깊은 생성적 적대 신경망을 이용한 교차 모달리티 의료 이미지 생성

두이풍다오 [1], 양형정 [1], 정혜원 [2], 로저 데이비 [2]
[1] 전남대학교 인공지능융합학과
[2]AdelaideMRI, 호주

phuongdd.1997@gmail.com, hjyang@jnu.ac.kr, hwjung92@gmail.com, daviesroger55@gmail.com

Cross-modality Medical Image Generation using Deep Generative Adversarial Network

Duy-Phuong Dao[1], Hyung-Jeong Yang[1], Hye-Won Jung[2], and Roger Davies[2]
[1]Dept. of Artificial Intelligence Convergence, Chonnam National University, South Korea
[2]AdelaideMRI, Australia
Corresponding author: Hyung-Jeong Yang (hjyang@jnu.ac.kr)

요 약

Due to the advancement of deep learning techniques, the medical field is undergoing significant upheaval. One of the prominent applications is generating an imaging modality from another imaging modality. This application helps reduce the cost of taking multiple types of medical images for diagnostic imaging. Although many methods have been proposed for generating medical images, only a few studies focus on three-dimensional (3D) images. Therefore, in this paper, we propose a deep generative adversarial network (GAN) for generating a 3D target image from a 3D source image. The results have shown that our proposed approach can generate high-quality images and holds promise for practical use.

1. INTRODUCTION

Generative models [1,2] are now highly effective instruments in many fields, including medicine, where they can greatly improve therapy and diagnosis capacities. Cross-modality generation [3,4], in which a model is trained to synthesis one form of medical image from another, is one of the main uses of generative models in medical imaging. This methodology tackles the difficulty of acquiring all-encompassing multimodal data, including PET (positron emission tomography), CT (computerized tomography), and MRI (magnetic resonance imaging), which sometimes necessitate distinct and expensive imaging procedures. Through the use of generative models, multi-source data integration can be used to produce missing modalities, minimize the number of scans that patients must undergo, and provide more precise clinical insights.

A generative model is trained to map images from a source modality to a target modality in the context of cross-modality generation. For instance, generating PET images from MRI inputs allows for the high-resolution anatomical features that MRI gives, while also facilitating the non-invasive extraction of functional information from PET. Across various medical modalities, the application of deep learning models specifically, generative adversarial networks (GANs) [5,6] and denoising diffusion probabilistic model (DDPM) [7] has produced impressive results in the synthesis of realistic and diagnostically valuable images. This feature can facilitate clinical decision-making, improve illness identification, and improve image interpretation.

However, many studies [3,5, 6] have focused on generating two-dimensional (2D) medical images, which overlook the rich three dimensional (3D) structural information crucial for accurate diagnosis in modalities like MRI and CT. This restriction to 2D slices can lead to inconsistencies when reconstructing the full 3D anatomy, limiting the model's effectiveness in capturing spatial relationships across different planes. Another significant limitation in previous work [7] is the long inference time required for generating images. In medical settings, rapid decision-making is often crucial, and prolonged inference times can hinder timely diagnosis and treatment.

In this work, we propose a generative adversarial network that utilizes 3D images for cross-modality generation, addressing the limitations of previous approaches that rely on 2D images. By working directly with 3D volumetric data, our model is able to capture the full spatial relationships and anatomical structures, leading to more accurate and consistent results across slices.

The organization of the subsequent sections of this paper is as follows: Section II describes the proposed model. Section III discusses the dataset and experimental results, including a comparison with existing methods. Lastly, Section IV summarizes our findings and conclusions.

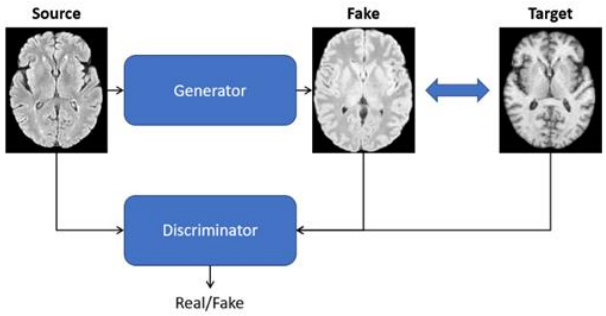

Figure 1. The overall architecture of the proposed method.

2. METHOD

In this section, we describe our model architecture in detail. The proposed method consists of two main networks: generator and discriminator networks, as illustrated in Figure 1. The generator network is used to generate fake images to fool the discriminator. While the discriminator aims to distinguish between real and generated images.

The architecture of generator network is based on UNet [8] model and residual block, as shown in Figure 2. The generator network consists of three main modules: encoder, decoder, and skip-connection modules. The encoder is responsible for capturing patterns in the source image at various levels of abstraction, starting with low-level features (e.g., edges) and progressing to higher-level, more abstract features (e.g., hippocampus region). The decoder progressively restores the spatial dimensions and transform features of source domain to those of the target domain by performing upsampling and directly connecting corresponding layers from the encoder to the decoder through the skip connections.

