

RESEARCH

Open Access



MTFN: multi-temporal feature fusing network with co-attention for DCE-MRI synthesis

Wei Li¹, Jiaye Liu², Shanshan Wang^{2*} and Chaolu Feng¹

Abstract

Background Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) plays an important role in the diagnosis and treatment of breast cancer. However, obtaining complete eight temporal images of DCE-MRI requires a long scanning time, which causes patients' discomfort in the scanning process. Therefore, to reduce the time, the multi temporal feature fusing neural network with Co-attention (MTFN) is proposed to generate the eighth temporal images of DCE-MRI, which enables the acquisition of DCE-MRI images without scanning. In order to reduce the time, multi-temporal feature fusion cooperative attention mechanism neural network (MTFN) is proposed to generate the eighth temporal images of DCE-MRI, which enables DCE-MRI image acquisition without scanning.

Methods In this paper, we propose multi temporal feature fusing neural network with Co-attention (MTFN) for DCE-MRI Synthesis, in which the Co-attention module can fully fuse the features of the first and third temporal image to obtain the hybrid features. The Co-attention explore long-range dependencies, not just relationships between pixels. Therefore, the hybrid features are more helpful to generate the eighth temporal images.

Results We conduct experiments on the private breast DCE-MRI dataset from hospitals and the multi modal Brain Tumor Segmentation Challenge2018 dataset (BraTs2018). Compared with existing methods, the experimental results of our method show the improvement and our method can generate more realistic images. In the meanwhile, we also use synthetic images to classify the molecular typing of breast cancer that the accuracy on the original eighth time-series images and the generated images are 89.53% and 92.46%, which have been improved by about 3%, and the classification results verify the practicability of the synthetic images.

Conclusions The results of subjective evaluation and objective image quality evaluation indicators show the effectiveness of our method, which can obtain comprehensive and useful information. The improvement of classification accuracy proves that the images generated by our method are practical.

Keywords DCE-MRI, Medical image synthesis, Co-attention, Multi-temporal feature fusing, Hybrid features

Introduction

In recent years, medical images, such as positron emission tomography (PET), computed tomography (CT) and magnetic resonance images (MRI), occupy an important position in clinical application, which provide a lot of information for disease diagnosis [1–3]. MRI is widely used in the diagnosis of diseases because

*Correspondence:

Shanshan Wang
2010686@stu.neu.edu.cn

¹ Key Laboratory of Intelligent Computing in Medical Image MIIC, Northeastern University, Shenyang, China

² School of Computer Science and Engineering, Northeastern University, Shenyang, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of its noninvasive as it can provide images with different parameters, such as dynamic contrast enhanced MRI (DCE-MRI), diffusion weighted imaging (DWI) and T1/T2 [4–6]. The MRI images with different parameters can provide different relevant information of the disease.

Due to the high resolution of DCE-MRI, it is often used as the main mode in the diagnosis and treatment of clinical diseases. DCE-MRI has a total of eight time series, the first of which is the images without contrast agent and the other time series are acquired at intervals after the addition of contrast agent. In medicine, the first, third and eighth temporal images of DCE-MRI are of great significance. These three temporal images have potential anatomical information, combining multi temporal images can provide more comprehensive information. The collection of eight temporal images of DCE-MRI requires a long scanning time [7–10]. The patients don't feel well during this complete scan. It is very meaningful to shorten the scanning time as much as possible and simplify the breast MR scanning sequence.

Medical image synthesis aims to automatically extract features from existing medical images and synthesize target images. It is essential for modern computer-aided diagnosis (CAD) applications, such as disease screening and diagnosis. It can synthesize another modality images from one modality and not only solve the problem that deep learning requires a large amount of data and medical image data is less or missing, but also solve the possible adverse effects in the imaging process [11]. For example, due to the problems of poor image quality and data loss, Cheng, et al. designed an excellent network structure (HiNet), which realizes the synthesis of multi-modal images [12]. Wolterink JM et al. synthesized CT images by using MRI images through the CycleGAN model [13], which solves the problems that CT scanning is expensive, increases the burden on patients and is easy to cause harm to human body in the process of CT scanning. Standard dose PET is synthesized through low dose PET, so as to reduce artifacts and noise in low dose pet and avoid harm to human body caused by standard dose PET [14].

However, most of the existing synthesis research methods are single temporal images synthesis target temporal images or use multi temporal images feature fusion to generate images, the relationship between corresponding pixels can be obtained by simple fusion, the information interaction between different modal images is ignored and the relationship between modes is not fully utilized. And in the existing multi-modal synthesis methods, the unique features of single modality are often ignored.

How to fully integrate the features of multi temporal images to generate more realistic real images is a challenge. Using multi temporal images to obtain

comprehensive information improves the accuracy of disease diagnosis. Therefore, to explore their rich pixel information and find the potential correlation between them to improve various medical tasks, we use the Co-attention module to obtain the information interaction of multi-scale features of multi temporal images and obtain the remote dependence between pixels, which is conducive to GAN to generate higher quality images. The contributions of this paper are as follows:

- 1) We propose a novel network (MTFN) to realize the multi temporal image fusion of DCE-MRI and generate the eighth temporal image. Experiments prove the effectiveness of our method, and verify that in the synthesis, DCE-MRI multi temporal image feature fusion generates eighth temporal images better than single temporal image generation.
- 2) The Co-attention module is applied to the feature fusion of multi temporal images. We use Co-attention module to obtain the shared features of the two temporal images, and fully integrate the unique and shared features of each temporal images to obtain comprehensive information.
- 3) We integrate the synthesis task of DCE-MRI multi temporal images and the classification task of DCE-MRI molecular typing of breast cancer, used two types of evaluation indicators to comprehensively evaluate the performance of our method.

The main structure of this paper is as follows: the second part mainly introduces the related work, the third part introduces our method and network components in detail, the fourth part is the experimental setup, including datasets, evaluation indicators and other related contents and the fifth part is the experimental results. Finally is the conclusion.

