Hindawi
Complexity
Volume 2021, Article ID 5624909, 7 pages
https://doi.org/10.1155/2021/5624909

Research Article

Cross-Model Transformer Method for Medical Image Synthesis

Zebin Hu,[1] **Hao Liu**,[1,2] **Zhendong Li**,[1,2] **and Zekuan Yu**[3]

[1]*School of Information Engineering, Ningxia University, Yinchuan 750021, China*
[2]*Collaborative Innovation Center for Ningxia Big Data and Artificial Intelligence Co-founded by Ningxia Municipality and Ministry of Education, Yinchuan 750021, China*
[3]*Academy for Engineering and Technology, Fudan University, Shanghai 200433, China*

Correspondence should be addressed to Hao Liu; liuhao@nxu.edu.cn

Received 14 August 2021; Revised 25 September 2021; Accepted 7 October 2021; Published 25 October 2021

Academic Editor: Long Wang

Acquiring complementary information about tissue morphology from multimodal medical images is beneficial to clinical disease diagnosis, but it cannot be widely used due to the cost of scans. In such cases, medical image synthesis has become a popular area. Recently, generative adversarial network (GAN) models are applied to many medical image synthesis tasks and show prior performance, since they enable to capture structural details clearly. However, GAN still builds the main framework based on convolutional neural network (CNN) that exhibits a strong locality bias and spatial invariance through the use of shared weights across all positions. Therefore, the long-range dependencies have been destroyed in this processing. To address this issue, we introduce a double-scale deep learning method for cross-modal medical image synthesis. More specifically, the proposed method captures locality feature via local discriminator based on CNN and utilizes long-range dependencies to learning global feature through global discriminator based on transformer architecture. To evaluate the effectiveness of double-scale GAN, we conduct folds of experiments on the standard benchmark IXI dataset and experimental results demonstrate the effectiveness of our method.

1. Introduction

Magnetic resonance imaging (MRI) is a versatile and non-invasive imaging technique widely used in clinical applications. Tailored MRI pulse sequences enable to capture specific characteristics of the underlying anatomical information. For instance, T1-weighted brain images clearly depict the gray matter and white matter tissue, while T2-weighted images depict the fluid in the cortical tissue. Hence, acquiring complementary information about tissue morphology from multimodal images enables to improve accuracy and confidence in clinical diagnosis [1]. Unfortunately, acquiring multimodal MR imaging is often challenging due to numerous factors, such as uncooperative patients, limited availability of scanning time, and the expensive cost of prolonged exams [2, 3]. To address this issue, cross-modal medical image synthesis has been widely used, as it enables to synthesis unattained images in multimodal protocols from the subset of available images [4–7].

Currently, deep learning-based synthesis demonstrates more promising performance, which is compared with the traditional registration-based method [8, 9] and intensity-transformation-based methods [10, 11]. For the image synthesis task, convolutional neural network (CNN) architectures produce significant performance through minimizing pixelwise losses between synthetic and real images. However, pixelwise losses ignore high-level features in the training step. Since generative adversarial networks (GAN) were introduced by Goodfellow et al. [12], this problem was gradually solved by adversarial loss functions, which designed a training strategy between generator and discriminator networks based on the game theory. In this case, GAN enables to capture high-frequency texture information of medical images. Therefore, GAN-based methods surpass

many synthesis tasks based on traditional architectures [13, 14]. To be specific, the generator and discriminator networks of GAN deploy compact convolution filters, whereas CNNs are plugged with spatial locality on the entire images by the sliding window. This makes the long-range dependencies between distant regions lost [15].

Moreover, CNNs not only exhibit a strong locality bias but also a bias towards spatial invariance through the use of shared weights across all positions [16]. This prevents the networks from fully understanding the local region of the input image. To guide networks towards critical image regions, Zhao et al. [17] proposed the attention mechanisms that strengthen the features of important regions by learning the weight map and multiplying it on the feature map. However, conventional attention mechanisms still do not explicitly model long-range dependencies. Recently, transformer architectures have been applied to language tasks and are increasingly adopted in other areas such as segmentation tasks [18] and classification tasks [19]. In contrast to the predominant vision architecture, the emergent transformer architectures are integrated to learn complex relationships among its inputs, since it contains no built-in inductive prior on the locality of interactions such as sliding window. Hence, we consolidate transformer into our model due to capture more global information and make a comprehensive understanding of the input [16].

