Computers in Biology and Medicine 160 (2023) 106959

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Deep learning methods for medical image fusion: A review

Tao Zhou [a,d], QianRu Cheng [a,d,*], HuiLing Lu [b,**], Qi Li [a,d], XiangXiang Zhang [a,d], Shi Qiu [c]

[a] School of Computer Science and Engineering, North Minzu University, Yinchuan, 750021, China
[b] School of Science, Ningxia Medical University, Yinchuan, 750004, China
[c] Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, 710119, China
[d] Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan, 750021, China

ARTICLE INFO

Keywords:
Deep learning
Medical image fusion
Convolutional neural network
Generative adversarial network
Encoder-decoder network

ABSTRACT

The image fusion methods based on deep learning has become a research hotspot in the field of computer vision in recent years. This paper reviews these methods from five aspects: Firstly, the principle and advantages of image fusion methods based on deep learning are expounded; Secondly, the image fusion methods are summarized in two aspects: End-to-End and Non-End-to-End, according to the different tasks of deep learning in the feature processing stage, the non-end-to-end image fusion methods are divided into two categories: deep learning for decision mapping and deep learning for feature extraction. According to the different types of the networks, the end-to-end image fusion methods are divided into three categories: image fusion methods based on Convolutional Neural Network, Generative Adversarial Network, and Encoder-Decoder Network; Thirdly, the application of the image fusion methods based on deep learning in medical image field is summarized from two aspects: method and data set; Fourthly, evaluation metrics commonly used in the field of medical image fusion are sorted out from 14 aspects; Fifthly, the main challenges faced by the medical image fusion are discussed from two aspects: data sets and fusion methods. And the future development direction is prospected. This paper systematically summarizes the image fusion methods based on the deep learning, which has a positive guiding significance for the in-depth study of multi modal medical images.

1. Introduction

With the development of medical imaging technology, medical images are becoming more and more important in clinical diagnosis. Different modality medical images have their own advantages and limitations. For example, Computed Tomography (CT) images have high resolution and can accurately locate lesions, but it is weak for soft tissue structures; Magnetic Resonance Imaging (MRI) can clearly reflect the structure of soft tissues, but the imaging information of bones is insufficient [1]; Positron Emission Tomography (PET) and Single Photon Emission CT (SPECT) are often used to detect vascular and tumor diseases. Therefore, single modality medical images can't fully and accurately reflect the information of medical tissues, which leads to complicated diagnosis process and diagnostic errors. These inspire researchers to integrate this complementary information into a single image. The purpose of medical image fusion is to improve the utilization rate of medical images, and help doctors to understand the contents of images. The fused images contain more information and accuracy than each source medical images [2]. It can help doctors diagnose and treat diseases more accurately, quickly and comprehensively [3].

Image fusion algorithms can be divided into two categories: transform domain algorithm and spatial domain algorithm. Algorithms based on transform domain are usually based on the theory of Multi-scale Transform (MST), such as Laplace Pyramid (LP) [4], Wavelet Transform (WT) [5], Curved Wave Transform (CVT) [6] and Nonsubsampled Contourlet Transform (NSCT) [7]. The steps of these methods are as follows: firstly, the source images are decomposed into coefficients, then coefficients are fused through fusion rules, and finally, the fused image are reconstructed through the inverse of the transform. Besides MST methods, some methods based on feature space are proposed in recent years, such as Independent Component Analysis (ICA) [8] and Sparse Representation (SR) [9]. However, there are some shortcomings among these methods: fusion rule is designed by the developer, fusion images need to be registered, and the image reconstruction also results in an image-quality degradation [10]. Spatial domain-based algorithms do not need to convert the source images into another feature domain, it

* Corresponding author. School of Computer Science and Engineering, North Minzu University, Yinchuan, 750021, China.
** Corresponding author. School of Science, Ningxia Medical University, China
E-mail addresses: chengqianru5@163.com (Q. Cheng), Lu_huiling@163.com (H. Lu).

https://doi.org/10.1016/j.compbiomed.2023.106959
Received 27 October 2022; Received in revised form 12 April 2023; Accepted 17 April 2023
Available online 20 April 2023

can be divided into block-based, region-based, and pixel-based fusion algorithms. The block-based algorithms usually segment the images into blocks, measure their spatial frequencies, sum and modify Laplace [11], then fusing the image blocks, in these algorithms, the size of the image blocks has a great influence on the results and it is difficult to segment; The region-based algorithms decomposed the input images into regions according to some criteria, and then the saliency of the corresponding regions is measured, finally, the most salient regions are combined to form the fused image. but the accuracy of image segmentation has a great influence on the efficiency of the algorithm. The pixel-based algorithms generate the fused decision graph by the activity level measurement strategy directly, and some pixel-based spatial domain methods are proposed, such as Multi-scale Weighted Gradient Fusion (MWGF) [12], image fusion with Guided Filtering (GFF) [13] and dense SIFT, these methods based on a single pixel in the process of fusion, which ignoring the similarity of information. In recent years, the development of deep learning promotes the progress of image fusion, the powerful feature extraction and data expression capabilities of deep learning make the development of image fusion very promising. Deep learning methods learn the fusion model with good generalization ability from a large amount of data [14], which can make the fusion process more robust and overcomes the shortcomings of manual feature selection, such as time-consuming, expensive and prone to human errors, and show the strong development potential. Since 2012, many deep convolution neural network models are proposed, such as VGG [15], Residual neural network (Res-Net) [16], Dense neural network (Dense-Net) [17] and U-Net [18], these deep learning models provide useful theoretical support and practical experience for image fusion. The image fusion model based on deep learning has more obvious advantages compared with traditional image fusion algorithms.

(1) Stronger ability of feature extraction and expression. The deep learning model learns the feature information of the source images by training on a larger data set, and enhances the expression ability of the network for image feature information through continuous iterative process, but this also makes it more dependent on large data sets.
(2) More flexible network architecture. Image fusion methods based on deep learning can continuously adjust the quality of fused images through the training process, while traditional fusion methods can only improve the quality of fused images by manually adjusting the algorithm rules.
(3) The end-to-end fusion process. The image fusion method based on deep learning can overcomes the drawback of traditional fusion that requires manual setting fusion rules. The processes of feature extraction, feature fusion, and feature reconstruction are implicitly, which avoids information loss and produce better quality and more effective fusion images.

In this paper, the image fusion methods based on deep learning are summarized, and the existing problems and development directions are found out from the existing methods, aiming at providing clues and theoretical support for future research. The remainder of this paper is organized as follows. In section 2, the principle and advantages of image fusion based on deep learning are expounded. In section 3, the existing image fusion methods are introduced from Non-End-to-End. In section 4, the existing image fusion methods are introduced from End-to-End. In section 5, the application of different medical image fusion methods based on deep learning are summarized. Section 6 sort out 14 evaluation metrics commonly used in medical image fusion field. In section 7, the main challenges faced and the existing problems in the field of medical image fusion are discussed, and the future development directions are prospected.

2. Image fusion based on deep learning

There are three stages in the image fusion based on deep learning: feature processing stage, feature fusion stage and feature reconstruction stage. The processes are as follows: firstly, feature information or decision maps are obtained through deep learning networks; then, they are fused through the fusion strategy. finally, the fused image is obtained by feature processing inverse transformation. Due to the strong ability of deep learning networks in feature extraction and information expression, the quality of fused images can be significantly improved. In this paper, the fusion methods based on deep learning are divided into two categories: non-end-to-end image fusion methods and the end-to-end image fusion methods. In the non-end-to-end image fusion methods, the deep learning network is always applied to the feature processing stage before the fusion rules, the input of the network are the source images or image blocks, and the output are the feature information or decision maps, these processes are shown in Fig. 1 (a). Liu et al. applies CNN to the image fusion for the first time [19], and they regard the image fusion task as a classification task, CNN is used to classify image areas and generate decision maps. In the end-to-end image fusion methods, the deep learning network is applied to the whole process from source images to fused images, the input of network are the source images, and the output are the fused images, these processes are shown in Fig. 1 (b). The IFCNN proposed by Zhang et al. is an end-to-end image fusion network [20], two convolutional layers are used to extract features from the source images, and then fusion rules are used to fuse the features, the fused features are reconstructed by two convolutional layers to obtain the fused image.

In this paper, image fusion methods based on the deep learning are divided into two categories: the non-end-to-end image fusion methods and the end-to-end image fusion methods. In the non-end-to-end image fusion methods, there are two categories: deep learning for decision mapping and deep learning for feature extraction; in the end-to-end image fusion methods, there are three categories: image fusion methods based on CNN, image fusion methods based on GAN and image fusion methods based on Encoder-Decoder Network. The following will be introduced from these aspects.

3. Non-end-to-end image fusion methods

The non-end-to-end image fusion methods refer to the application of deep learning network in the feature processing stage before the fusion stage. The processes are as follows: firstly, the source images are processed through the deep learning network to obtain feature information or decision maps; then, the features are fused by the fusion rules; finally, the fused features are reconstructed to obtain the final fused image. Effective feature processing methods are the prerequisite of high-quality fusion methods, and the development of image representation theory has a great influence on the progress of image fusion, which promotes the further improvement of fusion rules. This section will introduce from two aspects: decision mapping based on deep learning and feature extraction based on deep learning.

The non-end-to-end image fusion methods are summarized in Table 1. And the datasets adopted in each method are classified, in which MFI stands for Multi-Focus Image, MEI stands for Multi-Exposure Image, IR stands for Infrared image, VIS stands for Visible Image, CT stands for Computed Tomography, PET stands for Positron Emission Tomography, SPECT stands for Single Photon Emission Computed Tomography, MRI stands for Magnetic Resonance Image, HSI stands for Hyper-spectral Image, and MSI stands for Multi-spectral Image. The abbreviation names in Table 2 are the same as in Table 1.

3.1. Decision mapping based on deep learning

The image fusion methods of deep learning for decision mapping are regarded as a classification problem, these methods are usually based on

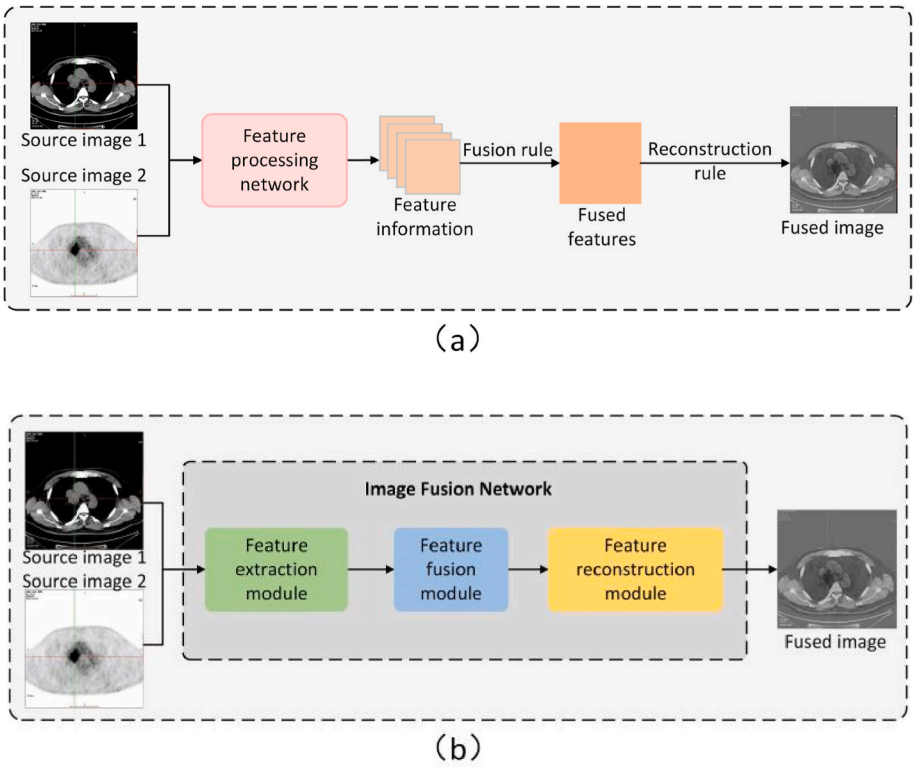

Fig. 1. Image fusion framework based on deep learning: (a) The Non-End-to-End image fusion framework; (b) The End-to-End image fusion framework.