Related work

In 2014, GAN was first proposed by Goodflow et al. [15], an excellent generative model trained by the method of confrontation learning. GAN is composed of generator network and discriminator network. The generator network can generate sample data approximately distributed with the input sample, and the discriminator network can judge the authenticity of the sample, which will correctly judge the real samples and the samples generated by the generator as much as possible. The discriminator network is essentially a binary classification network. Both of them can improve the performance of the generator and discriminator through confrontation learning. GAN is used in various fields of medicine such as classification, denoising, reconstruction and synthesis [16–18] and has achieved

great success [19, 20]. In recent years, researchers have proposed many improved GAN models [21].

LAN et al. proposed that 3D GAN synthesize MRI images through PET images [22], Bmgan can synthesize PET from brain, which can preserve brain structure well [23]. Wang et al. proposed Ea-GANs based on edge perception, which adds additional constraints to the similarity between the real edge map and synthesized edge map. Ea-GANs pays more attention to the edge information of the image and generates high-quality images [24]. Wu et al. develop an encoder-decoder neural network with residual inception block to synthesize breast images [25]. M. Yurt et al. synthesized images with Multi-Contrast MRI through cycleGAN and pix2pix model to obtain more diagnostic information [26]. Multi-modality glioma MRI synthesis method is designed, which can obtain the common feature space and lesion-specific representations [27]. Zhou et al. proposed HiNet to synthesize the target modality images and Mixed Fusion Block is used to fuse features, which contains element-wise summation, element-wise product and element-wise maximization to fuse the features of each mode [28]. Luo et al. proposed EP_IMF-GAN with iterative multi-scale feature fusion and the MRI synthesis method can preserve edge information by generating the edge image of the target modality as an auxiliary condition [29]. Shen et al. use the channel-wise attention and spatial attention to get the target feature map, which makes the effective learning of the image synthetic model [30].

Method

The network structure (MTFN) can be divided into three parts: feature extraction network of single temporal image, Co-attention module and eighth temporal image synthesis network. The single temporal image feature extraction network is mainly used to obtain the unique features of the first phase and the third phase. To obtain more comprehensive mixed features, Co-attention module is used to fully fuse the features of the first and the third temporal images, which are extracted in different layers of the feature extraction network. The features include both the unique features of single temporal images and the shared features of multi temporal images, which are used as the input of the synthesis network. Finally the eighth temporal images are synthesized. The flow chart of this architecture is shown in Fig. 1.

Feature extraction network of single temporal image

The first and the third temporal image feature extraction network have the same network structure. The feature extraction network of single temporal image is composed of three encoder blocks and decoder, which is similar to the Unet structure without skip connection. An encoder block consists two 3×3 convolutional layers and one maxpool operation. The decoder consists of 3×3 convolution layer. The activation function used in the last layer of the network is Tanh, others are ReLU. After each encoder block in the network, we will get the feature maps of the first and third temporal images. The feature maps of the two temporal images will be fused through

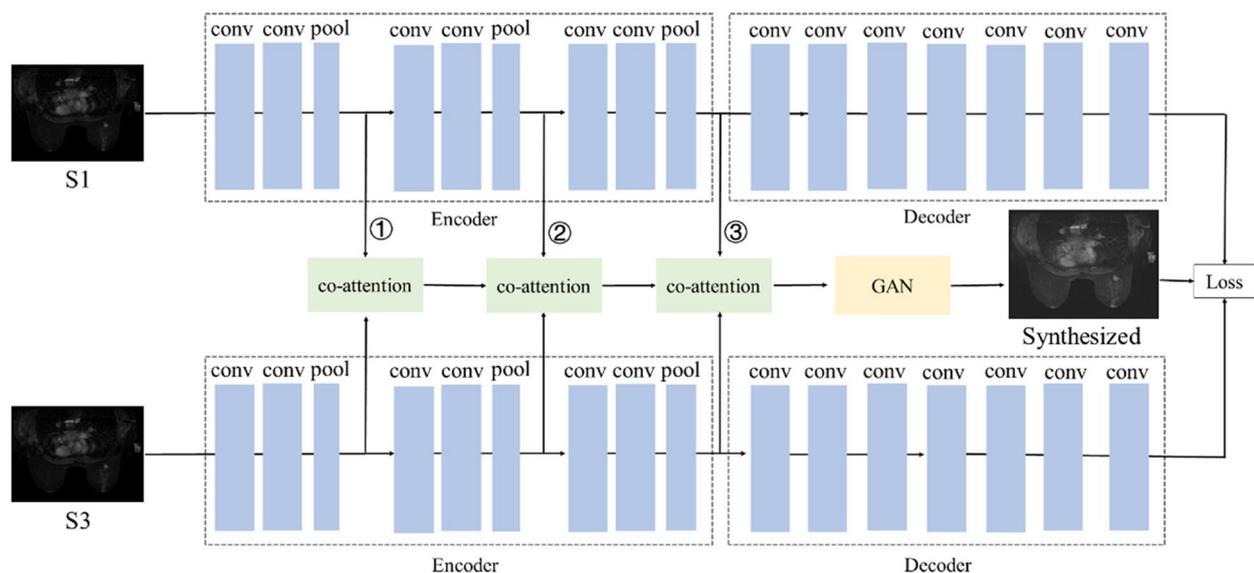


Fig. 1 The network structure of MTFN

Co-attention module, and the feature maps will also be added to the generator.

In the process of image generation, in addition to the common features of multiple temporal images, the unique features of each temporal images are also very important. Therefore, we have reconstructed the images, and the reconstruction loss value is a part of the total loss. The objective function of reconstruction is as follows, where S_i represents the original image and \hat{S}_i represents the reconstructed image.

$$\mathcal{L}_{R_i} = \sum \|S_i - \hat{S}_i\|_1 \tag{1}$$

Co-attention module

In the existing fusion operations, it is often only to fuse the features at the pixel level, such as simple addition and averaging of the corresponding pixels, which can not obtain the remote dependency between the pixels. Therefore, in order to explore the position relationship of pixels between two different temporal images, we introduce Co-attention module into the feature fusion of two temporal images, capture the dependency between the features of two temporal images and the hybrid features are obtained. The architecture of the model is shown in Fig. 2.