In this paper, we propose a double-scale deep learning method for cross-modal medical image synthesis. Motivated by the fact that low-level image structure and high-level feature is equally important to cross-modal medical image synthesis we integrate the ability of transformer to efficiently seek long-range interactions inside our model, which enables to capture global feature as complementary information for CNNs. To achieve this, we carefully design double-scale discriminator GAN which specifically consists of the transformer-based global discriminator and CNN-based local discriminator.

The main contributions of this paper are listed as follows. (1) We introduce a double-scale discriminator GAN for medical image synthesis. (2) The global discriminator of our model is designed on vision transformer that utilizes long-range dependencies between distant patches and captures global features.

2. Related Works

2.1. Medical Image Synthesis. Recently, GAN-based models have been successfully applied to kinds of tasks including data augmentation [20–22] and image synthesis tasks [23–25]. For example, Nie et al. [5] utilized MR images to synthesize computed tomography (CT) images with a context-aware GAN model; Wolterink et al. [7] utilized GAN to generate low-dose CT from routine-dose CT images. Nevertheless, as the traditional GAN has failed to meet the gradually higher application requirements, pix2pix [26] has recently begun to attract the attention of researchers, which utilizes paired data to enhance the pixel-to-pixel similarity between the real and the synthesized images, and then, Olut et al. [27] developed a CycleGAN-based method to synthesis MRA from T1-MRI and T2-MRI. These methods are unable to capture the features of critical image regions. Therefore, Zhao et al. [17] used a self-attention in the generator of GAN to enhance the feature of tumour and improve the performance of tumour detection. Isola et al. [26] used a patch-based discriminator to refine the extraction of features. However, these methods cannot solve the problem that the strong prior position information introduced by the sliding window in the convolution operation, which destroys the modelling of the distant dependence relationship, so that all the local information cannot be better captured.

2.2. The Transformer Architecture. The transformer architecture is designed to handle complicated interactions between inputs regardless of their relative position to one another through modelling interactions between its inputs solely through attention mechanism. Transformer is originally applied to language tasks, Floridi and Chiriatti [28] introduced GPT to use language modelling as its pretraining task. Recently, this method also can be used in computer vision. Esser et al. [16] proposed a VQGAN which represents images as a composition of perceptually rich image constituents and thereby overcomes the infeasible quadratic complexity when modelling images directly in pixel space. However, the codebook of VQGAN requires numerous datasets to fit, which is impractical in the medical image field. Meanwhile, the increased expressivity of transformers comes with quadratically increasing computational costs, because all pairwise interactions are taken into account. Finally, our method is based on a vision transformer which crops interactions between inputs based on nonoverlapping patch-level.

3. Approach

3.1. Overview of Our Method. The overview of double-scale GAN is illustrated in Figure 1. Our method is comprised of three main components: generator network, global discriminator network, and local discriminator network. In the remainder of this section, we explain the detailed composition of each network component and the loss functions.

3.2. Generator Network. The first component of our method is a deep encoder network that contains a series of convolutional layers to capture a hierarchy of localized features of source images. To learn a meaningful and effective high-level representation, we adopt an autoencoder structure as our main framework. In order to reduce the use of upsampling layer, deconvolution operation is used instead.

The detail of generator is illustrated in Figure 1. In the downsampling process, our method uses two convolutional layers of kernel size with 3 and stride with 2. In the upsampling process, our method uses two deconvolutional layers of kernel size with 3 and stride with 2. Besides, we also introduce instance normalization after each convolutional layer. After the instance normalization, the activation function ReLU is used in the encoder and decoder. For spatial and depth feature extraction, our method also adds 9 ResNet blocks between downsampling and upsampling.

