temos Informação Mútua I(X, Y) (Eq. 4.9).

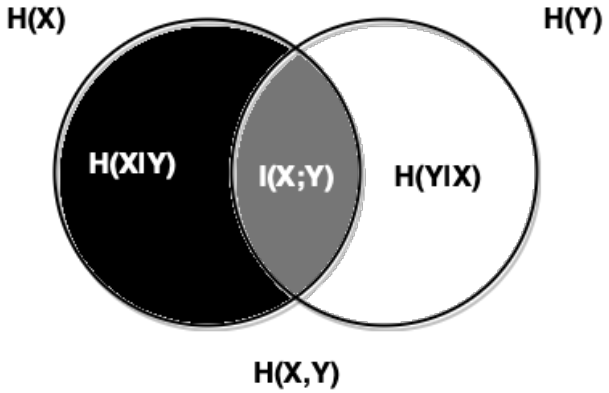


Figura 4.7: Gráfico de informação mútua. Fonte: elaborado pelo autor.

$$I(X,Y) = H(X,Y) - H(X|Y) - H(Y|X) \tag{4.9}$$

SSIM foi usado para avaliar a semelhança entre a imagem traduzida e sua verdade fundamental, usando características estruturais, luminância e contraste para quantificar essa diferença. SSIM é descrito pela Eq. **??**, onde μ_x, μ_y, σ_x, σ_y e σ_{xy} são os médias locais, desvios padrão e covariância cruzada para as imagens x (imagem traduzida), y (*ground truth*). As constantes C_1 e C_2 são usadas para estabilizar a divisão.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.10}$$

Para melhor avaliar nossos resultados, utilizamos o teste não paramétrico de Kruskal Wallis para determinar se havia diferença significativa nos três grupos (MMIGAN, Pix2pix e Cyclegan), testamos todas as métricas utilizadas neste trabalho (MAE, PSNR , MI e SSIM). Após o teste de Kruskal Wallis, realizamos o teste de Dunn para múltiplas comparações entre os pares

de grupos, para descobrir quais diferiam significativamente (p <0.05). Se houvesse diferença significativa entre os grupos, seria feita uma análise descritiva com base na mediana, uma vez que a distribuição não segue a distribuição normal, para saber se os resultados do MMI-GAN foram significativamente maiores ou menores. As análises estatísticas foram realizadas usando R (versão 4.0.2) com pacote dplyr.

5 Resultados e Discussão

Este capítulo foi dividido em duas seções principais: Resultados do Banco de Dados de Chest (5.1) e Resultados do Banco de Dados de Cabeça (5.1). Em cada seção, apresentamos os resultados obtidos pelos GANs utilizados neste trabalho (MMI-GAN, Pix2pix e Cyclegan), onde cada métrica foi apresentada individualmente. Por fim, analisamos as características mais relevantes obtidas no processo de tradução entre as GANs avaliadas neste trabalho.

5.1 Resultados do Banco de Dados Tórax

A Fig. 5.1 mostra seis exemplos de imagens de entrada TC, T1 ou T2, suas respectivas traduções e as respectivas imagens com as quais as imagens traduzidas devem se parecer (target). Na Tabela 5.1, apresentamos a avaliação quantitativa usando MAE, PSNR, MI e SSIM para comparar os diferentes métodos no conjunto de teste.

	Mean Absolute Error (MAE)			Peak Signal-to-Noise Ratio (PSNR)			Mutual information (MI)			Structural Similarity Index Measure (SSIM)		
Translation	MMI-GAN	Pix2pix	Cyclegan	MMI-GAN	Pix2pix	Cyclegan	MMI-GAN	Pix2pix	Cyclegan	MMI-GAN	Pix2pix	Cyclegan
T1-T2	24.41±4.50	**23.06±4.0**	31.20±6.70	16.25±1.06	**16.50±1.03**	13.62±1.39	**0.98±0.06**	0.91±0.06	0.82±0.09	0.50±0.07	**0.52±0.06**	0.44±0.06
T1-TC	**26.49±4.41**	27.22±4.05	31.80±5.02	**13.45±0.65**	13.27±0.56	12.41±0.65	**0.74±0.04**	0.69±0.05	0.65±0.05	**0.48±0.06**	0.46±0.06	0.38±0.061
T2-T1	**24.37±8.36**	26.24±4.06	29.19±4.79	16.83±2.83	**17.00±1.26**	16.02±1.33	**0.84±0.08**	0.74±0.07	0.67±0.09	**0.28±0.04**	0.17±0.03	0.12±0.03
T2-TC	**26.86±4.20**	28.05±4.03	51.75±10.65	**13.40±0.71**	13.14±0.57	9.68±1.18	**0.73±0.06**	0.69±0.05	0.63±0.05	**0.47±0.06**	0.45±0.05	0.29±0.04
TC-T1	**23.31±6.19**	23.88±3.73	29.33±4.25	17.16±1.96	**17.71±1.21**	16.38±1.25	**0.77±0.06**	0.70±0.06	0.65±0.07	**0.24±0.03**	0.16±0.02	0.13±0.01
TC-T2	26.75±5.52	**26.05±5.81**	34.21±4.62	15.26±1.47	**15.40±1.70**	12.84±0.84	**0.89±0.08**	0.80±0.06	0.78±0.05	**0.47± 0.08**	**0.47±0.07**	0.43±0.07

Tabela 5.1: Comparação da média e dos desvios padrão obtidos no banco de dados do tórax.

MAE. Encontramos diferenças significativas ($p < 0.05$) ao comparar as traduções de Pix2pix e Cyclegan, Cyclegan e MMI-GAN, Pix2pix e MMI-GAN. Para Pix2pix e Cyclegan em todas as traduções, houve diferenças significativas, um comportamento que se repete ao comparar Cyclegan e MMI-GAN também para todas as traduções e ambos, Pix2pix e MMI-GAN, com medianas abaixo de Clyclegan (Fig. 5.2). Portanto, assumimos que, para a métrica MAE, as técnicas de tradução mais adequadas foram MMI-GAN e Pix2pix. Quando observamos se havia diferenças estatísticas significativas entre o MMI-GAN e Pix2pix, os resultados foram variados (Fig. 5.2), notamos que houve diferença estatística nas traduções T2-T1 e T2-TC , em que ao fazermos uma análise descritiva, notamos que a mediana do MMI-GAN foi inferior à do Pix2pix, portanto, nessas traduções o MMI-GAN apresentou melhores resultados. Porém, para os casos T1-TC e TC-T1, não foram encontradas diferenças estatisticamente significativas, o que nos leva a crer que o MMI-GAN é pelo menos tão eficiente quanto o Pix2pix.

PSNR. Assim como para MAE, Pix2pix e MMI-GAN obtiveram diferenças significativas em relação ao Cyclegan em todas as traduções, com medianas sempre maiores (Fig. 5.3). Na comparação do MMI-GAN com o Pix2pix, os resultados foram variados, sendo que para as traduções T1-TC e T2-TC, houve diferença significativa. Em uma análise descritiva, notamos que a mediana do MMI-GAN foi superior à do Pix2pix, portanto nessas traduções o MMI-GAN se mostrou superior. Porém, para os casos T1-TC e T2-TC, não foram encontradas diferenças estatisticamente significativas, o que nos leva a acreditar que o MMI-GAN é pelo menos tão eficiente quanto o Pix2pix. Porém, apesar de apresentarem mediana menor nos casos T2-T1 e TC-T2, os testes estatísticos não mostraram diferença significativa entre o MMI-GAN e o Pix2pix, portanto o MMI-GAN e o Pix2pix podem ser considerados equivalentes para esses casos.