Table 1
Non-End-to-End image fusion methods.

Methods	Year	Backbone Network	Function of Deep Learning		Domain		Supervised		Data Set
			Decision mapping	Feature extraction	Spatial	Transform	yes	no	
Liu et al. [19]	2017	CNN		✓	✓		✓		MFI
Singh et al. [28]	2020	CNN	✓			✓	✓		CT-MRI、MRI-SPECT、MRI-PET
Wu et al. [29]	2021	CNN	✓			✓	✓		MS-PAN
Zhou et al. [24]	2021	CNN	✓			✓	✓		IR-VIS
Wang et al. [30]	2021	CNN	✓			✓	✓		CT-MRI
Li et al. [25]	2019	ResNet	✓			✓	✓		IR-VIS
DRPL [10]	2020	ResNet	✓		✓		✓		MFI
Fu et al. [27]	2021	ResNet	✓			✓	✓		MRI-CT、MRI-PET、MRI-SPECT
Li et al. [31]	2021	ResNet	✓			✓	✓		MRI-PET、MRI-SPECT
Gai et al. [21]	2020	DenseNet		✓	✓		✓		MFI
Zhang et al. [26]	2021	DenseNet	✓			✓	✓		CT-MRI
Ren et al. [32]	2021	DenseNet	✓			✓	✓		IR-VIS
GEU-Net [22]	2021	U-Net		✓	✓		✓		MFI
Fuse GAN [23]	2019	GAN		✓	✓			✓	MFI

spatial domain, and its processes are as follows: Firstly, the source images are divided into blocks, these blocks are used as the input of the network, and a classification task is constructed to judge the category of each block; Secondly, the feature maps of different stages are merged by performing linear convolution, nonlinear activation and spatial pooling on the feature maps, and the decision maps containing the feature information of the source images are output; Thirdly, the decision maps are processed; Finally, the decision maps are fused by using fusion rules to obtain the final fused image. These processes are shown in Fig. 2.

First class is decision mapping based on CNN. The decomposed high-frequency sub-bands are inputted into CNN by Wang et al. to generate decision maps, and CNN is used as the fusion rule of the frequency sub bands, which is not only adaptive, but also replaces the traditional rule that requires manual design, then the decision maps of low-frequency sub bands and high-frequency sub bands are fused respectively, and finally they carry out the inverse transformation on the fused coefficients to obtain the final fused image. This method improves the quality of the fused image [1].

Second class is decision mapping based on ResNet. In order to solve the difficulty of boundary blur level estimation, DRPL is proposed that the source images are input into the CNN composed of convolution blocks and residual blocks to extract the shallow and deep features, then obtaining their corresponding weighted maps, performing dot product and weighted sum operation on them to obtain the fused image. This method can make use of the complementary information existing in the source images [10].

Third class is decision mapping based on DenseNet. The source image blocks are input into the DenseNet by Gai et al. to get the score maps, then the decision maps are gotten through binarization, finally the fused images are obtained by using the fusion rules. In the feature processing stage, DenseNet can make full use of the feature information of the images, and it can effectively solve the classification problem of the fused image decision maps [21].

Fourth class is decision mapping based on U-Net. GEU-Net regards

Table 2
End-to-End image fusion methods.

Methods	Year	Backbone Network	Domain		Supervised		Code	Data Set
			Spatial	Transform	yes	no		
IFCNN [20]	2019	CNN		✓	✓		✓	MFI、IR-VIS、CT-MRI
Wang et al. [60]	2019	CNN		✓	✓			MFI、IR-VIS、MRI-PET、MRI-SPECT
HPCFNet [35]	2020	CNN		✓	✓			Natural pictures
UFA-FUSE [33]	2021	CNN		✓	✓			MFI
AE-Netv2 [61]	2021	CNN		✓		✓		MEI、MFI、CT-MRI、IR-VIS
MMF [34]	2021	CNN		✓		✓	✓	IR-VIS
Li et al. [37]	2021	CNN		✓	✓			MFI
RXDNFuse [36]	2020	ResNet		✓		✓		IR-VIS
STDFusionNet [39]	2021	ResNet		✓	✓		✓	IR-VIS
Wang et al. [62]	2021	ResNet		✓	✓			MFI
DMDN [38]	2021	ResNet		✓	✓			MFI
MLDNet [41]	2020	DenseNet		✓		✓		MFI
VIF-Net [45]	2020	DenseNet		✓		✓		IR-VIS
Xu et al. [44]	2020	DenseNet		✓		✓		IR-VIS
Deng et al. [63]	2021	DenseNet		✓	✓		✓	MEI
Li et al. [64]	2021	DenseNet		✓	✓		✓	IR-VIS
FusionGAN [47]	2018	GAN		✓		✓		IR-VIS
Ma et al. [65]	2019	GAN		✓		✓		IR-VIS
LBP-BEGAN [66]	2019	GAN		✓		✓		IR-VIS
RCGAN [53]	2019	GAN		✓		✓		IR-VIS
DDcGAN [52]	2020	GAN		✓		✓	✓	IR-VIS
MGMDcGAN [54]	2020	GAN		✓	✓			MRI-PET、MRI-SPECT、CT-SPECT
D2WGAN [67]	2020	GAN		✓		✓		IR-VIS
DSAGAN [68]	2020	GAN		✓	✓		✓	MRI-PET、MRI-SPECT
Liu et al. [69]	2020	GAN		✓	✓			MRI(T1)-MRI(T2)-MRI(DWI)-MRI(Flair)
MFF-GAN [70]	2020	GAN	✓			✓		MFI
MgAN-Fuse [71]	2020	GAN		✓	✓			IR-VIS
DiCyc [51]	2020	GAN		✓	✓			CT-MRI
FLGC-FusionGAN [72]	2020	GAN		✓		✓		IR-VIS
Fu et al. [73]	2021	GAN		✓		✓		IR-VIS
MFIF-GAN [48]	2021	GAN		✓	✓			MFI
LatRAIVF [74]	2021	GAN		✓	✓			IR-VIS
DFPGAN [75]	2021	GAN		✓	✓			IR-VIS
Chartsias et al. [59]	2017	U-Net		✓	✓		✓	MRI(T1)-MRI(T2)-MRI(DWI)
DenseFuse [55]	2018	U-Net		✓		✓		IR-VIS
MCFNet [2]	2019	U-Net		✓	✓			CT-MRI、MRI-SPECT、MRI-MRI
Kumar et al. [76]	2019	U-Net		✓	✓			CT-PET
An et al. [77]	2020	U-Net		✓	✓			IR-VIS
PFAF-Net [56]	2020	U-Net		✓	✓			MFI、IR-VIS、MEI、CT-MRI
Mustafa et al. [41]	2020	U-Net		✓	✓			IR-VIS
RFN-Nest [78]	2021	U-Net		✓	✓		✓	IR-VIS
DMC-Fusion [79]	2021	U-Net		✓		✓		MRI-MRI、MRI-SPECT、MRI-PET
IR-MSDNet [80]	2021	U-Net		✓	✓			IR-VIS
Hou et al. [81]	2021	U-Net		✓		✓		MEI
SEDRFuse [57]	2020	U-Net		✓	✓			IR-VIS
SMFuse [58]	2021	U-Net	✓		✓			MFI
Ren et al. [82]	2021	U-Net		✓		✓		IR-VIS
UMAG-Net [83]	2021	U-Net		✓		✓		HSI-MSI

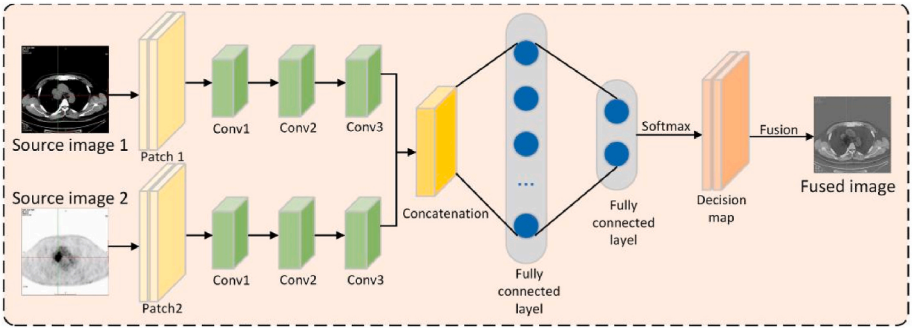

Fig. 2. Deep learning for decision mapping.

the generation of fused image as a global binary segmentation task. In order to improve the global feature encoding ability of U-Net, a Global Feature Pyramid Extraction module (GFPE) and a Global Attention Connection Up-sampling module (GACU) are introduced to effectively extract and utilize global semantic and edge information, the final decision map is estimated by the contextual relationship among the pixels in the feature map, finally the fused image is obtained using the pixel-wise weighted-average strategy [22].

Fifth class is decision mapping based on GAN. Guo et al. proposes an image fusion method based on cGAN, which is called Fuse GAN. In this method, the task of image fusion is regarded as the conversion problem from source images to decision maps, and the least square GAN objective is employed to enhance the training stability of Fuse GAN, resulting in an accurate confidence map for focus region detection [23].

3.2. Feature extraction based on deep learning

The processes of feature extraction in deep learning are as following: firstly, the source images are inputted into deep learning network to extract features, secondly, the feature information of each output layer are fused by fusion rules, finally, the fused image is obtained by the reconstruction process. The inputs of the network are the source images and the output are the feature information. These processes are shown in Fig. 3. Deep learning methods have stronger feature extraction ability than traditional methods, and it is widely used in the field of image fusion.

First class is featuring extraction based on CNN. VGG-19 is used by Zhou et al. as a feature extractor to extract the low-level and high-level features of the source images and obtain the candidate fusion image, the maximum strategy was used to generate the final fusion image from the candidate fusion image and the reconstruction of the fused image is completed. VGG-19 has strong ability to extract the feature information of the source images, and the fusion processing has less noise and artifacts [24].

