Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 5, No. 3, 2023

Multimodal Medical Image Fusion: The Perspective of Deep Learning

Mingyang Wei*, Mengbo Xi, Yabei Li, Minjun Liang, Ge Wang

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P R China
*Corresponding author: Mingyang Wei (wmy@home.hpu.edu.cn)

Abstract: Multimodal medical image fusion involves the integration of medical images originating from distinct modalities and captured by various sensors, with the aim to enhance image quality, minimize redundant information, and preserve specific features, ultimately leading to increased efficiency and accuracy in clinical diagnoses. In recent years, the emergence of deep learning techniques has propelled significant advancements in image fusion, addressing the limitations of conventional methods that necessitate manual design of activity level measurement and fusion rules. This paper initially presents a systematic description of the multimodal medical image fusion problem, delineating the interrelationships between different fusion modalities while summarizing their characteristics and functions. Subsequently, it reviews the theories and enhancement approaches associated with deep learning in the medical image fusion domain, striving for a comprehensive overview of the state-of-the-art developments in this field from a deep learning perspective. These developments encompass multimodal feature extraction methods based on convolutional techniques, adversarial learning-based methods, convolutional sparse representation and stacked autoencoder-based signal processing methods, and unified models. Lastly, the paper summarizes the enhancement techniques for multimodal medical image fusion methods, highlighting the pressing issues and challenges encountered by deep learning approaches in this domain.

Keywords: Medical image, Multimodal fusion, Deep learning.

1. Introduction

Image fusion entails merging two or more images from the same or various modalities, aiming to enhance image content and retain significant details. As a vital subdomain of image fusion, multimodal medical image fusion can augment image quality while maintaining specific characteristics, thereby boosting the clinical utility of images for diagnosing and evaluating medical issues. For instance, Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) deliver high-resolution anatomical information about human organs, while Positron Emission Computed Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT) offer lower spatial resolution but yield metabolic function data of organs[1]. Combining these imaging modalities allows for the acquisition of more comprehensive information on human anatomy, physiology, and pathology, thereby effectively assisting physicians in clinical diagnostics and subsequent decision-making processes for treatment decisions[2].

The deep learning approach to multimodal medical image fusion has emerged as a prominent research area in the field. In this paper, we provide a comprehensive description of the multimodal medical image fusion methodology based on deep learning. We analyze the fusion relationships and characteristics of various modal medical images, summarize the primary improvement concepts for deep learning methods in multimodal medical image fusion, and offer a thorough review of deep learning applications in this domain. Furthermore, we discuss the current state and future prospects of deep learning techniques for multimodal medical image fusion, emphasizing their potential impact and significance in advancing the field.

2. Multimodal Medical Image Fusion Based on Deep Learning

The fusion method grounded in deep learning focuses on addressing the essential challenges of feature extraction, fusion, and reconstruction in medical image fusion. This primarily encompasses methods such as convolution-based feature extraction, adversarial learning techniques, signal processing approaches (including CSR and SAE), and fusion techniques that rely on unified models. By tackling these key issues, deep learning-based fusion methods aim to improve the overall performance and applicability of medical image fusion across various modalities.

2.1. Multimodal feature extraction based on convolution technology

Both CNN and U-Net employ convolution techniques for feature extraction from source images. CNN achieves functional enhancement of activity level measurement by extracting image features through convolutional and max-pooling layers, while utilizing fully connected layers for target classification. U-Net, on the other hand, can be considered a variant of the fully convolutional neural network. Although its primary application is in image segmentation, it has demonstrated remarkable potential in multimodal fusion research[3]. In summary, by training with large amounts of data, convolutional neural networks can facilitate the automatic optimization of activity level measurement and fusion rules. This, in turn, helps circumvent the complexities and shortcomings associated with manual design in conventional image fusion processes.