MI. Como com as outras métricas, houve diferenças significativas (p <0.05) entre Pix2pix e Cyclegan e entre MMI-GAN e Cyclegan em todas as traduções (Fig. 5.4), com Pix2pix e MMI-GAN sempre com medianas acima de Cyclegan. Na comparação entre MMI-GAN e Pix2pix, notamos que houve diferenças estatísticas em todas as traduções, exceto na tradução TC-T2, que evidenciou a superioridade do MMI-GAN sobre Pix2pix nesta métrica.

SSIM. Assim como nas demais métricas, existe diferença significativa para Pix2pix e Cyclegan em todas as traduções, enquanto para MMI-GAN, apenas na tradução TC-T2, não houve diferença estatística (p <0.05). Nos casos em que houve diferenças, o MMI-GAN e o Pix2pix sempre obtiveram medianas superiores ao Cyclegan. Portanto, presumimos que para a métrica SSIM, as técnicas de tradução mais adequadas foram MMI-GAN e Pix2pix. Por fim, observamos diferença estatística entre Pix2pix e MMI-GAN (Fig. 5.5) em quatro traduções: T1-TC, T2-T1, T2-TC e TC-T1.

Cyclegan foi a técnica que obteve os menores valores quantitativos em todas as métricas, apesar de ter maior liberdade, pois pode ser treinada com dados ininterruptos. Apesar desse ganho de liberdade, os resultados abaixo de Pix2pix e MMI-GAN podem ser atribuídos ao fato de o treinamento ter sido realizado sem par. A perda cíclica, em que Cyclegan se traduz no domínio de destino e depois se traduz de volta no domínio de entrada. Este fato pode ter limitado seu poder de tradução, visto que as traduções entre diferentes domínios não são lineares, tanto que para qualquer tradução temos que sua tradução reversa tem resultados diferentes em todas as métricas.

Acreditamos que a superioridade quantitativa do MMI-GAN sobre o Cyclegan foi devido ao efeito implícito do aumento de dados em um ambiente de múltiplos domínios. O banco de dados do tórax contém um número de amostras relativamente pequeno, 2.973 imagens por domínio. Quando treinado em dois domínios, o CycleGAN pode usar apenas 2 * 2973 imagens de treinamento por vez, mas o MMI-GAN pode usar todas as imagens de todos os domínios, neste caso 3 * 2973. Isso permitiu que o MMI-GAN aprendesse adequadamente enquanto mantinha a qualidade e a nitidez da imagem traduzida.

Tanto Pix2pix quanto MMI-GAN obtiveram resultados com diferença estatística em todas

as métricas, ora Pix2pix superior, ora MMI-GAN superior, sendo comparável em algumas traduções. Entretanto, um MMI-GAN se destaca pelo importante diferencial de usar apenas um MMI-GAN para todas as traduções, enquanto um Pix2pix usa seis Pix2pixs, um para alguns domínios envolvidos. Apesar da diferença serem cinco GANs, quando projetamos um cenário onde existem K domínios, podemos ver que Pix2pix não é robusto o suficiente, sendo necessários K * (K-1) Pix2pixs para realizar todas as traduções, enquanto um MMI-GAN continuaria precisando de apenas um.

5.2 Resultados da Tradução de Cabeça

A Fig. 5.6 mostra seis exemplos de imagens de entrada PD, T1 ou T2, suas respectivas traduções e respectivas imagens com as quais as imagens traduzidas devem se parecer (alvo). Na Tabela 5.2, temos uma avaliação quantitativa usando MAE, PSNR, MI e SSIM para comparar os diferentes métodos de conjunto de teste.

	Mean Absolute Error (MAE)			Peak Signal-to-Noise Ratio (PSNR)			Mutual information (MI)			Structural Similarity Index Measure (SSIM)		
Translation	MMI-GAN	Pix2pix	Cyclegan	MMI-GAN	Pix2pix	Cyclegan	MMI-GAN	Pix2pix	Cyclegan	MMI-GAN	Pix2pix	Cyclegan
T1-T2	**10.54±4.76**	10.72±5.01	17.50±6.11	**21.09±2.31**	20.79±2.47	17.33±2.87	1.17±0.24	**1.19±0.23**	0.82±0.21	0.74±0.07	**0.74±0.09**	0.51±0.08
T1-PD	**9.01±3.82**	9.11±5.04	15.80±4.05	**22.51±2.43**	22.39±2.62	18.28±1.88	1.27±0.27	**1.30±0.26**	0.87±0.20	0.78±0.07	**0.78±0.10**	0.50±0.08
T2-T1	6.71±2.79	**6.47±5.92**	13.17±3.48	24.78±2.31	**25.61±3.15**	19.33±2.05	1.16±0.29	**1.26±0.28**	0.83±0.16	0.83±0.08	**0.85±0.10**	0.64±0.07
T2-PD	5.79±3.30	**5.16±2.85**	15.34±5.22	27.39±2.84	**28.54±3.22**	18.83±3.59	1.43±0.29	**1.50±0.27**	0.87±0.20	0.90±0.06	**0.91±0.08**	0.59±0.08
PD-T1	6.307±2.73	**5.80±2.79**	12.76±4.36	25.10±2.34	**25.69±2.86**	20.19±2.01	1.16±0.30	**1.23±0.28**	0.81±0.23	0.83±0.08	**0.85±0.08**	0.61±0.09
PD-T2	6.73±4.77	**5.51±2.88**	12.06±3.25	25.57±2.89	**26.94±2.93**	21.46±2.21	1.31±0.27	**1.36±0.25**	1.1±0.19	0.86±0.06	**0.88±0.03**	0.69±0.05

Tabela 5.2: Comparação da média e dos desvios padrão obtidos na base de cabeça.

MAE. Com a realização dos testes estatísticos, observou-se diferença estatisticamente significante (p <0.05) quando relacionado ao Pix2pix e Cyclegan em todas as traduções, sempre com Pix2pix com medianas menores, e essa relação também foi encontrada quando relacionada ao MMI-GAN e Cyclegan para todas as traduções (Fig. 5.7)). Este resultado nos leva a crer que para a métrica MAE, MMI-GAN e Pix2pix foram as técnicas mais adequadas para traduções. Ao comparar o MMI-GAN e o Pix2pix, não foi encontrada diferença significativa para a tradução T1-T2. Portanto, assumimos que neste tipo de tradução os GANs são equivalentes.

PSNR. Assim como na métrica MAE, notamos que havia diferenças estatísticas entre Pix2pix e Cyclegan e entre MMI-GAN e Cyclegan em todas as traduções (Fig. 5.8), o que nos leva a acreditar que MMI-GAN e Pix2pix foram as técnicas mais adequadas para traduções. O MMI-GAN foi superior ao Pix2pix no caso de T1-T2 e T2-PD.

MI. Apesar de ter uma mediana menor no caso T1-T2, os testes estatísticos não mostraram diferenças significativas entre Pix2pix e MMI-GAN (Fig. 5.9), ou seja, Pix2pix e MMI-GAN podem ser considerados equivalentes . Entre Pix2pix e Cyclegan e entre MMI-GAN e Cyclegan, diferenças significativas foram encontradas em todas as traduções, assim como nas métricas MAE e PSNR.

