Network Modeling Analysis in Health Informatics and Bioinformatics (2024) 13:13
https://doi.org/10.1007/s13721-024-00449-2

ORIGINAL ARTICLE

CT and MRI image fusion via multimodal feature interaction network

Wenhao Song[1] · Xiangqin Zeng[2] · Qilei Li[3] · Mingliang Gao[1] · Hui Zhou[1] · Junzhi Shi[1]

Received: 23 December 2023 / Revised: 26 January 2024 / Accepted: 24 February 2024

Abstract
Computed tomography (CT) and magnetic resonance imaging (MRI) image fusion is a popular technique for integrating information from two different modalities of medical images. This technique can improve image quality and diagnostic efficacy. To effectively extract and balance complementary information in the source images, we propose an end-to-end multimodal feature interaction network (MFINet) to fuse CT and MRI images. The MIFNet consists of a shallow feature extractor, a feature interaction (FI), and an image reconstruction. In the FI, we design a deep feature extraction module, which consists of a series of gated feature enhancement units (GFEUs) and convolutional layers. To extract key features from images, we introduce a gated normalization block in the GFEU, which can achieve feature selection. Comprehensive experiments demonstrate that the proposed end-to-end fusion network outperforms existing state-of-the-art methods in both qualitative and quantitative assessments.

Keywords Image fusion · CT/MRI image · Gated mechanism · Healthcare

1 Introduction

Medical imaging can provide rich information about human tissues and structures. Therefore, medical imaging is widely used in diagnosis, treatment planning, and surgical navigation (Azam et al. 2022). Computed tomography (CT) and magnetic resonance imaging (MRI) are two widely used imaging modalities in medical diagnosis. CT images can capture the dense structures of the human body, such as bones and organs, with high spatial resolution and contrast. MRI images can reveal soft tissue information, such as brain tissues and tumors, with high contrast and sensitivity (Li et al. 2023). However, single-modality medical images have the limitation of insufficient information, which cannot meet the needs of medical diagnosis (Zhang et al. 2023). To address this limitation, image fusion technology integrates the complementary information from both modalities and enhances the image quality (Jian et al. 2020).

In recent years, the research on image fusion has made significant progress, with the emergence of various fusion methods. The different fusion methods broadly categorized into traditional methods (Huang et al. 2018; Faragallah et al. 2022; Anu and Khanaa 2023) and deep learning-based methods (Song et al. 2023; Zhai et al. 2023; Gao et al. 2023). Traditional image fusion methods extract features from source images using different decomposition methods. Then, the decomposed images are fused according to manually designed fusion rules, and the fusion results are finally generated. In most deep learning-based methods, features of different modality images are extracted by neural networks, and the fusion image is directly generated by the end-to-end model, which avoids the need for manual design of fusion rules. Nonetheless, the absence of ground truth presents a significant challenge in training an end-to-end model to efficiently extract and combine the complementary features of the source images.

To address this problem, we proposed a modal interaction network (MFINet) to fuse the CT and MRI images. The MFINet is composed of a shallow feature extractor (SFE), a feature interactor (FI), and an image reconstructor (IR). As for FI, we designed a deep feature extraction module (DFEM), which is constructed by a series of gated

✉ Mingliang Gao
mlgao@sdut.edu.cn

✉ Junzhi Shi
shijz@sdut.edu.cn

[1] School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China

[2] Zibo Central Hospital, Zibo 255020, China

[3] School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK

Published online: 27 March 2024

feature enhancement units (GFEU). To efficaciously extract crucial features from images, a Gated Normalization Block (GNB) is introduced in the GFEU that allows for feature selection. The contributions of this work can be summarized as follows,

- A multimodal feature interaction fusion network is introduced for CT and MRI image fusion. The framework achieves end-to-end image fusion by fully extracting the complementary information of the two modalities.
- We design a feature extractor to extract and interact features from two modalities. The feature extractor consists of deep feature extraction and channel attention modules.
- Extensive experimental results demonstrate that the proposed method outperforms state-of-the-art (SOTA) methods in both qualitative and quantitative evaluations.