2.1.1. Medical image fusion based on convolutional neural network

As depicted in Figure 1, the weight map is generated using

a trained CNN model. To achieve this, an array of multi-scale transformation methods are employed to decompose the high-frequency and low-frequency components of the source image. By applying different fusion rules for merging these decomposed components, the combined features are obtained, further enhancing the resulting image. In the final step, an inverse reconstruction process is carried out, yielding the composite fused image. In this architecture, both weight branches are identical, ensuring consistency in the feature extraction and activity level measurement approaches used for the source image. By implementing a fusion strategy grounded in local similarity, the decomposed coefficients can be adaptively adjusted, optimizing the image fusion process. This approach not only facilitates more accurate and efficient fusion but also reduces the overall difficulty in training the model, making it more practical for a broader range of applications. Consequently, this method can significantly improve the quality and usability of the fused image while minimizing computational complexity and training challenges.

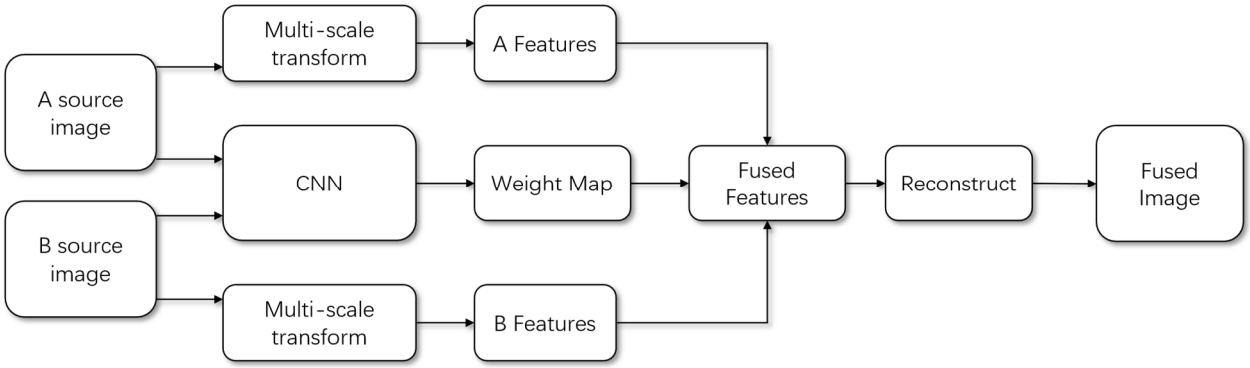

Figure 1. Medical image fusion framework based on CNN

Hermess[4] et al. proposed an MRI and CT image fusion method based on CNN. They utilized a Non-subsampled Contourlet Transform (NSST)[5] in the shear domain to decompose images into multiple sub-images. For the high-frequency sub-bands, they employed a fusion approach based on the weighted normalized cross-correlation between the feature images obtained through CNN extraction. In the case of low-frequency sub-bands, local energy was used for merging. Compared to the CNN fusion method proposed by Liu[3], the resulting fused image contained more edge information and offered a stronger visual perception effect, highlighting the impact of knowledge transfer on fusion outcomes.

Zhou[6] et al. introduced a CT and MRI image fusion algorithm that combined a Convolutional Neural Network (CNN) with a Dual-Channel Sub-band Cortex Model (DCSCM). They replaced the single Non-subsampled Contourlet Transform (NSCT)[7] with a method that combined NSST and NCSCM. This approach maintained the translation invariance of the non-downsampling process, inherited the primary characteristics of shearlet and wavelet transformations, such as anisotropy and computational speed, and resolved the issue of high computational complexity associated with NSCT, resulting in more robust weight maps.

In contrast to Hermess et al., Singh[8] implemented an adaptive biomimetic neural network model for fusing low-frequency NSST sub-bands. This model was activated by focusing on the features of source images based on weighted and modified Laplacian operators (WSML). By incorporating these various techniques, these researchers were able to develop advanced fusion methods for medical images, enhancing the quality and utility of the resulting fused images.