SSIM. Assim como na métrica MI, a única tradução em que o MMI-GAN não obteve diferença estatística com Pix2Pix foi T1-T2 (Fig. 5.10). No caso do Cyclegan, houve diferenças

significativas com o MMI-GAN e Pix2pix em todas as traduções, assim como nas outras métricas, tornando MMI-GAN e Pix2pix superiores nesta métrica.

As imagens traduzidas pelo MMI-GAN conseguiram obter MAE de 5.792, PSNR de 27.398, MI de 1.430 e SSIM de 0,900. Esses resultados foram superiores aos obtidos com a base de dados do tórax, atribuímos isso ao fato de que na base de dados da cabeça as imagens foram obtidas utilizando o mesmo equipamento de ressonância magnética de forma consecutiva, o que facilitou o processo de registro. Além disso, a complexidade de registrar as imagens do tórax é muito maior do que a da cabeça, uma vez que o pulmão se move e a cabeça não. A RM do pulmão também é um exame que apresenta muito ruído devido à presença de ar nos pulmões, já que a RM é baseada em átomos de hidrogênio.

O efeito da complexidade do registro de imagens na base de dados de tórax e cabeça pode ser melhor visto quando olhamos os mapas de erros nas traduções das duas bases de dados, esses mapas mostram os locais onde mais erros ocorreram durante a tradução (Fig. 5.11). Os erros na Cabeça foram mais distribuídos e homogêneos, enquanto no Tórax os erros mais proeminentes concentraram-se na região óssea e no contorno corporal, isso pode ser atribuído a pequenos desalinhamentos causados durante o processo de registro, conforme citado anteriormente, consequentemente causando desalinhamento em as imagens traduzidas.

Assim como na avaliação do banco de dados do tórax, o Cyclegan foi o que obteve os menores resultados em todas as métricas, o que reforça que a perda utilizada para o treinamento do Cyclegan limitou as traduções. Apesar de ter trazido a liberdade de trabalhar com imagens desemparelhadas, fazendo com que você aprenda informações gerais sobre o domínio, ao passo que no treinamento emparelhado, usamos informações da imagem de *ground truth* para auxiliar no treinamento (Eq. 4.2).

Assim como no Chest, acreditamos que a superioridade do MMIGAN em testes estatísticos sobre os resultados quantitativos se deve ao efeito implícito causado pelo aumento de dados em um ambiente de múltiplos domínios. No entanto, tanto Pix2pix quanto MMI-GAN alcançaram resultados muito próximos em algumas métricas. Porém, o MMI-GAN teve um diferencial importante pelo fato de utilizar apenas um MMI-GAN para todas as traduções, enquanto o Pix2pix utilizou seis Pix2pixs, um para cada par de domínios envolvidos. Embora a diferença inicialmente pareça pequena, quando projetamos um cenário onde existem K domínios, podemos ver que Pixpix não tinha robustez suficiente, sendo necessário K*(K-1) Pix2pixs para fazer todas as traduções, enquanto o MMI-GAN faria precisa apenas de um GAN para realizar todas as traduções K.

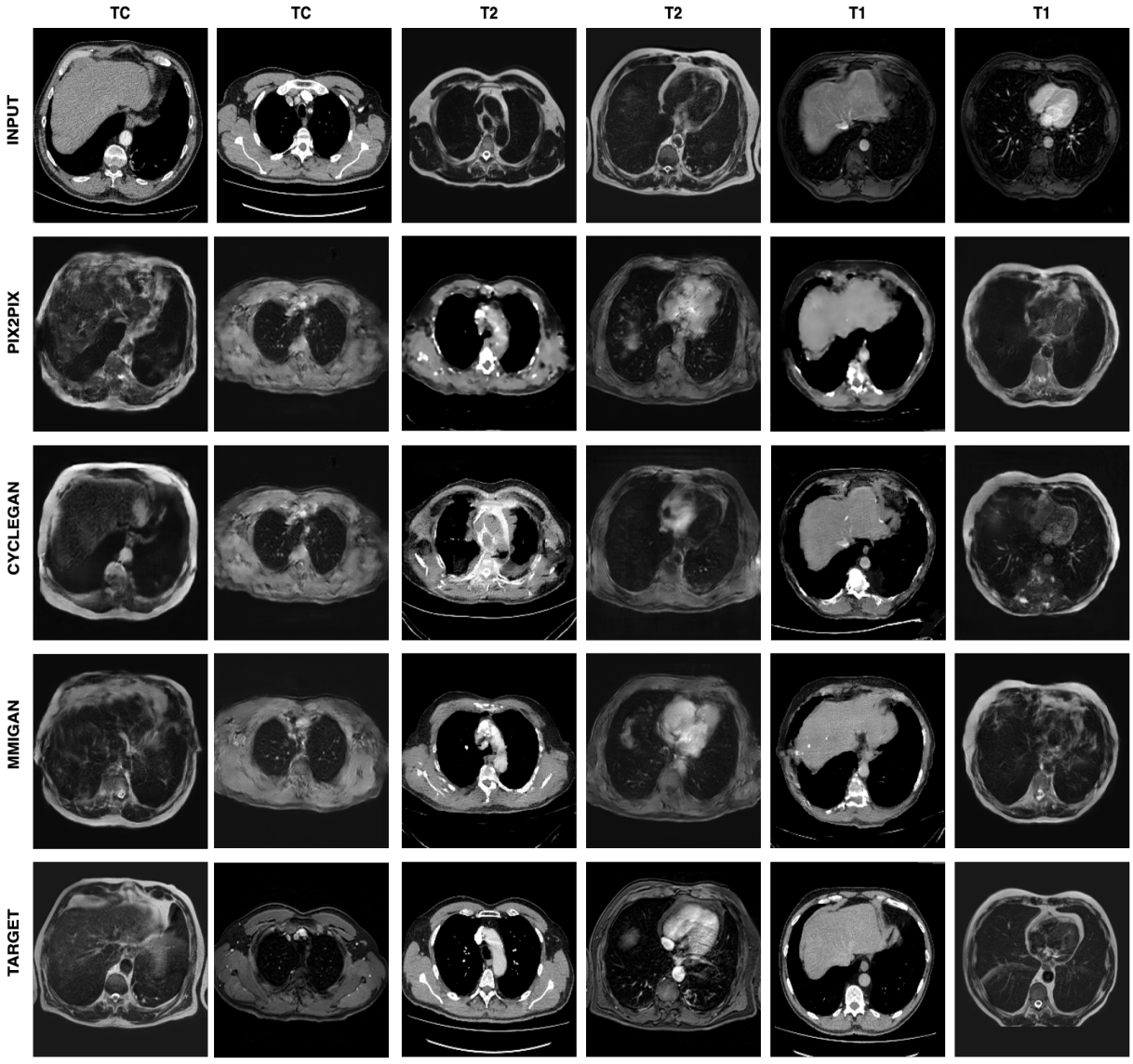


Figura 5.1: De cima a baixo por linha, temos as imagens de entrada, as imagens traduzidas pelo Pix2pix, as imagens traduzidas pelo MMI-GAN e, por fim, as imagens de verdade para cada tradução. As imagens de entrada são dos respectivos domínios: TC, TC, T2, T2, T1 e T1. As imagens de *ground truth* são: T2, T1, TC, T1, TC e T2. Fonte: elaborado pelo autor.

