CT and MRI image fusion via dual-branch GAN

Wenzhe Zhai, Wenhao Song, Jinyong Chen and Guisheng Zhang

School of Electrical and Electronic Engineering,
Shandong University of Technology,
Zibo, China
Email: wenzhezhai@163.com
Email: sdut_songwenhao@163.com
Email: sdut_jychen@163.com
Email: sdut_guisheng@163.com

Qilei Li

School of Electronic Engineering and Computer Science,
Queen Mary University of London,
London, UK
Email: qilei.li@outlook.com

Mingliang Gao*

School of Electrical and Electronic Engineering,
Shandong University of Technology,
Zibo, China
Email: mlgao@sdut.edu.cn
*Corresponding author

Abstract: CT and MRI image fusion is a popular research field that plays a vital role in clinical diagnosis. To retain more salient features and complementary information from source images, we propose a dual-branch generative adversarial network (DBGAN) to fuse the CT and MRI images. The proposed DBGAN is designed in a dual branching structure schema, which consists of a couple of generators and discriminators. The generators and discriminators establish a generative adversarial relationship so that the fused images generated by the generators are indistinguishable from the discriminators. Furthermore, we employ the multiscale extraction module (MEM) and self-attention module (SAM) in the generators to enhance the salient features and detailed information of the fused images. The subjective and objective evaluation demonstrate the superiority of the proposed method over the state-of-the-art methods.

Keywords: image fusion; generative adversarial network; CT/MRI image; healthcare.

Reference to this paper should be made as follows: Zhai, W., Song, W., Chen, J., Zhang, G., Li, Q. and Gao, M. (xxxx) 'CT and MRI image fusion via dual-branch GAN', *Int. J. Biomedical Engineering and Technology*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Wenzhe Zhai is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include crowd counting, image fusion, and deep learning.

Wenhao Song is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include computer vision and deep learning.

Jinyong Chen is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include crowd counting and deep learning.

Guisheng Zhang is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include fake face and deep learning.

Qilei Li is currently a PhD student with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK. He received an MS in Signal and Information Processing from the Sichuan University. His research interests are computer vision and deep learning.

Mingliang Gao received his PhD in Communication and Information Systems from the Sichuan University, Chengdu, China, in 2013. He is currently an Associate Professor at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His main research interests include computer vision and deep learning.

1 Introduction

Computed tomography (CT) images and magnetic resonance imaging (MRI) images are two most commonly used medical imaging modes for clinical diagnosis. The CT images preserve bone structures, and the MRI images can conserve tissue details. CT and MRI medical image fusion are of great helps in preserving more detailed information, and it is crucial in diagnosing diseases (Tawfik et al., 2021; Du et al., 2016). Nevertheless, a solitary medical imaging modality cannot conserve adequate information for medical diagnosis. As a result, the retention of salient features and complementary information is crucial in the field of multimodal medical image fusion.

A variety of image fusion algorithms have been proposed in the literature which can be divided into three classifications, i.e., traditional methods, CNN-based methods, and GAN-based methods (Li et al., 2013; Li and Wu, 2018; Li et al., 2019). The traditional methods usually adopt manual-designed strategies, and they are time-consuming due to the complicated fusion strategies (Li et al., 2013; Kumar, 2015; Li et al., 2018). In recent years, benefitting from the excellent performance of feature extraction, many

CNN-based methods have been put forward and have been proved to outperform the traditional methods (Liu et al., 2017; Li and Wu, 2018; Zhang et al., 2020; Xu et al., 2020b, 2022). However, it is hard to train a CNN model well in the absence of labelled samples. To address this problem, the generative adversarial network (GAN) (Goodfellow et al., 2014) has been broadly employed to generate images with favourable visual effects without needing the ground truth of fusion image. However, the fusion results of GAN-based methods are prone to lose valuable details, as they are liable to overrate one source image while underrating the another (Ma et al., 2020; Fu et al., 2021).