First, the feature maps of the first temporal images are passed through 1×1 convolution operation, and then multiply the obtained feature maps, and softmax operation to obtain attention maps, so we obtain the positional relationship between each position of the first temporal

feature map and the whole feature map, and capture the remote dependency. Second, to obtain feature dependency of the first and third temporal images, multiply attention maps and the feature maps which are the feature maps of the third temporal images through 1×1 convolution. Finally, after softmax, the final hybrid feature maps are obtained.

The features of single temporal images are captured through convolution layer, and the feature information between different scales of each phase is mined, and then feature fusion is carried out. According to the different location of fusion features, the usual feature fusion methods are divided into early fusion, middle fusion and late fusion. Early fusion is a basic fusion method, which directly fuses the signals output by each sensor, middle fusion is the direct fusion of feature extraction, and late fusion is the process of re fusion of feature recognition and classification results. In this paper, we experiment on different feature fusion methods to obtain the optimal fusion position.

Eighth temporal image generation network

We take the obtained hybrid features in the early stage as the input of the network, and use the important temporary image generation network to realize image generation. The network is a standard GAN network, which is composed of generator and discriminator. Its structure is shown in Figs. 3 and 4.

The generator generates as realistic images as possible, trying to deceive the discriminator. At the same time, the discriminator judges the generated image as

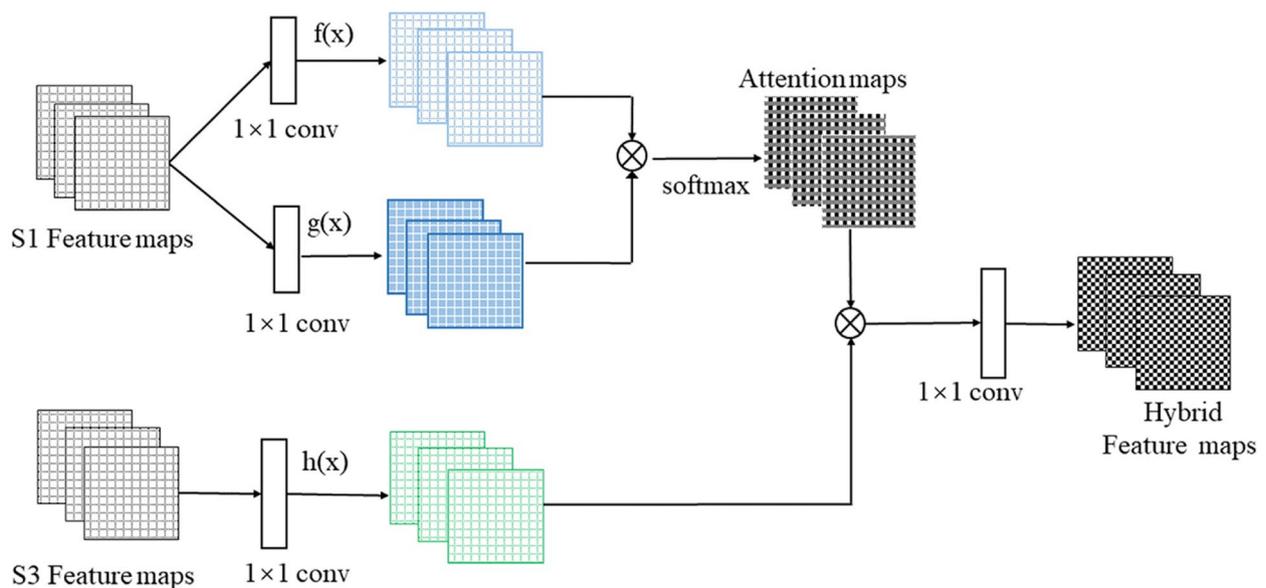


Fig. 2 The architecture of the Co-attention module

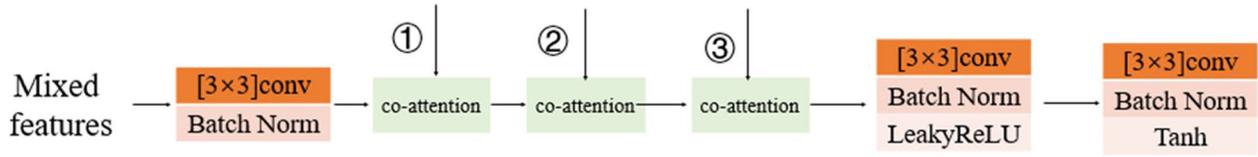


Fig. 3 The architecture of the generator

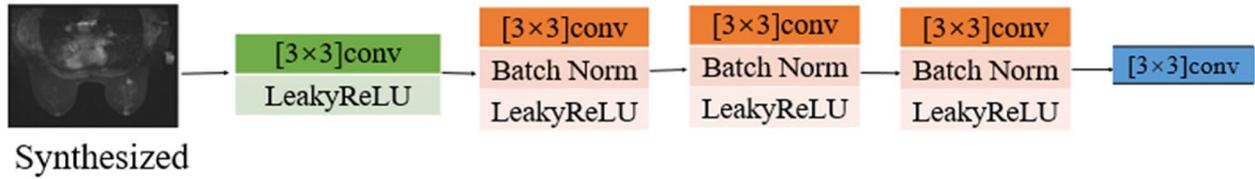


Fig. 4 The architecture of the discriminator

false as possible, and we use PatchGAN structure as discriminator that its output is $n \times n$ matrix to evaluate the image generated by the generator. In our method, we set $n=6$. In the process of network training, the generator and discriminator gamble each other, and finally generate realistic images. And loss function of the GAN and total objective function consists of three terms and can be written as

$$\begin{aligned} \mathcal{L}_{GAN} = & \mathbf{E}_{S_1, S_3 \sim p_{data}} [\log(1 - D(G(S_1, S_3)))] \\ & + \mathbf{E}_{S_1, S_3, S_8} [\|S_8 - G(S_1, S_3)\|_1] \\ & + \mathbf{E}_{S_8 \sim p_{data}} [\log(D(S_8))] \end{aligned} \quad (2)$$

$$\mathcal{L}_{Total} = \lambda_{R_1} \mathcal{L}_{R_1} + \lambda_{R_3} \mathcal{L}_{R_3} + \lambda_{GAN} \mathcal{L}_{GAN} \quad (3)$$

Experimental setup

Dataset

We use two datasets: the multi modal Brain Tumor Segmentation Challenge 2018 dataset (BraTs2018) and clinical breast DCE-MRI dataset from hospitals.