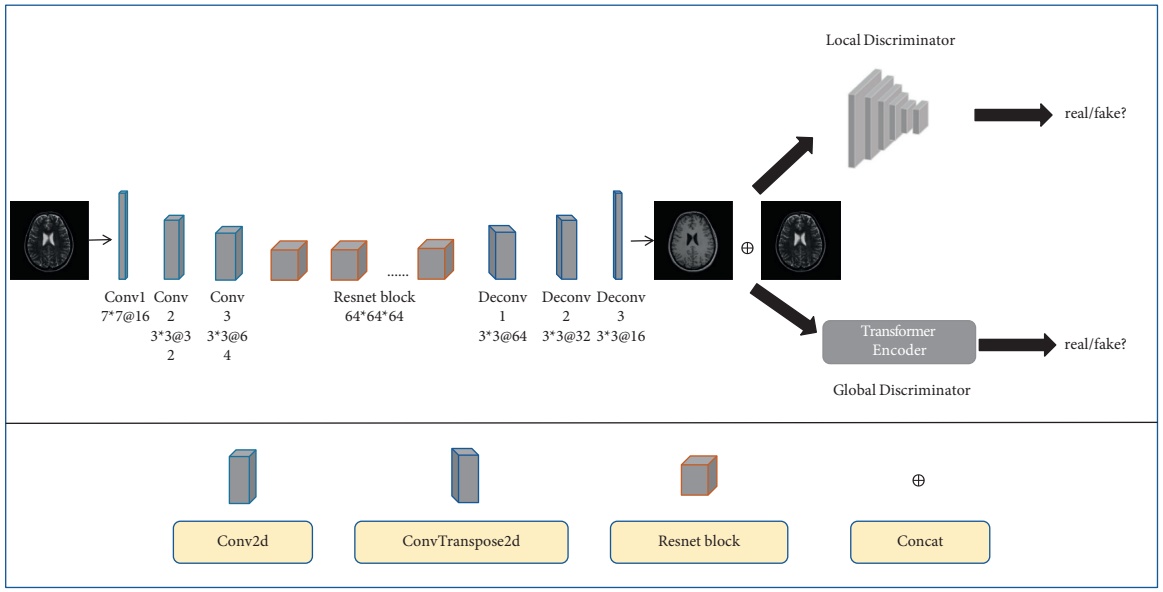

Figure 1: Schematic flow chart of the proposed algorithm for cross-modal medical image synthesis, which consists of generator, CNN-based local discriminator, and transformer-based global discriminator. The local discriminator guides the generator to learn structural representation with inductive bias. The global discriminator guides the generator to learn comprehensive features by utilizing long-range dependencies between patches of input image.

3.3. Local Discriminator Network. The local discriminator is based on a condition PatchGAN architecture [26]. It receives as input the concatenation of the source and target contrast images [29] and then obtains $30 * 30$ overlapped patches of $70 * 70$ size through sliding window for prediction to real or fake. Although this patch-based discriminant is more robust than the image-based discriminant in the extraction of local detail features, the overlapping patches it extracts destroy the long-range dependencies by introducing a strong prior position relationship, so as to have a comprehensive understanding of the input images.

3.4. Global Discriminator Network. In order to synthesize high-quality medical images, global and local features are equally important. Inspired by the DeblurGAN-v2 [30], we use a pure transformer method to replace convolutional network to capture long-range dependencies for a comprehensive understanding of the input image. The details of global discriminator network are depicted in Figure 2.

The input image is first split into $32 * 32$ nonoverlapping patches, in which kernel size is equal to stride:

$$P_1, P_2, \ldots, P_N = \text{split (input)}, \tag{1}$$

where P_i denotes the i-th patch of the input image; we set $N = 8^2$ to divide the input into 64 patches. Then, all patches are flattened to D dimension by a trainable linear projection. Similar to the class token in BERT [31], we also prepend a learnable embedding to the sequence of embedded patches. Position embeddings are added to the patch embeddings to retain positional information. Our method uses standard learnable 1D position embeddings because many studies have shown that using more advanced 2D-aware position embeddings not works [32], which can be therefore formulated as follows:

$$Z_0 = [x_{\text{class}}; P_1E; P_{2E}; \ldots; P_{\text{NE}}] + E_{\text{pos}}, \tag{2}$$

where Z_0 denotes the input of transformer encoder; E denotes embedding projection which maps patch image to vector; and E_{pos} denotes the learnable positional embedding that carries information about patch location.

