

文章编号 1004-924X(2024)02-0252-16

# 基于结构功能交叉神经网络的多模态 医学图像融合

邸 敬, 郭文庆\*, 任 莉, 杨 燕, 廉 敬  
(兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070)

**摘要:**针对多模态医学图像融合中存在纹理细节模糊和对比度低的问题,提出了一种结构功能交叉神经网络的多模态医学图像融合方法。首先,根据医学图像的结构信息和功能信息设计了结构功能交叉神经网络模型,不仅有效地提取解剖学和功能学医学图像的结构信息和功能信息,而且能够实现这两种信息之间的交互,从而很好地提取医学图像的纹理细节信息。其次,利用交叉网络通道和空间特征变化构造了一种新的注意力机制,通过不断调整结构信息和功能信息权重来融合图像,提高了融合图像的对比度和轮廓信息。最后,设计了一个从融合图像到源图像的分解过程,由于分解图像的质量直接取决于融合结果,因此分解过程可以使融合图像包含更多的细节信息。通过与近年来提出的 7 种高水平方法相比,本文方法的 AG, EN, SF, MI,  $Q^{AB/F}$  和 CC 客观评价指标分别平均提高了 22.87%, 19.64%, 23.02%, 12.70%, 6.79%, 30.35%, 说明本文方法能够获得纹理细节更清晰、对比度更好的融合结果,在主观视觉和客观指标上都优于其他对比算法。

**关键词:**多模态医学图像融合;结构功能信息交叉网络;注意力机制;分解网络  
**中图分类号:**TP391 **文献标识码:**A **doi:**10.37188/OPE.20243202.0252

## Multimodal medical image fusion method based on structural functional cross neural network

DI Jing, GUO Wenqing\*, REN Li, YANG Yan, LIAN Jing

(School of Electronic and Information Engineering, Lanzhou Jiaotong University,  
Lanzhou 730070, China)

\* Corresponding author, E-mail: 344385945@qq.com

**Abstract:** To solve the problems of texture detail blurring and low contrast in multimodal medical image fusion, a multimodal medical image fusion method with structural-functional crossed neural networks was proposed. Firstly, this method designed a structural and functional cross neural network model based on the structural and functional information of medical images. Within each structural-functional cross module, a residual network model was also incorporated. This approach not only effectively extracted the structural and functional information from anatomical and physiological medical images but also facilitated interaction between structural and functional information. As a result, it effectively captured texture details from multi-source medical images, creating fused images that closely align with human visual characteris-

收稿日期:2023-05-05;修订日期:2023-07-13.

基金项目:甘肃省科技计划资助项目(No. 22JR5RA360);国家自然科学基金资助项目(No. 62061023);甘肃省杰出青年基金资助项目(No. 21JR7RA345);甘肃省教育科技创新产业支撑项目(No. 2021CYZC-04)

tics. Secondly, a new attention mechanism module was constructed by utilizing the effective channel attention mechanism and spatial attention mechanism model (ECA-S), which continuously adjusted the weights of structural and functional information to fuse images, thereby improving the contrast and contour information of the fused image, and to make the fused image color more natural and realistic. Finally, a decomposition process from the fused image to the source image was designed, and since the quality of the decomposed image depends directly on the fusion result, the decomposition process could make the fused image contain more texture detail information and contour information of the source image. By comparing with seven high-level methods for medical image fusion proposed in recent years, the objective evaluation indexes of AG, EN, SF, MI,  $Q^{AB/F}$  and CC of this paper's method are improved by 22.87%, 19.64%, 23.02%, 12.70%, 6.79% and 30.35% on average, respectively, indicating that this paper's method can obtain fusion results with clearer texture details, higher contrast and better contours in subjective visual and objective indexes are better than other seven high-level contrast methods.

**Key words:** multimodal medical image fusion; structural and functional information cross-interacting network; attention mechanism; decomposition network

## 1 引言

现代医学影像设备提供了人体不同部位的病变图像,协助医生对疾病进行快速诊断和治疗。多模态医学图像通常分为两种类型:解剖医学图像和功能医学图像<sup>[1]</sup>。解剖医学图像有较高的分辨率,能够清晰地提供人体器官的结构信息,如计算机断层扫描(Computer Tomography, CT)和核磁共振成像(Magnetic Resonance Imaging, MRI)图像。功能医学图像主要反映人体器官的代谢信息,如单光子发射计算机断层扫描(Single-Photon Emission Computed Tomography, SPECT)和正电子发射计算机断层扫描(Positron Emission Tomography, PET)图像。由于单模态医学图像成像单一,不能提供病灶的准确信息。因此,多模态医学图像融合被广泛用于临床分析,实现对病灶部位准确、全面地描述<sup>[2]</sup>。

目前,图像融合分为特征级、决策级和像素级三个层次。特征级融合虽然降低了算法复杂度,但是融合图像容易丢失部分细节信息。决策级融合具有较好的实时性和自适应性,但会增加实验数据误差,导致图像信息损失最大。像素级融合是对图像的像素点进行信息融合的过程,也是当前主流研究方法。目前医学图像融合算法分为传统融合方法和深度学习融合方法。多尺

度变换(MultiScale Transform, MST)是一类典型的传统融合方法,常用的MST融合算法包括小波变换<sup>[3]</sup>、金字塔变换<sup>[4]</sup>、非下采样剪切波变换(Non-Subsampled Shearlet Transform, NSST)<sup>[5]</sup>、非下采样轮廓波变换(Non-Subsampled Contourlet Transform, NSCT)<sup>[6]</sup>。近年来,诸多学者提出了许多人工神经网络方法用于模拟动物神经系统原理。例如,脉冲耦合神经网络(Pulse Coupled Neural Network, PCNN)<sup>[7]</sup>,它正是根据小型哺乳动物视觉皮层神经系统模拟产生的,具有处理复杂信息的能力。PCNN模型最早由Eckhorn等人提出<sup>[8]</sup>,此方法无需训练测试数据,网络框架可以直接提取特征信息。目前,深度学习的热潮迅速席卷全球,基于深度学习的方法在图像配准<sup>[9]</sup>、病变分割<sup>[10]</sup>、图像融合<sup>[11-12]</sup>等高级视觉任务中取得了较好的成绩。

然而,基于MST融合算法需要多步分解操作。因此,它的适应性不好。基于PCNN的融合算法模型中有大量的参数,增加了复杂度。基于深度学习的医学图像融合算法虽然结构简单、易于实现、适应性强,但在提取图像纹理特征、对比度、多任务融合能力和效率方面都需要提高。WANG等提出一种模态合成和增强的弱监督医学图像融合方法<sup>[13]</sup>,能有效增强MRI图像的纹理和结构,但在此过程中忽略了CT图像的骨骼信息,导致边缘轮廓模糊。陈永等将注意力机制引

入图像融合中<sup>[14]</sup>,其融合图像获得了源图像较多信息,但在融合模型中使用了单分支神经网络,并没有信息之间的交互,因此纹理细节较差。FANG等提出了卷积神经网络和注意力机制的多模态医学图像融合<sup>[15]</sup>,通过减少冗余信息的影响来提高模型的整体性能,但减少模型复杂度的过程中纹理细节严重丢失,并且仅适用医学领域,实时性较差。TANG等人提出了多尺度自适应变换器来融合多模态医学图像<sup>[16]</sup>,从不同尺度的角度充分获取了图像的结构信息,但在此过程中忽略了图像的功能信息,导致软组织信息丢失严重。ZHANG等人提出了端到端的融合分解网络<sup>[17]</sup>,通过融合分解使图像细节纹理更加清晰,但此网络在损失函数中的梯度损失和强度损失通过调整权重比例会造成融合图像轮廓信息丢失、对比度低,从而降低了医学图像融合结果质量。

为此,针对融合图像纹理细节模糊、边缘轮廓缺失和对比度低等问题,本文提出了一种结构功能交叉神经网络的多模态医学图像融合方法,采用结构功能模块和功能结构模块构成的交叉网络实现信息交互,提升融合图像的纹理细节信息。使用有效通道注意力机制和空间注意力机

制 (Effective Channel Attention mechanism and Spatial attention mechanism model, ECA-S)结合的改进注意力机制进行图像融合,来增强融合图像对比度。设计了融合图像到源图像的分解过程,使融合图像包含更多的边缘轮廓信息。本文的多模态医学图像融合是基于结构和功能信息的提取与重建,并设计了由结构项和功能项组成的损失函数。大量实验结果表明,本文方法融合的图像不但纹理细节清晰,而且图像整体对比度更高。最后将本方法扩展到红外与可见光和多聚焦图像融合,也取得了较好的融合结果,说明本文方法具有适用性。

## 2 本文方法

为了解决多模态医学图像融合结果中出现纹理细节模糊、边缘轮廓缺失和对比度低的问题,本文提出了基于结构功能交叉神经网络的多模态医学图像融合模型,网络总体框架如图1所示。在融合网络模型中,两种单模态医学图像作为输入,通过结构功能模块和功能结构模块提取图像特征信息,ECA-S注意力机制融合出最终图像。在分解网络模型中,由残差网络组成的分解网络迫使融合图像包含更丰富的信息。