- 1) BraTs2018: The BraTs2018 dataset is used for multi modal brain tumor segmentation in MICCAI competition. It contains 285 patients. Each patient contains four modal datasets, namely T1, T1c, T2 and Flair. The image size is $240 \times 240 \times 155$. In the experiment, we use the T1 and T2 modes to generate Flair.
- 2) Clinical Breast DCE-MRI dataset: The clinical breast DCE-MRI dataset contains 232 patients.

Each patient contains a data set of eight time series. The image size of each time series is $512 \times 512 \times 48$. We

generate the eighth temporal images through the first and the third temporal images.

Evaluation metrics

Evaluate the quality of generated images to verify the effectiveness of the synthesis method that introduces the Co-attention module, we use the three commonly used evaluation indicators to measure the quality of synthetic images, and compare the performance of the method. The three evaluation indexes: PSNR (Peak signal-to-noise ratio), NMSE (Normalized Mean Squared Error) and SSIM (Structural Similarity Index Measurement).

PSNR represents the ratio of the maximum possible power of the signal to the destructive line noise power that affects its representation accuracy. The calculation formula is as follows, where m, n represents the size of the image, $T(i, j)$ represents the pixels in the real image, and $G(i, j)$ represents the pixels in the generated image.

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \|T(i, j) - G(i, j)\|^2 \quad (4)$$

$$PSNR = 10 * \log_{10} \left(\frac{MAX^2 [T(i, j), G(i, j)]}{MSE} \right) \quad (5)$$

Table 1 The results of the different way of feature fusion on DCE-MRI

Methods	DCE-MRI	PSNR↑	NMSE↓	SSIM↑
Early fusion	$S_1, S_3 \rightarrow S_8$	21.45	1.38	0.67
Late fusion	$S_1, S_3 \rightarrow S_8$	21.46	1.20	0.67
Medium-term fusion	$S_1, S_3 \rightarrow S_8$	21.67	1.04	0.68

Table 2 The results of the different way of feature fusion on BraTs2018

Methods	BraTs2018	PSNR↑	NMSE↓	SSIM↑
Early Fusion	T1, T2 → Flair	21.29	0.09	0.82
Late Fusion	T1, T2 → Flair	22.27	0.07	0.83
Medium-term fusion	T1, T2 → Flair	22.41	0.07	0.82

NMSE represents normalized mean square error. The smaller the value, the better. The calculation formula is as follows:

$$NMSE = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [T(i,j) - G(i,j)]^2}{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [T(i,j)]^2} \quad (6)$$

The similarity of two images can be measured by SSIM. The calculation formula is as follows, and μ_G is the average value of the generated image, μ_T is the average value of the real image, μ_G^2 is the variance of the generated image, μ_T^2 is the variance of the real image, σ_{GT} is the standard deviation between the generated image and the real image, c_1 and c_2 are constants respectively to avoid errors caused by denominator 0.

$$SSIM = \frac{(2\mu_G\mu_T + c_1)(\sigma_{GT} + c_2)}{(\mu_G^2 + \mu_T^2 + c_1)(\sigma_G^2 + \sigma_T^2 + c_2)} \quad (7)$$

The above three evaluation indexes are used to evaluate the quality of the generated images. In addition, to verify the usefulness of the images generated by our method, we also used real images and generated images to classify breast cancer molecular subtypes which is helpful for doctors to formulate individualized treatment plans for patients. The evaluation indexes used for classification are Accuracy, Precision, Recall and F1 score.

Implementation details

Each image in the BraTs2018 dataset and the private DCE-MRI dataset is divided into $4 \times 128 \times 128$ and $25 \times 128 \times 128$ patch images. All images are normalized and preprocessed. The preprocessed images are input into the network in pairs. The network parameters are updated through the total loss function value of a single phase network and the network synthesizing the final target image.