The remainder of this paper is structured as follows: Sect. 2 introduces the prior work. Section 3 provides details of MIFNet. Section 4 presents the experimental results and analysis. Section 5 concludes the paper.

2 Related work

Medical image fusion has gained widespread attention due to its practicality. Based on the fusion approaches, image fusion methods can be categorized into two main types, namely traditional methods and deep learning-based methods (Haribabu et al. 2023).

2.1 Traditional medical image fusion methods

Traditional methods typically use different mathematical transformations to decompose images to extract image features, e.g., Curvelet Transform (CVT) (Ali et al. 2010), Non-Subsampled Contourlet Transform (NSCT) (Zhu et al. 2019), Non-Subsampled Shearlet Transform (NSST) (Ganasala and Prasad 2018), and Daubechies complex Wavelet Transform (Singh and Khare 2014). Then, the features are fused using manually designed fusion strategies. Finally, the image is reconstructed using an inverse transformation. For example, Bhavana and Krishnappa (2015) designed a fusion method based on the Discrete Wavelet Transform (DWT) for combining medical images. Du et al. (2016) introduced a technique utilizing a union Laplacian pyramid and multiple features to fuse salient details from source images with enhanced accuracy. Maqsood and Javed (2020) introduced a multi-modal medical image fusion method. It utilizes two-scale decomposition and sparse representation to enhance detail visibility and improve clinical diagnosis accuracy.

2.2 Deep learning-based medical image fusion method

Most deep learning-based methods employ CNNs to learn image fusion strategies. These methods exhibit enhanced learning abilities and can automatically learn the optimal fusion strategy from data. For instance, Xu et al. (2020b) introduced FusionDN, an unsupervised and unified densely connected network. This network fuses source images utilizing data-driven weights, which are reflective of their feature quality and information content. Ma et al. (2020) developed a dual-discriminator architecture, termed DDcGAN. It fuses source images by simultaneously preserving thermal signatures and texture details. Xu et al. (2020a) presented U2Fusion, a unsupervised end-to-end image fusion network. U2Fusion employs feature extraction and information measurement techniques to assess the significance of source images, thereby producing a composite image that maintains adaptive similarity with the source images.

3 Proposed method

3.1 Overview

The overall framework of MFINet is shown in Fig. 1. Given CT $I_{ct} \in \mathbb{R}^{1\times H\times W}$ and MRI $I_{mri} \in \mathbb{R}^{1\times H\times W}$ images as input. These images are independently fed into the SFE based on the Restormer block (Zamir et al. 2022) for mapping the images to a feature space and extracting shallow information. Subsequently, the shallow information is input to the FI, which extracts deep semantic features through the interaction of three pairs of DFEMs. The final stage involves the IR module, which processes the outputs of the feature extractor to produce the final fused image.

3.2 Network architecture

The architecture of the deep feature extraction module (DFEM) is depicted in Fig. 2. This module adopts a residual dense connection structure to enhance feature reuse and gradient flow, thereby improving the capacity of the model for feature representation. In addition, to enhance the information transmission between the DFEMs, we adopt a channel attention mechanism. This technique assigns different weights to different feature channels based on their importance. It enables the feature extractor to focus on the most informative and salient features from the source images and suppress the irrelevant or