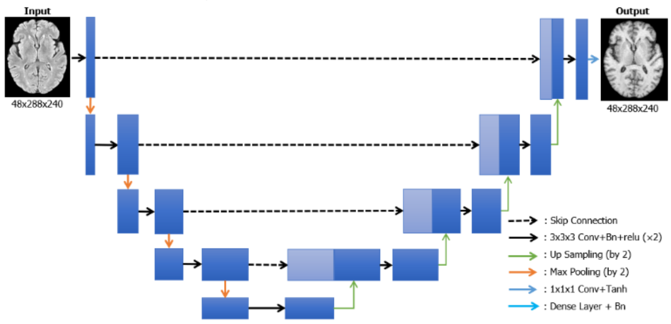

Figure 2. The architecture of the generator network.

To enhance the quality of generated images, the discriminator network takes a pair of images: one is the source image, and the other is either the real target image or the image generated by the generator, as shown in Figure 3. The concept of the discriminator network is based on PatchGAN [9], which determines whether each patch of the input looks realistic, meaning it enforces that the local features in the image (e.g., textures, patterns) are consistent with those found in real images. If the pair includes the generated image, it is labeled as fake (0), and if it includes the real target image, it is labeled as real (1).

To optimize the model's parameters, we use the loss function as follows:

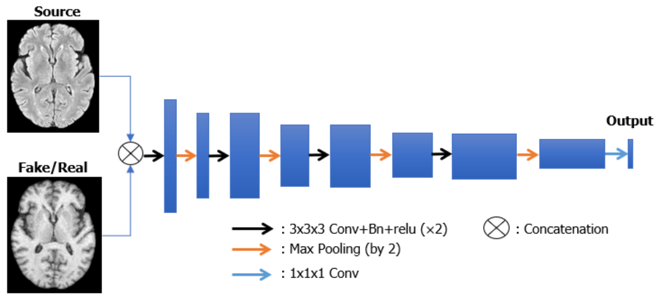

Figure 3. The architecture of the discriminator network.

$$L(G,D) = \mathbb{E}[logD(x,y)] + \mathbb{E}\left[\log\left(1 - D\left(x, G(x)\right)\right)\right] + \mathbb{E}[|y - G(x)|] \quad (1)$$

Where x and y are source and target images, respectively; G and D are generator and discriminator losses, respectively.

3. EXPERIMENT RESULTS

3.1 Dataset and Settings

In this paper, we trained and evaluated on 939 and 100 pairs of FLAIR and T1 MRI that collected from AdelaideMRI[1]. We use FLAIR modality as the source image and T1 modality as the target image. The size of images is 48×288×240. We use Adam optimizer [10] with a learning rate of 0.0001 for optimizing the model's parameters. We set the number of epochs as 100. We employed four common metrics to evaluate the model performance such as structural similarity (SSIM), fréchet inception distance (FID), mean absolute error (MAE), and root mean square error (RMSE).

3.2 Experiment results

We compared our proposed method with the two existing methods (DDPM and Pix2Pix). As can be seen in Table 1, the proposed method achieved the best performance across all evaluation metrics. We also provide several examples of the generated images, as shown in Figure 4. The generated images from our model are of good quality and are quite similar to the real images, demonstrating the effectiveness of our model.

Table 1. Performance comparison

	SSIM↑	FID↓	MAE↓	RMSE↓
DDPM[7]	0.8679	47.51	88.77	181.81
Pix2Pix[3]	0.9359	**15.68**	33.13	76.83
Ours	**0.9413**	16.11	**32.59**	**75.56**

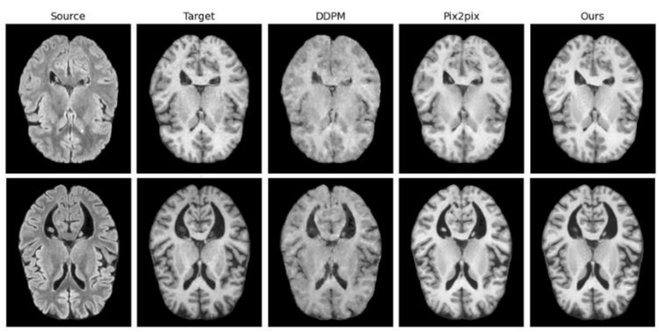

Figure 4. Several examples of generated images.

[1]https://adelaidemri.com/

4. Conclusion

In this work, we proposed a deep generative adversarial network for generating 3D images from a source modality to a target modality. By leveraging 3D volumetric data, our model effectively captures spatial relationships and anatomical structures, resulting in more accurate and consistent cross-modality image synthesis. The experiment results demonstrated that our proposed method provides a good solution for generating missing or complementary medical imaging modalities. Future work will focus on further improving the model's performance by applying advanced deep learning techniques such as attention mechanisms.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) grant funded by the Korea government(MSIT).

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (RS-2023-00208397).

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-RS-2024-00437718) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[2] Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., & Krishnan, D. (2017). Unsupervised pixel-level domain adaptation with generative adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3722-3731).

[3] Aljohani, A., & Alharbe, N. (2022). Generating synthetic images for healthcare with novel deep pix2pix gan. Electronics, 11(21), 3470.

[4] Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E. I. C., & Xu, Y. (2020). MRI cross-modality image-to-image translation. Scientific reports, 10(1), 3753.

[5] Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., & Jin, M. (2022). Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1815-1824).

[6] Yan, S., Wang, C., Chen, W., & Lyu, J. (2022). Swin transformer-based GAN for multi-modal medical image translation. Frontiers in Oncology, 12, 942511.

[7] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

[8] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

[9] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

[10] Kingma, D. P. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.