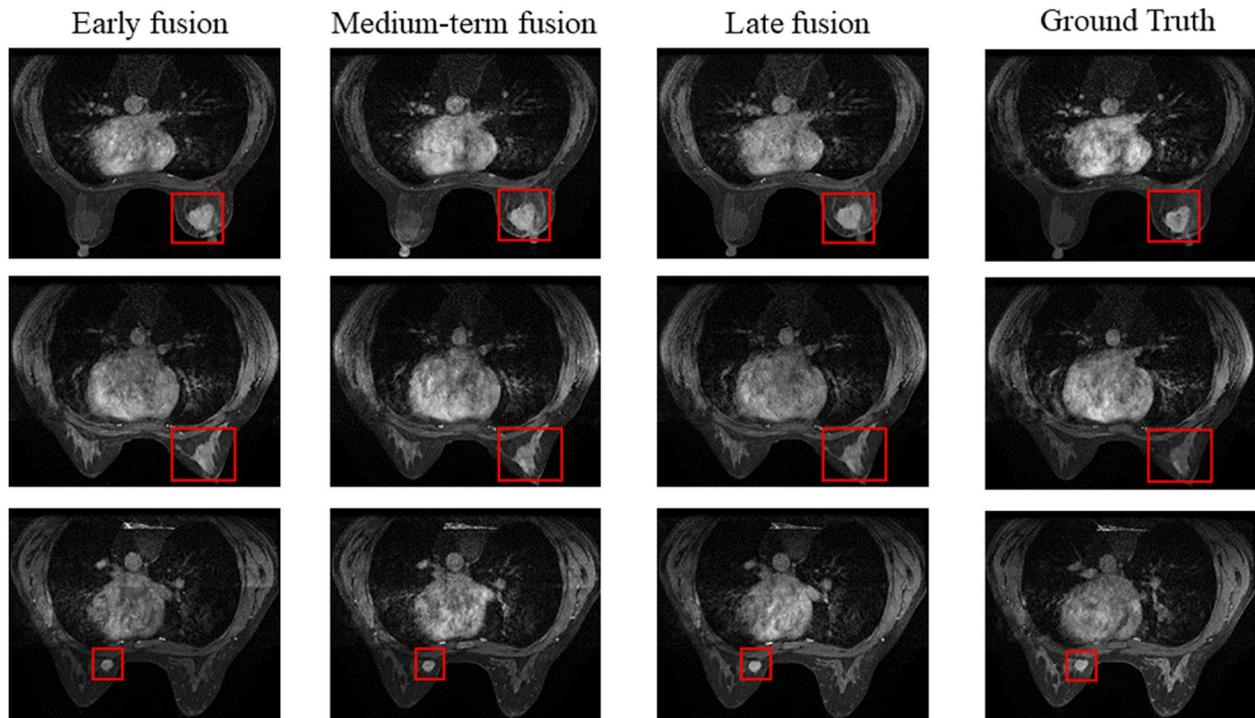


Fig. 5 The generated eighth images of DCE-MRI using different way of feature fusion

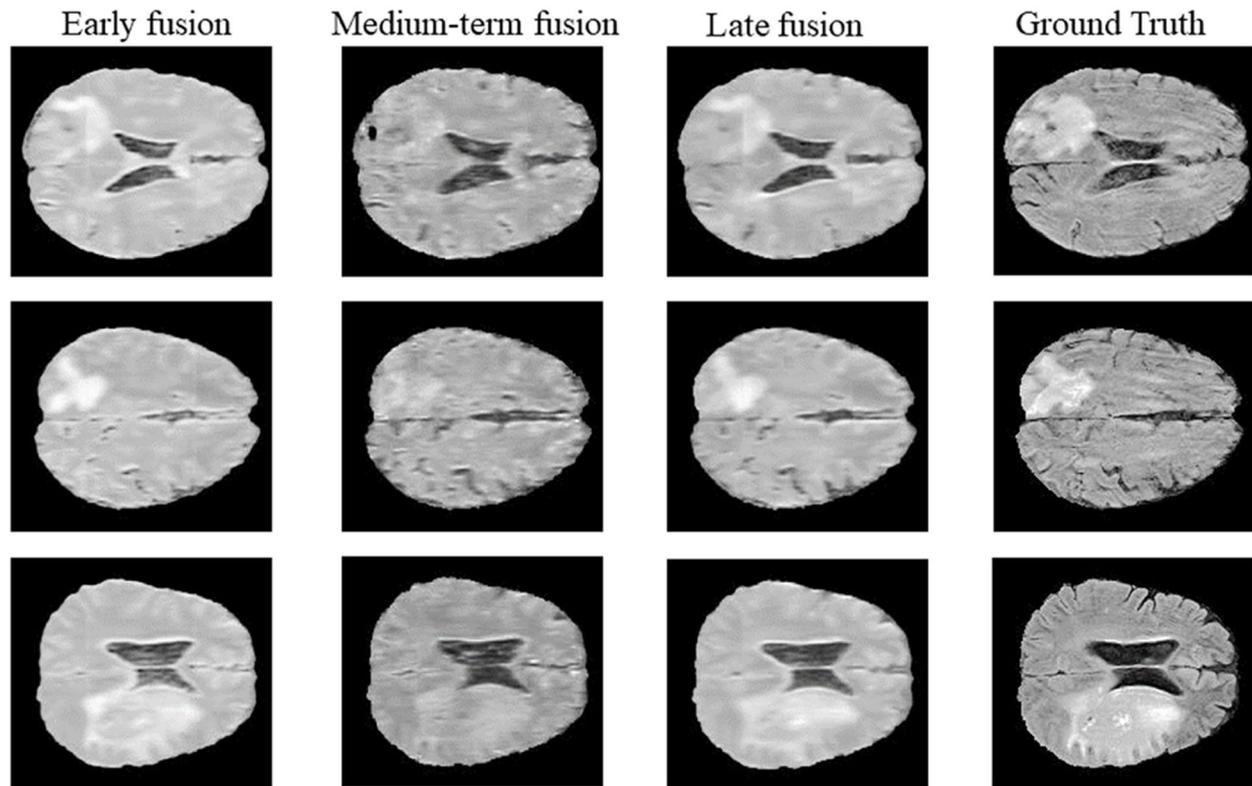


Fig. 6 The generated Flair images of BraTs2018 using different way of feature fusion

We use the pytorch framework for network construction, select Adam optimizer, the learning rate is set to 0.001, the trained batch size=4. All experiments are carried out in NVIDIA GeForce GTX1080 Ti GPU.

Result

In this paper, we have carried out three different tasks: explore the location of Co-attention fusion module in the network, so as to obtain the best fusion method. Through the advanced fusion method obtained from the previous implementation, we synthesize the eighth temporal images of DCE-MRI and compare it with the existing methods to verify the effectiveness of our method. The classification experiment of DCE-MRI breast molecular typing is carried out by using the generated image to further verify the practicability of the generated image.

Different fusion positions

First we explore the way of feature fusion using private dataset and public dataset, and conduct early fusion, medium-term fusion and late fusion experiments respectively. The early fusion is the fusion operation before the first layer of the network and only one

co-attention fusion. The middle fusion is the fusion operation in the process of extracting image features, and the late fusion is the fusion after the last layer of the network.

The experimental results are shown in Table 1 and Table 2, the generated images are shown in Figs. 5 and 6. Experiments using our DCE-MRI dataset show that the generated images are similar to the real images, and the results of medium-term fusion are the best from the analysis of objective indicators. On BraTs2018 dataset, the images generated by early fusion and late fusion are fuzzy and smooth, and the images generated by medium-term fusion contain more texture details. Medium-term fusion is the best fusion way. Therefore, our method adopts the feature fusion method based on Co-attention and medium-term fusion.

