

一种多尺度残差注意力 Unet-Like 网络的医学图像融合方法^{*}

杨卫明, 张伟豪, 余 磊

(重庆师范大学 计算机与信息科学学院, 重庆 401331)

摘要:针对目前医学图像融合任务中融合策略难以制定及模态重要信息挖掘不够的问题,提出了一种多尺度残差注意力 Unet-Like 网络(multiscale residual attention unet-like network, MRAUnet)框架用于医学图像融合。首先,传统的融合算法在处理多模态医学图像时存在特征学习能力不足的问题,无法更好地保留互补的多模信息。因此,为了增强模型的特征提取能力与稳定性,MRAUnet 将残差注意力机制注入到特征提取过程中。其次,Unet 架构在理解医学图像的语义上有着天然的优势,而且它的 U 型结构使得 Unet 对浅层简单特征与深层抽象特征有更强的感知能力和重用能力,所以 MRAUnet 采用了 Unet 架构来增强模型的语义感知能力和特征重用,同时整体网络采用多尺度的方式进行特征处理以强化 Unet 架构的特征抽取能力。另外,为了丰富在融合图像中的模态信息,还设计了一种基于结构信息与细节信息保留的损失函数用于训练 MRAUnet。最后,MRAUnet 是一个端到端的融合网络,无需人工制定融合策略,可以有效地利用多模态图像的浅层结构特征与深层抽象特征,以平衡融合图像中的解剖信息和功能信息,在保持较高清晰度的同时可以显示出元素放射量与血液流动性等功能特征。在哈佛医学院全脑图集上的实验结果表明,MRAUnet 与近期较为先进的算法相比,在主观和客观评判 2 个方面都取得了较大进步,在基于人类视觉感知的保真度、基于特征保留的互信息等指标上均有较大优势。

关键词:Unet; 残差注意力机制; 医学图像融合; 多尺度特征

中图分类号:TP391

文献标志码:A

文章编号:1672-6693(2024)04-0126-13

1 研究背景

医疗影像在现代医学临床上有着重要作用,在疾病的筛查、诊断和治疗阶段都需要医疗影像的辅助。在临床上,超过 70% 的诊断都依赖于医疗影像,多模态的医疗影像检查手段得到的医疗影像包含了不同的病理信息,为了消除单幅图像显示信息的局限性,图像融合技术开始应用于医疗影像。医疗影像一般分为解剖图像和功能图像 2 种类型。通常解剖图像拥有较高的分辨率并且可以显示出结构信息,但是它不能显示解剖部位新陈代谢的功能变化。例如磁共振成像(magnetic resonance imaging, MRI)显示的解剖结构越逼真,病理信息也就越清楚,但是它无法显示被检部位的新陈代谢等功能信息。而功能图像则显示了血流和新陈代谢等功能信息,但分辨率不高,不能准确描述被检部位的解剖学细节。例如,正电子发射断层扫描(positron emission tomography, PET)图像是一种反映分子代谢的显像,可以显示脑组织中的放射性物质的放射量,并以此来测量各个部位对葡萄糖的利用和局部血流量;单光子发射计算机断层扫描(single-photon emission computed tomography, SPECT)图像可以突出组织和器官的病变。由于多模态医学图像包含了丰富的解剖和功能信息,所以如果利用图像融合技术将多模态医学图像中的互补信息整合,则可以更好地表示其中的信息,让医生更加高效地利用这些信息进行病理的临床诊断和精确治疗^[1]。

多尺度变换是模仿人类视觉的一种方法,它被广泛应用于图像融合,经常使用的尺度变换算法包括 Laplacian 金字塔变换^[2-3]、小波变换^[4-6]、非下采样轮廓波变换(non-subsampled contourlet transform, NSCT)^[7]

^{*} 收稿日期:2023-11-17 修回日期:2024-06-07 网络出版时间:2024-10-11T10:35

资助项目:国家自然科学基金面上项目(No. 72071019);重庆市自然科学基金面上项目(No. cstc2021jcyj-msxmX0185);重庆市高等教育教学改革研究项目(No. 233178)

第一作者简介:杨卫明,男,讲师,研究方向为图像处理和排序算法,E-mail:20130956@cqu.edu.cn;通信作者:余磊,男,教授,E-mail:ylcqu@163.com

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20241010.1930.002

和非下采样剪切波变换(non-subsampled shearlet transform, NSST)^[8]等,这些都是非常经典的图像分解工具。其中,Laplacian 金字塔变换具有最好的鲁棒性;NSCT 和 NSST 具有平移不变性、多向性和各向异性的特点。因此,它们最常被推广和应用。在获得低频和高频特征后,使用适当的融合规则来获得融合的高频和低频子频段,然后使用逆变换来得到融合后的图像。但是多尺度方法需要确定尺度的数量,过多的尺度可能让融合结果对噪声的处理能力下降,过少的尺度又可能没法完整的保留原始数据的细节信息。另外,融合策略难以与多种多样的特征表征方式完全匹配,所以如何制定合适的融合策略是一个有挑战的任务。