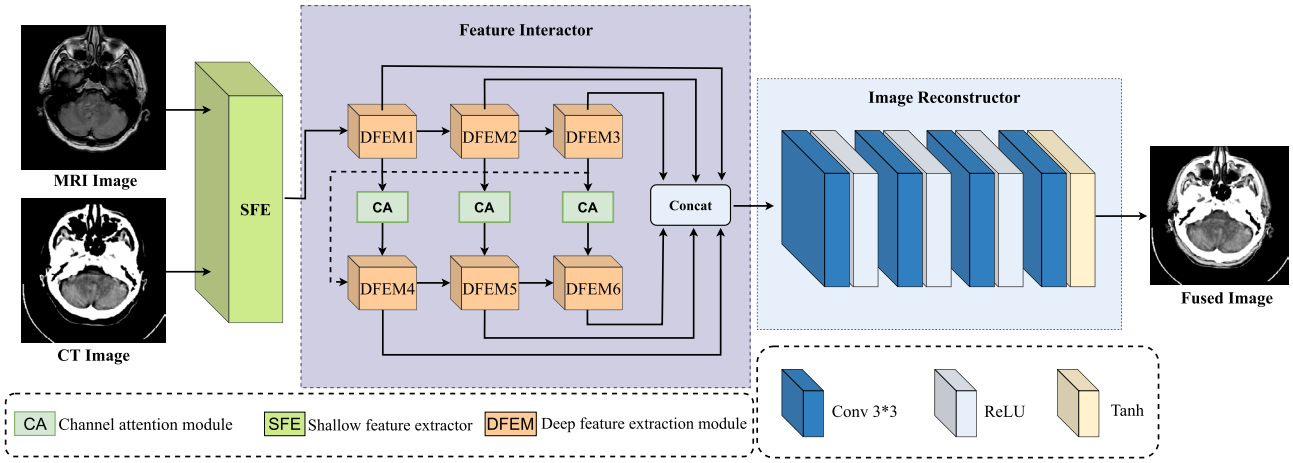

Fig. 1 The framework of MFINet for CT and MRI image fusion

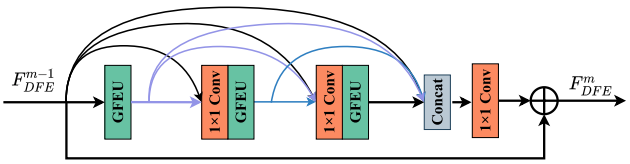

Fig. 2 Architecture of the deep feature extraction module (DFEM). The GFEU means the gated feature enhancement unit

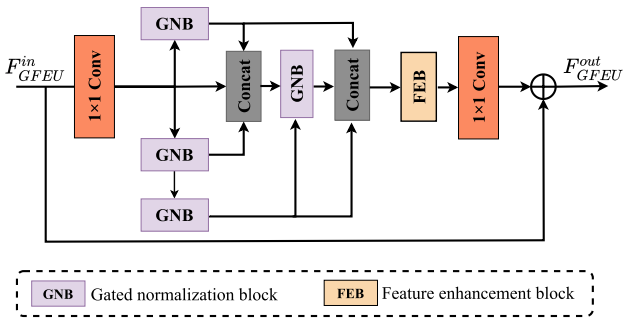

Fig. 3 Architecture of the gated feature enhancement unit (GFEU)

redundant ones. Channel attention also enhances the feature representation ability of the feature extractor, thereby improving the fusion performance. The DEF module comprises three 1×1 convolution layers and three GFEUs.

In Fig. 3, the GFEU is depicted as consisting of three parts, namely gated normalization blocks (GNBs), a feature enhancement block (FEB) based on convolutional block attention module (Woo et al. 2018), and 1×1 convolution layers. The gated normalization block is used to extract important features. The feature enhancement block is used to refine and enhance these features across both channel and spatial dimensions. Specifically, the GFEU adopts a multi-branch structure, with each branch employing a different

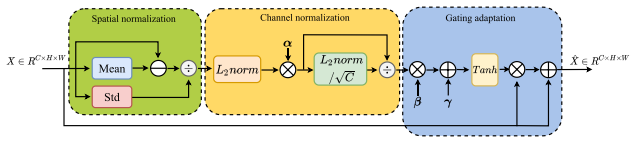

Fig. 4 Architecture of the gated normalization block (GNB)

number of GNBs to extract features at different levels. Therefore, the GFEU can selectively enhance or reduce features at various levels. This allows the GFEU to balance the features of source images. Then, the feature enhancement block refines and enhances these features separately in both the channel and spatial dimensions.

To further improve the generalization ability of the network and the efficiency of feature utilization, we introduce GNB in the GFEU. The comprehensive framework of GNB is depicted in Fig. 4. The normalization operation of GNB is non-parametric. To achieve the trainability of GNB, three learnable parameters α, β, and γ are introduced in channel normalization and gate adaptation to adjust the weights of each channel. Finally, the gate adaptation operator is used to adjust the input feature channels according to the normalization output.