Table 3 Compared with other methods on DCE-MRI breast images

Methods	DCE-MRI	PSNR↑	NMSE↓	SSIM↑
Pix2pix	S1, S3 → S8	18.53	0.29	0.66
HiNet	S1, S3 → S8	20.01	1.08	0.64
Ours	S1, S3 → S8	21.67	1.04	0.68

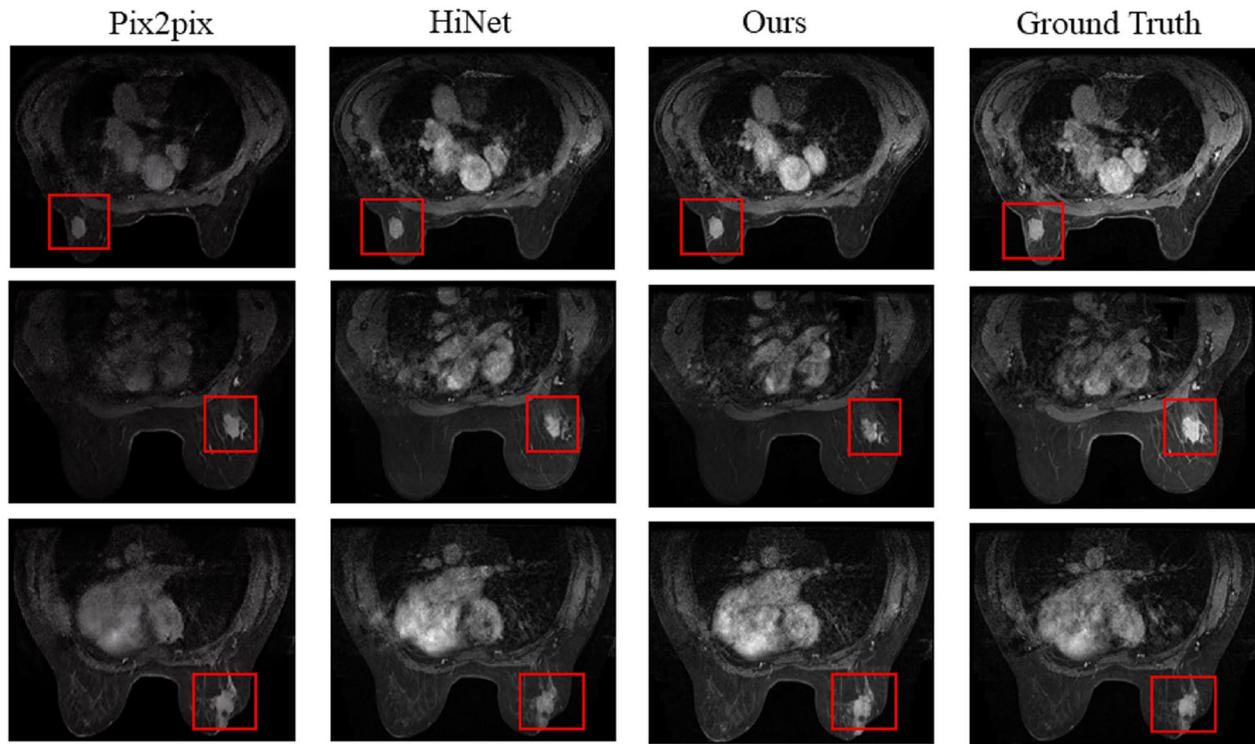


Fig. 7 The generated eighth temporal images of DCE-MRI with three methods

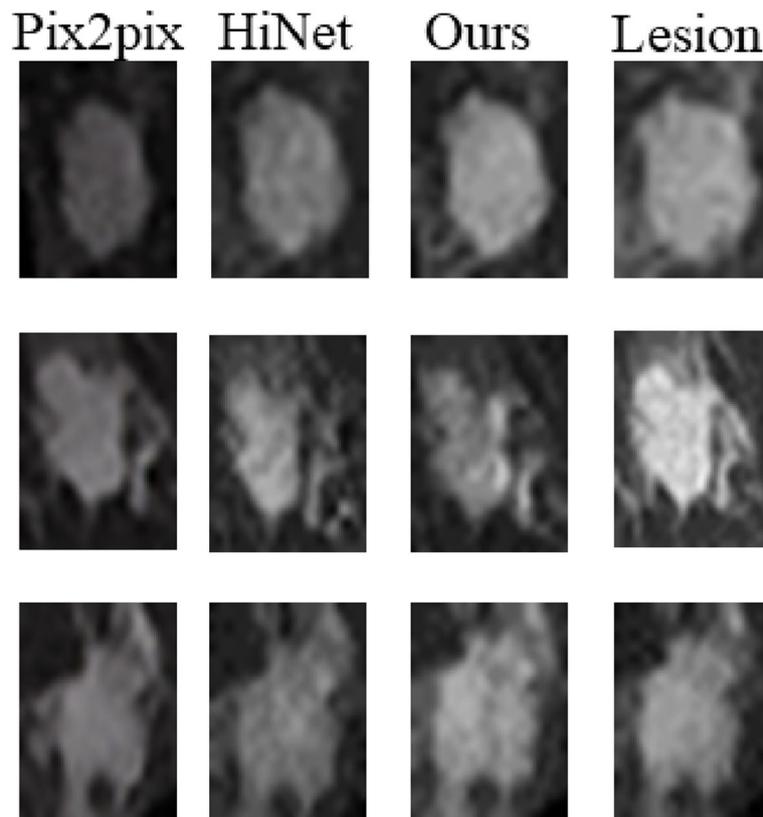


Fig. 8 The generated lesion with three methods, compared to other methods, the generated lesion with our methods is more complete and realistic

Table 4 The result of fusion of single temporal images and multi temporal image to generate the eighth temporal images