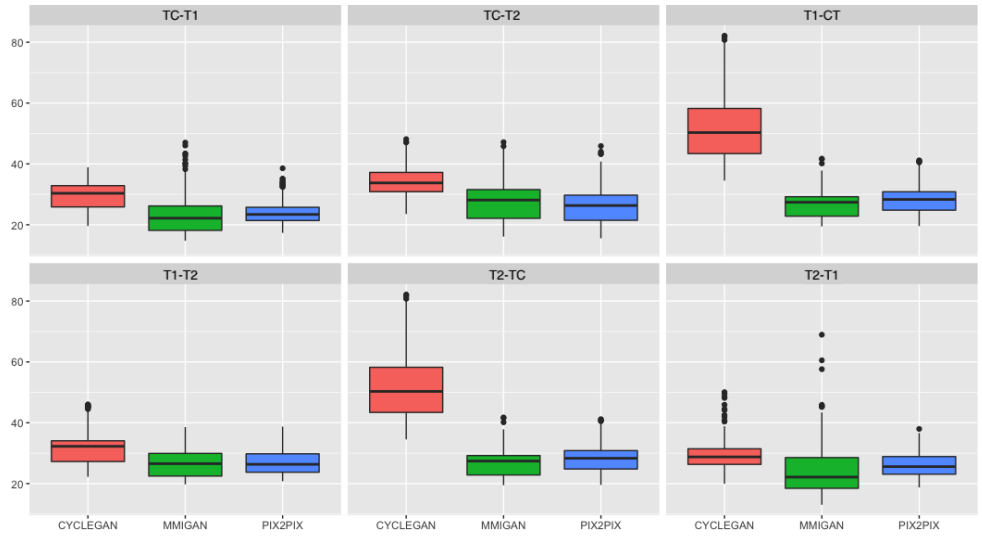


Figura 5.2: Boxplots with all the MAE metric results, where the red boxplot represents the results of Cyclegan, the green of MMI-GAN and the blue of Pix2pix. Translations with statistical differences between MMIGAN and Pix2pix: T2-T1 and T2-TC. No differences: T1-TC and TC-T1. Fonte: elaborado pelo autor.

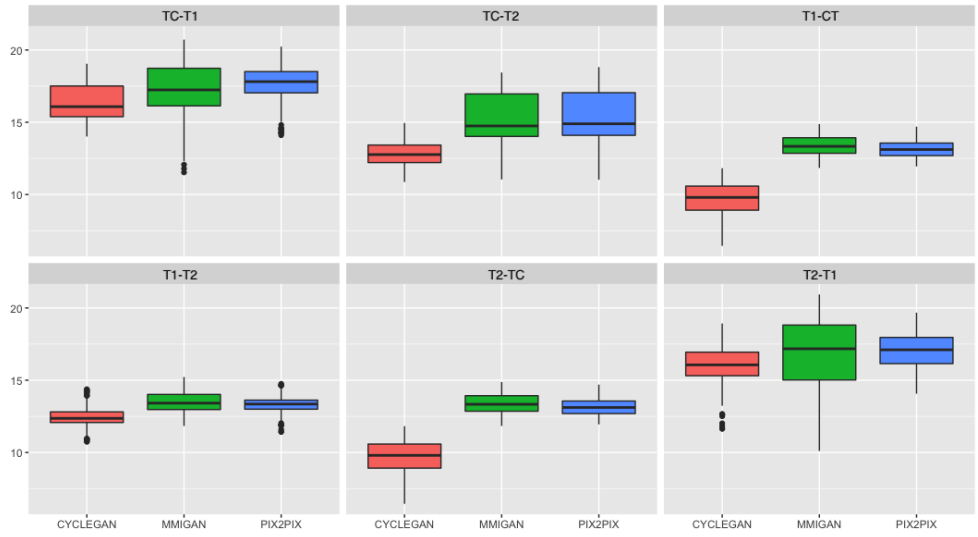


Figura 5.3: Boxplots com todos os resultados da métrica PSNR, onde o boxplot vermelho representa os resultados do Cyclegan, o verde do MMI-GAN e o azul do Pix2pix. Traduções com diferenças estatísticas entre MMIGAN e Pix2pix: T1-TC e T2-TC. Sem diferenças: T2-T1 e TC-T2. Fonte: elaborado pelo autor. Fonte: elaborado pelo autor.

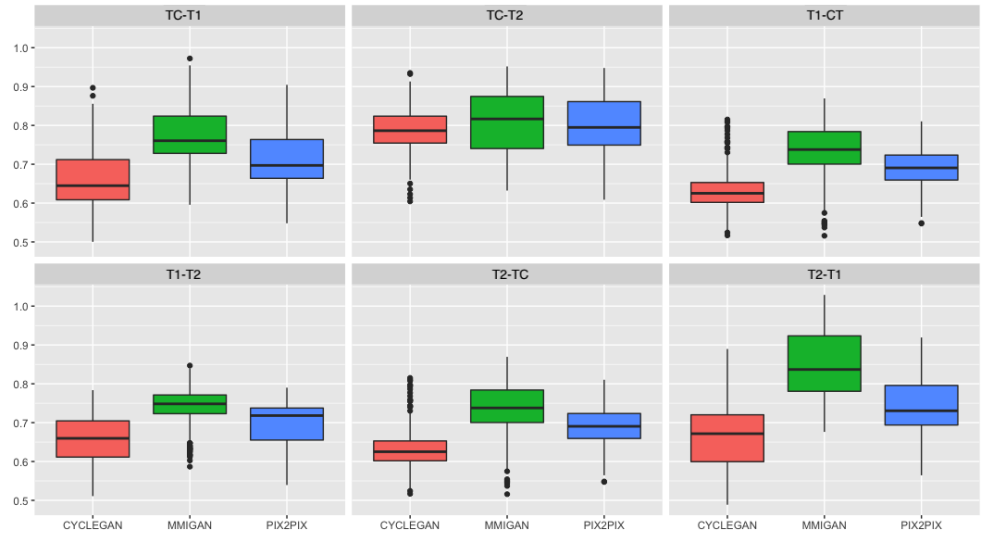


Figura 5.4: Boxplots com todos os resultados MI métricos, onde o boxplot vermelho representa os resultados do Cyclegan, o verde do MMI-GAN e o azul do Pix2pix. Traduções com diferenças estatísticas entre MMIGAN e Pix2pix: T1-T2, T1-TC, T2-T1, T2-TC e TC-T1. Sem diferenças: TC-T2. Fonte: elaborado pelo autor.

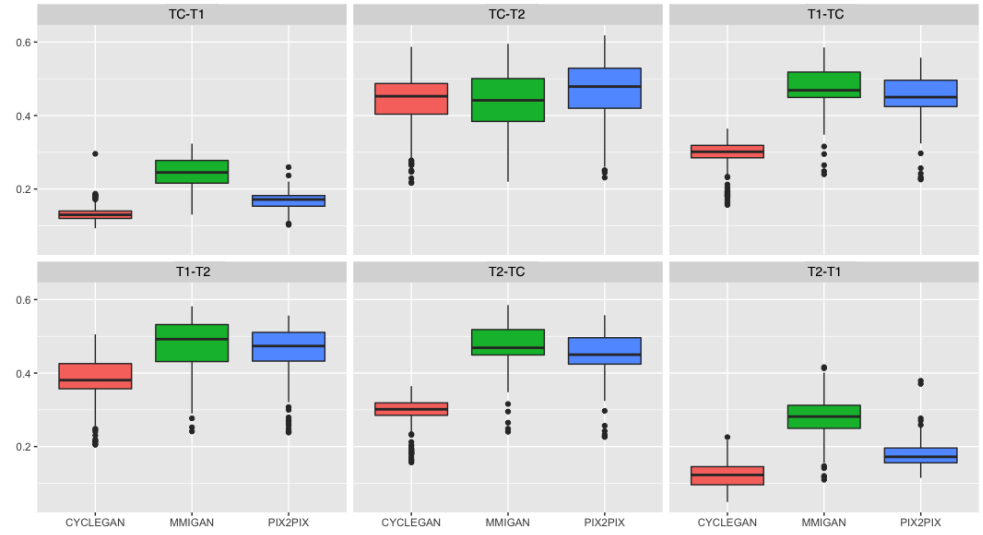


Figura 5.5: Boxplots com todos os resultados SSIM métricos, onde o boxplot vermelho representa os resultados do Cyclegan, o verde do MMI-GAN e o azul do Pix2pix. Traduções com diferenças estatísticas entre MMIGAN e Pix2pix: T1-TC, T2-T1, T2-TC e TC-T1. Fonte: elaborado pelo autor.