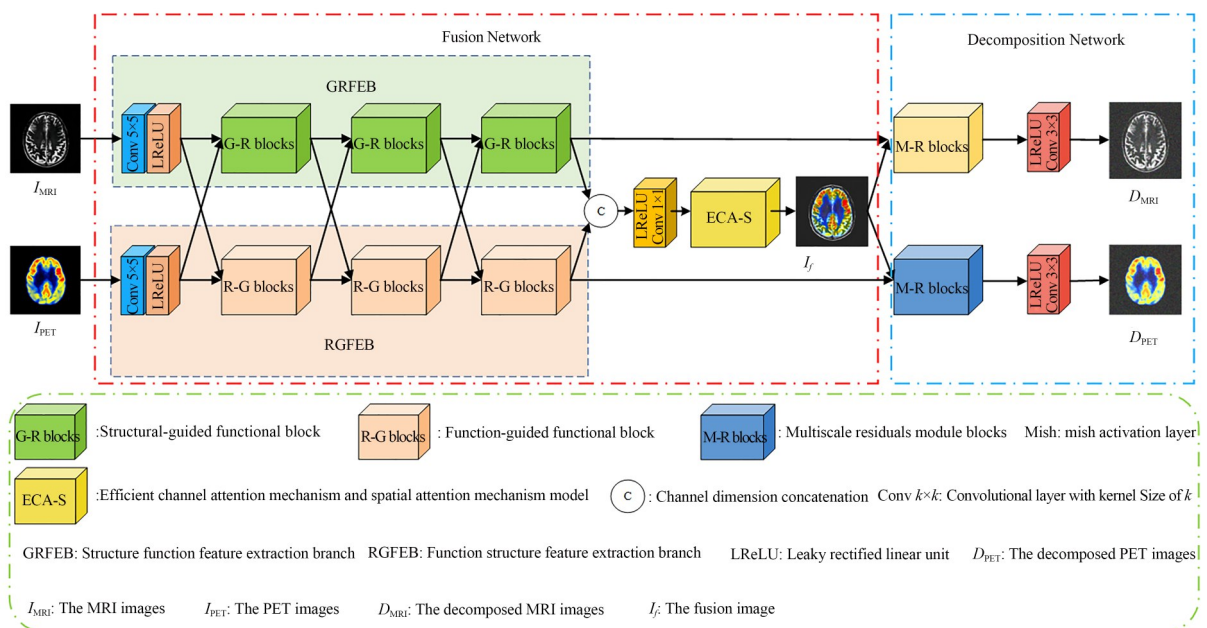


图1 网络总体框架图

Fig. 1 General network framework diagram

### 2.1 结构功能交叉神经网络和注意力机制融合图像

为了使细节信息被充分提取,本文将交互神经网络引入图像融合中。将源图像分别馈送到结构功能特征提取分支(GRFEB)和功能结构特征提取分支(RGFEB)模块。GRFEB包括一个核为 $5 \times 5$ 的卷积和激活函数为 Leaky ReLU (LReLU)的低级特征提取块以及三个结构引导功能块(G-R块)组成。RGFEB由低级特征提取块和三个功能引导结构块(R-G块)组成,如图1所示。在G-R块和R-G块上设计了多个交叉连接,以实现GRFEB和RGFEB之间的结构和功能信息交互。每个G-R块的输入不仅包括GRFEB中前一块的输出,还包括RGFEB中对应的前一块的输出。R-G块也是如此。第三个G-R块和第三R-G块的输出在通道维度上级联,并馈送到ECA-S注意力机制,得到最终的融合图像。表1列出了所提出方法模型的融合网络参数和分解网络参数,其中卷积层在每个块或模块中

从左到右、从下到上编号。

#### 2.1.1 G-R 模块和 R-G 模块

G-R模块的网络架构如图2所示。G-R块的输入包括GRFEB和RGFEB中先前块的输出,即 $I_{n-1}^G$ 和 $I_{n-1}^R$ ,其中 $n \in [1, 3]$ 表示特征提取分支中的块索引。为了从MRI图像中充分提取特征,使用了两阶段残差结构。具体来说, $I_{n-1}^G$ 被送入第一级残差结构,该结构由两个 $3 \times 3$ 卷积层和LReLU层组成。输入 $I_{n-1}^G$ 与输出相加构成残差连接。 $I_{n-1}^R$ 由相同的残差结构处理,以加入从PET或SPECT图像中提取的特征。这两种残差结构的输出在通道中级联,然后用于特征融合和 $3 \times 3$ 卷积层进行通道降维。第二阶段残差结构用于增强MRI图像中的结构信息。原始输入 $I_{n-1}^R$ 与 $1 \times 1$ 卷积层的输出相加得到最终输出 $I_n^R$ 。R-G模块也采用了两级残差结构,如图3所示。不同之处在于第二级残差连接是在RGFEB中的先前块的输出上执行的(即 $I_{n-1}^G$ ),以增强PET或SPECT图像中的功能信息。

表 1 网络模型具体参数

Tab. 1 Specific parameters of network model

Module	Layer	$k$	$S$	$P$	$I$	$O$
Input	Conv	$5 \times 5$	1	1	1	32
	Conv1	$3 \times 3$	1	1	32	32
G-R Block	Conv2	$3 \times 3$	1	1	32	32
	Conv3	$3 \times 3$	1	1	32	32
	Conv4	$3 \times 3$	1	1	32	32
	Conv5	$3 \times 3$	1	1	64	32
	Conv1	$3 \times 3$	1	1	32	32
R-G Block	Conv2	$3 \times 3$	1	1	32	32
	Conv3	$3 \times 3$	1	1	32	32
	Conv4	$3 \times 3$	1	1	32	32
	Conv5	$3 \times 3$	1	1	64	32
	Conv1	$1 \times 1$	1	1	33	64
M-R Blocks	Conv2	$3 \times 3$	1	1	64	64
	Conv3	$1 \times 1$	1	1	64	16
	Conv4	$3 \times 3$	1	1	16	16
	Conv5	$1 \times 1$	1	1	16	4
	Conv6	$3 \times 3$	1	1	4	4
	Conv7	$1 \times 1$	1	1	68	4
	Conv8	$3 \times 3$	1	1	4	4
	Output2	Conv	$3 \times 3$	1	1	20

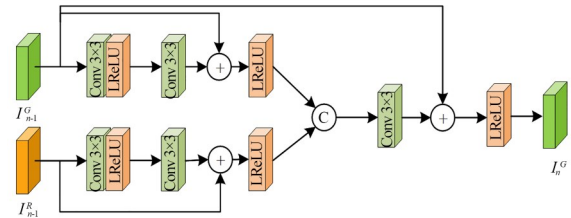


图 2 G-R Block 网络结构

Fig. 2 G-R Block network structure

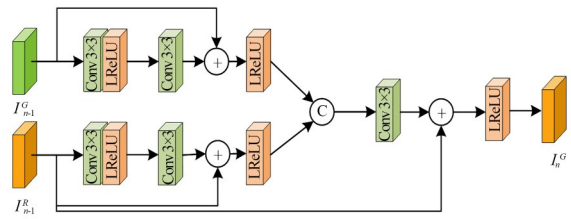


图 3 R-G Block 网络结构

Fig. 3 R-G Block network structure

#### 2.1.2 ECA-S 注意力机制模块

通过剖析有效通道注意力机制<sup>[18]</sup>(Efficient Channel Attention, ECA),发现避免维度缩减对于学习特征信息非常重要。同时注意到医学图像中只有部分部位对融合结果有贡献,而空间注意力机制<sup>[19]</sup>(Spatial Attention Mechanisms,

SAM)主要处理图像重要部位信息。因此,将 ECA 注意力机制和 SAM 注意力机制“并行连接”,使两种注意力模块都直接学习原始的输入特征信息,从而无需关注空间注意力和通道注意力顺序,形成一个新的 ECA-S 注意力机制模型,模型如图 4 所示。

ECA-S 注意力机制的一条分支由 SAM 构成。首先,对输入特征图进行最大池化和平均池化并在通道维度堆叠。其次,使用  $1 \times 1$  大小的卷积核对堆叠特征图特征提取。最后,将提取到的特征信息结果经过 sigmoid 激活函数空间权重归一化,再将输入特征图和权重相乘得到。