Second class is featuring extraction based on ResNet. ResNet50 are used as a feature extraction module by Li et al. to extract depth features from the source images, then the depth features are normalized to get the initial weight map, and finally the fused image are reconstructed by weighted average strategy. ResNet50 can achieve better fusion performance than VGG19 [25].

Third class is featuring extraction based on DenseNet. DenseNet is used by Zhang et al. to extract features and to reuse features through dense connection, which achieves better performance than the CNN with less parameters and computational cost, and finally the fused images are reconstructed by the average fusion strategy. The more details of the fusion image are preserved with fewer network layers [26].

Fourth class is featuring extraction based on Attention mechanism. Multi-scale Residual Pyramid Attention Network (MSRPAN) is proposed by Fu et al., which increase the multi-scale information compared with residual attention mechanism and enhanced the ability of feature extraction compared with pyramid attention mechanism, so it has better ability of feature extraction and expression [27].

4. End-to-end image fusion methods

In the non-end-to-end image fusion methods, the optimal feature in feature extraction stage is not the best final result in some times. Therefore, the end-to-end image fusion methods are proposed. End-to-end image fusion means that the inputs of the network are the source images and the output is the fused image. The whole learning process is not divided into sub-processes, and the deep learning model learns the mapping from the source images to the fused image. End-to-end image fusion methods include image fusion methods based on CNN, image fusion methods based on GAN and image fusion methods based on Encoder-Decoder Network. End-to-end image fusion methods are summarized in this paper. The methods are summarized in Table 2.

4.1. Image fusion methods based on Convolutional Neural Network (CNN)

The image fusion method based on CNN realizes implicit feature extraction, feature fusion and image reconstruction by designing network structure and loss function, and avoids the limitations of manually designing fusion rules. The processes of image fusion method based on CNN are: firstly, the source images are inputted into CNN for processing, then the processed features are fused, and finally the fused images are reconstructed through deconvolution. In the process, the output of intermediate results is not required, and the CNN learns the direct mapping from input to output. Compared with the traditional image fusion algorithms, the CNN can adapt to the image fusion tasks by learning the appropriate parameters of the convolutional filter, and the parameters of the CNN model can be optimized through end-to-end training. This section summarizes the image fusion methods based on CNN from four aspects: single-level feature fusion methods, multi-level feature fusion methods, image fusion methods based on ResNet and image fusion methods based on DenseNet, these processes are shown in Fig. 4.

4.1.1. Single-level feature fusion methods

The processes of single-level feature fusion methods (solid line in Fig. 4) are as follows: firstly, the features of the source images are extracted by several convolution blocks, then the features output by the last convolution layer are fused, and finally the features are reconstructed by several deconvolution blocks to obtain the final fused image.

UFA-FUSE uses convolution blocks to extract image features from the source images, then the features are fused by attention mechanism, and finally the fused image features are inputted into cascaded convolution blocks to reconstruct the fused images, this method refines the decision map through post-processing to avoid the generation of intermediate decision map and realize image fusion [33]. In order to improve the quality of image fusion, in Multi-scale MobileNet based Fusion (MMF), the high-dimensional features of the input images are extracted by Multi-scale Mobile Block (MMB), and the fused images are generated by combining the high-dimensional features [34].

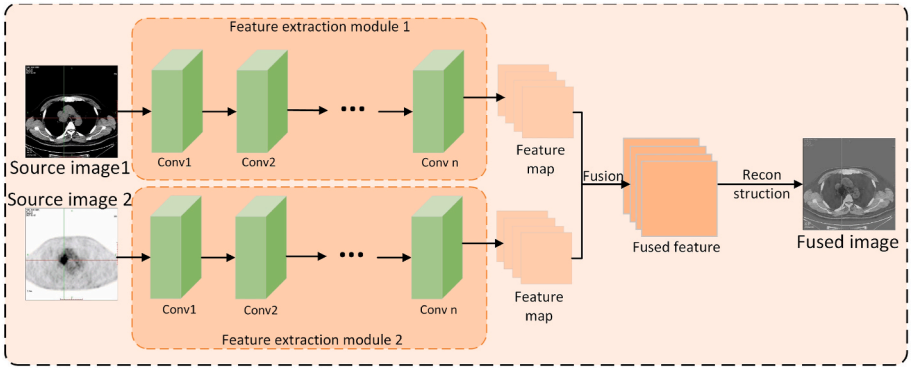

Fig. 3. Deep learning for feature extraction.

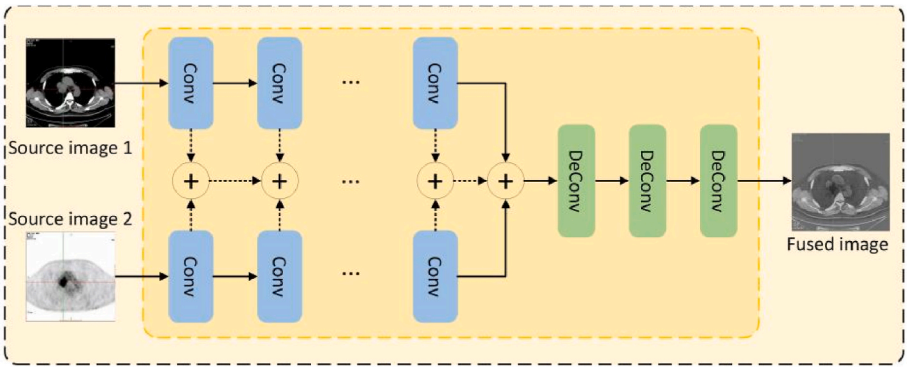

Fig. 4. Image fusion methods based on CNN.

4.1.2. *Multi-level feature fusion methods*

The processes of multi-level feature fusion (solid line and dotted line in Fig. 4) are as follows: firstly, the features of the source images are extracted by convolution blocks, then the features of the corresponding layers extracted by each convolution layer are fused, the fused features of each layer are fused to generate the final fused image. Multi-level feature fusion can make the image features more fully utilized.

In HPCFNet, the paired images are fed into Siamese CNN firstly, then the feature maps from convolutional layers are integrated by Paired Channel Fusion (PCF) module to generate channel-wise fused feature maps hierarchically, the fused feature maps are adjusted by the Reverse Spatial Attention (RSA) modules. PCF first combines the feature maps of the same level through Cross Feature Stack (CFS), and the channel pairs are fused by Parallel Atrous Group Convolution (PAGC) module to capture multi-scale feature representation [35].

4.1.3. *Image fusion methods based on residual neural network (ResNet)*

Among the fusion methods based on CNN, the shallow features are loosed with the increasing of the network layers, which reduces the fusion effects. The fusion methods based on ResNet can make better use of the extracted feature information, and the fused images can retain more details of the source images. These methods can be divided into two categories: global residual connection for image fusion and residual blocks for image fusion.

1) Global Residual Connection for Image Fusion

Global residual connection for image fusion refers to residual connection is established between different stages of fusion process to reuse features. Global residual connection can not only fuse global feature information, but also accelerate network convergence, these processes are shown in Fig. 5.

RXDNFuse uses global residual learning (GRL) to establish information residual connection between shallow feature extraction module and deep feature extraction module, so the final image features rely on the output of previous deep features and shallow features [36]. A full convolution image fusion network based on supervised learning is proposed by Li et al., the texture details of the shallow feature extraction module are fused into the spatial information of the deep feature extraction module by dense connection, the region, texture and edge details of the source images are effectively preserved, and the clarity of the image is further improved, and there is no artifact at the boundary of the fused image [37]. In DMDN, a long jump connection is added before and after the residual block group, which can preserve the shallow features and the useful details after many layers of convolutional operation of the residual block group [38].

2) Residual Blocks for Image Fusion

Residual blocks for image fusion refers to several residual blocks are used in different stages of the fusion process. The extracted features can be enhanced through residual blocks and the image information can be used more fully utilized. The processes of the method are shown in Fig. 6. According to the different structure of residual blocks, it can be divided into four categories: residual block, dense residual block, multi-scale residual block and residual attention block, which will be introduced from these four aspects.

First class is Residual Block. The Residual Block uses skip connection to alleviate the problem of gradient disappearance, and the batch normalization layer is removed to maintain the consistency of image contrast. Degradation Model-based Deep Network (DMDN) for image fusion are proposed by Xiao et al., a group of refined residual blocks without BN and ReLU layers are adopted, and short jump connection is adopted between adjacent residual blocks [38]; STFusionNET extracts image features from a common layer and three residual blocks, the residual blocks can enhance the extracted information [39].

Second class is Densely Residual Block. Densely Residual Block adaptively learns more efficient features through local feature fusion, which combines the structural advantages of ResNet and DenseNet to overcome the limitations of traditional fusion methods and fully exploit the features at different levels of the image. RXDNFuse is proposed by Long et al., the aggregated residual dense block (RXDB) is designed to

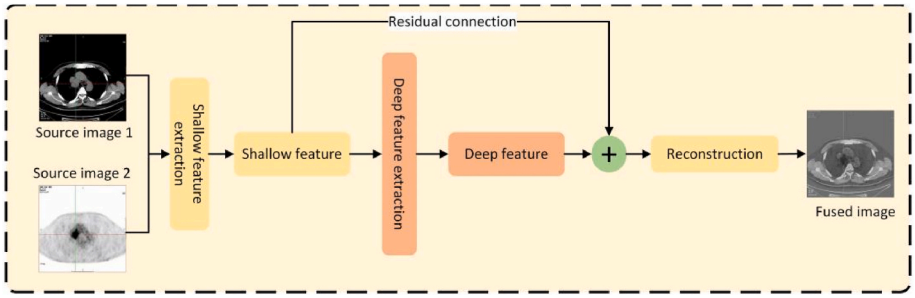

Fig. 5. Global residual connection for image fusion.

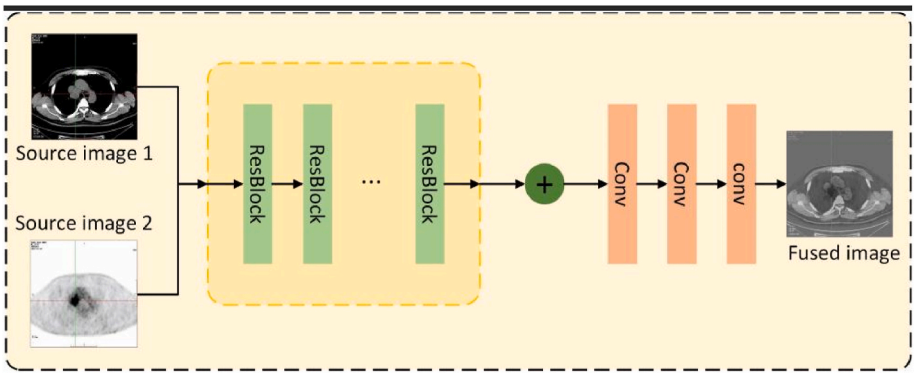

Fig. 6. Fusion network internal usage of residual block.

extract the features, which consists of residual blocks and residual dense blocks, there are six RXDBs staked in RXDNFuse, RXDBs increase the diversity of extracted features and improve the time efficiency of image fusion, reducing the amount of network computation [36].