3.3 Loss function

The total loss function comprises three loss functions i.e., pixel loss $\mathcal{L}_{pix}$, gradient loss $\mathcal{L}_{gra}$, and structural loss $\mathcal{L}_{ssim}$. The total loss can be formulated as,

$$\mathcal{L} = \lambda_1 \mathcal{L}_{ssim} + \lambda_2 \mathcal{L}_{gra} + \lambda_3 \mathcal{L}_{pix}, \tag{1}$$

where λ_1, λ_2, and λ_3 serve as weights to regulate the relative significance of each loss term.

The pixel loss function is critical for preserving information fidelity in image fusion. It enforces a close match between the intensity distributions of the fused image and the source images. This strategy effectively retains the rich dense information in the CT image while preserving the soft tissue information in the MRI image. The pixel loss function is defined as,

$$\mathcal{L}_{\text{pix}} = \frac{1}{HW}\left(\left\|I_f - I_{ct}\right\|_F^2 + \left\|I_f - I_{mri}\right\|_F^2\right), \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of the matrix. I_f denotes the fused image. H and W are the height and width of the image, respectively.

The gradient loss is a commonly used loss function in image fusion. Its purpose is to constrain the fused image to retain the important detail information of the source images.

$$\mathcal{L}_{\text{gra}} = \|\nabla I_f - \max\left\{\nabla I_{ct}, \nabla I_{mri}\right\}\|_2, \tag{3}$$

where $\|\cdot\|_2$ is the ℓ_2-norm of the matrix. ∇ represents the gradient operation, and $max\{,\}$ denotes maximum selection.

The structural loss constrains the fused image in brightness, contrast, and structure by introducing the structural similarity index measurement (SSIM) (Wang et al. 2004). This constraint mechanism guarantees that the fused image maintains structural similarities with the source images. The structural loss is formulated as,

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}\left(I_f, \max\left\{I_{ct}, I_{mri}\right\}\right), \tag{4}$$

where $SSIM(\cdot)$ means the structural similarity measurement.

4 Experimental results and analysis

4.1 Datasets and implementation details

In this work, a total of 184 pairs of CT and MRI images were obtained from the Harvard Medical School Whole Brain Atlas database.[1] These image pairs are from different patients with various brain diseases, and they cover different regions and angles of the brain. These images were partitioned into training (160 pairs) and test (24 pairs) sets randomly. We resized the image size to 256×256.

The proposed model was optimized using the Adam optimizer with an initial learning rate of 0.0002, decaying by 5% every 5 epochs for a total of 100 epochs. The weight factors λ_1, λ_2, and λ_3 were specified as 10, 100, and 1, respectively. The experiments were executed using the PyTorch framework on an NVIDIA GeForce RTX 3090 Ti GPU.

4.2 Compared methods and quantitative evaluation metrics

To assess the efficacy of the proposed method, we conducted performance comparisons with nine SOTA approaches, i.e., CSF (Xu et al. 2021), DensFuse (Li and Wu 2019), FusionGAN (Ma et al. 2019), PMGI (Zhang et al. 2020), RFN-Nest (Li et al. 2021), SDNet (Zhang and Ma 2021), STDFusionNet (Ma et al. 2021), U2Fusion (Xu et al. 2020a) and UMF-CMGR (Di et al. 2022).

We employed four metrics to quantitatively assess the performance of the method, namely mutual information (MI), spatial frequency (SF), visual information fidelity (VIF), and Q_{abf}. The MI metric evaluates the information transfer from the source image to the fused image by measuring their correlation. The SF captures the variations in different scales and frequencies in the fused image. It reflects the sharpness, clarity, and fine details of the fused image. The VIF quantifies the fidelity of the fused image based on human visual perception. The Q_{abf} estimates the amount of edge information transferred from the source image to the fused image, indicating the integration of edge information in the fused image. Higher values for all four metrics indicate better model performance.

4.3 Qualitative evaluation

Figure 5 displays the qualitative fusion results of two typical CT and MRI image pairs on different methods. Notably, STDFusionNet and the proposed method both preserve the dense structures, while the dense structures in other methods are greatly weakened. Furthermore, the results obtained through STDFusionNet exhibit a noticeable blurring of soft tissue details derived from the MRI image. Thus, the proposed method outperforms other methods by preserving the dense structures from the CT images and enhancing the detailed information from the MRI. Figure 5 presents two examples in the first and second rows. In the first row, our method preserves more details from the MRI image even when the CT image has less information. In the second row, our method retains both the dense structures and the texture information from the source images. Other methods reduce the information from the CT image and blur the information from the MRI image.