The transformer encoder consists of two parts: multihead self-attention (MSA) and multilayer perceptrons (MLP). MSA enables to learn different levels of features benefit from multihead attention. In addition, layer norm (LN) is applied before every block, and residual connections after every block. At the end of these blocks, the output is taken by the classification head to output the real/fake prediction. The output of the l-th layer in the transformer encoder can be formulated as

$$Z'_l = \text{MSA}(\text{LN}(Z_{l-1})) + Z_{l-1}, \tag{3}$$

$$Z'_l = \text{MLP}(\text{LN}(Z'_l)) + Z'_l \quad , \tag{4}$$

where Z_{l-1} represents the feature extracted from the previous layer.

3.5. Loss Function. The first component of the loss function in our method is a pixelwise loss as inspired by the pix2pix architecture [26]:

$$L_1 = E_{x,y}|y - G(x)|_1, \tag{5}$$

where x denotes the source image and y denotes the target image.

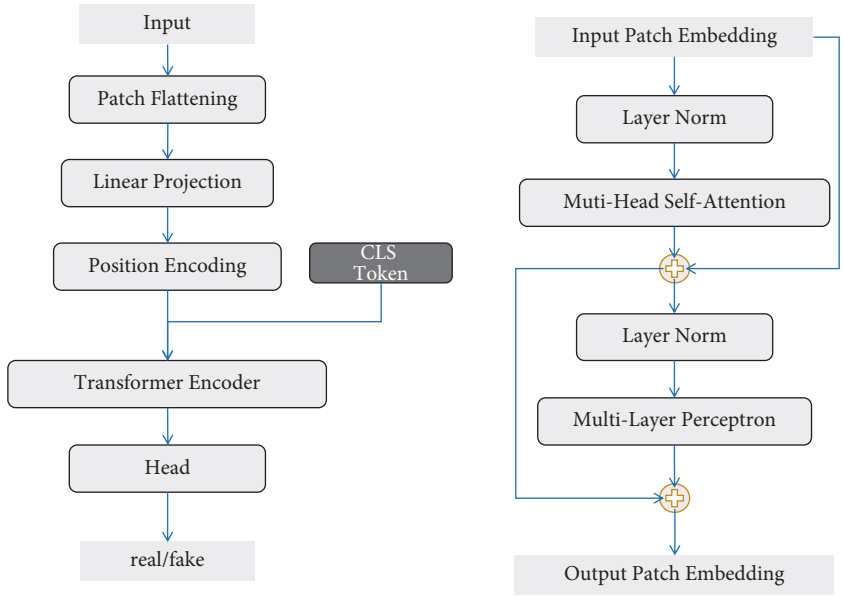

Figure 2: Detailed chart of global discriminator. The left side shows the overall computational flow of the global discriminator, and the right side shows the details of the transformer encoder on the left.

Unlike loss functions based on pixelwise differences, perceptual loss relies on differences in higher feature representations that are often extracted from networks pretrained for more generic tasks [33]. A commonly used network is VGGNet which trained on the ImageNet [34] dataset for object classification. Here, following [33], we extracted feature maps right before the second max-pooling operation of VGG16 pretrained on ImageNet:

$$L_{\text{per}} = E_{x,y}|V(y) - V(G(x))|_1, \tag{6}$$

where $V(\cdot)$ denotes pretrained VGG16.

The local discriminator is based on the conditional discriminator; its loss function can be formulated as

$$L_{\text{Local}}(G,D) = -E_{x,y}\left[(D(x,y)-1)^2\right] - E_{x,z}\left[D(x,G(x,z))^2\right], \tag{7}$$

where z denotes the synthesis image from generator.