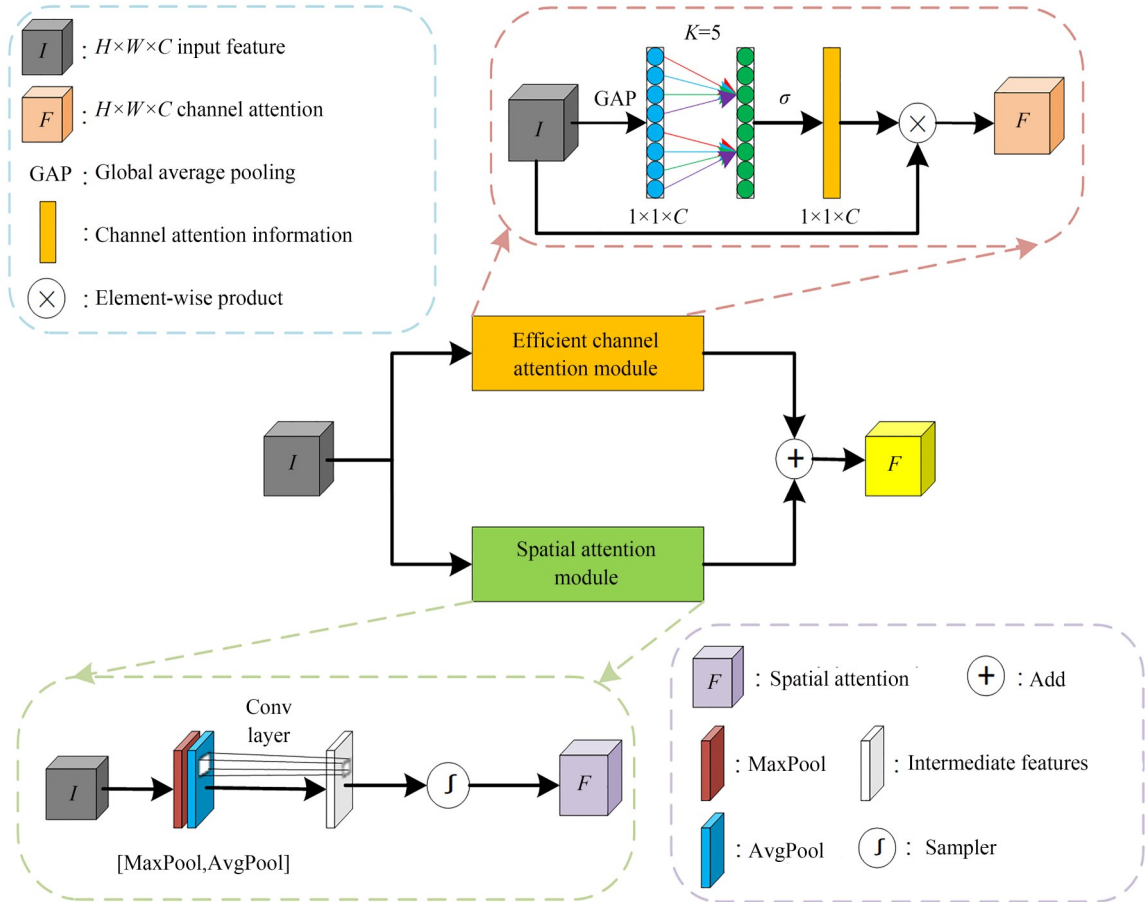


图 4 ECA-S 注意力机制模块

Fig. 4 ECA-S attention mechanism module

另一条分支由 ECA 构成。首先,针对输入的特征层进行全局平均池化。其次,将池化后的特征长条进行卷积提取,通过 sigmoid 激活函数生成每一个特征点的权值。最后,权值与原始特征层结合,便可以得到  $M_{ECA}$ 。将其计算的  $M_S$  与  $M_{ECA}$  相加得到  $M_{ECA,S}$ 。其全部计算过程可以表示为式(1)~式(3):

$$M_S(F) = I * \sigma(f^{1 \times 1} [AvgPool(I); MaxPool(I)]), \quad (1)$$

$$M_{ECA}(F) = I * \sigma(f^{k \times k} (GlobalAvgPool(I))), \quad (2)$$

$$M_{ECA,S}(F) = M_S(F) \oplus M_{ECA}(F), \quad (3)$$

其中:  $M_S$  为 SAM 输出特征信息,  $M_{ECA}$  为 ECA 输出特征信息,  $M_{ECA,S}$  为 ECA-S 输出特征信息,  $I$  为输入特征信息,  $\sigma$  为 sigmoid 激活函数,  $AvgPool$  表示平均池化,  $MaxPool$  表示最大池化,  $GlobalAvgPool$  表示全局平均池化,  $\oplus$  表示逐元素相加。

## 2.2 基于残差网络的分解模型

为了提取到更丰富的图像信息,本文引入分解网络模型。具体来说,不仅考虑了从源图像到融合结果的融合过程,而且考虑了从融合结果到源图像的分解过程,可以使融合结果包含更多的纹理细节。本文的分解网络模型如图 5 所示,每

一块先经过  $1 \times 1$  的卷积层,再经过  $3 \times 3$  的卷积层,组成多尺度结构,充分提取空间信息,最后再通过 LReLU 激活函数输出特征信息。为了充分分解融合图像的特征信息,分解网络的输入分为两部分,一部分输入为通过 ECA-S 注意力机制融合后的图像,另一部分为 G-R Block 和 R-G Block 的第三层的输出,其目的是经过 ECA-S 注意力机制提取到的是图像的显著信息,而一些不重要的信息却容易被忽略,因此直接将 G-R Block 和 R-G Block 的第三层作为分解网络的一部分输出让分解过程中信息提取更充分。

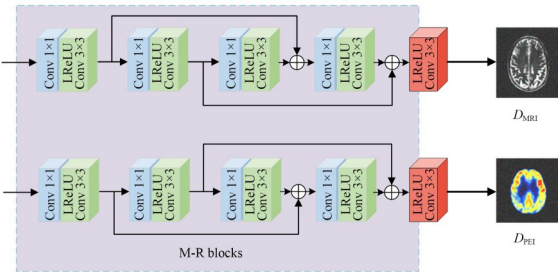


图 5 分解网络流程

Fig. 5 Flowchart of decomposition network

### 2.3 损失函数

本文的网络框架分为两个部分,融合网络特征提取生成单个融合图像。分解网络对结构和功能信息进行分解和重构,分解网络专用于从融合结果分解近似源图像的结果。相应地,损失函数也由两部分组成:融合损失  $L_{sf}$  和分解损失  $L_{dc}$ 。

$$L = L_{sf} + L_{dc}. \quad (4)$$

#### 2.3.1 融合损失

图像中包含的信息可以分为结构信息和功能信息,因此,本文将多模态医学图像融合问题转化为提取和重建图像中的结构信息和功能信息的问题,融合损失也由结构损失项和功能损失项组成,损失函数可表示为公式(5):

$$L_{sf} = \beta L_{stru} + L_{func}, \quad (5)$$

其中,  $\beta$  用于平衡结构损失项和功能损失项。

本文设计了一个自适应结构损失函数  $L_{stru}$  来增加融合图像的纹理特征。自适应决策块通过高斯低通滤波器滤除噪声对图像的影响,尽可能减少噪声对决策块的影响。利用 Scharr 算子求出结构图,结构图包含大量的纹理细节信息以及清晰的轮廓信息。之后根据结构的大小生成决

策图。其原理图如图 6 所示。

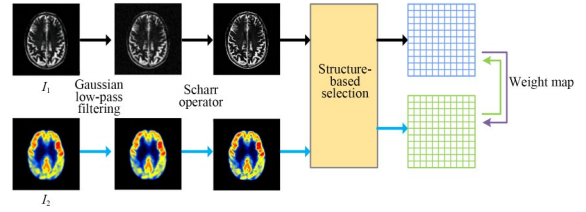


图 6 自适应决策块

Fig. 6 Adaptive weight block

$$L_{stru} = \left( \frac{1}{HW} \sum_i \sum_j S_{1i,j} \left\| \nabla I_{fused,i,j} - \nabla I_{1,i,j} \right\|_F^2 + S_{2i,j} \cdot \left\| \nabla I_{fused,i,j} - \nabla I_{2,i,j} \right\|_F^2 \right)^{\frac{1}{2}}, \quad (6)$$

$$S_{1i,j} = \text{sign} \left( \left| \nabla (L(I_{1i,j})) \right| - \min \left( \left| \nabla (L(I_{1i,j})) \right|, \left| \nabla (L(I_{2i,j})) \right| \right) \right) \quad (7)$$

$$S_{2i,j} = 1 - S_{1i,j}, \quad (8)$$

其中:  $i$  和  $j$  表示像素在第  $i$  行和第  $j$  列权重图,  $H$  和  $W$  代表图像的高度和宽度,  $I_1$  和  $I_2$  为源图像,  $I_{fusion}$  为融合图像,  $|\cdot|$  表示绝对值函数,  $\nabla(\cdot)$  为 Scharr 算子,  $S(\cdot)$  是由决策块基于源图像的结构水平生成的决策图,  $L(\cdot)$  表示高斯低通滤波器函数,  $\min(\cdot)$  表示最小函数,  $\text{sign}(\cdot)$  为符号函数。

功能损失函数  $L_{func}$  影响融合图像的对比度。功能损失函数通过对 Frobenius 范数进行改进,改进后的 Frobenius 范数与欧式距离有所不同,它是矩阵中每一个位置的元素的平方和取平均数再开方,其作用是衡量融合图像像素和源图像像素的距离,有效调整融合图像和两个源图像之间的权重,迫使网络选择更有效的信息。功能损失的公式可以表示为式(9):

$$L_{func} = \sqrt{\frac{1}{HW} \sum_i \sum_j \left\| I_{fused,i,j} - I_{1,i,j} \right\|_F^2 + \alpha \left\| I_{fused,i,j} - I_{2,i,j} \right\|_F^2}. \quad (9)$$

#### 2.3.2 分解损失

分解损失项  $L_{dc}$  要求融合图像的分解结果尽可能与源图像相似,可以表示为:

$$L_{dc} = \frac{1}{HW} \sum_i \sum_j (I_{1,de,i,j} - I_{1i,j})^2 + (I_{2,de,i,j} - I_{2i,j})^2, \quad (10)$$

其中:  $I_{1,de}$  和  $I_{2,de}$  为融合图像的分解结果,  $I_1$  和  $I_2$  为源图像。由于分解后的图像与输入图像之间

