

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

Eduardo Felipe de Souza, Prof. Dr. Marcelo Costa Oliveira

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
CEP – 57072-970 – Maceió – AL – Brasil

eduardo@ic.ufal.br, oliveiramc@ic.ufal.br

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. To solve these limitations, we developed MMI-GAN, a new approach for translation between multiple image domains, capable of translating intermodal (CT and MR) and intramodal (PD, T1 and T2) images using only one generator and one discriminator. The images translated by MMI-GAN managed to obtain MAE of 5.79, PSNR of 27.39, MI of 1.43 and SSIM of 0.90. Their results were often statically comparable or superior to the state of the art.*

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. 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 gerador e um discriminador. As imagens traduzidas pela MMI-GAN conseguiram obter MAE de 5.79, PSNR de 27.39, MI de 1.43 e SSIM de 0.90. Os seus resultados foram por muitas vezes estaticamente equiparáveis ou superiores ao estado da arte.*

1. Introdução

Diversos métodos de imagens médicas são utilizados para capturar informações sobre órgãos e tecidos (ressonância magnética (RM), tomografia computadorizada (TC), etc.). Os princípios físicos usados na obtenção das diferentes modalidades de imagem possuem características distintas que resultam em exames de imagens anatomicamente específicos, com diferentes dimensões (2D e 3D), resoluções de contraste e resoluções espaciais [Armanious et al. 2020].

Os aspectos inerentes de cada modalidade permitem ao especialista definir o diagnóstico e o tratamento do paciente a partir de informações contidas em duas ou mais modalidades de imagem. 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. Consequentemente, utilizar as duas técnicas fornecem ao especialista uma melhor discriminação dos tecidos e auxiliam na avaliação dos limites da infiltração tumoral [Tanaka et al. 2011].

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, a RM não é adequada para pacientes com claustrofobia, portadores de marca-passos, próteses ortopédicas, etc; enquanto na TC, o especialista deve solicitar os exames respeitando o risco associado à exposição da radiação ionizante [Krupa and Bekiesińska-Figatowska 2015]. Nestes contextos é clinicamente desejável gerar imagens ausentes (e.g., TC) a partir de imagens existentes (e.g., RM), técnica essa conhecida como tradução de imagens.

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

Diversos modelos de *Generative Adversarial Networks* (GANs) [Yi et al. 2017, Liu 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 atualmente [Yi et al. 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.

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, RM-T1, RM-T2, densidade de próton (PD) e T2-FLAIR, seriam necessários 20 geradores diferentes, pois, quanto mais domínios envolvidos mais geradores são necessários. Outra limitação das redes Pix2pix e CycleGAN é o fato dos geradores não utilizarem 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].

Neste contexto, o objetivo principal deste trabalho foi desenvolver um modelo de tradução entre múltiplos domínios de imagens médicas utilizando apenas um gerador, que usa apenas a imagem no domínio de entrada e a indicação via *target image-label* para qual domínio de interesse a imagem deve ser traduzida, chamamos esse modelo de MMI-GAN (*Multi Medical Imaging Generative Adversarial Network*). Como objetivos secundários avaliamos como o pareamento intra/inter modalidade e o processo de registro afeta a tradução de imagens médicas e se a abordagem multidomínio é capaz de obter resultados tão bons quanto o estado da arte.

2. Multi Medical Imaging Generative Adversarial Network (MMI-GAN)

A principal contribuição deste trabalho foi o desenvolvimento da MMI-GAN, uma nova abordagem de tradução entre múltiplos domínios de imagens médicas. Em nossos experimentos a MMI-GAN foi capaz de traduzir imagens intermodais (TC e RM) e intramodais (PD, T1 e T2) usando apenas um gerador e um discriminador, treinados efetivamente a partir de imagens médicas de todos os domínios disponíveis. Além disso, a MMI-GAN foi desenvolvida de maneira a ser facilmente estendida a outras tarefas de tradução para beneficiar a comunidade de imagens médicas.

A Fig.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 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** e o **Discriminator**. O **Generator** (Fig. 1-B) recebe uma imagem da **training data** em conjunto com a **target image-label**, indicando o destino do domínio da tradução. O **Generator** então cria uma nova imagem que chamaremos de imagem traduzida. O **Discriminator** (Fig. 1-C) recebe um par de imagens, podendo ser um **real pair**, contendo a imagem de origem e sua correspondente no domínio indicado pela **target image-label** (Fig. 1-D), ou um par contendo a imagem de origem e a imagem traduzida para o domínio de destino (Fig. 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.