To fully exploit multimodal information, we propose a dual-branch generative adversarial network (DBGAN) model to fuse the CT and MRI images. The proposed DBGAN is designed in a dual branching structure schema, consisting of a couple of generators and discriminators. The generators extract details to fuse meaningful information by sharing the same high-level information and utilising the diverse underlying details. To enhance the salient features and detail information of the fused images, we employ the multiscale extraction module (MEM) and self-attention module (SAM) in the generators. The coupled discriminators pull each other on the distribution of the generated data attained by the generator so that the fused image saves the most prominent features from both CT and MRI images. Extensive experimental results verify the superiority of the proposed method over the state-of-the-art methods in both subjective and objective evaluations.

The remaining paper is described as below. Section 2 introduces the previous work. Section 3 presents the details of the DBGAN. Section 4 elaborates comparative results and discussions. Section 5 summarises the paper.

2 Related work

Image fusion is a broad domain that includes a range of algorithms and applications (Tawfik et al., 2021; Du et al., 2016). According to the different categories, the methods of image fusion can be roughly divided into three categories, i.e., traditional methods, CNN-based methods, and GAN-based methods (Li et al., 2019).

The early traditional methods usually utilise spatial domain-based to calculate the pixels of images. For example, Li et al. (2013) used the average filter to obtain the two-scales representations, and the details of different source images can be well conserved for image fusion. Kumar (2015) measured the strength of the horizontal and vertical details to fuse the source images. However, these methods depend on complex fusion rules, which may take a long time for decomposition (Li et al., 2019).

In the past few years, the emergence of CNN-based methods have provided a new branch for multi-modal medical image fusion. For instance, Zhang et al. (2020) proposed the IFCNN to generate rich fused images by the reconstructed fused features through convolutional layers. Xu et al. (2020b) presented the FusionDN to preserve the detail information in the fusion process. Xu et al. (2022) trained the U2Fusion network to maintain the similarity between the fusion results and source samples. Nevertheless, the CNN-based methods cannot be adequately trained, because the multimodal medical image pairs are fewer and the ground-truth can not be clearly defined.

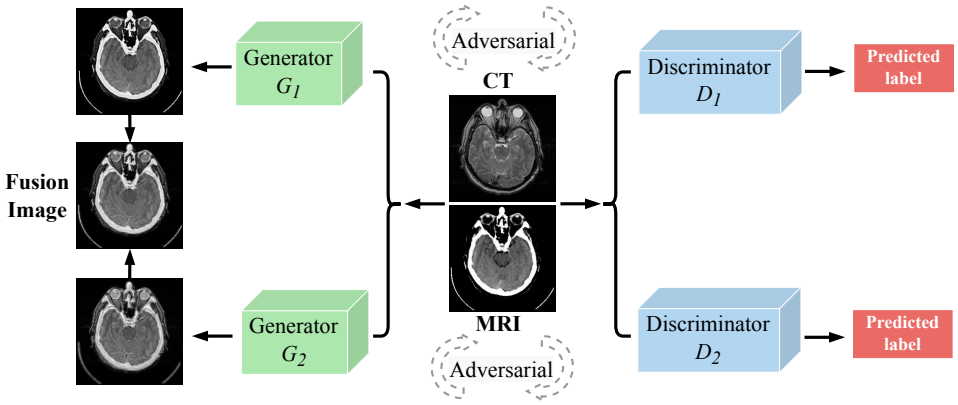
Recently, more and more multimodal medical images are fused using GAN-based methods to address the problem of training samples insufficiency. Ma et al. (2019) designed the FusionGAN to blend the pixel intensity of one source image with the texture detail of another source image. Ma et al. (2020) adopted a generator and two discriminators to build an adversarial game, which can adequately train and avoid information loss. Xu et al. (2020a) designed the MEF-GAN to boost the performance of the fused images by combining the attention mechanism. Fu et al. (2021) trained a fusion network based on GANs and dense-block to improve the fusion effect of two source images.

3 Proposed method

3.1 Overview

The schematic of the proposed DBGAN is depicted in Figure 1. It is designed in a dual branching structure schema, consisting of a couple of generators and a pair of discriminators. The generators aim to generate samples by capturing the data distribution, and the discriminators devote to estimating the probability in which the sample images come from the training data rather than generators. Eventually, the generators generate samples which the discriminators are unable to distinguish.