随着机器学习方法的不断发展,图像处理领域也开始使用机器学习的相关方法对图像进行处理且获得了很大的成功。稀疏表示(sparse representations, SR)由压缩感知衍生出^[9-20],Barba-J 等人^[13]提出了一种基于离散 Hermite 变换(discrete Hermite transformation, DHT)和 SR 的 CT(计算机断层扫描)-SPECT 图像融合方法,该方法使用离散 Hermite 变换将 CT 图和 SPECT 图分解,再使用基于 SR 的融合规则和基于最大值准则的融合规则进行融合,最后使用逆 DHT 变换恢复成融合图像。Chen 等人^[14]提出了一种基于滚动引导滤波的医学图像融合新方法,该方法使用滚动引导滤波(rolling guidance filter, RGF)将源图像分解成高频和低频 2 个部分,再对高频部分使用基于 Laplacian 金字塔和 SR(LP-SR)的方法融合,对低频部分使用基于能量的融合规则融合,最后重建图像。但是由于此类算法面对复杂图像时的学习能力相对不足,导致对医学图像中模态特征的提取不够充分,使得融合图像中的模态信息保留不完整。

伴随着深度学习的快速发展,卷积神经网络(convolutional neural networks, CNN)在图像处理领域得到了应用而且取得了明显成效。CNN 可以提取特征并且重建图像,Liu 等人^[21]将 CNN 用于医学图像融合,此项工作中基于卷积神经网络生成一个权重图,该权重图集成了来自源图像的像素活动信息。Wang 等人^[22]将 CNN 组用于提取由 Gabor 表示的 CT 和 MRI 图像对里的特征信息,再使用模糊神经网络(fuzzy neural networks, FNN)将特征融合重构成合成图像。Zhang 等人^[23]提出了一个通用的图像融合框架 IFCNN,该方法可用于融合多聚焦图像、红外-可见光图像以及医学图像。因 CNN 拥有极强的特征提取能力,基于深度学习的多模态医学图像融合算法相较于其它方法均有较好的表现,但也会存在一些不足,例如网络不易训练、模态信息保留不完整等问题。

近些年来,Transformer 模型在自然语言处理领域取得了惊人的效果。Transformer 最初被提出是用于机器翻译领域,不同于常用的循环神经网络(recurrent neural networks, RNN),Transformer 模型完全由自注意力机制组成,解决了 RNN 中的长距离依赖问题,同时还可以并行训练。随后,使用全自注意力代替传统神经网络的思想也被应用在计算机视觉领域,Vision Transformer(ViT)^[24]应运而生,并在各项任务上获得了令人满意的效果。在图像融合领域,ViT 常被用于特征提取或者学习合适的融合策略。目前,许多图像融合方法将 CNN 与 ViT 结合以同时建立长程依赖和保留细节特征^[25-29]。Wang 等人^[25]提出了一种简单有效的红外和可见光图像融合方法(SwinFuse),该方法使用 Swin Transformer 进行特征提取和融合以建立长程依赖,最后由一个简单的重建模块得到融合图像。Fu 等人^[26]提出了一种批次金字塔 Transformer 并应用于图像融合,首先将输入图片转换为一系列图像块,同时提取块中的局部特征和整幅图像的全局特征,最终用重建网络得到融合图像。Tang 等人^[27]提出了一种基于 Transformer 的多尺度融合框架(multimodal medical image fusion via multiscale adaptive transformer, MATR)用于医学图像融合,设计了一种 CNN 与 Transformer 结合的自适应网络用于获取不同模态间的互补语义信息,而且还将网络应用到类似的生物细胞图像融合任务中,都取得了不错的效果。与传统 CNN 相比,基于 ViT 的图像融合方法能够更高效地捕获全局上下文信息,具有更强的表示能力。然而,当图像分辨率较大时,对于一些密集的预测任务,基于 ViT 的融合方法则需要大量的计算资源。

针对融合策略难以制定以及模态关键信息保留不够的问题,本文提出了一种多尺度残差注意力 Unet-Like 网络(multiscale residual attention unet-like network, MRAUnet)用于多模态医学图像融合。首先, MRAUnet 采用端到端网络框架,无需手动制作融合策略或决策图。此外,在特征提取阶段利用了 Unet 网络架构,有效地提取了源图像中的多尺度信息。最后,使用残差注意力机制来挖掘医学图像中的重要视觉信息,使得融合结果包含了丰富的模态互补信息,同时保证了模型的收敛和梯度的稳定。大量实验结果表明,所提出的方法在主观与客观方面均优于一些近年来较为先进的医学图像融合方法。