Third class is Multi-Scale Residual Block. A convolutional layer with a small receptive field can extract low-frequency features but not high-frequency features, while a convolutional layer with a large receptive field can extract more significant image features. Multi-scale Dilated Residual Blocks (MDRB) is designed by Song et al., which extract the multi-scale features through two parallel convolution kernels, and the features are inputted into two convolution kernels with different dilation rates to expand the receiving field, with lower computational cost [40].

Fourth class is Residual Attention Block. The residual attention block is implemented by adding an attention mechanism to the residual block, which gives its weights according to the importance of the source feature map, residual connection enables the attention mechanism to learn the weights of each channel globally, which greatly enhancing the universality of the attention block. Mustafa et al. introduce a residual self-attentive block to fuse and refine features, the output of the residual self-attention block is the weighted sum of the original local features and the attention map, which also contains self-attention information and global contextual information [41].

4.1.4. *Image fusion methods based on dense neural network (DenseNet)*

Image fusion methods based on DenseNet refers to adding dense connections to the CNN, or replacing convolution blocks with dense blocks. These processes are shown in Fig. 7. DenseNet is able to obtain smoother decision functions by exploiting low-complexity information of shallow layers, and therefore it has better generalization performance [42]. Dense blocks have a stronger dense connection mechanism compared with residual blocks. The features of each layer can be fully utilized by using dense connection to ensure that the fused image contains more of the multi-scale, multi-level features of the source image, alleviating the problem of gradient disappearance.

In MLDNet, the features of different levels are firstly extracted from the source images by the feature extraction module, including global features and dense features, then the global features are learned in the process of dense feature fusion, and the feature mapping is globally fused through the global feature fusion module. The network has the advantages of wide structural depth, large computation, fast convergence, few parameters and more information flows [43]. Xu et al. use the DenseNet to extract and reconstruct features, which fully utilizes the features of each layer, the network uses a two-channel structure, and the fusion part treats training as an end-to-end mapping from input to output, designing an optimized SSIM loss function and a perceptual loss function, and the fused image retains rich information [44]. VIF-Net feeds the source images into two branches respectively, there is a convolution block and a dense block in each branch, and the same weight is used to extract the same type of depth features, the network has simple structure, high operational efficiency, and overcomes the ghosting artifacts around the boundary of targets [45].

4.2. *Image fusion methods based on Generative Adversarial Network (GAN)*

Since GAN [46] was proposed in 2014, it has been widely used in the field of imaging due to its flexibility and excellent performance. The process of image fusion based on GAN can be regarded as an adversarial game between the source images and the fused images, more specifically, the GAN-based image fusion methods use discriminators to force the generator to generate fusion results that are consistent with the target distribution in terms of probability distribution, thus implicitly enabling feature extraction, fusion and image reconstruction, which enables the fused image to obtain the feature information of two source images at the same time. These processes are shown in Fig. 8. GAN-based image fusion methods can be divided into three categories: image fusion methods based on classical GAN, image fusion methods based on dual-discriminators GAN, and image fusion methods based on multi-GAN.

4.2.1. *Image fusion methods based on classical GAN*

Image fusion based on classical GAN refers to a fusion network containing a generator and a discriminator. The generator is used to

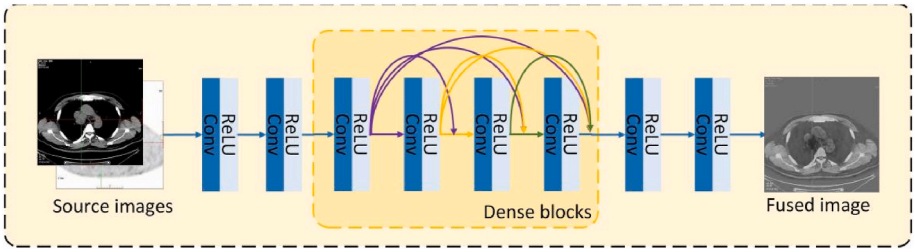

Fig. 7. Image fusion methods based on DenseNet.

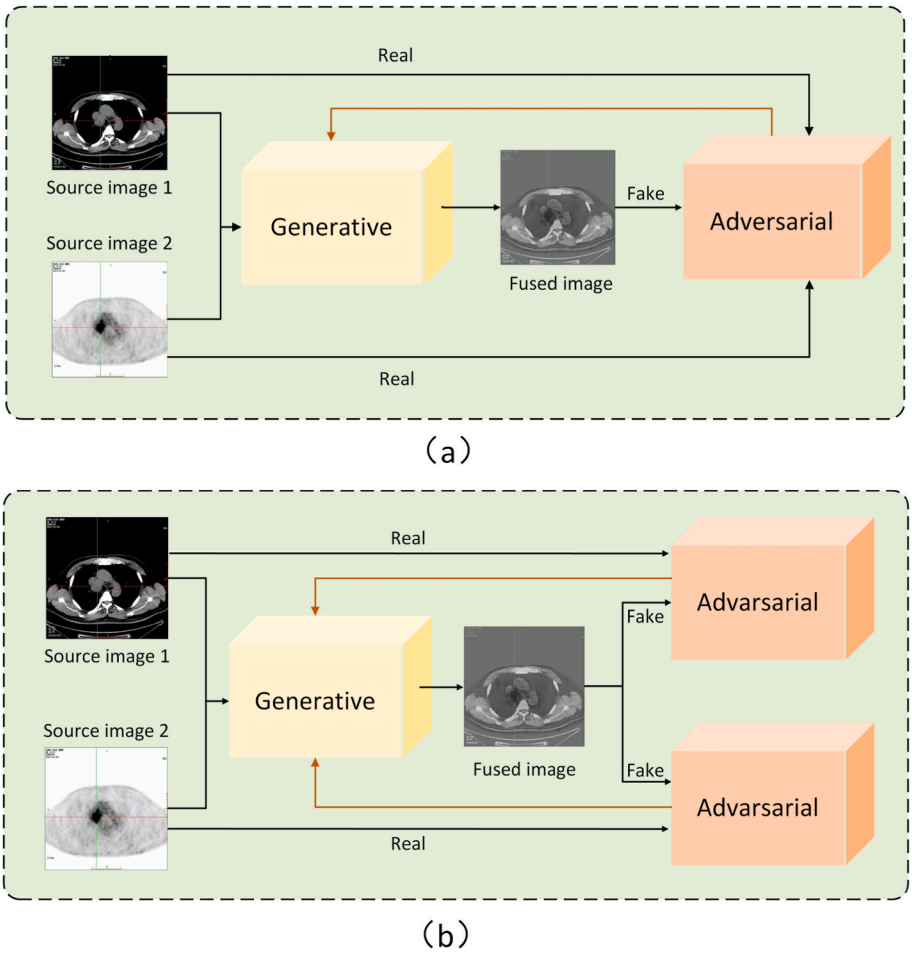

Fig. 8. Image fusion methods based on GAN: (a)Image fusion based on classical GAN; (b) Image fusion based on double discriminator GAN.

generate a fusion image with the feature information of two source images, and the discriminator is used to identify the authenticity of the generated images and the source images. The generator and the discriminator circularly play with each other until the generator gets the best result.

Ma et al. propose Fusion GAN, which is the first time that GAN is used for image fusion [47], two images are concated by channels and fed into generator to generates a fused image containing intensity information and gradient information, the discriminator forces the generator to retain as much detail as possible in the source images. Wang et al. propose MFIF-GAN to address the limitations of Fuse GAN, six parallel attention-based networks are used to extract features of all channels from color images, the limitation of color information loss in Fuse GAN is solved, the adversarial loss and gradient penalty in WGANGP [49] are used to avoid gradient disappearance or explosion, stabilize the training process, the L1 norm is used instead of binary cross entropy (BCE) as the reconstruction loss function, and simple small region removal (SRR) is used instead of convolutional conditional random fields (ConvCRF) as the post-processing technique [48].

4.2.2. *Image fusion methods based on dual-discriminators GAN*

The classical generation adversarial network achieves good results in the image fusion task, but it is insufficient in retaining different details of two source images. Source images of different modalities have different feature information, and the adversarial game between a single generator and discriminator leads to fused image that is more similar to one image. To solve this problem, researchers propose image fusion methods based on dual-discriminators GAN. The processes of these methods as follows: two source images are fed into the generator to generate a fused image with the feature information of the two source images, the purposes of these two discriminators are to respectively calculate the structural difference between the fused image and the two source images, as well as the content loss.

Zhu et al. propose Cycle GAN for unpaired image generation and transformation, which has been applied to CT and MRI image fusion of heart and brain with good fusion results [50]. To solve the problem of Cycle GAN cannot achieve good alignment between the synthesized images and data from the source domain, Wang et al. propose DiCyc, integrating a modified deformable convolutional layer into the network, which avoids the conflict between the CycleGAN loss and the image alignment losses, and the associated deformation-invariant cycle consistency loss and NMI-based alignment loss function are proposed, which demonstrated better robustness for synthesis of images from different domains [51]. Ma et al. proposed DDcGAN, which the Encoder-Decoder structure based on DenseNet is adopted in the generator, and fused images are generates based on a specially designed content loss function, thus deceiving two discriminators, the purposes of the two discriminators are to separately calculate the structural differences and content loss between the fused images and the two source images, and the method achieve good results in fusion tasks of brain medicine images [52].

4.2.3. *Image fusion methods based on Multi-GAN*

Multi-GAN-based methods are usually consisting of two or more GANs, which is used to solve the problem that the existing image fusion methods based on GAN can only fuse part of the information in the source images, resulting in the loss of other important information.

Two groups of generator and discriminator are designed in RCGAN,

the first generator generates an image with structure information based on the pre-fused image, the first discriminator measures the relative offset between the generated image and the visible image, the second generator enhances the gradient information based on the pre-fused image, and the second discriminator measures the offset of the second generated image relative to the infrared image [53]. Huang et al. propose MGMDcGAN, which contains two cGANs, in the first cGAN, the generator aims to generate the fused image and deceive two discriminators based on a specially designed content loss function, while the discriminator aims to calculate structural differences between the fused images and the source images; the second cGAN with a mask is used to enhance the dense information in the final fused images, while preventing the functional information being weakened, the final fused images retain both structural and functional information at different resolutions, and the network can be applied to different types of medical image fusion [54].

4.3. Image fusion methods based on encoder-decoder network

Image fusion methods based on Encoder-Decoder Network refers to: firstly, the encoder is pre-trained on the dataset to extract the features of the source image; then, the extracted features are fused by the fusion rules; finally, the fused features are reconstructed by the pre-trained decoder to obtain a fused images containing the feature information of the source images. In Generally, the encoder is used to learn the feature information of source images, and the decoder is used to recover the position information of images. The Encoder-Decoder Network used for image fusion include Single Encoder-Decoder Network, Double Encoder-Decoder Network and Multi Encoder-Decoder Network. These processes are shown in Fig. 9.