的差异程度决定融合结果的效果,因此分解损失函数将尽可能使融合结果获得更多图像信息,从而使融合质量达到最好。

### 3 实验结果和分析

本文模型包括融合网络和分解网络。无论是在训练阶段还是测试阶段,框架都使用 Tensorflow 实现。所有实验均在 Windows 11(64 位),硬件配置为 Intel(R) Core(TM) i7-12700H 2.30 GHz 笔记本电脑和 16 GB RAM 软件上运行。

本节证明本文算法具有更好的性能,验证所提出的融合模型的实用性。对来自公开数据集哈佛医学库(<http://www.med.harvard.edu/aanlib/home.html>)的 200 对  $256 \times 256$  的医学图像,共裁剪成 39 200 个大小为  $120 \times 120$  的图像块进行训练。实验设置 epoch=30, batch\_size=32、学习率为  $1 \times 10^{-4}$ 。在优化算法方面,选择适应性强的自适应优化算法 Adam。此外,在总损失中包含的梯度损失、强度损失、分解损失的比例经过重复实验后设置为 10,1,1。

#### 3.1 客观评价指标

为了进一步说明本文算法的有效性,选择了包括空间频率(SF)<sup>[20]</sup>、平均梯度(AG)<sup>[21]</sup>、融合质量( $Q^{AB/F}$ )<sup>[22]</sup>、相关系数(CC)<sup>[23]</sup>、熵(EN)<sup>[24]</sup>和互信息(MI)<sup>[20]</sup>等客观指标来评价融合图像的结果。

AG 和 EN 是衡量图像清晰度和信息量的指标,数值越大说明从输入图像中提取的信息越丰富和全面。MI 通过度量融合图像和源图像之间的相似程度,即融合图像获取了源图像信息量的多少。互信息越大,质量越好。两幅图像之间的相关性用 CC 表示,数值越接近 1 表示两幅图像之间的特征信息越接近。图像灰度的测量用 SF 表示, SF 值越大,表明融合结果的灰度值越接近输入图像的灰度值。 $Q^{AB/F}$  是对融合图像信息表现程度的一种估计,数值越大说明信息量越大,图像质量越好。

#### 3.2 方法比较

为了验证本文方法的有效性、实用性和时效性,与七种可比较的方法进行了主观视觉评价和客观数据比较: MLEPF<sup>[25]</sup>, NSST<sup>[26]</sup>, LEGFF<sup>[27]</sup>, MATR<sup>[16]</sup>, CNP<sup>[28]</sup>, SDNet<sup>[17]</sup> 和 CFL<sup>[29]</sup>。

##### 3.2.1 主观视觉评价比较

图 7 展示了“轻度阿尔茨海默病”MRI-PET 图像。图 7(a)和图 7(b)为 MRI 和 PET 源图像,图 7(f)和图 7(h)说明该实验结果对应的方法融合后的图像较模糊,存在纹理细节缺失。图 7(c)和图 7(d)实验结果对应的方法融合后的图像软组织中间区域出现紫色块效应。图 7(i)实验结果对应的方法在软组织处亮度过于饱和,丢失 MRI 软组织解剖信息。相比之下,图 7(e)和图 7(g)实验结果对应的方法融合图像获取了较多来自源 MRI 和 PET 图像中的细节信息。但是,目

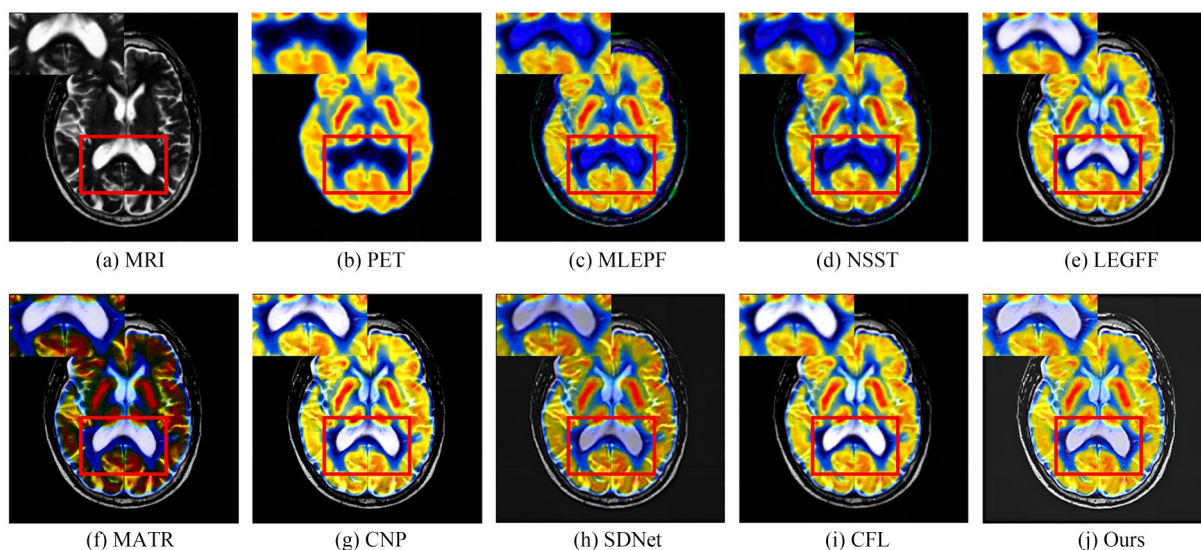


图 7 “轻度阿尔茨海默病”MRI-PET 图像融合比较

Fig. 7 Comparison of MRI-PET image fusion in "mild Alzheimer's disease".

标边缘处比较模糊。对比来看,本文方法结果图 7(j)保留了PET彩色信息,并且获取了较多MRI软组织亮度解剖信息,纹理细节更清晰。

图 8 展示了“转移性支气管癌”MRI-SPECT 图像(彩图见期刊电子版)。图 8(a)和图 8(b)为 MRI 和 SPECT 源图像,图 8(c)和图 8(h)说明该实验结果对应的方法融合后的图像纹路模糊,导致软组织分辨率低。图 8(d)说明该实验结果对应的方法融合后的图像软组织区域出现紫色块效应,无法正常观察病变区域。图 8(e)和图 8(i)说

明该实验结果对应的方法包含了 PET 图像较多的色彩信息,但没有较好得融入 MRI 软组织解剖信息。图 8(f)和图 8(g)说明该实验结果对应的方法内部出现伪影,能量信息严重丢失。观察本文实验结果图 8(j),不仅对比度增强,且软组织纹理细节和轮廓信息特征明显,融合图像色彩更自然。

图 9 展示了“脑膜瘤”MRI-CT 图像。图 9(a)和图 9(b)为 MRI 和 CT 源图像,图 9(c)说明该实验结果对应的方法融合结果图较模糊。图 9(d)说明该实验结果对应的方法融合结果中纹理骨

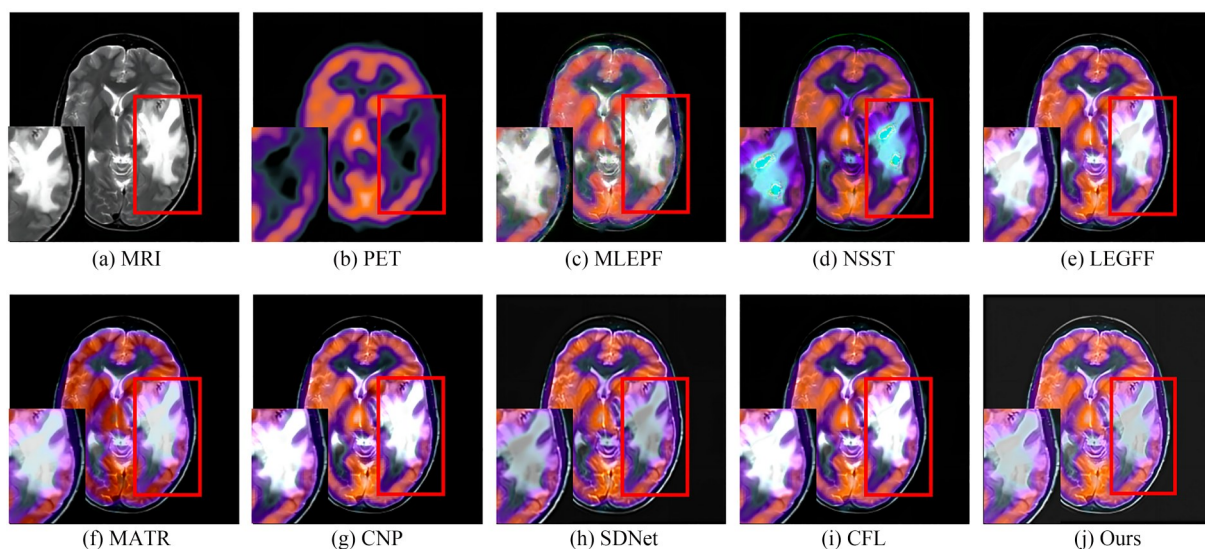


图 8 “转移性支气管癌”MRI-SPECT 图像融合比较

Fig. 8 Comparison of MRI-SPECT image fusion of "metastatic bronchogenic carcinoma".