The global discriminator uses hinge loss to optimize the generator; hinge loss can be formulated as

$$\begin{aligned} L_{\text{Gocal}}(G,D) = &-E_{x,y}[\min(0, D(x,y)-1)] \\ &- E_{x,z}[\min(0, -D(x,G(x,z)) \\ &- 1)] - \lambda_{\text{adv}} E_{x,y,z}[D(G(x,z),y)], \end{aligned} \tag{8}$$

By aggregating all the above losses, we can formulate our aggregate loss function as

$$L_{\text{aggregate}} = \lambda_{L_1} L_1 + \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{Local}} L_{\text{Local}} + \lambda_{\text{Global}} L_{\text{Global}}, \tag{9}$$

where λ_{L_1} denotes the weighing of the pixelwise loss; λ_{per} denotes the weighing of the perceptual loss; λ_{Local} denotes the weighing of the adversarial loss of local discriminator; and λ_{Global} denotes the weighing of the adversarial loss of global discriminator.

4. Experiments

In this section, we will first describe the information about the dataset used in our method and then introduce the implementation details of experiments. We present experimental results that compare with several state-of-the-art methods.

4.1. Dataset. The dataset used in the evaluation is provided by the IXI dataset. The experimental dataset we used totals 40 subjects, and each subject has corresponding T1-MRI and T2-MRI, where 30 subjects were used for training and 10 were used for testing. Acquisition parameters were as follows: T1-weighted images: TE = 4.603 ms, TR = 9.813 ms, and spatial resolution = $0.94 \times 0.94 \times 1.2\ \text{mm}^3$. T2-weighted images: TE = 100 ms, TR = 8178.34 ms, and spatial resolution = $0.94 \times 0.94 \times 1.2\ \text{mm}^3$. Since multicontrast images were unregistered, we use FSL [35] to register T1-MRI and T2-MRI. Finally, we use zero-padding to fill all images in axial cross-sections used in experiments to a consistent size of $256 * 256$.

4.2. Implementation Details. Our method is implemented in PyTorch. All methods were trained and tested on 1 NVIDIA Tesla V100 with 32 GB of memory for each GPU. In the stage of training of our method, we set the epoch as 100, learning rate as 0.0002, and batch size as 1 which causes the training time to increase to 5 hours. Model training was performed via the Adam optimizer with $\beta 1 = 0.5$ and $\beta 2 = 0.999$. In global discriminator, we use

Table 1: Comparisons of T2-weight MRI synthesis between our proposed method with different approaches of PSNR and SSIM (data in the table denote the average value and standard deviation of the test dataset).

Method	PSNR	SSIM
pix2pix	34.38 ± 0.84	0.775 ± 0.04
CycleGAN	34.75 ± 0.86	0.786 ± 0.03
PGAN (without global)	34.82 ± 0.98	0.892 ± 0.06
Ours (global and local)	34.91 ± 1.00	0.895 ± 0.07

Table 2: Comparisons of T1-weight MRI synthesis between our proposed method with different approaches of PSNR and SSIM (data in the table denote the average value and standard deviation of the test dataset).

Method	PSNR	SSIM
pix2pix	34.58 ± 0.84	0.758 ± 0.04
CycleGAN	34.73 ± 0.82	0.795 ± 0.04
PGAN (without global)	35.85 ± 1.09	0.887 ± 0.07
Ours (global and local)	35.34 ± 0.95	0.895 ± 0.07