2 相关工作

2.1 残差注意力机制

随着神经网络技术不断发展,网络变得越来越深,可以提取到更多层次的特征,但是在使用深层次的网络时,过多的参数量会产生梯度消失、梯度爆炸和网络退化等问题。一般会使用标准的初始化或采用正则化层来解决,但是这些方法仅可以保证几十层的网络能够收敛,随着网络参数量不断地增大,梯度消失和梯度爆炸的问题仍然是存在的。在 2016 年,He 等人^[30]提出了残差网络来应对梯度消失和爆炸的问题。残差网络的模型结构如图 1 所示,残差网络的核心就是跳跃连接,残差网络可以表示为:

$$F(x) = x + f(x), \quad (1)$$

其中: x 是网络的输入, $f(x)$ 代表残差跳跃的卷积操作, $F(x)$ 表示残差网络的输出。残差网络因能有效解决梯度消失和梯度爆炸现象而被广泛使用,跳跃连接的加入使得更深层次的网络也可以收敛。

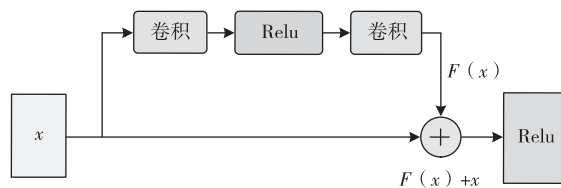


图 1 残差网络块

Fig. 1 Residual network block

注意力机制^[31]的灵感来自人类视觉系统,在认知科学中,由于瞬时能接收到的信息量有瓶颈,人们会自觉地选择某一部分作为注意力集中的区域。计算机领域的注意力机制根据原始特征图中特征的重要性来赋予权重以提高表征能力。残差注意力机制在文献[32]中被提出,它通过在残差网络中加入注意力机制来提高特征表达能力,它的结构如图 2 所示。可以看出残差注意力机制分为 2 个部分,上半部分是残差网络的主干分支,下半部分对输入下采样之后再上采样,再使用 Sigmoid 激活函数使之成为软掩膜分支。使用均衡方法来将 2 个部分合成形成残差注意力机制,具体表示为:

$$H(x) = (1 + M(x)) * V(x), \quad (2)$$

其中: x 是输入特征, $M(x)$ 是注意力机制的软掩膜参数, $V(x)$ 是残差部分的卷积操作, $H(x)$ 表示输出特征。

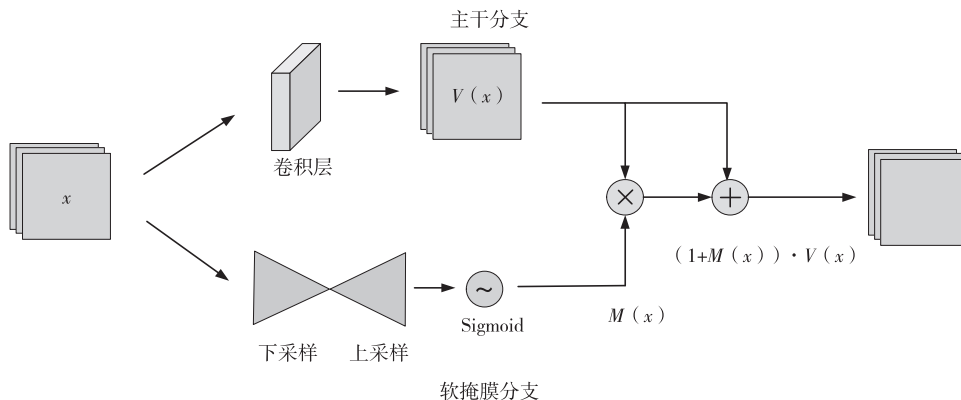


图 2 残差注意力模块

Fig. 2 Residual attention module

2.2 Unet

Unet 最初被提出并应用于医学图像分割任务^[33],取得了优异的效果。由于医学图像中的语义较为简单且结构较为固定,而且它所包含的低级特征和高级语义特征都很重要,所以 Unet 中的跳跃链接与多尺度的结构恰好完美的与医学图像的特点相符合。Unet 的网络结构由图 3 所示,由编码器、解码器与跳跃连接组成,而 Unet 也因 U 型的网络结构而得名。除了跳跃连接之外,Unet 如此有效的原因还在于它的解码器部分将浅层特征与深层特征拼接起来,这提高了模型对浅层简单特征与深层抽象特征的感知能力。