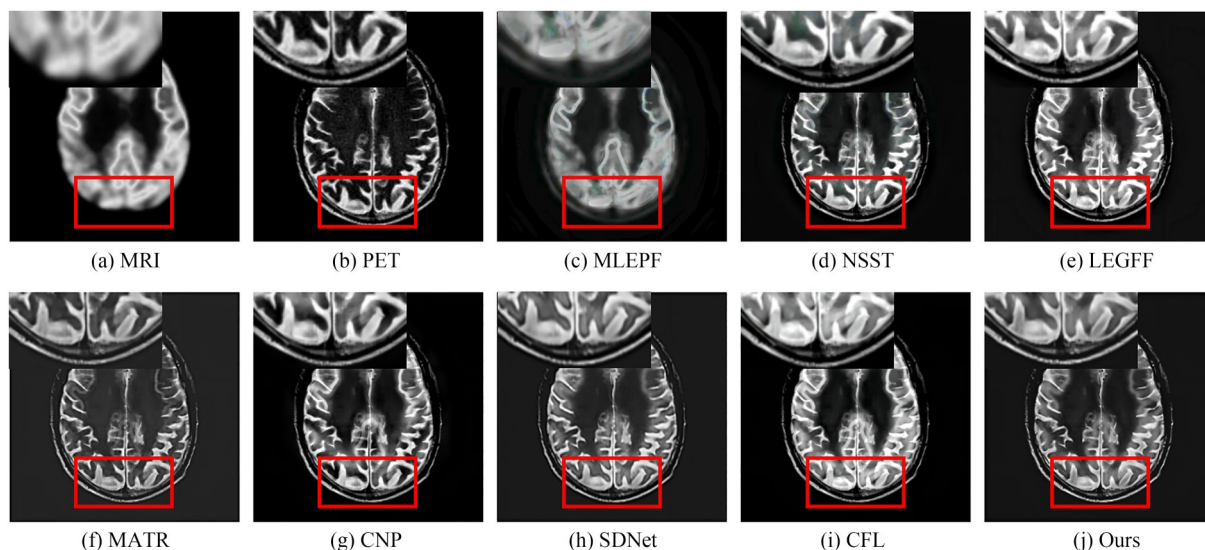


图 9 “脑膜瘤”MRI-CT 图像融合比较

Fig. 9 Comparison of MRI-CT image fusion of "meningioma".



骼出现失真。图 9(f)和图 9(h)说明该实验结果对应的方法融合结果整体偏暗,导致纹理模糊。图 9(e)和图 9(i)说明该实验结果对应的方法融合结果对比度高,但边缘轮廓平滑,无法正确显示骨骼细节特征。综合来看,本文算法融合结果图 9(j)保留了单模态医学图像的特征,能有效地提取软组织细节信息及灰度信息,能综合 MRI 与

CT 源图像丰富信息。

### 3.2.2 客观评价指标

从三组医学图像的融合结果表明,本文提出的结构功能交叉神经网络的多模态医学图像融合策略得到的融合图像高度符合人眼视觉特性,细节提取更充分,色彩更自然。从表 2~表 4 可以看出,AG,EN,SF,MI, $Q^{AB/F}$ 和 CC 等指标均取得

表 2 “轻度阿尔茨海默病”MRI-PET 图像融合客观评价指标

Tab. 2 Objective evaluation indexes of MRI-PET image fusion in "mild Alzheimer's disease"

	AG	EN	SF	MI	$Q^{AB/F}$	CC	Time
MLEPF	8.221 8	4.311 0	40.430 8	2.653 4	0.353 6	0.585 7	7.751 4
NSST	7.962 6	4.205 1	39.369 8	2.514 4	0.367 5	0.276 5	10.595 1
LEGFF	10.986 9	4.793 1	49.882 2	2.303 3	0.573 0	0.284 0	0.572 6
MATR	9.118 2	3.434 9	45.902 4	1.841 9	0.577 8	0.287 0	9.758 7
CNP	10.947 6	3.711 8	51.913 4	2.126 8	0.546 4	0.775 9	6.529 8
SDNet	11.150 9	5.189 2	49.373 4	2.552 6	0.519 7	0.610 4	0.047 0
CFL	11.023 0	3.991 7	52.121 8	2.444 4	0.574 2	0.586 3	34.727 1
Ours	<b>12.763 0</b>	<b>5.590 2</b>	<b>58.207 9</b>	<b>2.683 6</b>	<b>0.580 5</b>	<b>0.783 7</b>	0.116 6

表 3 “转移性支气管癌”MRI-SPECT 图像融合客观评价指标

Tab. 3 Objective evaluation indexes of MRI-SPECT image fusion in "metastatic bronchogenic carcinoma"

	AG	EN	SF	MI	$Q^{AB/F}$	CC	Time
MLEPF	5.692 3	5.058 0	26.656 3	3.457 8	0.506 7	0.860 8	7.553 1
NSST	6.160 9	5.057 0	29.595 9	<b>3.495 7</b>	0.561 6	0.864 8	49.169 4
LEGFF	6.498 0	5.142 9	31.044 2	2.948 8	0.620 4	0.871 9	0.654 8
MATR	5.963 4	4.146 0	30.029 0	2.610 8	<b>0.652 9</b>	0.533 9	0.343 9
CNP	6.616 7	3.975 2	34.040 3	2.599 5	0.604 8	0.635 7	8.965 4
SDNet	7.868 7	5.337 5	36.571 3	2.708 8	0.531 2	0.787 5	0.046 5
CFL	6.492 4	4.442 0	32.997 0	2.948 1	0.621 5	0.840 8	28.659 7
Ours	<b>8.476 5</b>	<b>5.748 2</b>	<b>42.331 2</b>	3.188 5	0.603 5	<b>0.877 5</b>	0.236 5

表 4 “脑膜瘤”MRI-CT 图像融合客观评价指标

Tab. 4 Objective evaluation indexes of "meningioma" MRI-CT image fusion

	AG	EN	SF	MI	$Q^{AB/F}$	CC	Time
MLEPF	4.009 2	5.124 5	17.820 1	1.531 4	0.265 7	0.659 4	9.632 5
NSST	8.486 1	5.232 2	41.925 7	2.819 9	<b>0.644 0</b>	0.891 0	8.734 9
LEGFF	8.885 1	5.223 5	42.716 3	2.949 1	0.617 1	0.849 0	0.438 3
MATR	8.296 7	4.896 8	40.518 5	3.314 2	0.607 2	0.647 3	0.288 3
CNP	<b>8.950 9</b>	5.019 5	41.972 8	3.329 5	0.625 2	0.707 7	3.587 8
SDNet	8.201 1	4.772 6	38.926 9	3.113 4	0.549 8	0.899 3	0.040 0
CFL	8.320 0	4.545 0	40.693 3	3.038 7	0.580 2	0.799 0	14.614 2
Ours	8.578 4	<b>5.248 0</b>	<b>41.987 2</b>	<b>3.339 9</b>	0.564 5	<b>0.905 2</b>	0.224 0

了较好的性能,但 AG 和  $Q^{AB/F}$  在 MRI-PET 和 MRI-CT 图像中表现稍弱,这是因为在 LEGFF, NSST, PCNN 方法中过度去关注对比度,是以图像的亮度去换取图像的细节信息,如在 LEGFF 方法融合 MRI-CT 图像时,对比度指标 AG 较高,但 MI 却低于本文方法,说明信息提取较弱。在 NSST 方法融合 MRI-PET 图像中,整体亮度较高但软组织处出现失真,出现蓝色斑块,导致信息丢失。因此,与其他对比算法相比,本文的算法在融合过程中更注重 PET 图像和 SPECT 图像的色彩丰富度、MRI 图像的边缘轮廓信息和软组织信息,以及 CT 图像的骨骼信息,使融合后的图像轮廓边缘更清晰,色彩更自然,更符合人眼视觉特性,可以辅助医生快速诊断和治疗疾病。

### 3.3 消融实验

为了验证本文方法中结构功能交叉网络、改进注意力机制、分解网络的有效性,设计了四组消融实验。如图 12 所示,所提出的方法被降级为三种不同的网络结构用于测试。

第一组实验将交叉网络变为双通道网络,每一个 G-R Block 和 R-G Block 模块变为蓝色区域的残差网络,其他网络结构不变,记为 ECA\_S+DeNet。第二组实验将改进注意力机制取消,其余网络结构不变,记为 FuNet+DeNet。第三组

实验将分解网络取消,其他网络结构均不变,记为 ECA\_S+FuNet。第四组实验将结构功能交叉网络、改进注意力机制、分解网络都放入模型中去观察融合结果,记为 All。实验的具体流程如图 12 所示。

如图 13 四种不同网络结构 20 幅图像均值柱状图所示,在六个指标中完整架构在所有指标都获得最高值,这证明了在本文网络中采用的结构功能交叉神经网络、改进注意力机制和分解网络的重要性。图 14 展示了由不同模型生成的两组融合结果,可以看出,四幅图视觉质量差异显著,完整模型在细节保留、对比度和边缘轮廓方面都优于其他三个降级模型。

### 3.4 效率对比

方法的运行时间也是评价性能的一个重要指标。表 2~表 6 记录了全部算法的运行时间,可以看出本文方法的网络运行时间略高于 SDNet,但 SDNet 算法融合图像的纹理细节和对比度都较差。因此,综合来看,本文算法的效率在时间和性能方面更好。