Liang[9] et al. introduced an end-to-end deep learning network called MCFNet, which employs convolution and deconvolution processes for MRI and CT image fusion. MCFNet integrates spatial features of source images during the up-sampling stage, eliminating the need for manually designing intricate feature extraction and fusion rules. As a result, it effectively compensates for the low-frequency content loss that occurs due to down-sampling. Wang et al.[10] proposed a Gabor representation method that combines multiple CNNs and fuzzy neural networks to address the issue of inadequate representation of complex textures and edge information of lesions in fused multimodal medical images.Li et al.[11] presented a multimodal medical image fusion approach that combines CNN and supervised learning. This method can handle various types of multimodal medical image fusion problems through batch processing, enhancing the overall efficiency and applicability of the technique.Guo et al.[12] proposed an end-to-end CNN fusion framework based on a Pulse Coupled Neural Network (PCNN). In the feature fusion module, feature maps combined with a pulse coupled neural network were utilized to minimize information loss caused by convolution in the preceding fusion module, consequently improving computational efficiency.

These novel approaches demonstrate the potential of deep learning techniques, such as convolutional neural networks, to enhance the quality and effectiveness of multimodal medical image fusion.

2.1.2. Medical image fusion based on semantics

The U-Net methodology establishes a direct link between the encoding and decoding architectures, effectively addressing the issue of semantic loss that can occur during the image fusion process. In the context of medical imaging, semantic concepts pertain to the varying interpretations of brightness levels found in medical images originating from different modalities. By incorporating the direct connections between encoding and decoding structures, the U-Net method ensures that critical semantic information is preserved throughout the fusion process. This preservation is particularly important when dealing with medical images, as the distinct modalities used for capturing these images often have unique interpretations of brightness levels. Consequently, maintaining the integrity of these semantic concepts is essential for accurate and meaningful analysis of

the fused medical images. In summary, the U-Net approach serves as a highly effective method for overcoming the challenges associated with semantic loss during image fusion. By facilitating the preservation of vital semantic concepts related to brightness levels across different modalities, U-Net can significantly improve the overall quality and interpretability of fused medical images, enabling medical professionals to make better-informed diagnostic and treatment decisions.

Han et al.[13] proposed a deep convolutional neural network model tailored for practical applications, which generates CT images from MRI images for medical diagnosis, thereby avoiding the radiation exposure associated with CT scans. This method employs the U-Net framework to directly create a dense label map for object segmentation in two-dimensional images.

This work builds upon advancements in semantic image segmentation and incorporates a crucial innovation proposed by Ronneberger et al. (the creators of U-Net): the introduction of direct connections between the encoding and decoding sections. This approach enables the high-resolution features from the first section to be used as input for additional convolutional layers in the second section. This design facilitates the generation of high-resolution predictions in the decoding section, ultimately enhancing the overall quality and usefulness of the resulting images.

By leveraging the U-Net framework and its unique encoding-decoding connection design, Han et al.'s method significantly improves the process of generating CT images from MRI data. This not only aids in more accurate medical diagnosis but also reduces the reliance on traditional CT scans, which expose patients to radiation.

2.2. Medical image fusion based on adversarial learning

The fusion framework of GAN is depicted in Figure 2. Initially, the source image is input into two generators, which produce a fused image. This fused image is then compared with the real image by the discriminator. The generator and the discriminator engage in a continuous adversarial process to further enhance the quality of the fused image. Ultimately, this process results in a fused image that closely resembles the source image.

In this GAN-based framework, the generators and discriminator work together to iteratively refine the fused image, ensuring that it retains the essential characteristics and information of the original source images. The adversarial nature of this method promotes continuous improvement, leading to high-quality, accurate fused images that can be invaluable in a variety of applications, including medical imaging and diagnosis.