在图像融合领域,Unet 也有一些应用。Wang 等人^[34]提出了一种全尺度连接 Unet(FSL-Unet)用于高光谱和多光谱图像融合,并取得了不错的效果,该方法首先分别计算源图像的空间-光谱注意力,然后使用全连接 U 型融合网络将提取到的特征融合并生成融合图像。Wang 等人^[35]基于交叉卷积与注意力机制提出了 Cross-Unet 网络用于红外-可见光图像融合,采用交叉结构并结合 Inception 模块用于提取特征,取得了优秀的结果。目前,Unet 在医学图像融合领域的应用比较少,有着很大的发展空间。

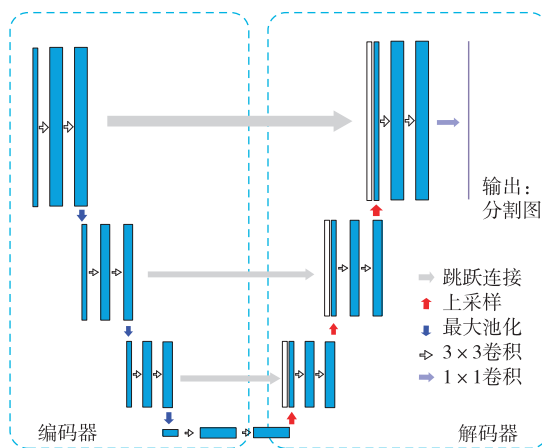


图 3 Unet 结构

Fig. 3 Structure of Unet

3 提出的算法

3.1 概述

本文算法的网络结构如图 4 所示,网络的输入为一对图像,即解剖图像与功能图像。其中,由于功能图像一般为彩色通道图像无法与单通道的解剖图像直接融合,所以先将功能图像的 RGB 色彩空间转换到 Y、Cb、Cr 空间,并将转换后的 Y 通道作为网络输入。为了更加充分地保留来自功能图像与解剖图像的特征,MRAUnet 首先将 2 个输入按通道拼接起来,随后通过特征增维得到 64 维的特征,高维的特征可以使得特征处理过程中更多的模态信息被保留。

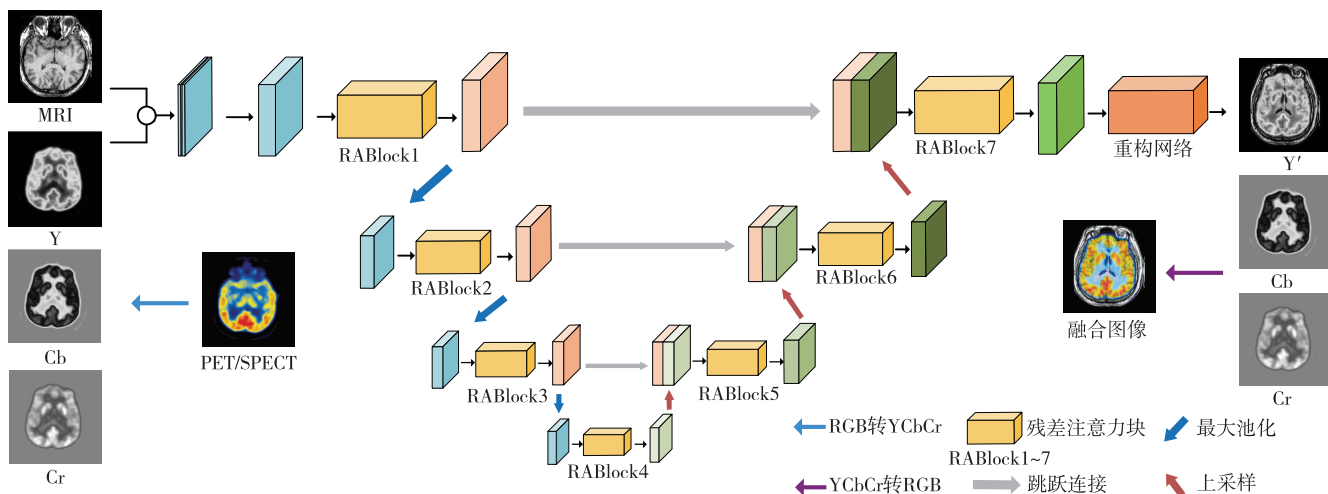


图 4 MRAUnet 网络结构

Fig. 4 Schematic of MRAUnet

接着将高维的原始特征输入到 Unet-Like 编码器中,在多模态医学图像中细节、结构和边缘信息等表层特征是融合任务需要首先保留的模态信息,而包含在图像中抽象的深层语义也是需要被保留的重要信息,所以编码器被设计成由 4 层不同尺度的特征提取层组成,每层的特征尺度随着层数增大而变小,尺度越小的特征就更能更好地提取原始特征中的抽象语义。各层中均有一个残差注意力块(RABlock)用于特征提取,特征提取过程可