Methods	DCE-MRI	PSNR \uparrow	NMSE \downarrow	SSIM \uparrow
Ours	S1 \rightarrow S8	21.58	1.06	0.67
Ours	S3 \rightarrow S8	21.33	1.20	0.66
Ours	S1, S3 \rightarrow S8	21.67	1.04	0.68

Eighth temporal image synthesis

We use Clinical dataset to do experiments with our method, and compare with the existing methods Pix2pix and HiNet. From the evaluation index of Table 3, the PSNR and SSIM obtained by our method are higher than those obtained by pix2pix and HiNet. NMSE is lower than that obtained by HiNet. On the whole, our method has also been improved. And the generated eighth temporal images are shown in Fig. 7, compared to real images, the eighth temporal images synthesized by pix2pix, HiNet and our method are basically consistent with the real images on the whole, but in terms of details, we truncate and enlarge the focus in the images, as shown in Fig. 8, the image generated by pix2pix and HiNet method is fuzzy and low contrast. The images generated by our method is more realistic and complete, such as the size and shape of the lesion.

Specifically, we use the single and multi temporal images of DCE-MRI to generate the eighth temporal

images respectively, as shown in Table 4. The experimental results show that multi temporal images feature fusion can obtain more information and contribute to the generation of the eighth temporal images which are shown in Fig. 9.

Molecular typing of breast cancer in DCE-MRI

The experiment fully confirms the advanced nature of our method, and at the same time, in order to further verify the usefulness of the generated images, we used the original eighth temporal images and the generated eighth temporal images to conduct breast cancer molecular typing experiments which respectively, that is, non LuminalB and LuminalB molecular typing two classification. The experimental results are shown in Table 5. We first carried out classification experiments using the first and third temporal images, and then added the eighth temporal images for experiments. The experimental results show the necessity of the existence of the eighth temporal images, which is an indispensable series. Finally, the generated eighth temporal images are used to replace the original images for experiments.

In molecular typing experiments, as classifier, we adopt the DenseNet [20] as the base classifier to predict. The classification results are shown in Table 5 and the generated eighth temporal image improves the classification accuracy which fully prove that the image generated by our proposed method is practical.

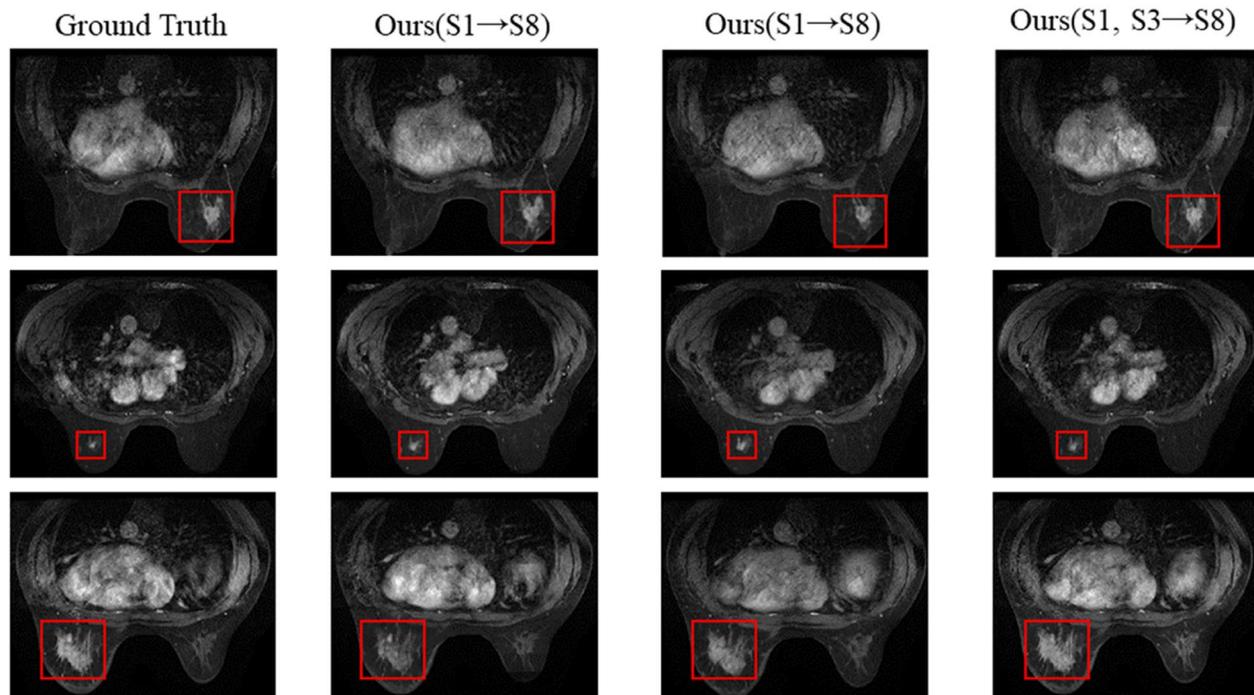
**Fig. 9** The generated eighth images with single temporal images and multi temporal images

Table 5 Results of molecular typing of DCE-MRI breast cancer on the First, Third, Eighth and Generated eighth temporal images by our method, which contains of two classes: non Luminal B and Luminal B

Data	Accuracy	Precision	Recall	F1-score
S1 + S3	83.55%	83.95%	83.70%	83.52%
S1 + S3 + S8	89.53%	90.44%	89.55%	89.46%
S1 + S3 + Generated_S8	92.46%	93.06%	92.33%	92.40%

Conclusions

In this paper, we have proposed the MTFN based GAN and attention mechanism to realize the multi temporal image fusion of DCE-MRI and generate the eighth temporal image. The effectiveness of the image synthesis method of feature fusion with improved attention mechanism is verified on the BraTs2018 public dataset and DCE-MRI dataset. It is worth noting that we not only retain the multi-scale features of single temporal images, but also explore the correlation between two temporal images of DCE-MRI by introducing attention mechanism. The results show that this method can obtain synthetic images with clear structure, which not only simplifies the breast MR scanning sequence, but also has application value in clinical diagnosis.