### 3.5 其他邻域应用

为了验证本文提出的算法在其他邻域的图像融合是否表现的同样有效,将其扩展到多焦点、红外和可见光图像。融合后的图像如图 10 和

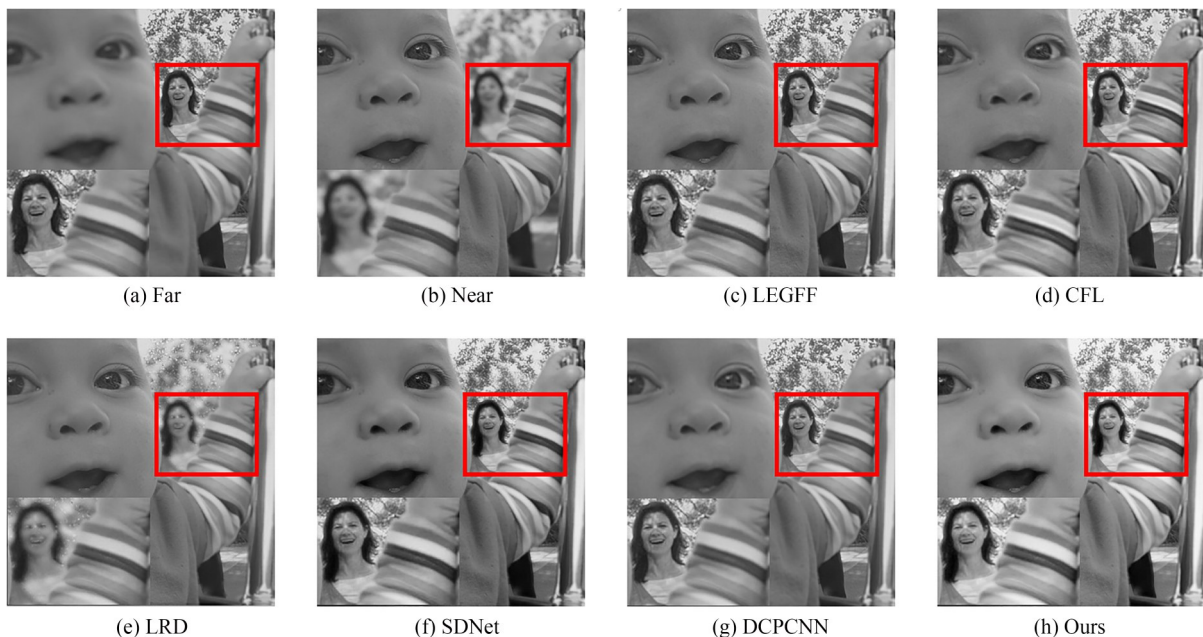


图 10 “人物”多聚焦图像融合比较

Fig. 10 "Person" multi-focus image fusion comparison

图 11 所示,表 5 和表 6 中提供了客观评价指标。可以看出,本文设计的结构功能交叉神经网络、改进注意力机制和基于残差网络的分解模型与人类视觉系统特点高度吻合。图 10(a)和图

10(b)显示了多聚焦输入图像,图 11(a)和图 11(b)显示了红外与可见光输入图像。本文融合方法与现有的 LEGFF<sup>[27]</sup>,SDNet<sup>[17]</sup>,CFL<sup>[29]</sup>,LRD<sup>[30]</sup>,DCPCNN<sup>[31]</sup>,SMVIF<sup>[32]</sup>,ResNet<sup>[33]</sup>和

表 5 “人物”多聚焦图像融合客观评价指标

Tab. 5 Objective evaluation index of multi-focus image fusion of "people"

	AG	EN	SF	MI	$Q^{AB/F}$	CC	Time
LEGFF	5.275 5	7.183 6	17.516 4	6.694 3	0.708 0	0.985 4	0.438 3
CFL	5.064 3	7.216 9	17.219 7	6.093 7	0.697 6	0.987 2	10.555 4
LRD	3.980 7	7.165 4	15.009 1	5.645 2	0.561 2	0.958 0	193.902 3
SDNet	4.763 3	7.445 8	<b>17.838 0</b>	6.822 6	0.633 4	0.985 9	0.040 0
DCPCNN	3.939 3	7.154 7	16.524 9	7.057 9	0.684 4	0.976 4	11.689 1
Ours	<b>5.931 6</b>	<b>7.496 6</b>	17.694 0	<b>7.485 8</b>	<b>0.719 1</b>	<b>0.987 9</b>	0.612 8

表 6 3种不同场景下红外与可见光图像融合客观评价指标

Tab. 6 Objective evaluation index of infrared and visible image fusion in three different scenes

Image	Method	AG	EN	SF	MI	$Q^{AB/F}$	CC	Time
helicopter	LEGFF	5.019 1	5.743 1	12.305 5	1.788 4	0.464 5	0.782 5	0.704 7
	CFL	1.916 9	5.604 3	5.842 8	2.490 5	0.417 4	0.429 0	15.712 8
	CCF	3.469 7	5.625 7	8.412 1	<b>2.678 2</b>	0.460 5	0.452 7	2.752 6
	SMVIF	3.205 5	5.024 2	7.829 2	1.796 4	<b>0.510 1</b>	0.753 5	6.923 1
	ResNet	1.792 5	4.859 4	4.502 1	2.114 7	0.397 7	0.783 3	11.689 1
	Ours	<b>5.634 8</b>	<b>6.359 3</b>	<b>15.662 6</b>	1.870 9	0.299 6	<b>0.914 4</b>	<b>0.195 6</b>
Nato camp	LEGFF	6.899 5	6.372 1	18.459 5	1.811 3	<b>0.441 0</b>	0.487 9	0.446 8
	CFL	6.812 7	4.860 5	15.291 8	2.143 5	0.414 4	0.668 3	17.998 4
	CCF	5.470 9	<b>6.793 4</b>	16.148 1	2.161 9	0.423 3	0.651 5	31.268 2
	SMVIF	6.330 8	5.278 2	15.644 2	1.523 5	0.425 2	0.516 6	6.521 4
	ResNet	6.250 5	3.286 8	9.928 9	1.622 6	0.345 1	0.517 4	1.423 6
	Ours	<b>6.920 9</b>	6.386 0	<b>18.601 5</b>	<b>2.189 7</b>	0.364 2	<b>0.706 3</b>	<b>0.356 8</b>
Movie_18	LEGFF	7.176 5	6.457 2	17.418 5	1.404 0	0.431 5	0.720 2	0.446 8
	CFL	2.946 4	6.720 0	7.824 8	<b>3.324 1</b>	0.389 6	0.759 7	17.998 4
	CCF	5.063 2	6.771 4	12.318 6	3.087 2	<b>0.453 4</b>	0.762 4	31.268 2
	SMVIF	4.576 1	5.654 1	10.878 4	1.281 4	0.448 8	0.691 8	6.521 4
	ResNet	2.585 9	5.524 9	6.315 5	1.546 5	0.340 1	0.727 5	1.423 6
	Ours	<b>8.632 5</b>	<b>6.998 1</b>	<b>21.409 5</b>	1.998 2	0.313 3	<b>0.766 4</b>	<b>0.356 8</b>