表示为:

$$X_n = P(B(X_{n-1})), \quad (3)$$

其中: X_n 是当前层输出, X_{n-1} 为当前层的输入同时也是上一层的输出, B 代表 RABlock 的特征提取过程, P 表示池化操作, 用于实现多尺度特征提取。

为了使 MRAUnet 有更好的特征挖掘与重用能力, 在 U 型结构的右半部分将小尺度的特征经过上采样后与跳跃连接的特征进行拼接再提取特征, 这样可以在尺度还原的同时最大程度的利用来自更小尺度特征中的抽象特征, 解码器被设置为 3 层用于处理编码器产生的多尺度特征, 解码器的解码过程表示为:

$$Y_n = B(U(Y_{n-1}) \odot X_n), \quad (4)$$

其中: Y_n 代表当前层的输出, Y_{n-1} 代表当前层的输入同时也是上一层的输出, X_n 代表了跳转连接的输出, B 与编码器中相同, 而 U 代表上采样, $\odot$ 表示拼接 (Concat 函数) 操作, 即当前层的输出进行上采样之后与跳转连接的输出拼接后作为下一层的输入。

最后一个 RABlock 的输出为包含着丰富模态信息的 64 维的多尺度特征, 为了得到最终的融合图像, MRAUnet 通过特征重构模块将高维特征重构为二维的图像尺寸后即可得到融合后的灰度图像, 再结合功能图像的 Cb、Cr 通道转换到 RGB 色彩空间即可获得最终的融合图像。

3.2 残差注意力块

残差注意力块的网络结构如图 5 所示, 为了提高模型的特征提取能力以及网络在训练阶段的稳定性和收敛速度, 在残差注意力块中加入了残差注意力机制。图 5 中的最上层为注意力掩膜分支, 由下采样层、上采样层和 Relu 激活函数组成, 将输出归一化到 0 到 1 之间再与主干分支进行点乘, 可以大大提高模型对特征的选择提取能力。

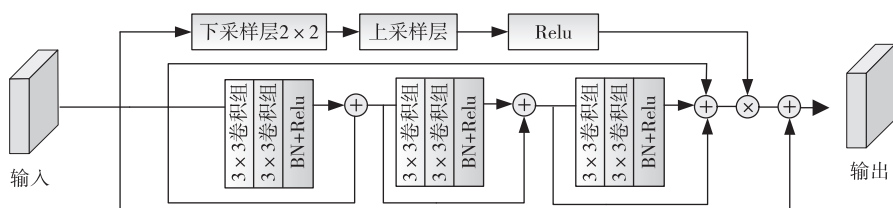


图 5 残差注意力块网络结构

Fig. 5 Schematic of RABlocks

主干分支由 3 组 3×3 卷积组与 3 个残差跳跃构成, 采用小卷积核的卷积层堆叠可以在减少参数量的同时达到与更大卷积核的卷积层相媲美的特征提取能力, 所以残差注意力块的主干分支采用了卷积层堆叠的方式来提高模型的特征提取能力, 同时还提升了模型的收敛速度, 另外每个卷积组后还加入了批归一化层与 Relu 激活函数用于保证模型的稳定, 防止梯度爆炸等情况。

为了尽可能的让网络中的特征在各个阶段的表征方式类似以提高模型的特征处理能力, MRAUnet 的 7 个残差注意力块被设计为相同的网络结构, 但由于 MRAUnet 是一个多尺度网络, 所以解码器和编码器各层的特征尺度不完全相同, 这也使得每个残差注意力块各自的参数有所不同, 参数对比如表 1 所示。

表 1 残差注意力块参数对比

Tab. 1 Comparison of RABlocks parameters

RABlock 序号	输入	输出	BatchNorm 通道数
1	$64 \times 256 \times 256$	$64 \times 256 \times 256$	64
2	$128 \times 128 \times 128$	$128 \times 128 \times 128$	128
3	$256 \times 64 \times 64$	$256 \times 64 \times 64$	256
4	$512 \times 32 \times 32$	$256 \times 32 \times 32$	512
5	$512 \times 64 \times 64$	$128 \times 64 \times 64$	512
6	$256 \times 128 \times 128$	$64 \times 128 \times 128$	256
7	$128 \times 256 \times 256$	$64 \times 256 \times 256$	128

3.3 损失函数

由于医学图像融合任务没有真实数据,所以网络训练时的损失函数对于融合结果至关重要。医学图像融合旨在将来自解剖图像丰富的结构信息和来自功能图像的功能信息融合到一起,为了防止单一优化目标可能带来的局限性,本文设计了由结构信息保留与细节信息保留 2 部分组成的损失函数,可表示为:

$$L_{\text{sum}} = \alpha L_{\text{structure}} + L_{\text{detail}}, \quad (5)$$

其中: L_{sum} 是总损失函数, $L_{\text{structure}}$ 为结构损失, L_{detail} 表示细节损失, α 是超参数被设置为 700。