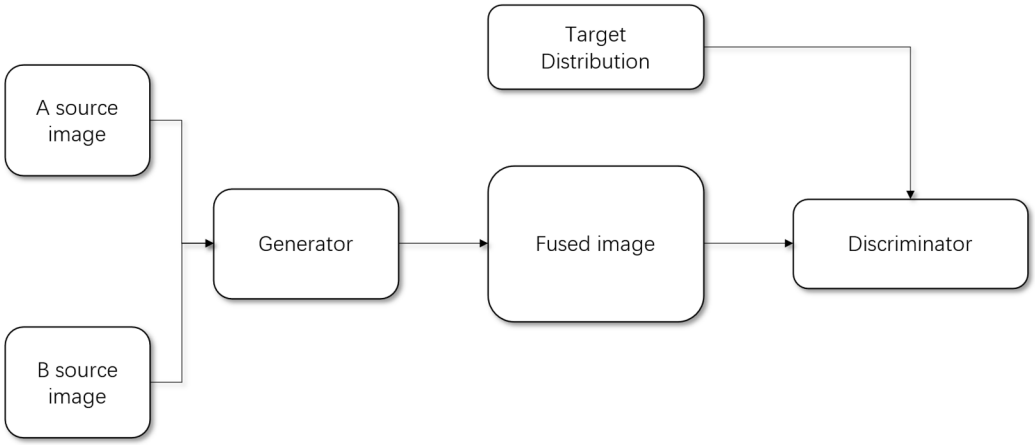

Figure 2. Medical image fusion framework based on GAN

The GAN-based medical image method improves the quality of the fused image generated by the generator by adjusting the number of discriminators. The Generative Adversarial Network (GAN) was first proposed by Goodfellow et al.[14] and later extended to the task of image analysis. GAN is a generative model consisting of two multilayer networks. The first network is a generator that creates synthetic data, while the second network is a discriminator that classifies images as either real or artificially generated.

Training through backpropagation enhances the ability of GAN to distinguish between real data and generated data. By adjusting the number of discriminators, the GAN-based medical image method can further refine the quality of the fused image produced by the generator. This process ensures that the resulting fused images retain essential characteristics and information from the original source images, ultimately leading to higher quality and more accurate fused images.

In recent years, GANs have been extensively utilized in the field of image fusion, demonstrating their potential for enhancing the quality and utility of fused images across various applications, including medical imaging and diagnosis.

Tang et al.[15] introduced GANs into the field of biological cytology and proposed the GFPPC and GAN methods to achieve the fusion of green fluorescent protein images (GFP) and phase contrast images (PC). This approach is capable of simultaneously extracting functional information from GFP images and structural information from phase images. The method demonstrates a high level of versatility, as it was also extended to MRI and PET image fusion, marking the beginning of GAN usage in medical image fusion.By incorporating GANs into medical image fusion, researchers can effectively combine valuable information from different imaging modalities, such as GFP and PC images or MRI and PET scans. This results in fused images that provide a more comprehensive view of the biological structures and functions in question. The use of GANs in medical image fusion has the potential to significantly improve the quality and utility of fused images, ultimately aiding medical professionals in

making better-informed diagnostic and treatment decisions.

In medical image fusion, Kang et al.[16] proposed a GAN-based method that uses tissue perception conditions to fuse brain PET and MRI images. Specifically, the process of fusing PET and MRI images of the brain is considered a game of confrontation between preserving color information from the PET and anatomical information from the MRI. In this approach, minimal and maximum optimization problems are modeled for generators and discriminators.

Le et al.[17] built upon Tang et al.'s work and proposed a new adversarial network, MGMDcGAN, which features multiple generators and discriminators. This method introduced a second adversarial game based on the calculated mask, further enhancing the preservation of bone density information. The MGMDcGAN method is applicable to various medical imaging modality combinations, including MRI and PET, MRI and SPECT, and CT and SPECT.By employing GAN-based methods for medical image fusion, researchers can effectively combine valuable information from different imaging modalities, leading to fused images that provide a more comprehensive understanding of the biological structures and functions being examined. The use of GANs in medical image fusion has the potential to significantly improve the quality and utility of fused images, ultimately aiding medical professionals in making better-informed diagnostic and treatment decisions.