At present, we realize the feature fusion of DCE-MRI multi-temporal images to generate the eighth-temporal images, and the accuracy of molecular classification of breast cancer using the generated eighth-temporal images is improved by 3%. Because the lesion information is more valuable for doctors' diagnosis, we will consider local enhancement of lesion location in the next step to further improve the accuracy of classification.

Acknowledgements

We thank everyone in the collaboration group for their efforts.

Authors' contributions

WL proposed experimental ideas, evaluated experimental data, and drafted manuscripts. JYL designs experimental procedures collects data and assists in manuscript writing. CLF proposed the overall structure of the article and supplemented the experimental chart. SSW revises the manuscript and evaluates the data. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key R&D Program of China (No.2021YFC2701003), the Fundamental Research Funds for the Central Universities (N2016006).

Availability of data and materials

The datasets used and analyzed during the current study available from the corresponding author on reasonable request. We use two datasets: the multi modal Brain Tumor Segmentation Challenge 2018 dataset (BraTs2018) and clinical breast DCE-MRI dataset from hospitals. The BraTs2018 is introduced by Bjoern H. Menze et al. in The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). The clinical data set involves patient privacy and is not convenient for disclosure.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 8 November 2022 Accepted: 15 January 2024

Published online: 19 February 2024

References

- Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shapeconsistency generative adversarial network. *Proc IEEE Conf Comput Vis Pattern Recognit.* 2018;9242–9251.
- Zhang J, He X, Qing L, et al. BPGAN: brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis. *Comput Methods Programs Biomed.* 2022;217:106676.
- Liu Y, Chen A, Shi H, et al. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. *Comput Med Imaging Graph.* 2021;91:101953.
- Ottens T, Barbieri S, Orton M R, et al. Deep learning DCE-MRI parameter estimation: application in pancreatic cancer. *Med Image Anal.* 2022: 102512.
- Si T, Patra D K, Mondal S, et al. Breast DCE-MRI segmentation for lesion detection using chimp optimization algorithm. *Expert Syst Appl.* 2022: 117481.
- Xu Z, Chen M, Zheng S, et al. Differential diagnosis of parotid gland tumours: application of SWI combined with DWI and DCE-MRI. *Eur J Radiol.* 2022;146:110094.
- Xie T, et al. Differentiation of triple-negative breast cancer from other subtypes through whole-tumor histogram analysis on multiparametric MR imaging. *Eur Radiol.* 2019;29(5):2535–44.
- Ab Mumin N, Ramli Hamid MT, Wong JHD. Magnetic resonance imaging phenotypes of breast cancer molecular subtypes: a systematic review. *Acad Radiol.* 2021;S1076–6332(21):00325–31.
- Huang Y, Shao L, Frangi AF. Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR).* 2017;6070–6079.
- Fan J, Cao X, Wang Q, Yap P-T, Shen D. Adversarial learning for mono-or multi-modal registration. *Med Image Anal.* 2019;58:101545.
- Luo Y, Zhou L, Zhan B, et al. Adaptive rectification based adversarial network with spectrum constraint for high-quality PET image synthesis. *Med Image Anal.* 2022;77:102335.
- Wolterink JM, Dinkla AM, Savenije MHF, et al. Deep MR to CT synthesis using unpaired data. *International workshop on simulation and synthesis in medical imaging.* Cham: Springer, 2017: 14–23.
- Ang SP, Phung SL, Field M, et al. An improved deep learning framework for MR-to-CT image synthesis with a new hybrid objective function. *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI).* IEEE. 2022: 1–5.
- Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Advances in neural information processing systems.* 2014. 27.
- Jeong JJ, Tariq A, Adejumo T. et al. Systematic review of Generative Adversarial Networks (GANs) for medical image classification and segmentation. *J Digit Imag.* 2022: 1–16.
- Wang C, Yang G, Papanastasiou G, et al. DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis. *Inform Fusion.* 2021;67:147–60.
- Fetty L, Bylund M, Kuess P, et al. Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Z Med Phys.* 2020;30(4):305–14.

18. Zhang Y, Yap PT, Chen G, et al. Super-resolution reconstruction of neonatal brain magnetic resonance images via residual structured sparse representation. *Med Image Anal.* 2019;55:76–87.
19. Kazemina S, Baur C, Kuijper A, et al. GANs for medical image analysis. *Artif Intell Med.* 2020;109:101938.
20. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017:4700–4708.
21. Lan H, et al. SC-GAN: 3D self-attention conditional GAN with spectral normalization for multi-modal neuroimaging synthesis. 2020: 1–35.
22. Hu S, Lei B, Wang S, et al. Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE Trans Med Imaging.* 2021;41(1):145–57.
23. Yu B, Zhou L, Wang L, et al. Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. *IEEE Trans Med Imaging.* 2019;38(7):1750–62.
24. Gao F, Wu T, Chu X, et al. Deep residual inception encoder–decoder network for medical imaging synthesis. *IEEE J Biomed Health Inform.* 2019;24(1):39–49.
25. Dar SU, Yurt M, Karacan L, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging.* 2019;38(10):2375–88.
26. Huang P, Li D, Jiao Z, et al. Common feature learning for brain tumor MRI synthesis by context-aware generative adversarial network. *Med Image Anal.* 2022;79:102472.
27. Luo Y, Nie D, Zhan B, et al. Edge-preserving MRI image synthesis via adversarial network with iterative multi-scale fusion. *Neurocomputing.* 2021;452:63–77.
28. Zhou T, Fu H, Chen G, Shen J, Shao L. Hi-net: hybrid-fusion network for multi-modal MR image synthesis. *IEEE Transactions on Imaging.* 2020;39(9):2772–81.
29. Shen J, Wu T. Learning inception attention for image synthesis and image recognition. *arXiv preprint arXiv:2112.14804*, 2021.
30. Ding S, Zheng J, Liu Z, et al. High-resolution dermoscopy image synthesis with conditional generative adversarial networks. *Biomed Signal Process Control.* 2021;64:102224.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.