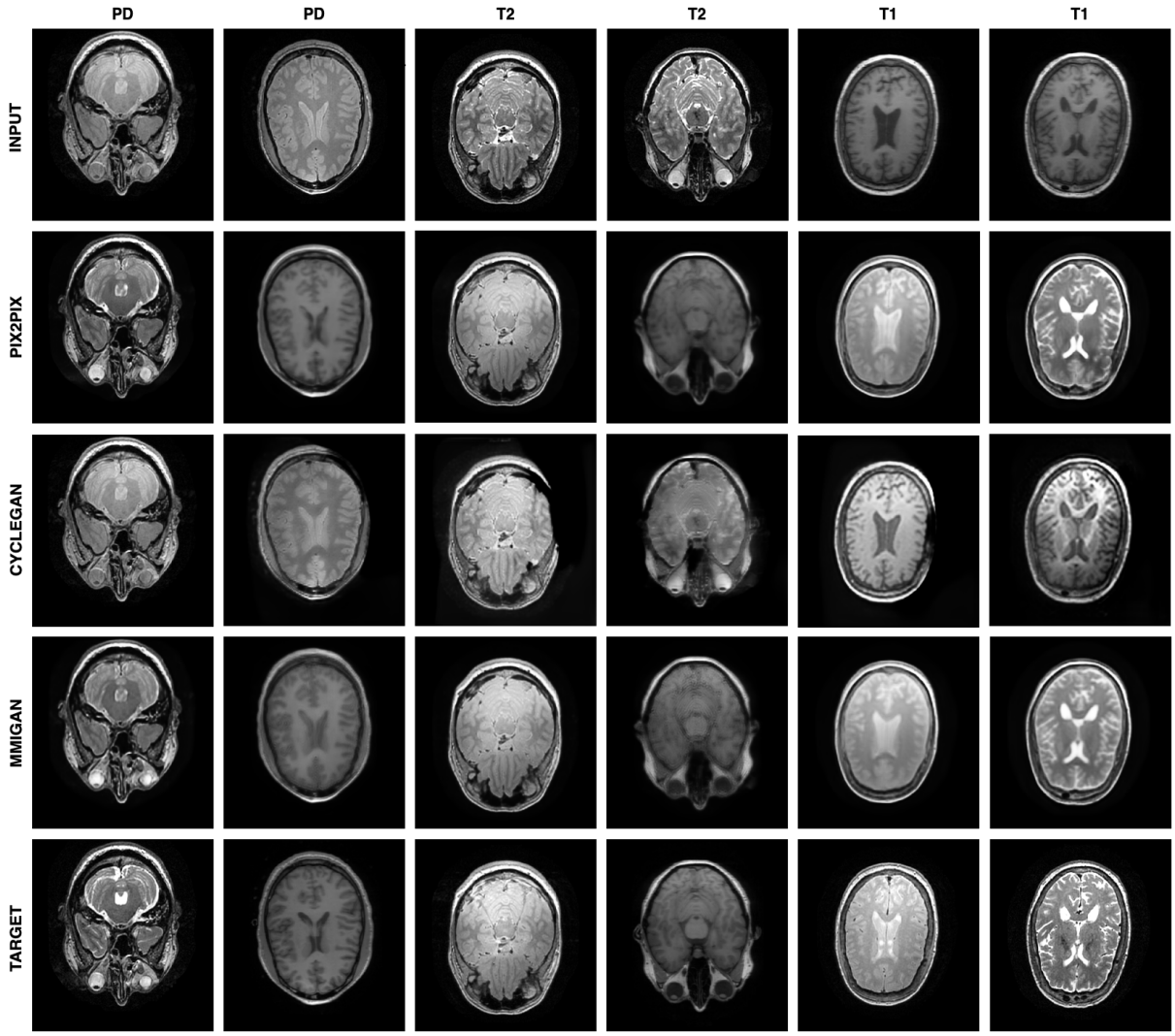


Figura 5.6: De cima a baixo por linha, temos as imagens de entrada, as imagens traduzidas pelo Pix2pix, as imagens traduzidas pelo MMI-GAN e, finalmente, as imagens verdadeiras para cada tradução. As imagens de entrada são dos respectivos domínios: PD, PD, T2, T2, T1 e T1. As imagens de *ground truth* são: T2, T1, PD, T1, PD e T2. Fonte: elaborado pelo autor.

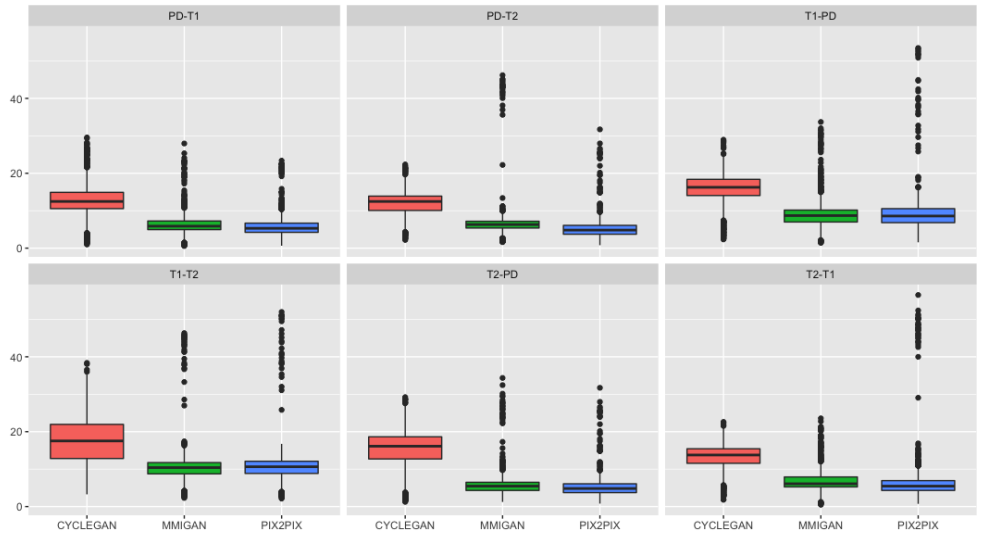


Figura 5.7: Boxplots with all the MAE metric results, where the red boxplot represents the results of Cyclegan, the green of MMI-GAN and the blue of Pix2pix. Translations without statistical differences between MMIGAN and Pix2pix: T1-T2. Fonte: elaborado pelo autor.