4.3.1. Image fusion methods based on single encoder-decoder network

Image fusion methods based on single encoder-decoder network refers to a fusion network containing one encoder, and its process are: the concated images are fed into the encoder to extract features; then, the encoded features are fused; and finally, the fused features are decoded to obtain the final fused images.

DenseFuse is a typical encoder-decoder image fusion network, in which the encoder consists of convolution blocks and dense blocks, the depth features of the encoder can be better retained by dense block, thus ensuring that more significant features can be used for the fusion rules, the output of the fusion layer is the input of the decoder, and the decoder is completed by four convolution layers [55]. Raza et al. propose PFAF-Net, firstly, different levels of features are obtained by encoder, and the high-level features are further generated pyramid features with rich multi-scale information; features are fused by using different fusion rules at the low and high levels; fused detail features are decoded to obtain the final fused image; the network retains a wealthy information in the fused images [56].

4.3.2. Image fusion methods based on Double Encoder-Decoder Network

The source images are concated at the input of the single encoder-decoder fusion network, but the images of the different modalities have different detailed information and require features to be extracted in different ways, hence the double encoder-decoder fusion network is proposed. The aim of each encoder is to extract the visual features of the image, with the advantage of ensuring that the different information of the two modal images can be fully utilized.

To address the problem of image detail loss during convolution, SEDRFuse adds a residual block to the encoder, which is decoded by deconvolution symmetrical to the encoder, a jump connection is used between the encoder and decoder to reduce the loss of texture detail in the images and to speed up the convergence of the network, the network reduces the computational complexity and can avoid artifacts in the fused images [57]. A classifier is used to determine whether each patch is in focus or out of focus, and a decision map is outputted, then the initial decision map is obtained by bootstrapping the filter, and finally a maximum gradient loss function is designed for optimization to learn a more accurate binary mask, the network avoids chromatic aberrations caused by patch decomposition and boundary blurring caused by post-processing [58].

4.3.3. Image fusion methods based on Multi Encoder-Decoder Network

As it is often necessary to fuse more than two images in an image fusion task, most of the existing methods target the fusion of two images, in order to fuse medical images from multiple modalities while retaining more detailed information about the various image types, multi encoder-decoder fusion networks are proposed to obtain better fusion results through information sharing.

multi modal MRI synthesis based on modality-invariant latent representation is proposed by Chartsias et al., each input image has an independent encoder, which embeds a single-channel image into a multi-

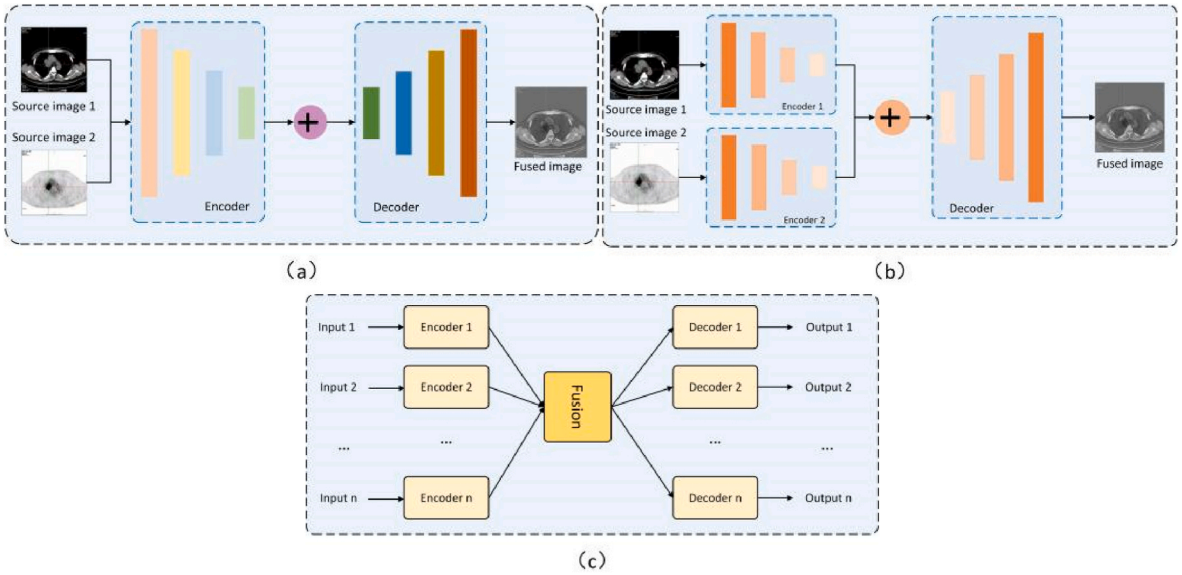

Fig. 9. Image fusion methods based on Encoder-Decoder Network: (a)Single Encoder-Decoder Network; (b) Double Encoder-Decoder Network; (c) Multi Encoder-Decoder Network.

channel latent space, this model combines information from multiple inputs to further improve the quality of fused images, and it is robust to missing data because it benefits from but does not rely on additional input modes [59].

5. Application in medical image fusion

5.1. Multi-modality medical image fusion methods

Medical images play an irreplaceable role in modern medical diagnosis and treatment. As each type of medical images have their own purposes, advantages and limitations, a single-modality medical image often does not provide sufficient information for the medical diagnostic process. Therefore, the image fusion is essential in multimodal medical image processing and it also plays an important role in the computer-aided diagnosis tasks, such as image visualization processing, disease detection, disease segmentation and disease classification.

The imaging methods for processing medical images of different lesion areas have their own characteristics: Computer Tomography (CT) imaging uses X-rays to detect the information of high-density structures such as bones; Magnetic resonance imaging (MRI) can provide more detailed information about soft tissues such as brain and reproductive system; Positron Emission Tomography (PET) and Single Photon Emission CT (SPECT) are commonly used in the detection tasks of vascular and tumor diseases [84]. Therefore, in order to obtain more accurate clinical diagnosis results, researchers apply various fusion technologies to the multi-modal medical image fusion tasks. This section will summarize the multi-modal medical image fusion methods of different modalities from four aspects.

First class is PET and CT image fusion. PET/CT scanning devices are an important tool for the diagnosis, staging and evaluation of many cancers [85]. PET/CT combines the sensitivity information of abnormal functional areas detected by PET images with the anatomical location information obtained by CT images [86]. Therefore, PET and CT image fusion technology can integrate the functional information of lesions and anatomical information of tissues and organs, and can also help doctors locate and qualitatively diagnose lung tumor areas more accurately. A collaborative learning feature fusion method is proposed by Kumar et al., in which two encoders are firstly used to extract the relevant visual features of PET and CT images respectively, then a fused image can be obtained by generating the feature information of specific modes, and finally the fused image can be obtained by the decoder. This is an end-to-end method, which can effectively fuse the complementary information of different modalities [76].

Second class is CT and MRI image fusion. Due to the high cost of PET/CT scanning devices, CT and MRI are the most widely used tomography techniques in clinical diagnosis. CT can better distinguish tissue regions with density differences compared to MRI which has a better ability to distinguish soft tissue regions, but the current medical conditions do not allow for CT and MRI to be scanned on the same device, therefore, the image fusion methods of CT and MRI are proposed. An image fusion method based on the combination of the G-CNN and fuzzy neural network is proposed by Wang et al., which is a non-end-to-end method: firstly, The CT and MRI images are represented by Gabor filter groups with different directions and scales, this method significantly outperforms other methods in characterizing the texture features and edge information of lesions in medical images [30].

Third class is MRI and PET/SPECT image fusion. MRI provides soft tissue information and functional information of blood flow and metabolism, but it has lower resolution and cannot accurately describe the anatomical details of organs; PET images reflect information of blood flow, oxygen and glucose metabolism in tissue, while SPECT images can highlight the lesion areas in tissues and organs. Therefore, fusion of MRI-PET or MRI-SPECT images can increase the complementarity of the images, facilitating clinical diagnosis and precise treatment [87]. A multi-scale dual-branch residual block is proposed by Li et al. to extract the feature information of the image, MRI-PET and MRI-SPECT images are used as the input of this module, the first branch uses a multi-scale mechanism and the second branch uses multiple convolutional kernels to obtain more informative and wider range of image features, this method is an end-to-end method, which has good fusion effect and low time cost [31].

Forth class is MRI image fusion. MRI can be used to obtain images with different contrast ratios by adjusting the imaging parameters, which can highlight different tissue areas under the same anatomical structure [88]. For example, T1-weighted images clearly delineate gray and white matter areas, while T2-weighted images provide more information of cortical fluid regions. It is difficult to obtain a complete MRI sequence due to the long imaging time, and some of the acquired image sequences are corrupted by noise or artifacts and cannot be used for medical image processing tasks, therefore, multimodal MRI fusion methods are proposed. Multimodal MR synthesis via modality-Invariant latent representation is proposed by Chartsias et al. [59], that is an end-to-end method, each modal input images have an independent encoder, which embeds a single-channel image into a multi-channel potential space, and the information from multiple modalities can be combined to further improve the image synthesis quality.

5.2. Data set of medical image fusion

In order to facilitate researchers to carry out in-depth study of medical image fusion methods, this section will be a commonly used medical image fusion data sets are summarized in Table 3.

Whole Brain Altes is a medical image dataset released by Harvard Medical School in 1999. The dataset is a collection of medical images, UCI data, and biomedical literature, containing more than 13,000 pairs of CT-MRI, PET-MRI, and SPECT-MRI images, the dataset is available for download, please visit: http://www.med.harvard.edu/AANLIB/.

Image Large Scale Visual Recognition Challenge (ILSVRC) is an Ischemic stroke lesion Segmentation (ISLES) competition co-organized by MICCAI 2015, which provides MRI scans containing a large number of stroke samples and related clinical parameters to evaluate the stroke lesion area in MRI images and clinical prediction results. The dataset is available for download, please visit: https://image-net.org/challenges/LSVRC/.

Ischemic stroke segmentation (ISLES) is an ischemic stroke lesion segmentation (Isles) competition jointly organized by the 2015 MICCAI conference, provided MRI containing a large number of stroke samples and associated clinical parameters, used to evaluate stroke lesion areas on MRI scans and to predict clinical outcomes. The dataset is available for download, please visit: http://www.isles-challenge.org.

Brain Tumor Segmentation (BRATS) for the segmentation task of brain tumors that are inherently heterogeneous (appearance, shape, and histology) using multi-institutional preoperative magnetic resonance imaging. The MRI image set for each patient includes T1w, T2w, contrast-enhanced T1, and FLAIR, as well as voxel-level ground truth for edema, enhanced tumors, and nonenhanced tumors. The data set can be downloaded, please visit: http://braintumorsegmentation.org/.

IXI Nearly 600 MRI images of healthy subjects from 3 hospitals were collected. The MRI acquisition information of each subject included: T1, T2, and PD weighted images, MRA images, and diffusion weighted images in 15 directions. The dataset is available for download, please visit: http://brain-development.org/ixi-dataset/.

I2CVB provides multiparametric MRI datasets to aid in the development of computer-aided detection and diagnosis (CAD) systems. The dataset includes MRI of 17 prostate cancer patients who underwent biopsy testing, and the dataset is available for download, please visit: http://i2cvb.github.io/#properi-data.