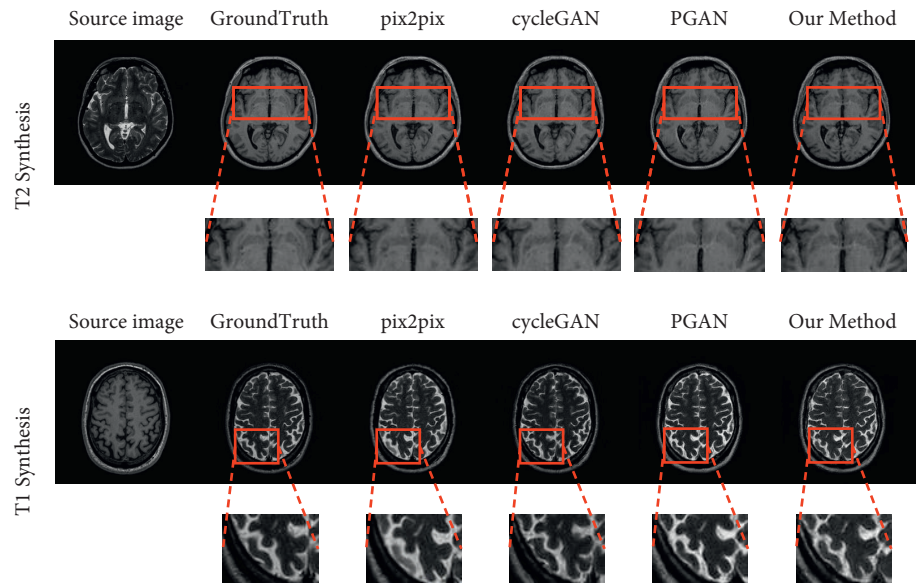

Figure 3: Synthesized images from all competing methods are shown along with the source images and the reference target image. Our method improves synthesis performance in regions that are depicted suboptimally in competing methods. Obviously, the composite images from our method have less noise and sharper tissue depiction.

multihead attention with 4 heads and set D as 64. In each multihead attention, we performed GeLu activation and set dropout as 0.1. Limited by the small size of the medical image dataset, we utilize pretrained model in global discriminator for object classification tasks on the ImageNet database. All weights were initialized using normal distribution with 0 mean and 0.02 std. We set the hyperparameter in the aggregate loss function as $\lambda_{L_1} = 1$, $\lambda_{\text{per}} = 1$, $\lambda_{\text{Local}} = 0.8$, and $\lambda_{\text{Global}} = 0.3$. For the fairness of the experiment, we designed 4-fold cross-validation by randomly sampling nonoverlapping training, validation, and testing sets in each fold.

4.3. Comparison Methods. To validate the effectiveness of the proposed synthesis method, we compare it with three state-of-the-art cross-modality synthesis methods:

(1) pix2pix [26]: this method is based on a convolutional GAN model and UNet backbone, which synthesizes the whole image by focusing on the pixelwise similarity.

(2) CycleGAN [27]: this method consists of two generators and two discriminators, which uses a cycle consistency loss to enable to train with unpaired data. In our comparison, we use the paired data to training this method and our method.

(3) PGAN [29]: this method is based on conditional GAN; its generator consists of a encoder, a decoder, and 9 ResNet blocks. Meanwhile, this method has shown superior performance in many cross-modal image synthesis tasks.

4.4. Results and Analysis. We employ two measurements to evaluate the synthesis performance of the proposed methods and our method in comparison: structural similarity index measurement (SSIM) and peak-signal-to-noise ratio (PSNR). The data in all tables are represented by the mean and standard deviation. Further details can be found in Tables 1 and 2.

To demonstrate the effectiveness of our double-scale discriminator method with regard to subjective quality, a demonstrated example is shown in Figure 3.

5. Conclusion

In this paper, we have proposed a double-scale discriminator GAN for cross-modal medical image synthesis. By compositing both CNN and transformer to design double-scale discriminator, our method has explicitly exploited the localization power of CNNs and the sensitivity of vision transformers to global context meanwhile. Experimental results have demonstrated the effectiveness of the proposed method. In the future, we will focus on the medical image generation method which integrated multiview and multimodal information through transformer, which solves the problem that 2D medical image generation cannot exploit 3D information and 3D medical image generation needs high computing power.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Science Foundation of China under Grants 61806104 and 62076142, in part by the West Light Talent Program of the Chinese Academy of Sciences under Grant XAB2018AW05, and in part by the Youth Science and Technology Talents Enrolment Projects of Ningxia under Grant TJGC2018028.

References

[1] B. J. Pichler, M. S. Judenhofer, and C. Pfannenberg, "Multimodal imaging approaches: pet/ct and pet/mri," *Molecular Imaging I*, vol. 185, pp. 109–132, 2008.

[2] K. Krupa and M. Bekiesińska-Figatowska, "Artifacts in magnetic resonance imaging," *Polish Journal of Radiology*, vol. 80, pp. 93–106, 2015.

[3] B. B. Thukral, "Problems and preferences in pediatric imaging," *Indian Journal of Radiology and Imaging*, vol. 25, no. 4, p. 359, 2015.

[4] Y. Huang, L. Shao, and A. F. Frangi, "Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 815–827, 2017.