2.3. Method based on signal processing

2.3.1. Medical image fusion based on convolutional sparse representation (CSR)

The medical image fusion method based on CSR (Convolutional Sparse Representation) aims to obtain the sparse representation of the entire image using deconvolution methods for medical image fusion in the context of signal processing. The concept of CSR is derived from the deconvolution network proposed by Zeiler et al.[18]. The fundamental idea of CSR is to achieve image convolution decomposition under sparse constraints.

The goal of deconvolution networks is to learn multi-stage feature representations of input images by constructing a decomposed hierarchy. Input images can be reconstructed from their decomposition in a hierarchical manner. As a result, deconvolution networks offer a promising image representation method for feature learning and reconstruction-based problems.

As the foundational architecture of deep deconvolution networks, CSR has been successfully applied to a wide range of visual problems, such as video background modeling[19], target detection [20], super-resolution[21], and more. Image representation methods based on signal processing involve image fusion through coding and filtering. One approach is based on convolutional sparse representation (convolutional sparse coding), while the other relies on stacked auto-encoders (SAE). These techniques can effectively represent and fuse medical images, ultimately enhancing the quality and utility of the resulting fused images.

In the field of image fusion, the sparse representation (SR) method was first proposed by Yang and Li[22] and originated from signal processing technology. In traditional SR-based image fusion methods, sparse decomposition is independently executed on a group of overlapping blocks extracted by sliding window technology. As a result, the representation is multi-valued and not optimal relative to the whole image.Liu et al.[23] applied CSR (Convolutional Sparse Representation) in multi-modal medical image fusion, decomposing the source image into a base layer and detail layer. The fundamental idea behind this approach is to model the entire source image Y as a set of convolution summations between sparse mapping X and dictionary filter D. By doing so, the CSR method can provide a more optimal representation of the source image, leading to improved fusion results.By incorporating CSR into multi-modal medical image fusion, researchers can create fused images that better capture the valuable information from different imaging modalities. This, in turn, allows medical professionals to gain a more comprehensive understanding of the biological structures and functions being examined, ultimately aiding in more accurate diagnoses and more effective treatment decisions.

2.3.2. Medical image fusion based on Stacked auto-encoder (SAE)

Image fusion based on stacked autoencoding employs stacking to train multiple basic units, addressing the issue of information loss during image decomposition. In existing medical image fusion approaches based on CNN, the feature dimension of the final output from convolution kernel subsampling is low. Thus, directly using the CNN model for image decomposition can lead to information loss and suboptimal fusion.

To tackle this problem, Xia et al.[24] comprehensively utilized the features of multi-scale transformation and deep convolutional neural networks. Laplace filters and Gaussian filters are employed to decompose the source image into sub-images, which then serve as the first layer of the network. The remaining convolution kernels are initialized based on the He method, and the back-propagation algorithm is used to train the basic units. Multiple basic units are trained using the stacked autoencoder (SAE) concept, and a deep stacked neural network (DSCNN) is created by stacking the CNNs.

This method enables adaptive decomposition and reconstruction of the fused image without the need for manual filter selection. Its performance surpasses that of NSCT and NSST methods; however, the fusion rules still require manual selection. Overall, the image fusion approach based on stacked autoencoding offers an effective solution for minimizing information loss during the fusion process and generating higher-quality fused images.

2.4. Unification based fusion method

The performance of medical image fusion methods based on deep learning surpasses that of traditional approaches. However, training these models necessitates substantial data support, and medical image data is often scarce and challenging to label. Some strategies address this issue through the use of artificially created real images, but there is no uniform standard for assessing the suitability of these artificial ground truth images.

Furthermore, the process of creating artificial images is not only time-consuming and demanding, but also limited in its effectiveness. It is typically useful only for specific image fusion tasks (such as multi-focus, multi-exposure, and multi-modal fusion) and can be difficult to generalize across different fusion tasks.

To overcome these challenges, future research may need to focus on developing more efficient labeling techniques, exploring unsupervised or semi-supervised learning methods, and designing more robust fusion models that can adapt to various tasks with minimal dependence on large-scale labeled data. Additionally, establishing uniform evaluation metrics

for artificial ground truth images could help standardize the assessment of fusion methods and enhance their applicability across a broader range of medical image fusion tasks.