Figure 1 The framework of DBGAN for CT and MRI image fusion (see online version for colours)

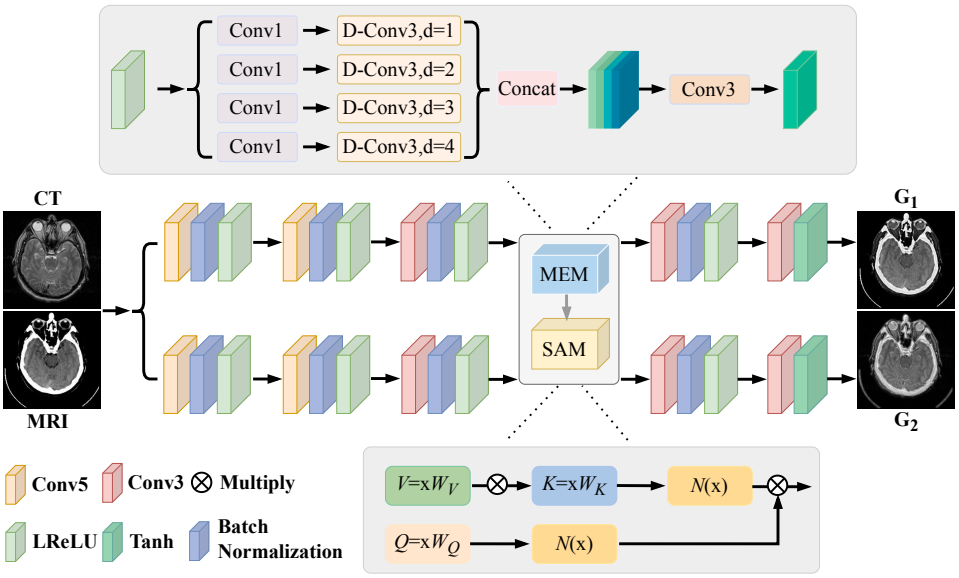


3.2 The architecture of generator

The structure of the generator for CT image is shown in Figure 2. It consists of three convolution blocks, a MEM, a SAM, and two convolution layers.

The three layers in the convolution blocks are convoluted by 5×5, 5×5, and 3×3, respectively. The large convolution filter can obtain large receptive fields from the input image pairs, and the small convolution filter can reduce the amount of computation. These convolution blocks combine with batch-normalisation (BN) layer and LReLU (Maas et al., 2013) activation function.

Figure 2 The overall structure of the designed generator (see online version for colours)



To increase the scale information, we adopt MEM to extract multiscale features. At first, the MEM squeezes the channels of the feature maps by 1×1 convolution layer. After that, the squeezed feature map is handled by dilated convolution with various dilation ratios of 1, 2, 3 and 4 to conserve the multiscale features of the source images. The feature maps are fused by channel-level concatenation operations and 3×3 convolution layer. The final feature map is the same size as the input.

To boost the capabilities of the DBGAN, we employ SAM to increase the attentional information. Given an input $\mathbf{x} \in \mathbb{R}^{N \times C}$, the proposed method aggregate the attention information by d_{model}-dimensional keys, values and queries. The attention $N(\mathbf{x})$ is formulated as,

$$Q = \mathbf{x}W_Q, K = \mathbf{x}W_K, V = \mathbf{x}W_V, \tag{1}$$

$$N(\mathbf{x}) = \frac{\gamma Q}{\sqrt{\sum_{i=0}^{h} \|Q\|^2}} \frac{\gamma K^T V}{\sqrt{\sum_{i=0}^{h} \|K^T V\|^2}}, \tag{2}$$

where Q, K, and V are queries matrices, keys matrices, and values matrices, respectively. W_i represents the i^{th} parameter metric. γ represents a learnable parameter and h is channel-dimension. Self-attention enables the model to focus on significant information from different locations.

Finally, the last two convolutional layers employ 3×3 and 1×1 with LReLU and Tanh activation function to generate the fusion image G_1.

Likewise, the structure of the generator G_2 for MRI image is arranged in the same scheme as CT generator, but they have different loss functions, which are detailed in Subsection 3.4. The generator G_2 produces the fusion image retaining tissue information.