[5] D. Nie, R. Trullo, J. Lian et al., "Medical image synthesis with context-aware generative adversarial networks," in *Proceedings of the International conference on medical image computing and computer-assisted intervention*, pp. 417–425, Springer, Quebec City, Quebec, Canada, September 2017.

[6] Y. Wang, L. Zhou, B. Yu et al., "3d auto-context-based locality adaptive multi-modality gans for pet synthesis," *IEEE Transactions on Medical Imaging*, vol. 38, no. 6, pp. 1328–1339, 2018.

[7] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[8] N. Burgos, M. J. Cardoso, K. Thielemans et al., "Attenuation correction synthesis for hybrid pet-mr scanners: application to brain studies," *IEEE Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2332–2341, 2014.

[9] J. Lee, A. Carass, A. Jog, C. Zhao, and J. L. Prince, "Multi-atlas based ct synthesis from conventional mri with patch-based refifinement for mri-based radiotherapy planning," in *Proceedings of the Medical Imaging 2017: Image Processing*, vol. 10133, International Society for Optics and Photonics, Orlando, Florida, US, February 2017.

[10] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Mr image synthesis by contrast learning on neighborhood ensembles," *Medical Image Analysis*, vol. 24, no. 1, pp. 63–76, 2015.

[11] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Medical Image Analysis*, vol. 35, pp. 475–488, 2017.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[13] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, "Mustgan: multi-stream generative adversarial networks for mr image synthesis," *Medical Image Analysis*, vol. 70, Article ID 101944, 2021.

[14] D. Lee, J. Kim, W.-J. Moon, and J. C. Ye, "Collagan: collaborative gan for missing image data imputation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2487–2496, Long Beach, CA, USA, June 2019.

[15] T. Roughgarden, "Algorithmic game theory," *Communications of the ACM*, vol. 53, no. 7, pp. 78–86, 2010.

[16] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12 873–912 883, Nashville, TN, USA, June 2021.

[17] J. Zhao, D. Li, Z. Kassam et al., "Tripartite-gan: synthesizing liver contrast-enhanced mri to improve tumor detection," *Medical Image Analysis*, vol. 63, Article ID 101667, 2020.

[18] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "Transbts: multimodal brain tumor segmentation using transformer," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 109–119, Springer, Strasbourg, France, October 2021.

[19] I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3286–3295, Seoul, Korea, October 2019,.

[20] C. Bermudez, A. J. Plassard, L. T. Davis, A. T. Newton, S. M. Resnick, and B. A. Landman, "Learning implicit brain mri manifolds with deep learning," in *Proceedings of the Medical Imaging 2018: Image Processing*, vol. 10574, International Society for Optics and Photonics, Houston, Texas, US, February 2018.

[21] Z. Xu, C. Qi, and G. Xu, "Semi-supervised attention-guided cyclegan for data augmentation on medical images," in *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 563–568, IEEE, San Diego, CA, USA, November 2019.

[22] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *Proceedings of the International conference on artificial neural networks*, pp. 626–634, Springer, Alghero, Italy, September 2017.

[23] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[24] Z. Han, B. Wei, A. Mercado, S. Leung, and S. Li, "Spine-gan: semantic segmentation of multiple spinal structures," *Medical Image Analysis*, vol. 50, pp. 23–35, 2018.

[25] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Medical Image Analysis*, vol. 49, pp. 14–26, 2018.

[26] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, Honolulu, HI, USA, July 2017.

[27] S. Olut, Y. H. Sahin, U. Demir, and G. Unal, "Generative adversarial training for MRA image synthesis using multi-contrast MRI," in *Proceedings of the International workshop on predictive intelligence in medicine*, pp. 147–154, Springer, Granada, Spain, September 2018.

[28] L. Floridi and M. Chiriatti, "GPT-3: its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, no. 4, pp. 681–694, 2020.

[29] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[30] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8878–8887, Seoul, Korea, October 2019.

[31] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the NAACLHLT*, Minneapolis, USA, June 2019.

[32] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16×16 words: transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations*, Addis Ababa, Ethiopia, April 2020.

[33] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proceedings of the European conference on computer vision*, pp. 694–711, Springer, Amsterdam, Netherlands, October 2016.

[34] O. Russakovsky, J. Deng, H. Su et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[35] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, 2001.