4.4 Quantitative evaluation

We conducted quantitative experiments on 24 pairs of CT and MRI images to demonstrate the effectiveness of the proposed method. Table 1 shows that our method outperformed

[1] [Online]. Available online: http://www.med.harvard.edu/aanlib/home.html.

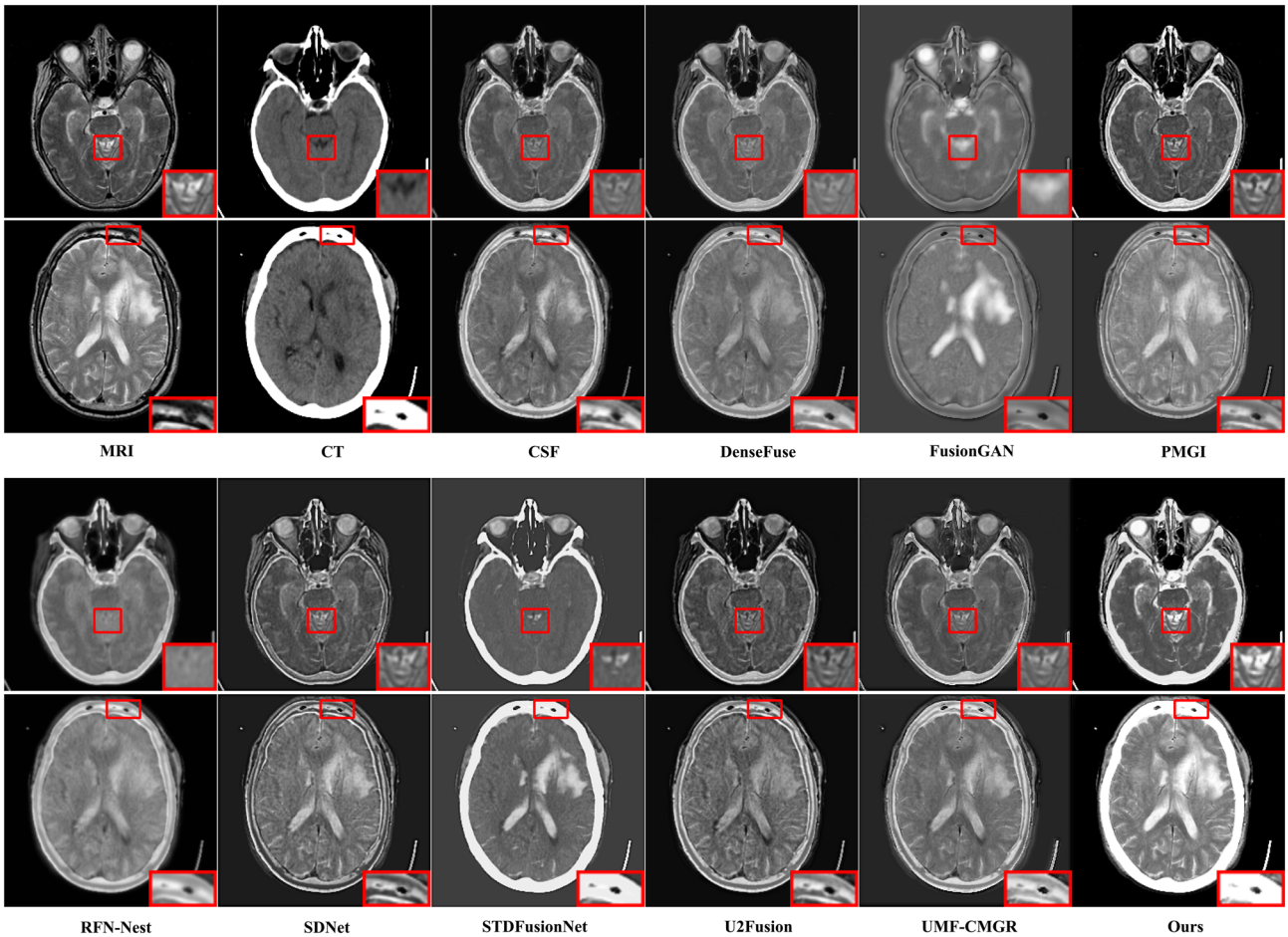

Fig. 5 Qualitative comparison of the proposed method with nine SOTA methods on two typical image pairs from the CT and MRI image pairs