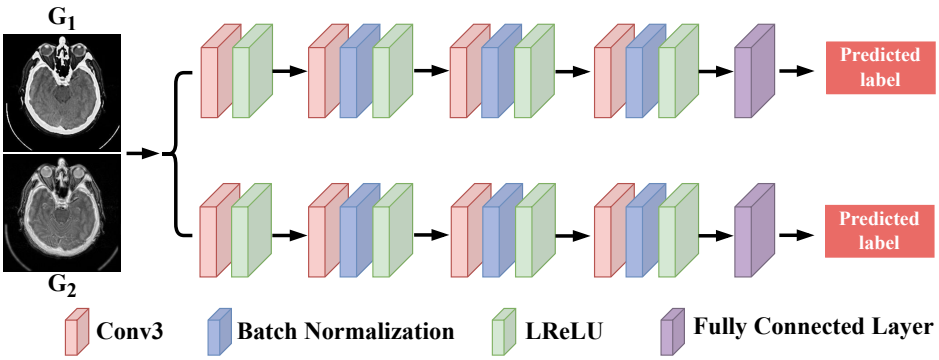
The final fusion image G preserves the bone structure information of the CT image and tissue information of the MRI image. It is formulated as follows,

$$G = \frac{1}{2}\left(G_1 + G_2\right). \tag{3}$$

3.3 The architecture of discriminator

The discriminator is also designed in a two-branches structure, which is depicted in Figure 3. For the first branch of the discriminator, the discriminator D_1 is intended to be adversarial to the generators. Wherein a discriminator comprises four convolutional layers and one fully connected layer. The activation function of the convolutional layer is LRelu. The predicted label is eventually generated. Identically, the prediction label is ultimately generated by the discriminator D_2.

Figure 3 The overall structure of the designed discriminator (see online version for colours)



3.4 Loss function

3.4.1 Loss function for generator

The first generator loss function L_{G1} is defined as,

$$L_{G_1} = \Psi\left(G_1\right) + \lambda L_{\text{Con1}}. \tag{4}$$

where $\Psi\left(G_1\right)$ denotes the adversarial loss between G_1 and D_1. L_{Con1} represents content loss, and λ denotes the weight parameter to balance the different loss of the generator. They are formulated as follows,

$$\Psi\left(G_1\right) = \frac{1}{N}\sum_{n=1}^{N}\left(D_1\left(G_1^{(n)}, I_{CT}^{(n)}\right) - a\right)^2, \tag{5}$$

$$L_{\text{Con1}} = \frac{1}{wh}\left(\rho\left\|G_1 - I_{CT}\right\|_2^2 + \left\|G_1 - I_{Pf}\right\|_2^2\right), \tag{6}$$

where w and h are the width and height of the input image, respectively. I_{CT} and I_{Pf} denote the CT image and pre-fused image, respectively. a and μ are the hyperparameters.

Likewise, the second generator loss function L_{G1} is formulated as,

$$L_{G_2} = \Psi\left(G_2\right) + \lambda L_{\text{Con2}}. \tag{7}$$

where $\Psi\left(G_2\right)$ denotes the adversarial loss between G_2 and D_2. L_{Con1} represents content loss, and λ denotes the weight parameter to balance the different loss of the generator. $\Psi\left(G_2\right)$ and L_{Con2} are formulated as follows,

$$\Psi\left(G_2\right) = \frac{1}{N}\sum_{n=1}^{N}\left(D_1\left(G_1^{(n)}, I_{MRI}^{(n)}\right) - a\right)^2, \tag{8}$$

$$L_{\text{Con2}} = \frac{1}{wh}\left(\beta\left\|\nabla G_2 - \nabla I_{MRI}\right\|_2^2 + \left\|G_2 - I_{Pf}\right\|_2^2\right), \tag{9}$$

where $\bigtriangledown$ is the gradient operator. w and h are the width and height of the input image, respectively. I_{MRI} and I_{Pf} denote the MRI image and pre-fused image, respectively. a and μ are the hyperparameters.