Alzheimer's Disease Neuroimaging Initiative (ADNI) was established in 2003 to test whether sequential magnetic resonance imaging (MRI), positron emission tomography (PET), other biomarkers, and clinical and neuropsychological assessments can be combined to

Table 3
Medical image fusion dataset.

Data Set	Year	Modality	Disease	Quantity	Download Address
The Whole Brain Alta's of Harvard Medical School	1999	CT、MRI、PET、SPECT	Normal brain, Cerebrovascular disease, Brain tumor, Alzheimer's disease	13000	http://www.med.harvard.edu/AANLIB/
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)	2010–2017	CT、MRI、PET、SPECT	Hypertensive encephalopathy of brain	1200000	https://image-net.org/challenges/LSVRC/
Ischemic Stroke Lesion Segmentation (ISLES)	2015	MRI	Ischemic stroke	/	www.isles-challenge.org
Brain Tumor Segmentation (BRATS)	2015–2021	MRI	Ischemic stroke	8000	braintumorsegmentation.org
IXI	2015	MRI	Normal brain	600	http://brain-development.org/ixi-dataset/
I2CVB	2016	MRI	Prostate	/	http://i2cvb.github.io/#properi-data
Alzheimer's Disease Neuroimaging Initiative (ADNI)	2003	MRI、PET	Alzheimer's disease	/	www.adni-info.org

measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The dataset is available for download, please visit: www.adni-info.org.

6. Evaluation metrics of medical image fusion

Image quality evaluation refers to designing an algorithm to automatically evaluate the image quality in a perceptually consistent way [89], which is very important in task of medical image fusion. This section will introduce the evaluation metrics of medical image fusion from two aspects: evaluation metrics with reference image and evaluation metrics without reference image.

6.1. Reference image

The evaluation metrics with reference image refers to the algorithm applied to both the source image and the fused image. It is used to measure the amount of information transmitted from the source image to the fused image or to measure the correlation of information between the source image and the fused image, where the reference image can be the source image or the truth image. This section introduces seven evaluation metrics, which are: nonlinear correlation information entropy (Q_{NCIE}), mutual information (*MI*), characteristic mutual information (*FMI*), normalized mutual information (*NMI*), Structural similarity Measure (*SSIM*), mean square error (*MSE*) and root mean square error (*RMSE*). Where, *MI*, *FMI*, *NMI*, *MSE*, and *RMSE* are measures based on information theory, and SSIM is a measure based on image structural similarity.

(1) Nonlinear Correlation Information Entropy (Q_{NCIE})

Q_{NCIE} measures the nonlinear correlation between source images *A* and *B*, and fusion image *F* [90]. The expression is:

$$Q_{NCIE} = 1 + \sum_{l=1}^{3} \frac{\lambda_i}{3} \sum_{l=1}^{3} \frac{\lambda_i}{3} \log_{256} \frac{\lambda_i}{3} \log_{256} \frac{\lambda_i}{3} \tag{1}$$

Where λ_i is the eigenvalue of the nonlinear correlation matrix R. Based on the nonlinear correlation coefficient (*NCC*) between the source image and the fused image, *R* is expressed as:

$$R = \begin{pmatrix} 1 & NCC_{A,B} & NCC_{A,F} \\ NCC_{B,A} & 1 & NCC_{B,F} \\ NCC_{F,A} & NCC_{F,B} & 1 \end{pmatrix} \tag{2}$$

The larger the value of Q_{NCIE}, the greater the nonlinear correlation between the fused image and the source images, and the better the fusion effect.

(2) Mutual Information (*MI*)

MI measures the amount of information transmitted from the source images to the fused image [91]. The expression is:

$$MI = MI_{A,F} + MI_{B,F} \tag{3}$$

Where $MI_{A,F}$ and $MI_{B,F}$ represent information transmitted from image A and image B to the fused image, The expression of $MI_{X,F}$ is:

$$MI_{X,F} = \sum_{x,f} P_{X,F}(x,f) \log_2 \frac{P_{X,F}(x,f)}{P_X(x) P_F(f)} \tag{4}$$

Where $P_X(x)$ and $P_F(f)$ are the edge histograms of source images *X* and fused image *F*, respectively. $P_{X,F}(x,f)$ is the joint histogram of the source images *X* and the fused image *F*. The larger the *MI* value, the more information is transmitted to the fused image, and the more information it contains, the better the fusion effect.

(3) Feature Mutual Information (*FMI*)

The *FMI* measures the amount of characteristic information transmitted from the source images according to the *MI* and characteristic information [92]. The expression is:

$$FMI = MI_{\hat{A},\hat{F}} + MI_{\hat{B},\hat{F}} \tag{5}$$

Where $\hat{A}$, $\hat{B}$, $\hat{F}$ are feature map of source images *A*, *B* and fused image *F*. The larger the *FMI* value, the more features are transmitted to the fused image, and the better the fusion effect.

(4) Normalized Mutual Information (*NMI*)

The expression of *NMI* [93] is:

$$NMI = 2\left[\frac{MI_{A,F}}{H(A) + H(F)} + \frac{MI_{B,F}}{H(B) + H((F)}\right] \tag{6}$$

$H(\cdot)$ represents the entropy of the image, and the larger the *NMI* value, the better the fusion effect.

(5) Structural Similarity Index Measure (*SSIM*)

SSIM is used to model information loss and distortion in the fusion process, which can reflect the structural similarity between images. It consists of three components: correlation loss, brightness distortion and contrast distortion [94]. The *SSIM* between the source image *X* and the fused image *F* is defined as the product of these three parts, and its expression is as follows.

$$SSIM_{X,F} = \sum_{x,f} \frac{2\mu_x\mu_f + C_1}{\mu_x^2 + \mu_f^2 + C_1} \cdot \frac{2\sigma_x\sigma_f + C_2}{\sigma_x^2 + \sigma_f^2 + C_2} \cdot \frac{\sigma_{x,f} + C_3}{\sigma_x^2\sigma_f^2 + C_3} \tag{7}$$

Where $\sigma_{x,f}$ is the covariance of the source image and the fused image, σ_x

and σ_f represent the standard deviation, μ_x and μ_f are the average values of the source image and the fused image, respectively. C_1, C_2, and C_3 are constants used to prevent the divisor from being 0. The structural similarity between the fused image and the two source images is expressed as:

$$SSIM = SSIM_{A,F} + SSIM_{B,F} \tag{8}$$

The larger the *SSIM* value, the smaller the information loss and distortion in the fusion process, and the better the image fusion effect.

(6) Mean Square Error (*MSE*)

MSE measures the degree of difference between the reference image and the fused image. The expression is:

$$MSE = \frac{MSE_{A,F} + MSE_{B,F}}{2} \tag{9}$$

$$MSE_{XF} = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{i=0}^{N-1}(X(i,j) - F(i,j))^2 \tag{10}$$

MSE_{AF} and MSE_{BF} represent the differences between the fused image and the two source images, respectively. The smaller the *MSE* value, the closer the fused image is to the source image, and the better the fusion effect is.

(7) Root Mean Square Error (*RMSE*)

RMSE is similar to *MSE*, and its expression is:

$$RMSE = \frac{RMSE_{AF} + RMSE_{BF}}{2} \tag{11}$$

$$RMSE = \sqrt{\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(X(m,n) - F(m,n))^2} \tag{12}$$

The smaller the *RMSE* value, the better the image fusion effect.

6.2. No-reference image

The evaluation metrics of no reference image refers to the algorithm only used for fusion image, which is used to measure the information contained in the fused image. This section introduces seven evaluation indicators: Entropy (*EN*), mean gradient (*AG*), edge-based similarity measure ($Q^{AB/F}$), standard deviation (*SD*), spatial frequency (*SF*), edge intensity (*EI*) and peak signal-to-noise ratio (*PSNR*), among which *EN* and *PSNR* are evaluation metrics based on information theory, and *AG*, $Q^{AB/F}$, *SD*, *SF* and *EI* are evaluation metrics based on image features.

(1) Entropy (*EN*)

EN calculates the amount of information contained in the fused image [95], and its expression is:

$$EN = -\sum_{l=0}^{L-1} pl \log_2 pl \tag{13}$$

Where L represents the number of gray levels, and pl represents the normalized histogram of the corresponding gray levels in the fused image. The larger the value of *EN*, the better the fusion effect.

(2) Average Gradient (*AG*)

The gradient information of the fused image is measured by *AG*, and its details and texture are expressed [96]. The expression is:

$$AG = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}} \tag{14}$$

$\nabla F_x(i,j) = F(i,j) - F(i+1,j)$, $\nabla F_y(i,j) = F(i,j) - F(i,j+1)$,The larger the *AG* value, the more gradient information contained in the fused image, and the better the fusion effect.

(3) Edge based Similarity Measurement ($Q^{AB/F}$)

$Q^{AB/F}$ represents the amount of edge information transmitted from the source images to the fused image [97]. The expression is:

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(Q^{A,F}(i,j)\omega^A(i,j) + Q^{B,F}(i,j)\omega^B(i,j)\right)}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\omega^A(i,j) + \omega^B(i,j)\right)} \tag{15}$$

Where $Q^{X,F}(i,j)$ is the reserved value of edge information, and its expression is:

$$Q^{X,F}(i,j) = Q_g^{X,F}(i,j)Q_a^{X,F}(i,j) \tag{16}$$

$Q_a^{X,F}(i,j)$ represents the edge intensity and orientation retention value at $(i\ ,\ j)$ position, and w^X represents the importance of each source image to the fused image. The larger the value of $Q^{AB/F}$ is, the more edge information of the source image is retained in the fused image, and the better the fusion effect is.

(4) Standard Deviation (*SD*)

SD reflects the distribution and contrast of the fused image [98]. The expression is:

$$SD = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - \mu)^2} \tag{17}$$

Where the average value of the fused images is represented. The larger the *SD* value, the higher the contrast of the fused image and the better the visual effect of the fused image.

(5) Spatial Frequency (*SF*)

SF measure the gradient distribution of *SF* fusion image [99]. The expression is:

$$SF = \sqrt{RF^2 + CF^2} \tag{18}$$

$$RF = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - F(i,j-1))^2} \tag{19}$$

$$CF = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - F(i-1,j))^2} \tag{20}$$

The larger the *SF* value, the richer the edge and texture information of the fused image, and the better the fusion effect.

(6) Edge Intensity (*EI*)

EI measuring edge intensity information of *EI* fused image [100]. The expression is:

$$EI = \sqrt{S_x^2 + S_y^2} \tag{21}$$

$$S_x = F * h_x, S_y = F * h_y \tag{22}$$

$$h_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, h_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{23}$$

The larger the *EI* value, the clearer the fused image and the better the

fusion effect.

(7) Peak Signal-to-Noise Ratio (*PSNR*)

PSNR represents the ratio of peak power and noise power in the fused image. It can measure the distortion degree in the process of image fusion [101], and its expression is

$$PSNR = 10\log_{10}\frac{r^2}{MSE} \tag{24}$$

Where r is the peak value of the fused image and *MSE* is the mean square error. The larger the *PSNR*, the closer the fused image is to the source image, the smaller the distortion, and the better the fusion effect.