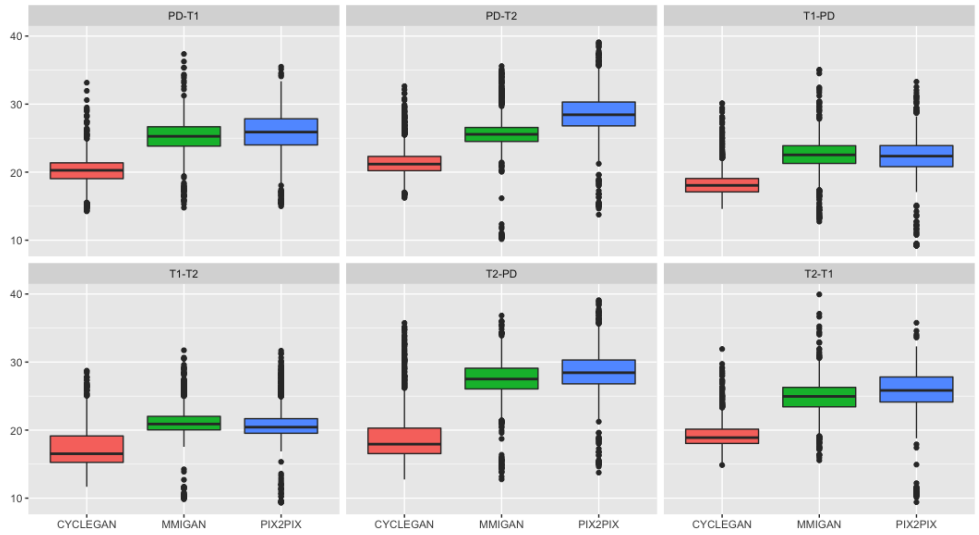


Figura 5.8: Boxplots com todos os resultados da métrica PSNR, onde o boxplot vermelho representa os resultados do Cyclegan, o verde da MMI-GAN e o azul da Pix2pix. Traduções com diferenças estatísticas entre MMIGAN e Pix2pix: T1-T2 e T2-PD. Fonte: elaborado pelo autor.

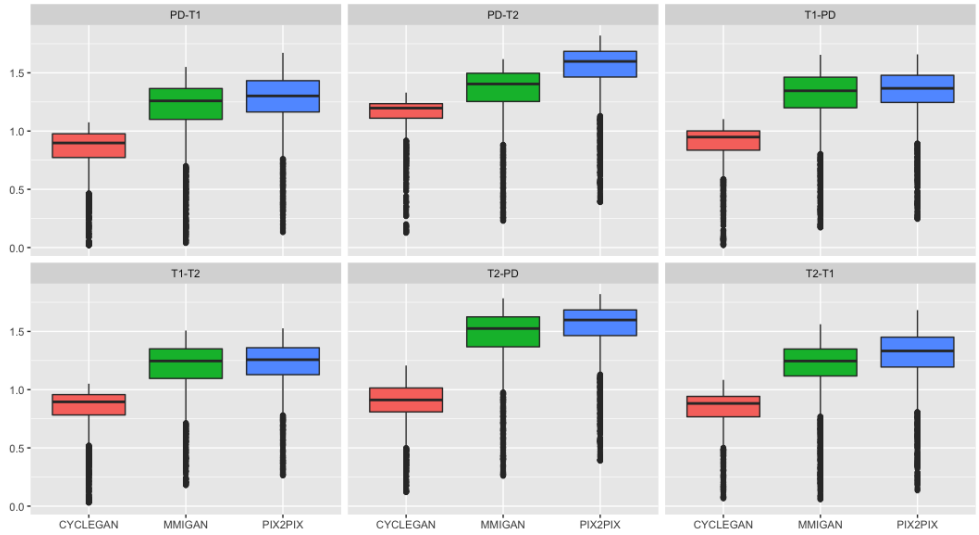


Figura 5.9: Boxplots com todos os resultados MI métricos, onde o boxplot vermelho representa os resultados do Cyclegan, o verde do MMI-GAN e o azul do Pix2pix. Traduções sem diferenças estatísticas entre MMIGAN e Pix2pix: T1-T2.

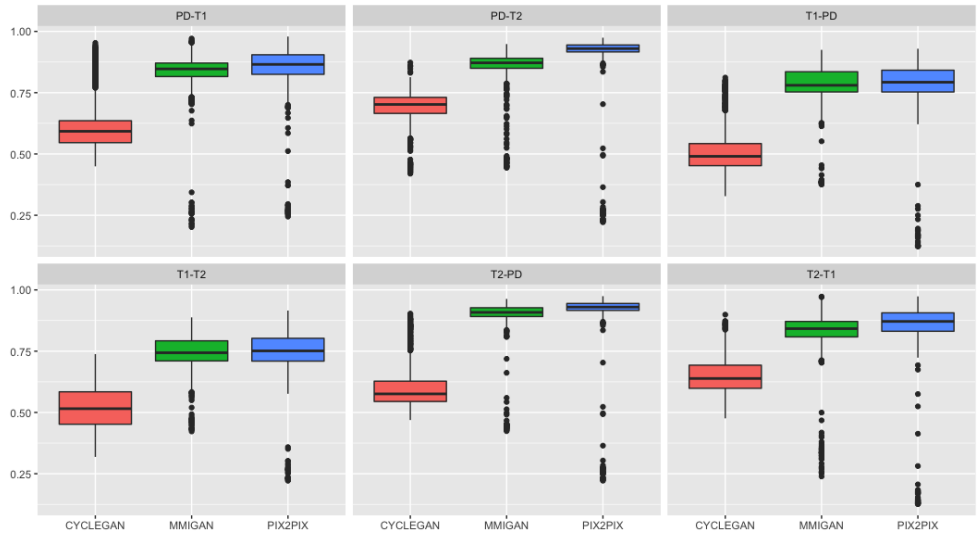


Figura 5.10: Boxplots com todos os resultados da métrica SSIM, onde o boxplot vermelho representa os resultados do Cyclegan, o verde do MMI-GAN e o azul do Pix2pix. Traduções sem diferenças estatísticas entre MMIGAN e Pix2pix: T1-T2. Fonte: elaborado pelo autor. Fonte: elaborado pelo autor.

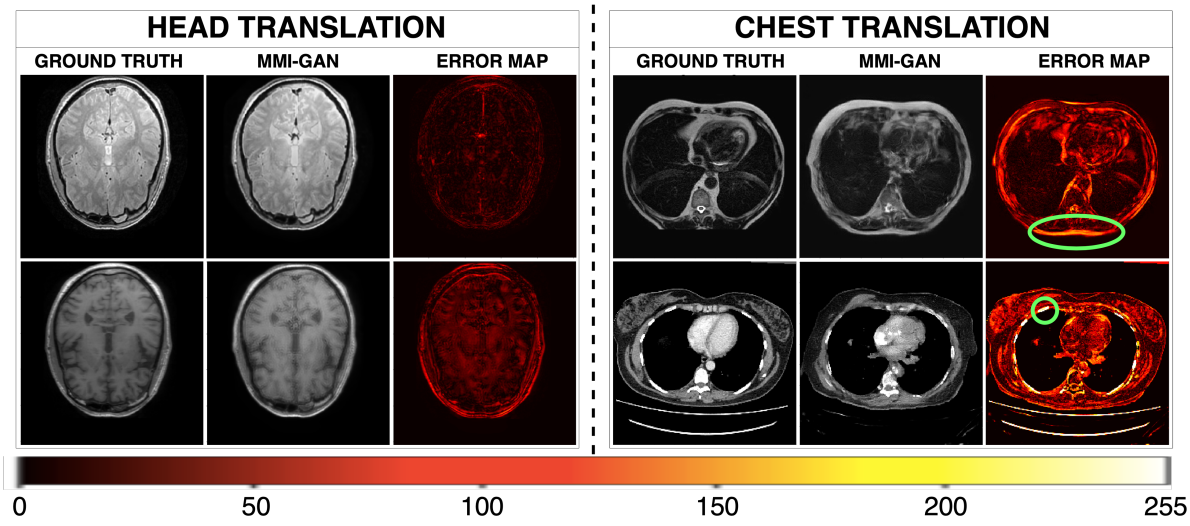


Figura 5.11: Exemplo da distribuição de erro nas traduções de tórax e cabeça. Fonte: elaborado pelo autor.