3.4.2 Loss function for discriminator

The first discriminator D_1 in DBGAN is utilised to discriminate between the CT image and the source image. The loss function of the first discriminator L_{G1} is formulated as,

$$L_{D_1} = \mathbb{E}\left[-\log D_1(I_{CT})\right] + \mathbb{E}\left[-\log\left(1 - D_1(G_1)\right)\right]. \tag{10}$$

The second discriminator D_2 in DBGAN is used to discriminate between the MRI image and the source image. The loss function of the second discriminatorL_{G2} is defined as,

$$L_{D_2} = \mathbb{E}\left[-\log D_2(I_{mri})\right] + \mathbb{E}\left[-\log\left(1 - D_2(G_2)\right)\right]. \tag{11}$$

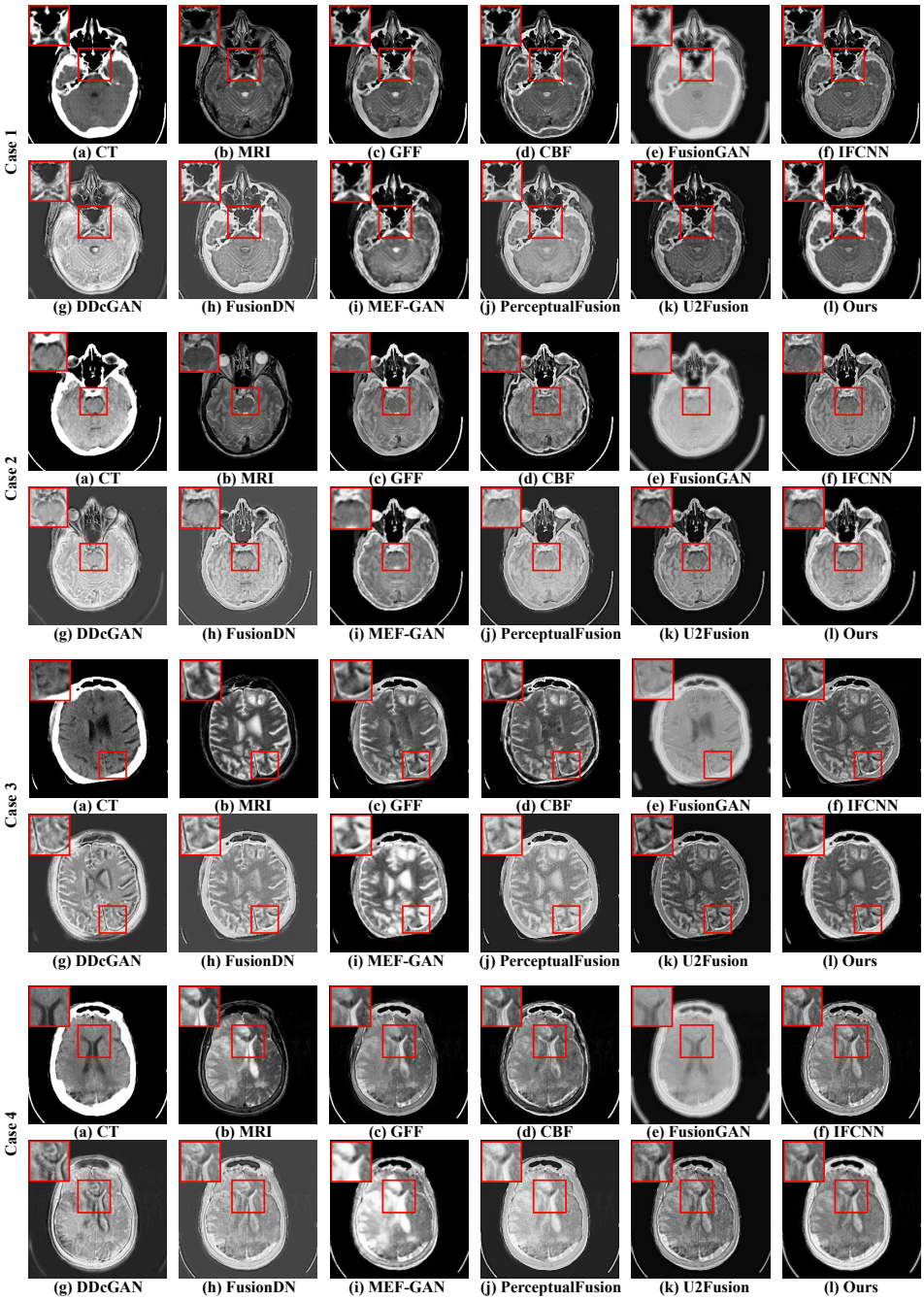
4 Experimental results and analysis

4.1 Datasets and training details

In the experiments, the CT and MRI images are acquired from the Whole Brain Atlas database of Harvard Medical School (http://www.med.harvard.edu/aanlib/home.html). The images are resized to 256 $\times$ 256 and delivered in JPG files. In the training process, 85 pairs of CT and MRI images from the trained database are cropped to size 120 $\times$ 120. In this paper, we utilise four typical cases to verify the effectiveness of the proposed method. Therein, case 1 preserves the information of acute stroke, case 2 reflects the acute stroke in the brain, case 3 is the visualisation of multiple embolic infarctions, and case 4 employs the fatal stroke in the brain.

The learning rates of the generators and discriminators are initialised as 10^{-4} and the Adam is used as the optimiser. The epoch number is 100 and the batch is set to 16. We utilise the fusion results of IFCNN (Zhang et al., 2020) as pre-fusion images. The code is based on the Pytorch framework within an Ubuntu server system. The configuration is equipped with Intel Core i7-9700K CPU @3.60 GHz and NVIDIA GeForce GTX 3090 GPU.

Figure 4 Subjective results on four typical cases, (a)–(b) illustrates the CT and MRI images (c)–(k) shows the fused results of the methods (l) is the result of DBGAN (see online version for colours)



4.2 Subjective evaluation

To evaluate the intuitive results, we perform the comparison of the proposed method with nine SOTA methods, including the GFF (Li et al., 2013), CBF (Kumar, 2015), FusionGAN (Ma et al., 2019), IFCNN (Zhang et al., 2020), DDcGAN (Ma et al., 2020), FusionDN (Xu et al., 2020b), MEF-GAN (Xu et al., 2020a), PerceptualFusion (Fu et al., 2021), and U2Fusion (Xu et al., 2022). The subjective results are depicted in Figure 4.