7. Conclusions

This paper firstly expounds the basic principle of image fusion based on deep learning; secondly, the research progresses of image fusion are summarized from the following two aspects: Non-End-to-End and End-to-End, and then the image fusion methods based on the deep learning in medical image applications are summarized from the following two aspects: fusion methods and datasets; thirdly, 14 commonly medical image fusion evaluation metrics were sorted out. This paper also has positive significance for the development of more image fusion methods based on deep learning in the future.

Although medical image fusion based on deep learning has made great progresses, there are still faced many challenges. Here, these challenges are discussed in this paper and the future development directions are pointed out from the following four aspects.

(1) Datasets. The image fusion method based on deep learning requires a large number of medical image datasets to train the model and improve the quality of the fusion image. At present, there are few public datasets in the field of medical image fusion due to the patient privacy and the brain anatomy and brain pathology tomography images of Harvard University still occupy the main position. Therefore, more high-quality public datasets can be created for medical image fusion in the future.
(2) Image principle: Different types of sensors or sensors under different settings usually have different imaging principles, these differences in imaging principles provide more priori information for the design of fusion algorithms. It is helpful to further improve the fusion performance to deeply analyze the imaging principles of different types of sensors or sensors in different imaging settings and to model them into the fusion process.
(3) Lightweight network. Deep learning networks with excellent performance usually have high computational cost and require large-scale data for training, so computing power resources and datasets size are also key factors limiting their development. Therefore, in the future, the lightweight fusion network can be improved to reduce the computational overhead of the network and the dependence on large-scale data sets.
(4) Evaluation metrics. Most fusion methods still use traditional reference indicators for the fused images quantitative evaluation, but the values of the evaluation metrics are not consistent with human perception sometimes, especially when fused images are used for the task of medical diagnosis, the validity of evaluation metrics for medical images remains to be explored compared with subjective measures from experienced human observers. Therefore, it is also the future development direction to propose evaluation metrics that are more in line with human visual perception.

Therefore, in-depth research from the above four aspects is of great significance for improving the quality of fused images, and assisting in the diagnosis of clinical diseases, medical image fusion methods based on deep learning will be more widely used in the future, which will greatly promote the progress and development of medical image fusion and make greater contributions to the improvement of medical diagnosis level.

Author contributions

Conceptualization, T.Z., QR.C.; Methodology, T.Z., QR.C.; Investigation, Q.L., XX.Z.; writing—review and editing, QR.C.; Supervision, HL. L, S.Q; resources, QR.C., T.Z., Q.L., XX.Z.; data curation, QR.C.; All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China, grant no.62062003, Key Research and Development Program of Ningxia, grant no.2020BEB04022, Natural Science Foundation of Ningxia Province, grant no.2022AAC03149, and Key Research and Development Program of Ningxia, grant no.2020KYQD08.

Declaration of competing interest

The authors declare that they have no known competing fifinancial interests or personal relationships that could have appeared to inflfluence the work reported in this paper.

Acknowledgments

This project was carried out by Tao Zhou's Laboratory of Computer Science and Engineering School of North Minzu University. It was guided and supported by the research direction of Tao Zhou's team from North Minzu University of science and technology. We sincerely thank them for their help in revising the paper.

References

[1] J. Wang, X.F. Li, Y. Zhang, Adaptive decomposition method for multi-modal medical image fusion, IET Image Process. 12 (2018) 1403–1412.
[2] X.C. Liang, P.Y. Hu, L.G. Zhang, J. Sun, G. Yin, MCFNet: multi-layer concatenation fusion network for medical images fusion, IEEE Sensor. J. 19 (2019) 7107–7119.
[3] B. Palkar, D. Mishra, Fusion of multi-modal lumbar spine images using Kekre's hybrid wavelet transform, IET Image Process. 13 (2019) 2271–2280.
[4] C.O. Ancuti, C. Ancuti, C.D. Vleeschouwer, A.C. Bovik, Single-scale fusion: an effective approach to merging images, IEEE Trans. Image Process. 26 (2017) 65–78.
[5] P.F. Chai, X.Q. Luo, Z.C. Zhang, Image fusion using quaternion wavelet transform and multiple features, IEEE Access 5 (2017) 6724–6734.
[6] B. Ahmadreza, S. Susanne, B. Ali, M.F. Fathi, R.M.D. Souza, V.L. Rayz, Curvelet Transform-based volume fusion for correcting signal loss artifacts in Time-of-Flight Magnetic Resonance Angiography data, Comput. Biol. Med. 99 (2018) 142–153.
[7] Z.Y. Wang, X.F. Li, H.R. Duan, Y.C. Su, X.L. Zhang, X.J. Guan, Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform, Expert Syst. Appl. 171 (2021), 114574.
[8] N. Mitianoudis, T. Stathaki, Pixel-based and region-based image fusion schemes using ICA bases, Inf. Fusion 8 (2007) 131–142.
[9] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Medical image fusion via convolutional sparsity based morphological component analysis, IEEE Signal Process. Lett. 26 (2019) 485–489.
[10] J.X. Li, X.B. Guo, G.M. Lu, B. Zhang, D. Zhang, DRPL: deep regression pair learning for multi-focus image fusion, IEEE Trans. Image Process. 29 (2020) 4816–4831.
[11] W. Huang, Z.L. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recogn. Lett. 28 (2006) 493–500.
[12] Z.Q. Zhou, S. Li, B. Wang, Multi-scale weighted gradient-based fusion for multi-focus images, Inf. Fusion 20 (2014) 60–72.
[13] S.T. Li, X.D. Kang, J.W. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (2013) 2864–2875.
[14] J. Du, W.S. Li, K. Lu, B. Xiao, An overview of multi-modal medical image fusion, Neurocomputing 215 (2015) 3–20.