首先,为了使融合图像能够拥有较高的分辨率和视觉保真度,本文设计了基于结构信息保留的目标函数。因为解剖图像的结构信息、纹理信息和边缘特征丰富,所以对解剖图像使用结构相似性度量可以最大程度保留来自解剖图像的结构特征,结构损失函数 $L_{\text{structure}}$ 表示为:

$$L_{\text{structure}} = 1 - S(F, X_{\text{MRI}}), \quad (6)$$

其中: F 为网络输出, X_{MRI} 为输入的解剖图像, $S(\cdot)$ 为结构相似性度量。

其次,为了使融合图像中同时包含解剖学特征与元素放射量、血液流动等功能学信息,还设计了基于细节信息保留的目标函数,对功能图像和解剖图像同时计算细节信息损失,细节损失函数 L_{detail} 表示为:

$$L_{\text{detail}} = \|F - (\beta \cdot X_{\text{MRI}} + \gamma \cdot X_{\text{func}})\|_F^2, \quad (7)$$

其中: X_{func} 为输入的功能图像, β 和 γ 是用于平衡融合图像中解剖图像与功能图像之间的信息强度,分别被设置为 6.0 和 3.0。

4 实验结果与分析

4.1 实验说明与评价指标

本文的实验环境是一台带有 Windows 10 (64 位) 系统的计算机,硬件配置为 AMD Ryzen 7 5800X, NVIDIA GeForce RTX 3080Ti GPU 和 32 GB RAM。为了验证本文提出方法的有效性,实验选用了哈佛医学院全脑图集 (<http://www.med.harvard.edu/aanlib/home.html>) 中的 MRI、PET 及 SPECT 各 20 对作为测试图像来进行验证,所采用的源图像尺寸均为 256×256 且已精确配准。

为了证明本文方法的优越性,使用 9 个近些年来较为经典或者先进的方法作为对比,分别是 CNN 方法^[21]、U2Fusion 方法^[36]、RFN-Nest 方法^[37]、EMFusion 方法^[38]、DSAGAN 方法^[39]、NSST-PAPCNN 方法^[40]、CSMCA 方法^[41]、MATR 方法^[27] 和 GDFusion 方法^[42],上述所有算法的模型参数均来源于相应文献。

为了客观度量本算法和其它融合算法效果,再引入 7 种被广泛使用的度量方法,分别为 SD^[43]、MI^[44]、VIF^[45]、AG、 $Q_{\text{AB/F}}$ ^[46]、SF^[47] 和 EN^[48]。其中,所有指标值越大代表融合性能越好,表 2 给出了各种评价指标的说明。

表 2 7 种评价指标的说明

Tab. 2 Description of seven evaluation indicators

评价指标	说明	评价指标	说明
SD	基于图像特征的标准差	$Q_{\text{AB/F}}$	基于梯度的融合性能度量
MI	计算融合图像和源图像的互信息量	SF	计算梯度分布情况以揭示细节与纹理信息
VIF	基于人类视觉感知的视觉保真度	EN	计算融合图像和源图像的信息熵
AG	计算融合图像的平均梯度信息		

4.2 MRI-PET 融合结果对比

在来自 20 对测试数据中的 3 组 MRI-PET 融合任务上,本文方法与其它 9 种方法的融合结果对比如图 6 所示,为了更好地进行的比较,将局部区域放大在左下角。总的来说,所有方法都能在一定程度上实现融合目标,但与所提出的方法相比,其它方法都存在一定的不足。例如,基于 CNN 的方法降低了融合图像中功能信息的占比,使得融合图像中的功能特征不清晰。U2Fusion 可以很好地显示出代谢信息,但丢失了部分解剖信息。RFN-

Nest、CSMA 和 MATR 总体表现良好,但 RFN-Nest 的功能细节和 CSMA 的一些结构细节需要改进。EMFusion 无法完整地保留功能信息,而且还产生了颜色失真。NSST-PAPCNN 对功能图像边缘的处理很难令人满意,与源图像相比有较明显的像素失真。与 U2Fusion 类似,DSAGAN 和 GDFusion 捕获软组织信息的能力有限。总体而言,本文提出的 MRAUnet 方法很好地平衡了融合图像中的功能信息和结构信息,从而获得了更好的主观结果。