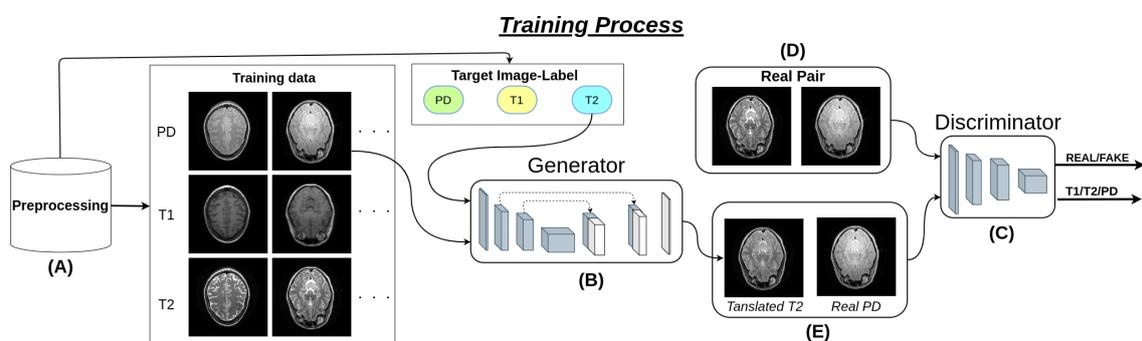


Figura 1. Visão geral do pipeline da MMI-GAN.

3. Resultados e Discussões

Os experimentos foram realizados em duas bases de dados diferentes com o objetivo de avaliar a MMI-GAN e compará-la ao estado da arte (Pix2pix e CycleGAN) em diferentes situações (Fig. 2): na tradução intermodal entre RM-TC (Tabela 1) e intramodal entre T1, T2 e PD (Tabela 2), usando diferentes regiões anatômicas (TÓRAX e CABEÇA). As métricas de avaliação utilizadas foram *Mean Absolute Error* (MAE), *Peak Signal-to-Noise Ratio* (PSNR), *Mutual information* (MI) e *Structural Similarity Index Measure* (SSIM). Por fim, analisamos as características mais relevantes obtidas no processo de tradução entre as GANs avaliadas neste trabalho.

3.1. Resultados do Banco de Dados TÓRAX

A CycleGAN foi a técnica que obteve os menores valores quantitativos em todas as métricas avaliadas. Mesmo com a capacidade de ser treinada com dados ininterruptos, os seus re-

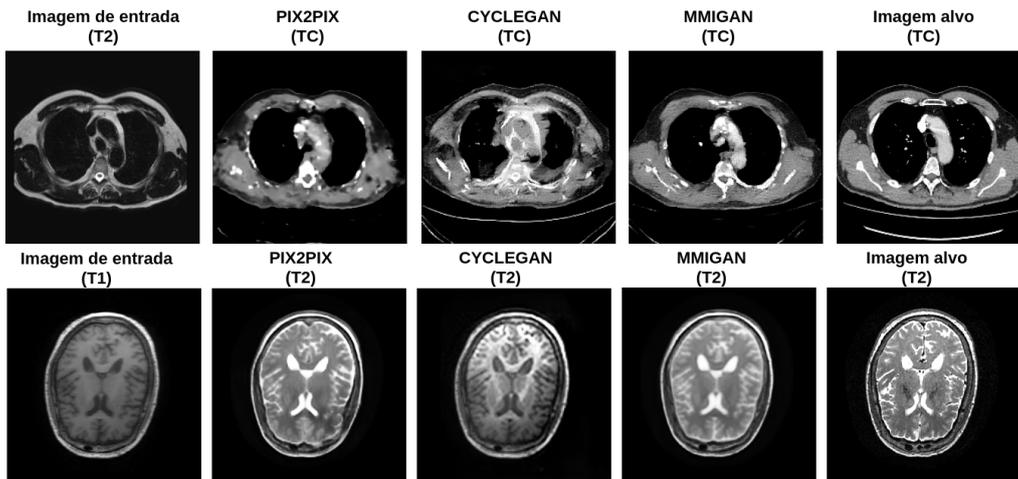


Figura 2. Da esquerda para a direita por coluna temos respectivamente: as imagens de entrada, as imagens traduzidas pela Pix2pix, CycleGAN e pela MMI-GAN, por ultimo as imagens alvo.