[15] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Comput. Sci. (2014) 1409–1556.
[16] K.M. He, X.Y. Zhang, S.Q. Ren, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[17] G. Huang, Z. Liu, L.V.D. Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.
[18] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.
[19] Yu Liu, Xun Chen, Hu Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (2017) 191–207.
[20] Yu Zhang, Yu Liu, Peng Sun, H. Yan, X.L. Zhao, L. Zhang, IFCNN: a general image fusion framework based on convolutional neural network, Inf. Fusion 54 (2020) 99–118.
[21] D. Gai, X.J. Shen, H.P. Chen, P. Su, Multi-focus image fusion method based on two stages of convolutional neural network, Signal Process. 176 (2020), 107681.
[22] B. Xiao, B.C. Xu, X.L. Bi, W. Li, Global-feature encoding U-net (GEU-Net) for multi-focus image fusion, IEEE Trans. Image Process. 30 (2020) 163–175.
[23] X.P. Guo, R.C. Nie, J.D. Cao, D. Zhou, L. Mei, K. He, FuseGAN: learning to fuse multi-focus image via conditional generative adversarial network, IEEE Trans. Multimed. 21 (2019) 1982–1996.
[24] J.W. Zhou, K. Ren, M.J.K. Wan, B. Cheng, Q. Chen, An infrared and visible image fusion method based on VGG-19 network, Optik 248 (2021), 168084.
[25] H. Li, X.J. Wu, T.S. Durrani, Infrared and visible image fusion with ResNet and zero-phase component analysis, Infrared Phys. Technol. 102 (2019), 103039.
[26] B. Zhang, C. Jiang, Y.X. Hu, Z.J. Chen, Medical image fusion based a densely connected convolutional networks, in: IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) vol. 5, 2021, pp. 2164–2170.
[27] Jun Fu, W.S. Li, J. Du, A multiscale residual pyramid attention network for medical image fusion, Biomed. Signal Process Control 66 (2021), 102488.
[28] S. Singh, R.S. Anand, Multimodal medical image fusion using hybrid layer decomposition with CNN-based feature mapping and structural clustering, IEEE Trans. Instrum. Meas. 69 (2020) 3855–3865.
[29] S.L. Wu, H.D. Chen, Smart city oriented remote sensing image fusion methods based on convolution sampling and spatial transformation, Comput. Commun. 157 (2020) 444–450.
[30] L.F. Wang, J. Zhang, Y. Liu, J. Mi, J. Zhang, Multimodal medical image fusion based on gabor representation combination of multi-CNN and fuzzy neural network, IEEE Access 9 (2021) 67634–67647.
[31] W.S. Li, X.X. Peng, J. Fu, G.F. Wang, Y.P. Huang, F.F. Chao, A multiscale double-branch residual attention network for anatomical–functional medical image fusion, Comput. Biol. Med. 141 (2021), 105005.
[32] K. Ren, D.W. Zhang, M.J. Wan, X. Miao, G.H. Gu, Q. Chen, An infrared and visible image fusion method based on improved DenseNet and mRMR-ZCA, Infrared Phys. Technol. 115 (2021), 103707.
[33] Y.S. Zang, D.M. Zhou, C.C. Wang, R. Nie, Y. Guo, UFA-FUSE: a novel deep supervised and hybrid model for multifocus image fusion, IEEE Trans. Instrum. Meas. 70 (2021) 1–17.
[34] Y. Liu, C.Y. Miao, J.H. Ji, X.G.M.M.F. Li, A Multi-scale MobileNet based fusion method for infrared and visible image, Infrared Phys. Technol. 119 (2021), 103894.
[35] Y.J. Lei, D. Peng, P.P. Zhang, Q. Ke, H. Li, Hierarchical Paired Channel Fusion network for street scene change detection, IEEE Trans. Image Process. 30 (2021) 55–67.
[36] Y.Z. Long, H.T. Jia, Y.D. Zhong, Y. Jiang, Y. Jia, RXDNFuse: a aggregated residual dense network for infrared and visible image fusion, Inf. Fusion 69 (2020) 128–141.
[37] H. Li, L.M. Zhang, M.R. Jiang, Y.L. Li, Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network, Pattern Recogn. Lett. 141 (2021) 45–53.
[38] Y.F. Xiao, Z.X. Guo, P. Veelaert, W.D.M.D.N. Philips, Degradation model-based deep network for multi-focus image fusion, Signal Process. Image Commun. 101 (2022), 116554.
[39] J.Y. Ma, L.F. Tang, M.L. Xu, H. Zhang, G. Xiao, STDFusionNet: an infrared and visible image fusion network based on salient target detection, IEEE Trans. Instrum. Meas. 70 (2021) 1–13.
[40] H.H. Song, W.J. Xu, D. Liu, B. Liu, Q. Liu, D.N. Metaxas, Multi-stage feature fusion network for video super-resolution, IEEE Trans. Image Process. 30 (2021) 2923–2934.
[41] H.T. Mustafa, J. Yang, H. Mustafa, M. Zareapoor, Infrared and visible image fusion based on dilated residual attention network, Optik 224 (2020), 165409.
[42] T. Zhou, X.Y. Ye, H.L. Lu, Dense convolutional network and its application in medical image analysis, BioMed Res. Int. 2022 (2022) 22.
[43] H.T. Mustafa, M. Zareapoor, Jie Yang, MLDNet: multi-level dense network for multi-focus image fusion, Signal Process. Image Commun. 85 (2020), 115864.
[44] D.D. Xu, Y.C. Wang, X. Zhang, N. Zhang, S. Yu, Infrared and visible image fusion using a deep unsupervised framework with perceptual loss, IEEE Access 8 (2020) 206445–206458.
[45] R.C. Hou, D.M. Zhou, R.C. Nie, VIF-net: an unsupervised framework for infrared and visible image fusion, IEEE Trans. Comput.Imag. 6 (2020) 640–651.
[46] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generat. Advers.Netw. 1406 (2014) 2661.
[47] J.Y. Ma, W. Yu, P.W. Liang, L. Chang, J.FusionGAN. Jiang, A generative adversarial network for infrared and visible image fusion, Inf. Fusion 48 (2019) 11–26.
[48] Y.C. Wang, S. Xu, J.M. Liu, Z.X. Zhao, C.X. a Zhang, J.S. Zhang, MFIF-GAN: a new generative adversarial network for multi-focus image fusion, Signal Process. Image Commun. 96 (2021), 116295.
[49] Hao Yin, Z.H. Ou, Z.B. Zhu, A novel asexual-reproduction evolutionary neural network for wind power prediction based on generative adversarial networks, Energy Convers. Manag. 247 (2021), 114714.
[50] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: IEEE International Conference on Computer Vision (ICCV) vol. 1109, 2017, pp. 2242–2251.
[51] C.J. Wang, G. Yang, G. Papanastasiou, S.A. Tsaftaris, D.E. Newby, C. Gray, G. Macnaught, T.J. MacGillivray, DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis, Inf. Fusion 67 (2021) 147–160.
[52] J.Y. Ma, H. Xu, J.J. Jiang, X. Mei, X.P. Zhang, DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion, IEEE Trans. Image Process. 29 (2020) 4980–4995.
[53] Q. Li, L. Lu, Z. Li, W. Wu, X. Yang, Coupled GAN with relativistic discriminators for infrared and visible images fusion, IEEE Sensor. J. 21 (2021) 7458–7467.
[54] J. Huang, Z.l. Li, Y. Ma, F. Fan, H. Zhang, L. Yang, MGMDcGAN: medical image fusion using multi-generator multi-discriminator conditional generative adversarial network, IEEE Access 8 (2020) 55145–55157.
[55] H. Li, X.j Wu, DenseFuse: a fusion approach to infrared and visible images, IEEE Trans. Image Process. 28 (2018) 2614–2623, 2019.
[56] A. Raza, H. Huo, Tao Fang, PFAF-net: pyramid feature network for multimodal fusion, IEEE Sens. Lett. 4 (2020) 1–4.
[57] L.H. Jian, X.M. Yang, Z. Liu, G. Jeon, M. Gao, D. Chisholm, SEDRFuse: a symmetric encoder–decoder with residual block network for infrared and visible image fusion, IEEE Trans. Instrum. Meas. 70 (2021) 1–15.
[58] J.Y. Ma, Z.L. Le, X. Tian, J. Jiang, SMFuse: multi-focus image fusion via self-supervised mask-optimization, IEEE Trans. Comput.Imag. 7 (2021) 309–320.
[59] A. Chartsias, T. Joyce, M.V. Giuffrida, Multimodal MR synthesis via modality-invariant latent representation, IEEE Trans. Med. Imag. 37 (2018) 803–814.
[60] M. Wang, X.W. Liu, H.P. Jin, A generative image fusion approach based on supervised deep convolution network driven by weighted gradient flow, Image Vis Comput. 86 (2019) 1–16.
[61] A.Q. Fang, X.B. Zhao, J.Q. Yang, B.B. Qin, Y.N. Zhang, A light-weight, efficient, and general cross-modal image fusion network, Neurocomputing 463 (2021) 198–211.
[62] C.C. Wang, D.M. Zhou, Y.S. Zang, R. Nie, Y. Guo, A deep and supervised atrous convolutional model for multi-focus image fusion, IEEE Sensor. J. 21 (2021) 23069–23084.
[63] X. Deng, Y.T. Zhang, M. Xu, S. Gu, Y. Duan, Deep coupled feedback network for joint exposure fusion and image super-resolution, IEEE Trans. Image Process. 30 (2021) 3098–3112.
[64] H.F. Li, Y.L. Cen, Yu Liu, X. Chen, Z. Yu, Different input resolutions and arbitrary output resolution: a meta learning-based deep framework for infrared and visible image fusion, IEEE Trans. Image Process. 30 (2021) 4070–4083.
[65] J.Y. Ma, P.W. Liang, W. Yu, C. Chen, X.J. Guo, J. Wu, J.J. Jiang, Infrared and visible image fusion via detail preserving adversarial learning, Inf. Fusion 54 (2020) 85–98.
[66] J.T. Xu, X.P. Shi, S.Z. Qin, K.G. Lu, H. Wang, J.G. Ma, LBP-BEGAN: a generative adversarial network architecture for infrared and visible image fusion, Infrared Phys. Technol. 104 (2020), 103144.
[67] J. Li, H.T. Huo, K.J. Liu, C. Li, Infrared and visible image fusion using dual discriminators generative adversarial networks with Wasserstein distance, Inf. Sci. 529 (2020) 28–41.
[68] Y. Fu, X.J. Wu, T. Durrani, Image fusion based on generative adversarial network consistent with perception, Inf. Fusion 72 (2021) 110–125.
[69] X.M. Liu, A.H. Yu, X.K. Wei, Z.F. Pan, J.S. Tang, Multimodal MR image synthesis using gradient prior and adversarial learning, IEEE J.Select.Topic Sig.Process. 14 (2020) 1176–1188.
[70] H. Zhang, Z.L. Le, Z.F. Shao, H. Xu, J.Y. Ma, MFF-GAN: an unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion, Inf. Fusion 66 (2021) 40–53.
[71] J. Li, H.T. Huo, C. Li, R. Wang, C. Sui, Z. Liu, Multigrained attention network for infrared and visible image fusion, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.
[72] C. Yuan, C.Q. Sun, X.Y. Tang, FLGC-fusion GAN: an enhanced fusion GAN model by importing fully learnable group convolution, Math. Probl Eng. 11 (2020) 1–13.
[73] Fu Yu, J.W. Xiao, T. Durrani, Image fusion based on generative adversarial network consistent with perception, Inf. Fusion 72 (2021) 110–125.
[74] X.Q. Luo, A.Q. Wang, Z.C. Zhang, X.G. Xiang, X.J. Wu, LatRAIVF: an infrared and visible image fusion method based on latent regression and adversarial training, IEEE Trans. Instrum. Meas. 70 (2021) 1–16.
[75] S. Yi, J.J. Li, X.S. Yuan, DFPGAN: dual fusion path generative adversarial network for infrared and visible image fusion, Infrared Phys. Technol. 119 (2021), 103947.
[76] A. Kumar, M. Fulham, D. Feng, J. Kim, Co-learning feature fusion map from PET-CT images of lung cancer, IEEE Trans. Med. Imag. 39 (2020) 204–217.
[77] W.B. An, H.M. Wang, Infrared and visible image fusion with supervised convolutional neural network, Optik 219 (2020), 165120.
[78] H. Li, X.J. Wu, J. Kittler, RFN-Nest: an end-to-end residual fusion network for infrared and visible images, Inf. Fusion 73 (2021) 72–86.

[79] Q. Zuo, J.P. Zhang, Y. Yang, DMC-fusion: deep multi-cascade fusion with classifier-based feature synthesis for medical multi-modal images, IEEE J. Biomedi. Health Inf. 25 (2021) 3438–3449.
[80] A. Raza, J.D. Liu, Y.F. Liu, Z. Li, T. Fang, IR-MSDNet: infrared and visible image fusion based on infrared features and multiscale dense network, IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14 (2021) 3426–3437.
[81] X.L. Hou, J.C. Zhang, P.P. Zhou, Reconstructing a high dynamic range image with a deeply unsupervised fusion model, IEEE Photon. J. 13 (2021) 1–10.
[82] L. Ren, Z.B. Pan, J.Z. Cao, J. Liao, Infrared and visible image fusion based on variational auto-encoder and infrared feature compensation, Infrared Phys. Technol. 117 (2021), 103839.
[83] Shuaiqi Liu, Siyu Miao, Jian Su, UMAG-net: a new unsupervised multiattention-guided network for hyperspectral and multispectral image fusion, IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14 (2021) 7373–7385.
[84] T. Zhou, H.L. Lu, Z.L. Yang, B.Q. Huo, The ensemble deep learning model for novel COVID-19 on CT images, Appl. Soft Comput. 98 (2021), 106885.
[85] S. Kligerman, S. Digumarthy, Staging of non–small cell lung cancer using integrated PET/CT, Ajr Am. J.Roentgenol. 193 (2009) 1203–1211.
[86] T.M. Blodgett, C.C. Meltzer, D.W. Townsend, PET/CT: form and function, Radiology 242 (2007) 360–385.
[87] A. James, B.V. Dasarathy, Medical image fusion: a survey of the state of the art, Inf. Fusion 19 (2014) 4–19.
[88] T. Zhou, X.Y. Chang, H.L. Lu, X.Y. Ye, Y.C. Liu, Pooling in deep learning : from “invariable” to “variable”, BioMed Res. Int. 2022 (2022) 17.
[89] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2006) 430–444.
[90] Q. Wang, Y. Shen, Performance evaluation of image fusion techniques, Imag. Fusion 19 (2008) 469–492.
[91] G.H. Qu, D.L. Zhang, P.F. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (2002) 313–315.
[92] M.B.A. Haghighat, A. Aghagolzadeh, H. Seyedarabi, A nonreference image fusion metric based on mutual information of image features, Comput. Electr. Eng. 37 (2011) 744–756.
[93] M. Hossny, S. Nahavandi, D. Creighton, Comments on information measure for performance of image fusion, Electron. Lett. 44 (2008) 1066–1067.
[94] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2003) 600–612.
[95] J.W. Roberts, A.J.A. Van, F.B. Ahmed, Assessment of image fusion procedures using entropy image quality and multispectral classification, J. Appl. Remote Sens. 2 (2008) 1–28.
[96] G.M. Cui, H.J. Feng, Z.H. Xu, Q. Li, Y. Chen, Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition, Optics 341 (2014) 199–209.
[97] C.S. Xydeas, V. Petrovic, Objective image fusion performance measure, Military Tech. Cour. 56 (2000) 181–193.
[98] Y.J. Rao, In-fibre Bragg grating sensors, Meas. Sci. Technol. 8 (1988) 355.
[99] A.M. Eskicioglu, P.S. Fisher, Image quality measures and their performance, IEEE Trans. Commun. 43 (1995) 2959–2965, https://doi.org/10.1109/26.477498.
[100] B. Rajalingam, R. Priya, Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis, Int. J. Eng. Sci. Invent. 2 (2018) 52–60.
[101] P. Jagalingam, A.V. Hegde, A review of quality metrics for fused image, Aquatic Proc. 4 (2015) 133–142.