6 Conclusão

Neste trabalho, apresentamos a MMI-GAN, uma nova abordagem para tradução entre múltiplos domínios de imagem, capaz de traduzir imagens intermodais (TC e RM) e intramodais (PD, T1 e T2). Demonstramos a eficiência do condicional em aprender a mapear entre múltiplos domínios usando apenas um único gerador e um discriminador, treinando efetivamente a partir de imagens de todos os domínios.

O processo de tradução usando GANs se mostrou variante ao pareamento intra/inter modalidade de exames de imagens médicas, com a Cyclegan, que utiliza dados não pareados, sempre obtendo resultados quantiativos inferiores a Pix2pix e a MMI-GAN, que utilizavam dados pareados, essa diferença ficou comprovada com o teste estatísto de Dunn. Também pudemos notar que devido a maior complexidade no registro de imagens na base de dados de tórax, os erros na tradução foram maiores se comparados aos erros da base de cabeça. Os erros na Cabeça foram mais distribuídos e homogêneos, enquanto no Tórax os erros foram mais concentrados na região óssea e no contorno corporal.

Propomos uma arquitetura GAN que pode ser facilmente estendida a outras tarefas de tradução para beneficiar a comunidade de imagens médicas. As imagens traduzidas pelo MMI-GAN conseguiram obter MAE de 5.792, PSNR de 27.398, MI de 1.430 e SSIM de 0.900. Os seus resultados se mostraram, por muitas vezes estatisticamente equiparáveis ou superiores a Pix2pix e em quase todas as traduções foi superior a Cyclegan.

6.1 Trabalhos Futuros

Com o objetivo de melhorar o modelo de classificação para nódulos pulmonares, e ainda, incluir nódulos semi-sólidos e não sólidos, os planejamentos futuros são os seguintes:

- incorporar mais domínios de imagens médicas como PET, MRA, etc;
- incorporar uma perda cíclica para criar um modelo que seja capaz de ser treinado com dados não pareados também;
- Ampliar a avaliação, fazendo um teste qualitativo de turing visual com um grupo de radiologistas experientes;

- Explorar técnicas de otimização.

Referências Bibliográficas

ĐAN, I. et al. Radiotherapy treatment planning: benefits of ct-mr image registration and fusion in tumor volume delineation. *Vojnosanitetski pregled*, v. 70, n. 8, p. 735–739, 2013.

ARMANIOUS, K. et al. Medgan: Medical image translation using gans. *Computerized Medical Imaging and Graphics*, Elsevier, v. 79, p. 101684, 2020.

BEN-COHEN, A. et al. Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 78, p. 186–194, 2019.

BI, L. et al. Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans). In: *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*. [S.l.]: Springer, 2017. p. 43–51.

BRAIN-IXI. *IXI Dataset*. 2020. Accessed August 12, 2020, <http://brain-development.org/ixi-dataset/>.

BRANT, W. E.; HELMS, C. A. Fundamentos de radiologia e diagnóstico por imagem. *Editora Guanabara*, v. 1, 2008.

BRISLIN, R. W. Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, Sage Publications Sage CA: Thousand Oaks, CA, v. 1, n. 3, p. 185–216, 1970.

BULLITT, E. et al. Vessel tortuosity and brain tumor malignancy: a blinded study1. *Academic radiology*, Elsevier, v. 12, n. 10, p. 1232–1240, 2005.

CHARTRAND, G. et al. Deep learning: A primer for radiologists. *Radiographics : a review publication of the Radiological Society of North America, Inc*, v. 37 7, p. 2113–2131, 2017.

CHARTSIAS, A. et al. Adversarial image synthesis for unpaired multi-modal cardiac data. In: SPRINGER. *International Workshop on Simulation and Synthesis in Medical Imaging*. [S.l.], 2017. p. 3–13.

CHARTSIAS, A. et al. Multimodal mr synthesis via modality-invariant latent representation. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 3, p. 803–814, 2017.

CHOI, Y. et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 8789–8797.

CIERNIK, I. F. et al. Radiation treatment planning with an integrated positron emission and computer tomography (pet/ct): a feasibility study. *International Journal of Radiation Oncology* Biology* Physics*, Elsevier, v. 57, n. 3, p. 853–863, 2003.

COHEN, J. P.; LUCK, M.; HONARI, S. Distribution matching losses can hallucinate features in medical image translation. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2018. p. 529–536.

CORDTS, M. et al. The cityscapes dataset for semantic urban scene understanding. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 3213–3223.

CRUZ-ROA, A. et al. Automatic Detection of Invasive Ductal Carcinoma in Whole Slide Images with Convolutional Neural Networks. In: *SPIE Medical Imaging*. [s.n.], 2014. v. 9041, p. 904103–904103–15. Disponível em: <http://dx.doi.org/10.1117/12.2043872>.

DAR, S. U. et al. Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE transactions on medical imaging*, IEEE, v. 38, n. 10, p. 2375–2388, 2019.

FLORKOW, M. C. et al. The impact of mri-ct registration errors on deep learning-based synthetic ct generation. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2019: Image Processing*. [S.l.], 2019. v. 10949, p. 1094938.

FUKUSHIMA, K. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. v. 36, p. 193–202, 02 1980.

GATYS, L.; ECKER, A. S.; BETHGE, M. Texture synthesis using convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2015. p. 262–270.

GATYS, L. A.; ECKER, A. S.; BETHGE, M. Image style transfer using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 2414–2423.

GOODFELLOW, I. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680.

HAACKE, E. M. et al. *Magnetic resonance imaging: physical principles and sequence design*. [S.l.]: Wiley-Liss New York:, 1999. v. 82.

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

HIASA, Y. et al. Cross-modality image synthesis from unpaired data using cyclegan. In: SPRINGER. *International Workshop on Simulation and Synthesis in Medical Imaging*. [S.l.], 2018. p. 31–41.

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. *science*, American Association for the Advancement of Science, v. 313, n. 5786, p. 504–507, 2006.

HORE, A.; ZIOU, D. Image quality metrics: Psnr vs. ssim. In: IEEE. *2010 20th international conference on pattern recognition*. [S.l.], 2010. p. 2366–2369.

IGLEHART, J. K. The new era of medical imaging-: Progress and pitfalls. *The New England journal of medicine*, v. 354, n. 26, p. 2822–2828, 2006.

INITIATIVE, A. D. N. *ADNI Dataset*. 2020. Accessed August 12, 2020, <http: //adni.loni.usc.edu/data-samples/access-data>.

IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

ISOLA, P. et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1125–1134.

JIANG, J. et al. Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2018. p. 777–785.

JIN, C.-B. et al. Deep ct to mr synthesis using paired and unpaired data. *arXiv preprint arXiv:1805.10790*, 2018.

JIN, C.-B. et al. Deep ct to mr synthesis using paired and unpaired data. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 19, n. 10, p. 2361, 2019.

JOG, A. et al. Random forest regression for magnetic resonance image synthesis. *Medical image analysis*, Elsevier, v. 35, p. 475–488, 2017.

JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 694–711.

KLEIN, S. et al. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, IEEE, v. 29, n. 1, p. 196–205, 2010.

KRASKOV, A.; STÖGBAUER, H.; GRASSBERGER, P. Estimating mutual information. *Physical review E*, APS, v. 69, n. 6, p. 066138, 2004.

KRUPA, K.; BEKIESIŃSKA-FIGATOWSKA, M. Artifacts in magnetic resonance imaging. *Polish journal of radiology*, Termedia Publishing, v. 80, p. 93, 2015.

LAM, D. L. et al. Communicating potential radiation-induced cancer risks from medical imaging directly to patients. *American Journal of Roentgenology*, Am Roentgen Ray Soc, v. 205, n. 5, p. 962–970, 2015.

LARSEN, A. B. L. et al. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015.

LECUN, Y. et al. Gradient-based learning applied to document recognition. v. 86, p. 2278 – 2324, 12 1998.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional Networks and Applications in Vision. p. 253–256, 05 2010.

LEE, D.; MOON, W.-J.; YE, J. C. Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks. *Nature Machine Intelligence*, Nature Publishing Group, v. 2, n. 1, p. 34–42, 2020.

LEI, Y. et al. Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks. *Medical physics*, Wiley Online Library, v. 46, n. 8, p. 3565–3581, 2019.

LI, C.; WAND, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2016. p. 702–716.

LIMA, L. L.; JUNIOR, J. R. F.; OLIVEIRA, M. C. Toward classifying small lung nodules with hyperparameter optimization of convolutional neural networks. *Computational Intelligence*, Wiley Online Library, 2020.

LIU, F. Susan: segment unannotated image structure using adversarial network. *Magnetic resonance in medicine*, Wiley Online Library, v. 81, n. 5, p. 3330–3345, 2019.

LIU, M.-Y.; BREUEL, T.; KAUTZ, J. Unsupervised image-to-image translation networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2017. p. 700–708.

MCDONALD, W. I. et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, Wiley Online Library, v. 50, n. 1, p. 121–127, 2001.

MENDRIK, A. M. et al. Mrbrains challenge: online evaluation framework for brain image segmentation in 3t mri scans. *Computational intelligence and neuroscience*, Hindawi, v. 2015, 2015.

Menze, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, v. 34, n. 10, p. 1993–2024, 2015.

MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

MOURÃO, A. P.; OLIVEIRA, F. A. de. *Fundamentos de radiologia e imagem*. [S.l.]: Difusão Editora, 2018.

NIE, D. et al. Medical image synthesis with context-aware generative adversarial networks. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2017. p. 417–425.

NIE, D. et al. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, IEEE, v. 65, n. 12, p. 2720–2730, 2018.

OLUT, S. et al. Generative adversarial training for mra image synthesis using multi-contrast mri. In: SPRINGER. *International Workshop on PRedictive Intelligence In MEdicine*. [S.l.], 2018. p. 147–154.

PATHAK, D. et al. Context encoders: Feature learning by inpainting. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 2536–2544.

PEIXOTO; CáMARA-CHáVEZ; MENOTTI. *Brazilian License Plate Character Recognition using Deep Learning*. 2015.

RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

RIREDATASET. *Retrospective Image Registration Evaluation Project*. 2020. Accessed August 12, 2020, <https://www.insight-journal.org/rire/>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.

SAILER, A. M. et al. Cost-effectiveness of cta, mra and dsa in patients with non-traumatic subarachnoid haemorrhage. *Insights into imaging*, Springer, v. 4, n. 4, p. 499–507, 2013.

Mean absolute error. In: SAMMUT, C.; WEBB, G. I. (Ed.). Boston, MA: Springer US, 2010. p. 652–652. ISBN 978-0-387-30164-8. Disponível em: <https://doi.org/10.1007/978-0-387-30164-8_525>.

SILVA, G. L. F. da; PAIVA, A. C. de; SILVA, A. C. Lung Nodules Diagnosis Based on Evolutionary Convolutional Neural Network. *Multimedia Tools and Applications*, v. 76, n. 18, p. 19039–19055, 2017. ISSN 1573-7721. Disponível em: <https://doi.org/10.1007/s11042-017-4480-9>.

TANAKA, H. et al. Usefulness of ct-mri fusion in radiotherapy planning for localized prostate cancer. *Journal of radiation research*, Journal of Radiation Research Editorial Committee, p. 1109280230–1109280230, 2011.

TU, Z. Auto-context and its application to high-level vision tasks. In: IEEE. *2008 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.], 2008. p. 1–8.

WANG, L. et al. Benchmark on automatic six-month-old infant brain segmentation algorithms: The iseg-2017 challenge. *IEEE transactions on medical imaging*, IEEE, v. 38, n. 9, p. 2219–2230, 2019.

WANG, X.; GUPTA, A. Generative image modeling using style and structure adversarial networks. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2016. p. 318–335.

WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

WEI, W. et al. Learning myelin content in multiple sclerosis from multimodal mri through adversarial training. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2018. p. 514–522.

WOLTERINK, J. M. et al. Deep mr to ct synthesis using unpaired data. In: SPRINGER. *International workshop on simulation and synthesis in medical imaging*. [S.l.], 2017. p. 14–23.

YANG, Q. et al. Mri image-to-image translation for cross-modality image registration and segmentation. *arXiv preprint arXiv:1801.06940*, 2018.

YANG, Q. et al. Mri cross-modality image-to-image translation. *Scientific Reports*, Nature Publishing Group, v. 10, n. 1, p. 1–18, 2020.

YI, X.; WALIA, E.; BABYN, P. Generative adversarial network in medical imaging: A review. *Medical image analysis*, Elsevier, v. 58, p. 101552, 2019.

YI, Z. et al. Dualgan: Unsupervised dual learning for image-to-image translation. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2849–2857.

YU, B. et al. 3d cgan based cross-modality mr image synthesis for brain tumor segmentation. In: IEEE. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. [S.l.], 2018. p. 626–630.

ZHANG, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 586–595.

ZHAO, M. et al. Craniomaxillofacial bony structures segmentation from mri with deep-supervision adversarial learning. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2018. p. 720–727.

ZHOU, T. et al. Learning dense correspondence via 3d-guided cycle consistency. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 117–126.

ZHU, J.-Y. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2223–2232.

A Descrição do Projeto

Neste trabalho, construímos uma *Generative Adversarial Network* chamada MMIGAN, uma abordagem para tradução entre múltiplos domínios de imagem, capaz de traduzir imagens intermodais (TC e RM) e intramodais(PD, T1 e T2). Como a proposta foi desenvolver uma pesquisa reprodutível, o código fonte foi disponibilizado no github: https://github.com/eduardofelipe95/Multi-Medical-Imaging-Translation-using-Generative-Adversarial-Network.

O projeto da MMI-GAN, consiste primeiramente do pré-processamento da base de imagem, eliminação de dados ruidosos, alinhamento e pareamento das imagens. Após o pré-processamento, é realizada a etapa de treinamento dos dois componentes principais da rede: Generator e o Discriminator. o Gerador tem como objetivo traduzir imagem de um dos domínios (TC, MR, PD, T1 e T2) trabalhados neste trabalho, mas que pode ser expandido para outros domínios. o Discriminador tenta prever se a imagem dada como entrada é uma imagem real ou se é um imagem traduzida pelo gerador, além disso avalia o domínio da imagem de entrada, assim como o gerador pode ser expandido para outros domínios além dos utilizados neste trabalho.