Table 1 Quantitative comparison results of the MIFNet with nine SOTA methods on the CT and MRI image fusion

Method	MI↑	SF↑	VIF↑	Q_{abf} ↑
CSF (Xu et al. 2021)	2.732	18.314	0.368	0.345
DenseFuse (Li and Wu 2019)	3.362	18.526	0.404	0.329
FusionGAN (Ma et al. 2019)	2.363	12.297	0.227	0.116
PMGI (Zhang et al. 2020)	2.656	17.577	0.389	0.304
RFN-Nest (Li et al. 2021)	2.605	12.406	0.330	0.209
SDNet (Zhang and Ma 2021)	2.562	26.746	0.357	0.480
STDFusionNet (Ma et al. 2021)	3.254	25.969	0.479	0.458
U2Fusion (Xu et al. 2020a)	2.585	23.313	0.337	0.458
UMF-CMGR (Di et al. 2022)	2.690	28.893	0.353	0.431
Ours	**4.600**	**29.536**	**0.544**	**0.536**

The best results are highlighted in bold

other methods on four evaluation metrics i.e., MI, SF, VIF, and Q_{abf}. The highest MI suggests that our method transferred the information of the source images to the fusion image effectively. The highest SF and VIF imply that our fusion image was clear, detailed, and visually pleasing. The highest Q_{abf} means that our method preserved more edge detail information in the fusion results.

5 Conclusion

We propose an end-to-end CT and MRI image fusion network, termed MFINet. MFINet maps the source images to the feature space using an SFE module. Then, the FI consisting of three pairs of DFEMs is employed to extract semantic features. Finally, the MFINet generates the fused image employing the IR. Furthermore, the DFEM contains a GFEU that enhances the prominent features and detailed information of the source image. The GFEU adopts the GNBs to highlight important features and suppress useless features. Extensive experiments show that the MFINet surpasses other SOTA methods in both subjective and objective evaluations.

Acknowledgements This work is supported by the National Natural Science Foundation of China (no. 62101310).

Data Availability The data used in this work is available at http://www.med.harvard.edu/aanlib/home.html.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

Ali FE, El-Dokany I, Saad A, Abd El-Samie F (2010) A curvelet transform approach for the fusion of mr and ct images. J Mod Opt 57(4):273–286

Anu PS, Khanaa V (2023) Multimodality brain tumor image fusion using wavelet and contourlet transformation. In: Joseph, F.J.J., Balas, V.E., Rajest, S.S., Regin, R. (eds) Computational intelligence for clinical diagnosis. Springer, pp 201–214

Azam MA, Khan KB, Salahuddin S, Rehman E, Khan SA, Khan MA, Kadry S, Gandomi AH (2022) A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. Comput Biol Med 144:105253

Bhavana V, Krishnappa H (2015) Multi-modality medical image fusion using discrete wavelet transform. Procedia Comput Sci 70:625–631

Di W, Jinyuan L, Xin F, Liu Risheng (2022) Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. In: International joint conference on artificial intelligence (IJCAI) 3508-3515

Du J, Li W, Xiao B, Nawaz Q (2016) Union Laplacian pyramid with multiple features for medical image fusion. Neurocomputing 194:326–339

Faragallah OS, El-Hoseny H, El-Shafai W, El-Rahman WA, El-sayed HS, El-Rabaie ES, El-Samie FA, Mahmoud KR, Geweid GG (2022) Optimized multimodal medical image fusion framework using multi-scale geometric and multi-resolution geometric analysis. Multimed Tools Appl 81(10):14379–14401