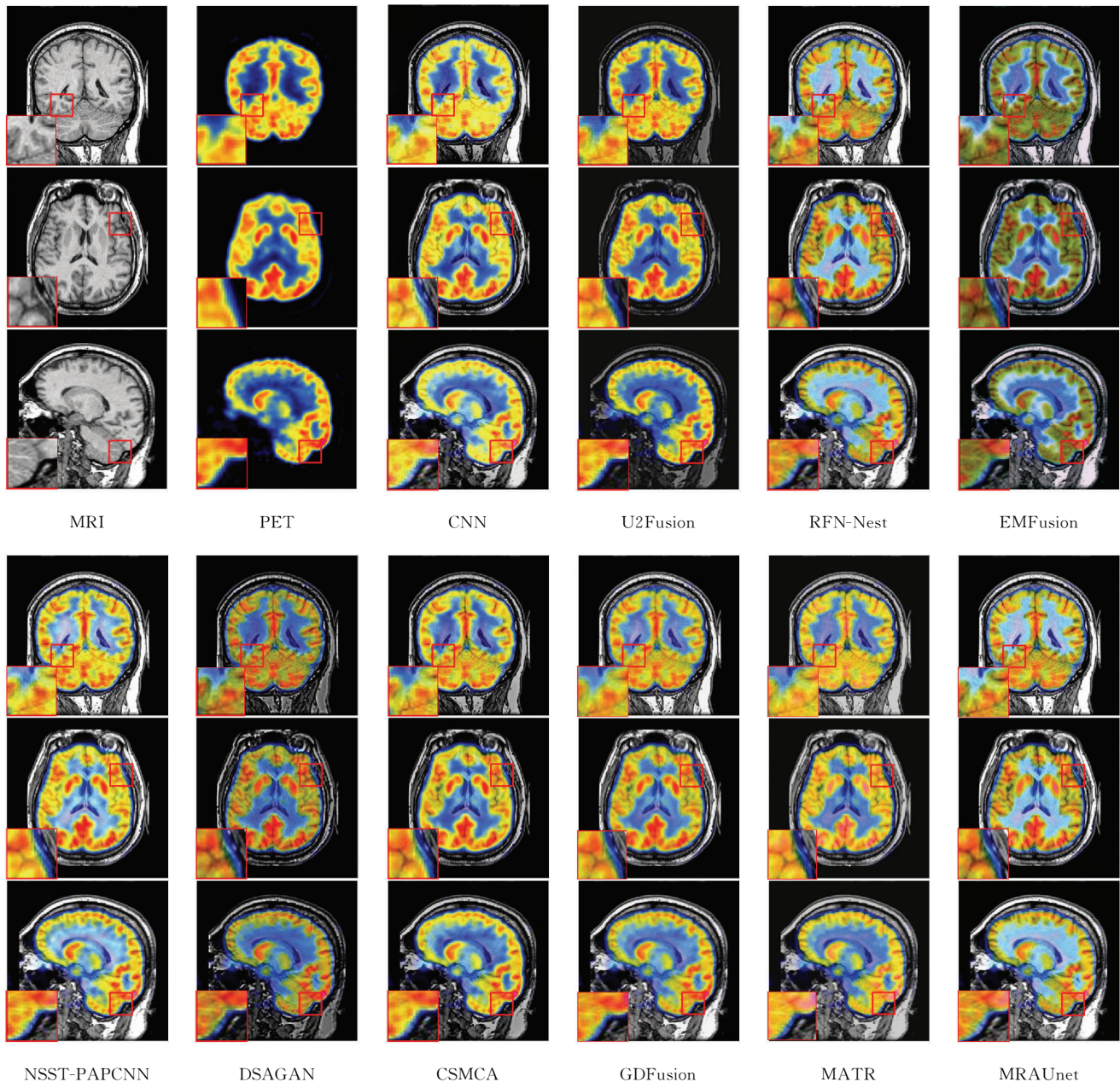


图 6 MRI-PET 融合任务上本文方法与其它 9 种方法的融合结果对比

Fig. 6 Comparisons with nine methods on MRI-PET fusion

MRI-PET 融合任务上的客观质量评价指标如表 3 所示,各指标的得分由 20 组实验结果对应指标的平均值得到。MRAUnet 在 AG 和 SF 的得分略差于 NSST-PAPCNN,但是 AG 和 SF 均是对单幅图像的梯度进行计算,而融合任务的目标是保留源图像中的信息,所以需要结合其它指标进行判断。不难看出,MRAUnet 在 SD、MI、VIF、EN 和 $Q_{AB/F}$ 均取得了最优的结果,再结合主观的对比可以看出本文的方法综合最优,可以有效地保留源图像的关键特征而且平衡了源图像对中互补的模式信息。