The unified fusion method aims to train a versatile, unified model on various types of images (multi-exposure, multi-focus, visible light, medical images) by adjusting the loss function and incorporating flexible weights, allowing the model to learn features from different image types and apply them to diverse fusion tasks. By doing so, the fused images produced by this method exhibit higher resolution compared to those generated by specific image fusion techniques.

This approach offers several advantages, such as reducing the need for multiple specialized models for different tasks, saving time and resources in model development, and improving the overall performance of image fusion. However, it may also face challenges in finding the optimal balance between learning diverse features and retaining specific characteristics of each image type, as well as in generalizing the model to new and unseen tasks.

To further enhance the performance of unified fusion methods, researchers may explore ways to improve the adaptability of the model to various tasks, develop more effective loss functions and weighting schemes, and investigate techniques for better feature extraction and fusion from diverse image types. This could ultimately lead to more robust and versatile models for a wide range of image fusion applications.

Zhang Yu et al.[25] proposed IFCNN, a unified fusion framework based on the fully convolutional network, which posits that any image fusion task essentially involves the effective selection of information. They introduced perceptual loss to optimize the fusion model for the ground truth fusion image, which helps the model generate fused images with more detailed textures. This approach is applicable to various fusion tasks. By incorporating perceptual loss, the IFCNN framework focuses on optimizing the model to produce visually appealing results that retain important texture details and information from the source images. This method has the potential to be more versatile and efficient than task-specific fusion methods, as it can be adapted to different types of image fusion tasks with less manual intervention. Moreover, the IFCNN framework can be further improved by exploring other optimization techniques, incorporating additional loss functions, or combining it with other deep learning architectures. This could lead to even more powerful and flexible fusion models, capable of tackling a wide range of image fusion challenges in various domains.

Han Xu et al.[27] introduced a unified unsupervised dense connection network FusionDN for image fusion to overcome the issue of catastrophic forgetting. Unlike previous methods, this approach does not require real fusion images for all fusion tasks, and the model is not tailored exclusively for a specific fusion task. Instead, by utilizing elastic weight consolidation (EWC), a single model capable of handling multiple distinct fusion tasks was achieved. This method effectively prevents the model from forgetting the knowledge acquired from prior tasks while training multiple tasks in sequence.

Based on the previous work, Han Xu[28] proposed a new unified fusion method called U2Fusion. This method improved the information protection strategy to prevent the network from forgetting the features of previous tasks and incorporated the intensity-based loss function to reduce the brightness deviation in the fusion image. In addition, a new loss function was introduced to measure the similarity of the fused image with the ground truth, which further improved the quality of the fused image. The experimental results demonstrated that U2Fusion outperformed the single model approach, and achieved better information retention in the fused image. With the unified fusion method, the model can be trained on multiple image fusion tasks, providing a more flexible and efficient solution for various image fusion scenarios.

3. Improvement of Multimodal Medical Image Fusion

The first way to improve medical image fusion based on deep learning is to focus on the improvement of medical images themselves and the deep learning models. Data enhancement techniques, such as rotation, flip, scaling, and noise addition, can be used to expand the data set and improve the generalization ability of the model. DenseNet module introduction can enhance feature learning and prevent over-fitting. Transfer learning can take advantage of pre-trained models to accelerate model convergence and improve performance.

The second way to improve medical image fusion is to focus on image decomposition and fusion strategy. For example, non-subsampled contour wave transform can replace traditional wavelet and pyramid transform for more accurate image decomposition. Adaptive fusion strategies can be developed based on different image features, such as local similarity and spatial frequency. Moreover, attention mechanisms can be used to improve the model's ability to focus on important information in the source images. Overall, the improvement of medical image fusion based on deep learning requires the joint efforts of medical imaging experts, machine learning experts, and domain experts to develop more effective methods and achieve better fusion results.