Ganasala P, Prasad A (2018) Medical image fusion based on frei-chen masks in nsst domain. In: 2018 5th international conference on signal processing and integrated networks (SPIN). IEEE, pp 619–623

Gao M, Zhou Y, Zhai W, Zeng S, Li Q (2023) Saregan: a salient regional generative adversarial network for visible and infrared image fusion. Multimed Tools Appl 1–13

Haribabu M, Guruviah V, Yogarajah P (2023) Recent advancements in multimodal medical image fusion techniques for better diagnosis: an overview. Curr Med Imaging 19(7):673–694

Huang Y, Li W, Gao M, Liu Z (2018) Algebraic multi-grid based multi-focus image fusion using watershed algorithm. IEEE Access 6:47082–47091. https://doi.org/10.1109/ACCESS.2018.2866867

Jian L, Yang X, Liu Z, Jeon G, Gao M, Chisholm D (2020) Sedrfuse: a symmetric encoder-decoder with residual block network for infrared and visible image fusion. IEEE Trans Instrum Meas 70:1–15

Li H, Wu XJ (2019) Densefuse: a fusion approach to infrared and visible images. IEEE Trans Image Process 28(5):2614–2623

Li H, Wu XJ, Kittler J (2021) Rfn-nest: an end-to-end residual fusion network for infrared and visible images. Inf Fusion 73:72–86

Li W, Zhang Y, Wang G, Huang Y, Li R (2023) Dfenet: a dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion. Biomed Signal Process Control 80:104402

Ma J, Yu W, Liang P, Li C, Jiang J (2019) Fusiongan: a generative adversarial network for infrared and visible image fusion. Inf Fusion 48:11–26

Ma J, Xu H, Jiang J, Mei X, Zhang XP (2020) Ddcgan: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Trans Image Process 29:4980–4995

Ma J, Tang L, Xu M, Zhang H, Xiao G (2021) Stdfusionnet: an infrared and visible image fusion network based on salient target detection. IEEE Trans Instrum Meas 70:1–13

Maqsood S, Javed U (2020) Multi-modal medical image fusion based on two-scale image decomposition and sparse representation. Biomed Signal Process Control 57:101810

Singh R, Khare A (2014) Fusion of multimodal medical images using daubechies complex wavelet transform-a multiresolution approach. Inf Fusion 19:49–60

Song W, Zhai W, Gao M, Li Q, Chehri A, Jeon G (2023) Multiscale aggregation and illumination-aware attention network for infrared and visible image fusion. Concurr Comput Pract Exp e7712

Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

Woo S, Park J, Lee JY, Kweon IS (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), Springer, Cham, pp. 3–19

Xu H, Ma J, Jiang J, Guo X, Ling H (2020a) U2fusion: a unified unsupervised image fusion network. IEEE Trans Pattern Anal Mach Intell 44:502–518

Xu H, Ma J, Le Z, Jiang J, Guo X (2020b) Fusiondn: a unified densely connected network for image fusion. Proceedings of the AAAI conference on artificial intelligence, New York, USA. vol 34, pp 12484–12491

Xu H, Zhang H, Ma J (2021) Classification saliency-based rule for visible and infrared image fusion. IEEE Trans Comput Imaging 7:824–836

Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH (2022) Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, New Orleans, Louisiana, USA. pp 5728–5739

Zhai W, Song W, Chen J, Zhang G, Li Q, Gao M (2023) Ct and mri image fusion via dual-branch gan. Int J Biomed Eng Technol 42(1):52–63

Zhang H, Ma J (2021) Sdnet: a versatile squeeze-and-decomposition network for real-time image fusion. Int J Comput Vis 129:1–25

Zhang H, Xu H, Xiao Y, Guo X, Ma J (2020) Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity. In: Proceedings of the AAAI conference on artificial intelligence, New York, USA. pp 12797–12804

Zhang G, Nie R, Cao J, Chen L, Zhu Y (2023) Fdgnet: a pair feature difference guided network for multimodal medical image fusion. Biomed Signal Process Control 81:104545

Zhu Z, Zheng M, Qi G, Wang D, Xiang Y (2019) A phase congruency and local Laplacian energy based multi-modality medical image fusion method in nsct domain. IEEE Access 7:20811–20824