Tabela 1. Comparação da média e desvios padrão dos resultados que obtiveram resultados estatisticamente relevantes no banco de dados do TÓRAX para as métricas MAE, PSNR, MI e SSIM.

Translation	MAE		PSNR		MI		SSIM	
	MMI-GAN	Pix2pix	MMI-GAN	Pix2pix	MMI-GAN	Pix2pix	MMI-GAN	Pix2pix
T1-T2	24.41±4.50	23.06±4.0	16.25±1.06	16.50±1.03	0.98±0.06	0.91±0.06	0.50±0.07	0.52±0.06
T1-CT	26.49±4.41	27.22±4.05	13.45±0.65	13.27±0.56	0.74±0.04	0.69±0.05	0.48±0.06	0.46±0.06
T2-T1	24.37±8.36	26.24±4.06	16.83±2.83	17.00±1.26	0.84±0.08	0.74±0.07	0.28±0.04	0.17±0.03
T2-CT	26.86±4.20	28.05±4.03	13.40±0.71	13.14±0.57	0.73±0.06	0.69±0.05	0.47±0.06	0.45±0.05
CT-T1	23.31±6.19	23.88±3.73	17.16±1.96	17.71±1.21	0.77±0.06	0.70±0.06	0.24±0.03	0.16±0.02
CT-T2	26.75±5.52	26.05±5.81	15.26±1.47	15.40±1.70	0.89±0.08	0.80±0.06	0.47±0.08	0.47±0.07

sultados inferiores a Pix2pix e a MMI-GAN podem ser atribuídos ao fato de o treinamento ter sido realizado sem par, tendo limitado seu poder de tradução.

Acreditamos que a superioridade dos resultados da MMI-GAN sobre o CycleGAN foi devido ao aumento do número de imagens propiciado pelo ambiente de treinamento de múltiplos domínios da MMI-GAN. O banco de dados do TÓRAX contém um número de amostras relativamente pequeno, 2973 imagens por domínio. Quando treinado em dois domínios, a CycleGAN pode usar apenas $2 * 2973$ imagens de treinamento por vez, mas a MMI-GAN pode usar todas as imagens de todos os domínios, neste caso $3 * 2973$. Isso permitiu que a MMI-GAN aprendesse adequadamente mantendo a qualidade e a nitidez

Tabela 2. Comparação da média e dos desvios padrão dos resultados que obtiveram resultados estatisticamente relevantes no banco de dados de CABEÇA para as métricas MAE, PSNR, MI e SSIM.

Translation	MAE		PSNR		MI		SSIM	
	MMI-GAN	Pix2pix	MMI-GAN	Pix2pix	MMI-GAN	Pix2pix	MMI-GAN	Pix2pix
T1-T2	10.54±4.76	10.72±5.01	21.09±2.31	20.79±2.47	1.17±0.24	1.19±0.23	0.74±0.07	0.74±0.09
T1-PD	9.01±3.82	9.11±5.04	22.51±2.43	22.39±2.62	1.27±0.27	1.30±0.26	0.78±0.07	0.78±0.10
T2-T1	6.71±2.79	6.47±5.92	24.78±2.31	25.61±3.15	1.16±0.29	1.26±0.28	0.83±0.08	0.85±0.10
T2-PD	5.79±3.30	5.16±2.85	27.39±2.84	28.54±3.22	1.43±0.29	1.50±0.27	0.90±0.06	0.91±0.08
PD-T1	6.307±2.73	5.80±2.79	25.10±2.34	25.69±2.86	1.16±0.30	1.23±0.28	0.83±0.08	0.85±0.08
PD-T2	6.73±4.77	5.51±2.88	25.57±2.89	26.94±2.93	1.31±0.27	1.36±0.25	0.86±0.06	0.88±0.03

da imagem traduzida.