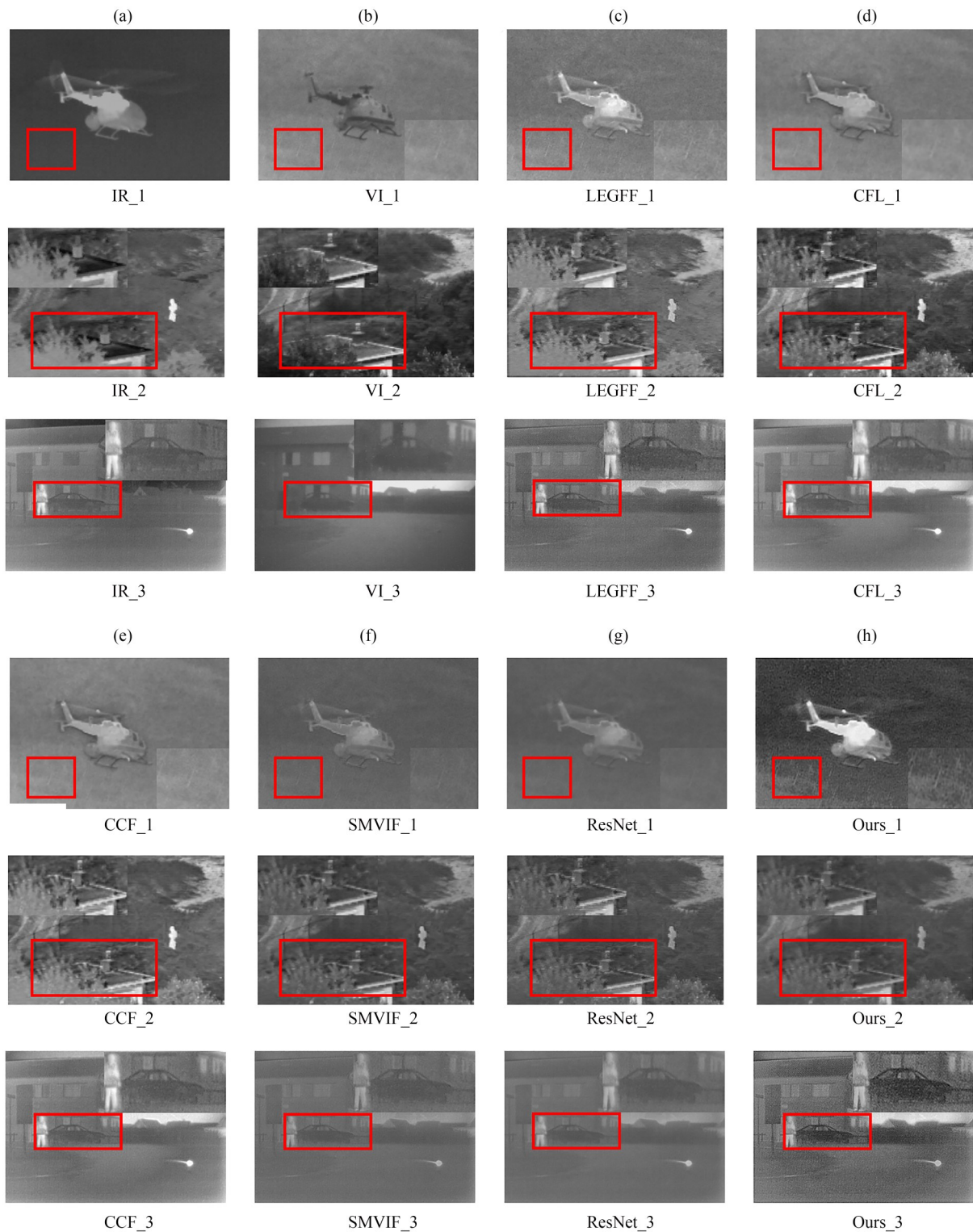


图 11 3 种不同场景下红外与可见光图像融合比较

Fig. 11 Comparison of infrared and visible image fusion in three different scenes

CCF<sup>[34]</sup>方法进行了比较,所得融合结果如图 10 和图 11 的(c)~(h)所示。从融合结果的主观和客观评价得出,本文方法的融合结果很好地保留了

输入图像的重要信息,并且在纹理、边缘轮廓和对比度方面处理都较好。充分证明了本文的方法在其他图像融合领域也具有很强的适应性。

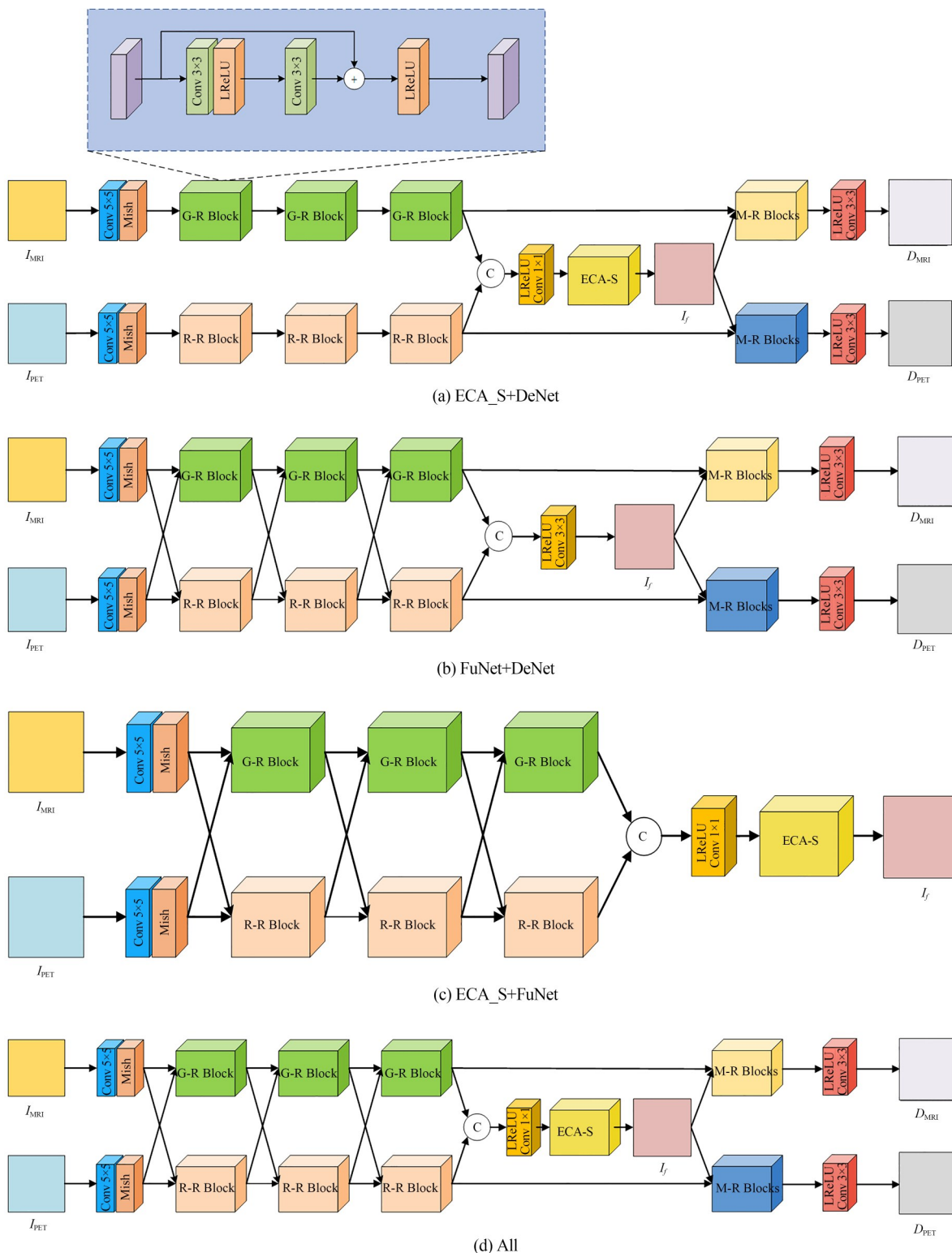


图 12 消融实验的四种不同网络结构

Fig. 12 Four different network structures for ablation experiments

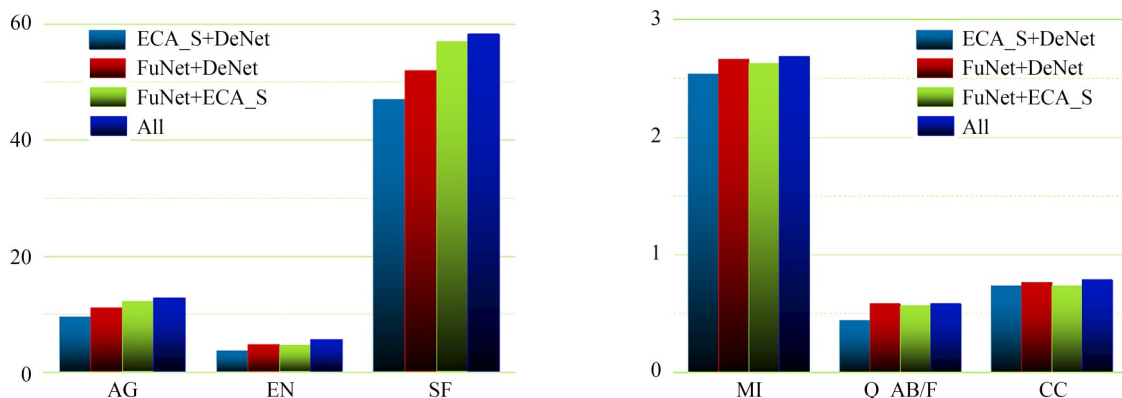


图 13 消融实验的四种不同网络结构 20 副图像均值柱状图

Fig. 13 Histogram of the mean values of 20 images of four different network structures for the ablation experiment

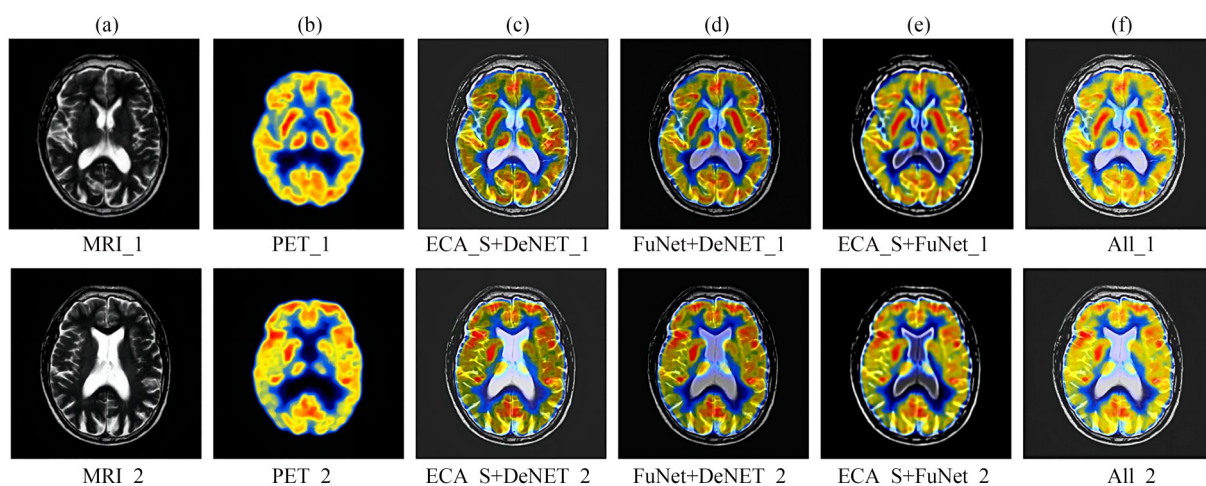


图 14 消融实验其中两组融合结果

Fig. 14 Fusion results of two groups of ablation experiments

## 4 结 论

为了提高医学图像的融合质量,本文提出了一种结构功能交叉神经网络的多模态医学图像融合方法。在融合阶段,通过结构功能交叉神经网络完成图像结构信息和功能信息交互,使其充分提取源图像中的信息,并将提取的特征信息通过 ECA-S 注意力机制得出融合图像。在分解阶段,为了提取到更多的图像信息,再将融合图像分解为两幅源图像,与输入图像比较尽可能减少其信息损失。实验结果表明,本文