For case 1, the CT image highlights bone information and the MRI image reflects soft tissue information. One can see that the GFF (Li et al., 2013) and CBF (Kumar, 2015) methods performs well in retaining details, but poorly in retaining essential information. The contrast of IFCNN (Zhang et al., 2020), MEF-GAN (Xu et al., 2020a) and U2Fusion (Xu et al., 2022) is lower in visualisation. The fusion images of FusionGAN (Ma et al., 2019) and DDcGAN (Ma et al., 2020) are not high, and they are indistinct. By comparison, FusionDN (Xu et al., 2020b), PerceptualFusion (Fu et al., 2021), and the proposed DBGAN can maintain the bone structures of CT image as well as retain soft tissues of MRI image in red box. For case 2, the fused images of the methods [e.g., GFF (Li et al., 2013), CBF (Kumar, 2015), IFCNN (Zhang et al., 2020), MEF-GAN (Xu et al., 2020a), U2Fusion (Xu et al., 2022) and DBGAN] maintain clear details and the methods [e.g., DDcGAN (Ma et al., 2020), FusionDN (Xu et al., 2020b), PerceptualFusion (Fu et al., 2021) and DBGAN] preserve appropriate brightness. For case 3, GFF (Li et al., 2013), CBF (Kumar, 2015), FusionGAN (Ma et al., 2019), IFCNN (Zhang et al., 2020) and U2Fusion (Xu et al., 2022) are incapable of preserving supplementary information of all regions. By comparison, DDcGAN (Ma et al., 2020), FusionDN (Xu et al., 2020b), and DBGAN can conserve the bone and tissue information of source images, as shown in red box. For case 4, CBF (Kumar, 2015), CBF (Kumar, 2015), DDcGAN (Ma et al., 2020), FusionDN (Xu et al., 2020b), U2Fusion (Xu et al., 2022) and DBGAN preserve more illumination information and texture preservation. Overall, the DBGAN is competitive with other competitors in terms of subjective maps.

4.3 Objective evaluation

Four evaluation metrics [structural similarity index measure (SSIM) (Wang et al., 2004), peak signal-to-noise ratio (PSNR) (Wang and Li, 2010), mutual information (MI) (Qu et al., 2002) and visual information fidelity (VIF) (Han et al., 2013)] are adopted for objective comparison. The objective results are illustrated in Table 1. For each of these four metrics, the higher value indicates a better performance.