Deep learning technology has achieved good results in the field of medical image fusion, but there are also many challenges, which can be summarized in the following two aspects:

(1) From the perspective of medical image itself and deep learning model. First of all, medical image data annotation requires professional medical personnel to operate, so data annotation is difficult and data sets are insufficient. Secondly, the fusion of data of different modes of the same patient can give play to the advantages of observation data of different instruments, but the extraction ability of deep learning model has higher requirements, and the lack of training labels leads to very difficult information extraction. Finally, deep learning is adopted for training. As the model deepens, the gradient will disappear or the gradient will explode, and effective features cannot be extracted.

(2) From the perspective of image decomposition, namely multi-scale transformation. The multi-scale decomposition methods commonly used in medical image fusion, such as wavelet transform and Laplacian pyramid, have deviation variance, aliasing and lack of direction.

Based on the characteristics of medical image, some improvement measures are introduced, such as introducing residual or dense link block to solve the problem of disappearing gradient or exploding gradient, using data enhancement to solve the problem of lacking medical image, and using transfer learning to solve the problem of difficult model training. From the perspective of multi-scale transformation optimization, the improvement measures are

proposed.

4. Conclusion

The multi-modal medical image fusion method based on deep learning demonstrates superior performance in extracting image features and effectively alleviates the shortcomings of traditional methods. The CNN-based approach addresses the deficiencies of activity level measurement and fusion rules in traditional methods. U-Net effectively mitigates the problem of semantic loss during CNN fusion. GAN method is able to solve the issue of manual design of fusion rules and demonstrates strong capabilities in data generation. However, as deep learning continues to evolve, multi-modal medical image fusion based on deep learning still faces the following challenges:

(1) In medical image fusion tasks, it is difficult to evaluate the quality of the fusion results due to the lack of real ground truth. Therefore, it is crucial to design reference-free measures with high representational power in this field. These measures can guide the fusion process to generate higher quality results when used in the loss function, and can also be used to fairly evaluate the fusion results to encourage further research. Distance measurement learning may be a good approach for comprehensive quality assessment.

(2) Deep learning has shown great potential in multi-modal medical image fusion, but it also faces challenges such as a single framework and limited training data. Annotating medical images for training data requires medical experts and is a time-consuming and costly process, which can lead to a lack of data and overfitting. In addition, incorporating feature information and lesion information into the dataset obtained through data augmentation can have a significant impact on the accuracy of the model. Therefore, designing an effective architecture through deep learning remains a challenge in multi-modal medical image fusion research.

(3) The current state-of-the-art medical image fusion methods have made significant progress. However, most of these methods still follow traditional problem-solving ideas and face some challenges, such as color distortion and incomplete feature information extraction in the fusion image. Therefore, there is a need for a more innovative and disruptive approach to medical image fusion algorithms to achieve even higher levels of performance.

(4) Medical image fusion of three or more modes is indeed a challenging problem. One possible solution is to use deep learning models with more complex architectures to extract and fuse features from multiple modalities. Another approach is to decompose each modality into different frequency bands and fuse the corresponding bands from different modalities. However, these methods are still in the experimental stage and require further research and development. Additionally, the evaluation of the fusion results in multi-modal medical image fusion is also a challenge, as it is difficult to establish a comprehensive evaluation standard due to the complexity and diversity of medical image data.

References

[1] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information fusion,* vol. 19, pp. 4-19, 2014.

[2] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *International conference on medical image computing and computer-assisted intervention*, 2016: Springer, pp. 212-220.

[3] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Information Fusion,* vol. 36, pp. 191-207, 2017.

[4] H. Hermessi, O. Mourali, and E. Zagrouba, "Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain," *Neural Computing and Applications,* vol. 30, no. 7, pp. 2029-2045, 2018.

[5] T.-y. Zhang, Q. Zhou, H.-j. Feng, Z.-h. Xu, Q. Li, and Y.-t. Chen, "Fusion of infrared and visible light images based on nonsubsampled shearlet transform," in *International Symposium on Photoelectronic Detection and Imaging 2013: Infrared Imaging and Applications*, 2013, vol. 8907: International Society for Optics and Photonics, p. 89071H.