Tanto a Pix2pix quanto a MMI-GAN obtiveram resultados com diferença estatística significativa em todas as métricas (Tabela 2), ora a Pix2pix superior, ora a MMI-GAN superior, sendo equivalentes em algumas traduções. Entretanto, a MMI-GAN se destaca pelo importante diferencial de usar apenas uma MMI-GAN para todas as traduções, enquanto são necessárias seis Pix2pixs, uma para cada domínio. Apesar da diferença ser de 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.

3.2. Resultados da Tradução de CABEÇA

As imagens traduzidas pela MMI-GAN conseguiram obter MAE de 5.792, PSNR de 27.398, MI de 1.430 e SSIM de 0.900 (Tabela 2). Esses resultados foram superiores aos obtidos com a base de dados do TÓRAX (Tabela 1), atribuímos isso ao fato de que na base de dados da CABEÇA as imagens foram obtidas utilizando o mesmo equipamento de RM 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. Ainda, a RM do pulmão também é um exame que apresenta muito ruído devido à presença de ar nos pulmões.

O efeito da complexidade do registro de imagens na base de dados de TÓRAX e CABEÇA pode ser melhor visto quando observamos os mapas de erros nas traduções das duas bases de dados, esses mapas mostram os locais onde mais erros ocorreram durante a tradução (Fig. 3). 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, consequentemente causando desalinhamento em as imagens traduzidas.

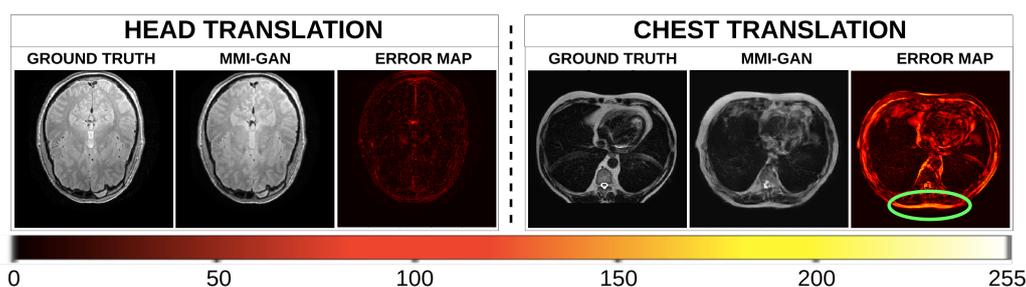


Figura 3. Exemplo da distribuição de erro nas traduções de TÓRAX e CABEÇA.

Assim como na avaliação do banco de dados do TÓRAX, a CycleGAN foi o que obteve os menores resultados em todas as métricas, o que reforça que a perda utilizada para o treinamento da CycleGAN limitou as traduções.

Assim como no TÓRAX, acreditamos que a superioridade da MMI-GAN em testes estatísticos sobre os resultados quantitativos deve-se ao efeito implícito causado pelo aumento de dados em um ambiente de múltiplos domínios. No entanto, tanto a Pix2pix quanto a MMI-GAN alcançaram resultados muito próximos em algumas métricas. Porém, a MMI-GAN teve um diferencial importante pelo fato de utilizar apenas uma MMI-GAN

para todas as traduções, enquanto a Pix2pix utilizou seis Pix2pixs, um para cada par de domínios envolvidos.

4. 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), cujos resultados se mostraram, por muitas vezes estatisticamente equiparáveis ou superiores a Pix2pix e em quase todas as traduções foi superior a CycleGAN.

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 quantitativos inferiores a Pix2pix e a MMI-GAN, essa diferença ficou comprovada com o teste estatístico 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.

Referências

- Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., and Yang, B. (2020). Medgan: Medical image translation using gans. *Computerized Medical Imaging and Graphics*, 79:101684.
- Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.
- Krupa, K. and Bekiesińska-Figatowska, M. (2015). Artifacts in magnetic resonance imaging. *Polish journal of radiology*, 80:93.
- Liu, F. (2019). Susan: segment unannotated image structure using adversarial network. *Magnetic resonance in medicine*, 81(5):3330–3345.
- Tanaka, H., Hayashi, S., Ohtakara, K., Hoshi, H., and Iida, T. (2011). Usefulness of ct-mri fusion in radiotherapy planning for localized prostate cancer. *Journal of radiation research*, pages 1109280230–1109280230.
- Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552.
- Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.