For case 1, the proposed DBGAN scores 0.6669, 18.7174, 17.4492 and 0.6669 in SSIM, PSNR, MI and VIF, respectively, all ranking the best place compared with other SOTA methods. Specifically, compared with DDcGAN (Ma et al., 2020) which adopts the GAN-based methods, the proposed DBGAN improves the SSIM by 134.4%, PSNR by 18.1%, MI by 17.5%, and VIF by 142.4%, respectively. Likewise, the DBGAN outperforms the other competitors in case 2. Especially, it improves the SSIM, PSNR, MI and VIF by 13.2%, 60.6%, 7.7%, and 41.6% compared with MEF-GAN (Xu et al., 2020a), which also adopts attention module. For cases 3 and 4, the proposed method ranks the first place in four evaluation indicators.

Table 1 Objective evaluation of the proposed method and the competitors in terms of SSIM, PSNR, MI and VIF

Method	*SSIM*				*PSNR*				*MI*				*VIF*			
	Case 1	*Case 2*	*Case 3*	*Case 4*	*Case 1*	*Case 2*	*Case 3*	*Case 4*	*Case 1*	*Case 2*	*Case 3*	*Case 4*	*Case 1*	*Case 2*	*Case 3*	*Case 4*
GFF	0.6374	0.7089	0.6845	0.6863	16.2305	15.9410	15.6453	16.1861	2.8087	2.7335	3.0338	2.9523	0.6374	0.6221	0.5472	0.5811
CBF	0.5401	0.7014	0.7097	0.6833	15.4376	14.8625	14.9885	15.2243	2.9698	2.7443	3.2341	3.2206	0.5401	0.4652	0.5103	0.4889
FusionGAN	0.4820	0.1782	0.1777	0.2418	15.0647	13.6398	13.7107	13.9243	2.8935	2.7466	3.1010	3.1129	0.4820	0.4517	0.4326	0.4990
IFCNN	0.5532	0.6850	0.6718	0.6780	12.2939	11.6880	13.5615	11.2934	2.9795	2.8248	3.1750	3.0655	0.5532	0.5271	0.5246	0.5668
DDcGAN	0.2845	0.1418	0.1530	0.1749	15.8520	15.7070	15.0202	15.7300	2.6399	2.5835	2.9022	2.9324	0.2845	0.2751	0.3147	0.2890
FusionDN	0.4795	0.2000	0.2160	0.2391	10.0455	10.4151	11.2482	12.4415	2.8352	2.6537	3.0886	2.9421	0.4795	0.4092	0.4399	0.4215
MEF-GAN	0.4726	0.6550	0.6333	0.6354	10.6287	10.0367	10.5251	11.2310	2.9247	2.8217	3.0830	3.0359	0.4726	0.4572	0.4881	0.4827
PerceptualFusion	0.5234	0.2183	0.2299	0.2583	15.3933	14.8140	13.0853	13.3388	3.0376	2.7561	3.1026	3.0096	0.5234	0.4480	0.5132	0.5019
U2Fusion	0.4747	0.2495	0.2692	0.3194	12.5682	12.9201	12.2827	12.9051	3.0094	2.7753	3.0895	2.9340	0.4747	0.4489	0.4381	0.4688
Ours	*0.6669*	*0.7417*	*0.7297*	*0.7219*	*18.7174*	*16.1211*	*17.1603*	*17.4492*	*3.1020*	*3.0578*	*3.4178*	*3.3341*	*0.6669*	*0.6472*	*0.6469*	*0.6599*

Note: the best results are highlighted in ital.

5 Conclusions

In this paper, we propose a DBGAN for CT and MRI image fusion. A pair of generators generate real-like fused images to fool a pair of discriminators, while the discriminators are designed to discriminate between structural differences between the fused image and the source image. In addition, the MEM and SAM are designed in generators to enhance the salient features and detail information of the fused images. Subjective evaluation results indicate that DBGAN can retain the bone structures of CT image and soft tissues of MRI image. Meanwhile, objective results prove that it outperforms other state-of-the-art methods in terms of SSIM, PSNR, MI and VIF compared with the state-of-the-art methods.