[6] R. Hou, D. Zhou, R. Nie, D. Liu, and X. Ruan, "Brain CT and MRI medical image fusion using convolutional neural networks and a dual-channel spiking cortical model," *Medical & biological engineering & computing,* vol. 57, no. 4, pp. 887-900, 2019.

[7] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: theory, design, and applications," *IEEE transactions on image processing,* vol. 15, no. 10, pp. 3089-3101, 2006.

[8] S. Singh and R. S. Anand, "Multimodal neurological image fusion based on adaptive biological inspired neural model in nonsubsampled shearlet domain," *International Journal of Imaging Systems and Technology,* vol. 29, no. 1, pp. 50-64, 2019.

[9] X. Liang, P. Hu, L. Zhang, J. Sun, and G. Yin, "MCFNet: Multi-layer concatenation fusion network for medical images fusion," *IEEE Sensors Journal,* vol. 19, no. 16, pp. 7107-7119, 2019.

[10] L. Wang, J. Zhang, Y. Liu, J. Mi, and J. Zhang, "Multimodal medical image fusion based on Gabor representation combination of multi-CNN and fuzzy neural network," *IEEE Access,* vol. 9, pp. 67634-67647, 2021.

[11] Y. Li, J. Zhao, Z. Lv, and Z. Pan, "Multimodal Medical Supervised Image Fusion Method by CNN," *Frontiers in Neuroscience,* vol. 15, 2021.

[12] K. Guo, X. Li, X. Hu, J. Liu, and T. Fan, "Hahn-PCNN-CNN: an end-to-end multi-modal brain medical image fusion framework useful for clinical diagnosis," *BMC Medical Imaging,* vol. 21, no. 1, 2021.

[13] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Medical physics,* vol. 44, no. 4, pp. 1408-1419, 2017.

[14] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems,* vol. 27, 2014.

[15] W. Tang, Y. Liu, C. Zhang, J. Cheng, H. Peng, and X. Chen, "Green fluorescent protein and phase-contrast image fusion via generative adversarial networks," *Computational and Mathematical Methods in Medicine,* vol. 2019, 2019.

[16] J. Kang, W. Lu, and W. Zhang, "Fusion of brain PET and MRI images using tissue-aware conditional generative adversarial network with joint loss," *IEEE Access,* vol. 8, pp. 6368-6378, 2020.

[17] Z. Le, J. Huang, F. Fan, X. Tian, and J. Ma, "A generative adversarial network for medical image fusion," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020: IEEE, pp. 370-374.

[18] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *2010 IEEE Computer Society*

Conference on computer vision and pattern recognition, 2010: IEEE, pp. 2528-2535.

[19] B. Wohlberg, "Endogenous convolutional sparse representations for translation invariant image subspace models," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014: IEEE, pp. 2859-2863.

[20] D. Carrera, G. Boracchi, A. Foi, and B. Wohlberg, "Detecting anomalous structures by convolutional sparse models," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015: IEEE, pp. 1-8.

[21] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1823-1831.

[22] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE transactions on Instrumentation and Measurement,* vol. 59, no. 4, pp. 884-892, 2009.

[23] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE signal processing letters,* vol. 23, no. 12, pp. 1882-1886, 2016.

[24] K.-j. Xia, H.-s. Yin, and J.-q. Wang, "A novel improved deep convolutional neural network model for medical image fusion," *Cluster Computing,* vol. 22, no. 1, pp. 1515-1527, 2019.

[25] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Information Fusion,* vol. 54, pp. 99-118, 2020.

[26] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "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*, 2020, vol. 34, no. 07, pp. 12797-12804.

[27] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "Fusiondn: A unified densely connected network for image fusion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 07, pp. 12484-12491.

[28] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 44, no. 1, pp. 502-518, 2020.