提出的方法客观评价指标 AG, EN, SF, MI,  $Q^{AB/F}$  和 CC 均提高了 22.87%, 19.64%, 23.02%, 12.70%, 6.79%, 30.35%, 因此本文方法不但对比度高,纹理细节清晰,且更符合人眼视觉特性。此外,通过客观实验数据显示,本文方法在多聚焦图像融合、红外和可见光图像融合领域的应用也有很大优势。本文方法的不足之处在于处理实时性任务,即没有在提取特征信息之前加入图像预处理。因此,下一步工作将在提取特征信息之前加入图像配准,并将其应用在实时性任务中。

### 参考文献:

[1] LI W S, ZHANG Y, WANG G F, *et al.*

DFENet: a dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion [J].

- Biomedical Signal Processing and Control*, 2023, 80: 104402.
- [2] LIU R S, LIU J Y, JIANG Z Y, *et al.* A bilevel integrated model with data-driven layer ensemble for multi-modality image fusion[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2021, 30: 1261-1274.
- [3] CHAO Z, DUAN X G, JIA S F, *et al.* Medical image fusion via discrete stationary wavelet transform and an enhanced radial basis function neural network [J]. *Applied Soft Computing*, 2022, 118: 108542.
- [4] YU N N, LI J J, HUA Z. Decolorization algorithm based on contrast pyramid transform fusion[J]. *Multimedia Tools and Applications*, 2022, 81 (11) : 15017-15039.
- [5] 林剑萍, 廖一鹏. 结合分数阶显著性检测及量子烟花算法的 NSST 域图像融合[J]. *光学精密工程*, 2021, 29(6): 1406-1419.
- LIN J P, LIAO Y P. A novel image fusion method with fractional saliency detection and QFWA in NSST [J]. *Opt. Precision Eng.*, 2021, 29 (6) : 1406-1419. (in Chinese)
- [6] TAWFIK N, ELNEMR H A, FAKHR M, *et al.* Multimodal medical image fusion using stacked auto-encoder in NSCT domain[J]. *Journal of Digital Imaging*, 2022, 35(5): 1308-1325.
- [7] 杨艳春, 裴佩佩, 党建武, 等. 基于交替梯度滤波器和改进 PCNN 的红外与可见光图像融合[J]. *光学精密工程*, 2022, 30(9): 1123-1138.
- YANG Y C, PEI P P, DANG J W, *et al.* Infrared and visible image fusion based on alternating gradient filter and improved PCNN[J]. *Opt. Precision Eng.*, 2022, 30(9): 1123-1138. (in Chinese)
- [8] ECKHORN R, REITBOCK H J, ARNDT M, *et al.* A neural network for feature linking via synchronous activity: results from cat visual cortex and from simulations[M]. 1989.
- [9] LARA-HERNANDEZ A, RIENMULLER T, JUAREZ I, *et al.* Deep learning-based image registration in dynamic myocardial perfusion CT imaging [J]. *IEEE Transactions on Medical Imaging*, 2023, 42(3): 684-696.
- [10] REHMAN M, ALI M, OBAYYA M, *et al.* Machine learning based skin lesion segmentation method with novel borders and hair removal techniques [J]. *PLoS One*, 2022, 17(11): e0275781.
- [11] ZHAO C, WANG T F, LEI B Y. Medical image fusion method based on dense block and deep convolutional generative adversarial network[J]. *Neural Computing and Applications*, 2021, 33(12) : 6595-6610.
- [12] 杨艳春, 高晓宇, 党建武, 等. 基于 WEMD 和生成对抗网络重建的红外与可见光图像融合[J]. *光学精密工程*, 2022, 30(3): 320-330.
- YANG Y C, GAO X Y, DANG J W, *et al.* Infrared and visible image fusion based on WEMD and generative adversarial network reconstruction [J]. *Opt. Precision Eng.*, 2022, 30(3) : 320-330. (in Chinese)
- [13] WANG L F, LIU Y, MI J, *et al.* MSE-Fusion: weakly supervised medical image fusion with modal synthesis and enhancement[J]. *Engineering Applications of Artificial Intelligence*, 2023, 119: 105744.
- [14] 陈永, 张娇娇, 王镇. 多尺度密集连接注意力的红外与可见光图像融合[J]. *光学精密工程*, 2022, 30(18): 2253-2266.
- CHEN Y, ZHANG J J, WANG Z. Infrared and visible image fusion based on multi-scale dense attention connection network [J]. *Opt. Precision Eng.*, 2022, 30(18) : 2253-2266. (in Chinese)
- [15] FANG M, PENG S Y, LIANG Y J, *et al.* A multimodal fusion model with multi-level attention mechanism for depression detection [J]. *Biomedical Signal Processing and Control*, 2023, 82: 104561.
- [16] TANG W, HE F Z, LIU Y, *et al.* MATR: multimodal medical image fusion via multiscale adaptive transformer[J]. *IEEE Transactions on Image Processing*, 1003, 31: 5134-5149.
- [17] ZHANG H, MA J Y. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion [J]. *International Journal of Computer Vision*, 2021, 129(10): 2761-2785.
- [18] WANG Q L, WU B G, ZHU P F, *et al.* ECA-net: efficient channel attention for deep convolutional neural networks[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 11531-11539.
- [19] ZHU X Z, CHENG D Z, ZHANG Z, *et al.* An empirical study of spatial attention mechanisms in deep networks[C]. 2019 *IEEE/CVF International*

- al Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019: 6687-6696.*
- [20] LI S T, YANG B. Multifocus image fusion using region segmentation and spatial frequency[J]. *Image and Vision Computing*, 2008, 26 (7) : 971-979.
- [21] ZHAO W D, WANG D, LU H C. Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29 (4) : 1102-1115.
- [22] PIELLA G, HEIJMANS H. A new quality metric for image fusion[C]. *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429). Barcelona, Spain. IEEE, 2003: III-173.*
- [23] DESHMUKH M, BHOSALE U. Image fusion and image quality assessment of fused images[J]. *International Journal of Image Processing (IJIP)*, 2010, 4(5) : 484.
- [24] VAN AARDT J. Assessment of image fusion procedures using entropy, image quality, and multi-spectral classification[J]. *Journal of Applied Remote Sensing*, 2008, 2(1): 023522.
- [25] TAN W, THITON W, XIANG P, *et al.* Multimodal brain image fusion based on multi-level edge-preserving filtering[J]. *Biomedical Signal Processing and Control*, 2021, 64: 102280.
- [26] TAN W, TIWARI P, PANDEY H M, *et al.* Multimodal medical image fusion algorithm in the era of big data[J]. *Neural Computing and Applications*, 2020: 1-21.
- [27] ZHANG Y, XIANG W H, ZHANG S L, *et al.* Local extreme map guided multi-modal brain image fusion[J]. *Frontiers in Neuroscience*, 2022, 16: 1055451.
- [28] LI B, PENG H, LUO X H, *et al.* Medical image fusion method based on coupled neural P systems in nonsubsampling shearlet transform domain[J]. *International Journal of Neural Systems*, 2021, 31 (1): 2050050.
- [29] VESHKI F G, VOROBYOV S A. Coupled feature learning via structured convolutional sparse coding for multimodal image fusion[C]. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, Singapore. IEEE, 2022: 2500-2504.*
- [30] LI X X, GUO X P, HAN P F, *et al.* Laplacian re-decomposition for multimodal medical image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(9): 6880-6890.
- [31] AGRAWAL C, YADAV S K, SINGH S P, *et al.* A simplified parameter adaptive DCPCNN based medical image fusion[C]. *Proceedings of International Conference on Communication and Artificial Intelligence. Singapore: Springer Nature Singapore, 2022: 489-501.*
- [32] CHEN J, WU K L, CHENG Z, *et al.* A saliency-based multiscale approach for infrared and visible image fusion[J]. *Signal Processing*, 2021, 182: 107936.
- [33] LI H, WU X J, DURRANI T S. Infrared and visible image fusion with ResNet and zero-phase component analysis[J]. *Infrared Physics and Technology*, 2019, 102: 103039.
- [34] VESHKI F G, OUZIR N, VOROBYOV S A, *et al.* Multimodal image fusion via coupled feature learning [J]. *Signal Processing*, 2022, 200: 108637.

#### 作者简介:



邸 敬(1979—),女,甘肃兰州人,副教授,硕士生导师,主要从事图像检测识别、信号处理技术和宽带无线通信方面的研究。E-mail: 46891771@qq.com

#### 通讯作者:



郭文庆(1998—),男,甘肃武威人,硕士研究生,2021年于兰州交通大学获得学士学位,主要研究医学图像融合、图像处理。E-mail: 344385945@qq.com