References

Du, J., Li, W., Lu, K. and Xiao, B. (2016) 'An overview of multi-modal medical image fusion', *Neurocomputing*, Vol. 215, pp.3–20.

Fu, Y., Wu, X-J. and Durrani, T. (2021) 'Image fusion based on generative adversarial network consistent with perception', *Information Fusion*, Vol. 72, pp.110–125.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) 'Generative adversarial nets', *Advances in Neural Information Processing Systems*, Vol. 63, No. 11, pp.139–144.

Han, Y., Cai, Y., Cao, Y. and Xu, X. (2013) 'A new image fusion performance metric based on visual information fidelity', *Information Fusion*, Vol. 14, No. 2, pp.127–135.

Kumar, B.S. (2015) 'Image fusion based on pixel significance using cross bilateral filter', *Signal, Image and Video Processing*, Vol. 9, No. 5, pp.1193–1204.

Li, H. and Wu, X-J. (2018) 'Densefuse: a fusion approach to infrared and visible images', *IEEE Transactions on Image Processing*, Vol. 28, No. 5, pp.2614–2623.

Li, S., Kang, X. and Hu, J. (2013) 'Image fusion with guided filtering', *IEEE Transactions on Image Processing*, Vol. 22, No. 7, pp.2864–2875.

Li, W., Xie, Y., Zhou, H., Han, Y. and Zhan, K. (2018) 'Structure-aware image fusion', *Optik*, Vol. 172, pp.1–11.

Li, Q., Lu, L., Li, Z., Wu, W., Liu, Z., Jeon, G. and Yang, X. (2019) 'Coupled GAN with relativistic discriminators for infrared and visible images fusion', *IEEE Sensors Journal*, Vol. 21, No. 6, pp.7458–7467.

Liu, Y., Chen, X., Peng, H. and Wang, Z. (2017) 'Multi-focus image fusion with a deep convolutional neural network', *Information Fusion*, Vol. 36, pp.191–207.

Ma, J., Yu, W., Liang, P., Li, C. and Jiang, J. (2019) 'FusionGAN: a generative adversarial network for infrared and visible image fusion', *Information Fusion*, Vol. 48, pp.11–26.

Ma, J., Xu, H., Jiang, J., Mei, X. and Zhang, X-P. (2020) 'DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion', *IEEE Transactions on Image Processing*, Vol. 29, pp.4980–4995.

Maas, A.L., Hannun, A.Y., Ng, A.Y. et al. (2013) 'Rectifier nonlinearities improve neural network acoustic models', in *Proceedings of the International Conference on International Conference on Machine Learning (ICML)*, Citeseer, Vol. 30, p.3.

Qu, G., Zhang, D. and Yan, P. (2002) 'Information measure for performance of image fusion', *Electronics Letters*, Vol. 38, No. 7, pp.313–315.

Tawfik, N., Elnemr, H.A., Fakhr, M., Dessouky, M.I., El-Samie, A. and Fathi, E. (2021) 'Survey study of multimodality medical image fusion methods', *Multimedia Tools and Applications*, Vol. 80, No. 4, pp.6369–6396.

Wang, Z. and Li, Q. (2010) 'Information content weighting for perceptual image quality assessment', *IEEE Transactions on Image Processing*, Vol. 20, No. 5, pp.1185–1198.

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) 'Image quality assessment: from error visibility to structural similarity', *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp.600–612.

Xu, H., Ma, J. and Zhang, X-P. (2020a) 'MEF-GAN: multi-exposure image fusion via generative adversarial networks', *IEEE Transactions on Image Processing*, Vol. 29, pp.7203–7216.

Xu, H., Ma, J., Le, Z., Jiang, J. and Guo, X. (2020b) 'FusionDN: a unified densely connected network for image fusion', in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, pp.12484–12491.

Xu, H., Ma, J., Jiang, J., Guo, X. and Ling, H. (2022) 'U2Fusion: a unified unsupervised image fusion network', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 1, pp.502–518.

Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X. and Zhang, L. (2020) 'IFCNN: a general image fusion framework based on convolutional neural network', *Information Fusion*, Vol. 54, pp.99–118.