表 3 MRI-PET 融合任务的客观评价指标对比
Tab. 3 Quantitative comparison on MRI-PET fusion

方法	SD	MI	VIF	AG	EN	$Q_{AB/F}$	SF
CNN	9.947 1	2.578 5	0.766 9	10.968 4	6.101 9	0.651 9	0.136 4
U2Fusion	9.112 4	2.889 7	0.603 4	7.120 5	5.699 9	0.384 8	0.083 2
RFN-Nest	10.318 0	3.385 6	0.907 2	11.133 8	5.867 7	0.688 9	0.135 7
EMFusion	10.195 5	3.018 2	0.696 9	9.735 9	5.870 3	0.669 5	0.115 7
NSST-PAPCNN	10.314 6	2.958 2	0.785 2	11.524 2	6.010 5	0.666 3	0.142 3
DSAGAN	9.849 4	2.323 1	0.532 7	9.101 4	5.483 9	0.465 8	0.111 2
CSMCA	9.878 4	2.560 9	0.708 7	10.388 9	5.727 8	0.648 0	0.130 5
MATR	9.984 4	2.644 1	0.685 3	9.559 4	5.746 2	0.644 1	0.114 3
GDFusion	10.128 2	2.935 2	0.838 5	10.291 4	5.983 7	0.681 3	0.130 6
MRAUnet	10.360 7	3.453 1	0.928 3	11.356 3	6.126 4	0.706 1	0.137 8

注:加粗的数字表示得分排名第 1,斜体的数字表示得分排名第 2。

4.3 MRI-SPECT 融合结果对比

MRI-SPECT 融合任务上本文方法与其它 9 种方法的融合结果对比如图 7 所示。同样在左下角对图像局部进行放大展示。观察结果可以发现 U2Fusion、DSAGAN 和 GDFusion 在功能图像的色彩上依然有着较好的保留,但是在软组织解剖图像上存在图像模糊和细节缺失。RFN-Nest 和 MATR 在功能源图像有着令人满意的色彩饱和度,在解剖图像上有着近似于源图像的对比度,没有出现软组织细节过暗的情况。对比其它算法 CSMA 的解剖图像中的对比度差强人意,丢失了软组织的细节侧重于功能信息的保留,而 EMFusion 则又侧重于软组织细节的保留,使得功能图像的色彩有着大范围的失真。NSST-PAPCNN 过多地保留了 SPECT 图像的功能信息,导致融合图像产生伪影。总体而言,本文提出的 MRAUnet 方法在 MRI-SPECT 任务上依然平衡了解剖和功能模态的信息,未出现明显的模态信息失衡,有着更好的主观对比结果。

表 4 给出了各种方法在 MRI-SPECT 融合任务上的客观评价指标对比结果。在 SD 和 SF 上,MRAUnet 排名第 3,超过了大部分的方法,而且与更高得分方法的差距比较小。MRAUnet 在 AG 和 EN 上取得了第 2 的得分,说明融合结果在梯度表现良好而且有充足的模态信息,同时在 MI、VIF 和 $Q_{AB/F}$ 上取得了最高的得分,结合以上对客观评价指标的分析以及主观表现说明 MRAUnet 拥有最高的性能,有较高的视觉保真度且符合人类的视觉感知,同时还能提取多模态医学图像中互补的模态信息,有望用于临床辅助治疗。

4.4 损失函数的消融实验

由于在医学图像融合任务中没有真实图像,所以为了增强模型的特征保留能力,本文采用了基于结构信息与细节信息保留的损失函数。为了证明采用的损失函数各部分的有效性,分别对结构信息损失与细节信息保留损失做了消融实验。

图 8 给出了消融实验不同分组在 MRI-PET 和 MRI-SPECT 融合任务上的 4 组融合结果对比。很明显,仅基于细节信息保留的损失函数得出的融合结果有很大的问题。首先是色彩严重失真;其次除了外部骨骼的部分解剖学特征,其余模态信息被过高的对比度完全覆盖了。而基于结构信息保留的损失函数可以基本上完成融合目标,在表观上与采用基于结构信息与细节信息保留的损失函数获得的融合结果差距不大,所以需要结合客观评价指标加以分析。

表 5 给出了这 2 组实验在 MRI-PET 与 MRI-SPECT 融合任务上各 20 对测试图像上的实验结果,客观评价指标得分由 20 组的各指标得分平均得到。其中实验组一和实验组二分别为基于结构信息保留的损失函数和基于结构信息与细节信息保留的损失函数在 MRI-PET 任务上的评价指标得分,实验组三和实验组四分别为基于结构信息保留的损失函数和基于结构信息与细节信息保留的损失函数在 MRI-SPECT 任务上的评价指标得分。从表 5 可以看出在所有的质量评价指标上,本文所采用的基于结构信息与细节信息保留的损失函数在 2 个医学图像融合任务的表现均优于仅使用基于结构信息保留的损失函数。结合对基于结构信息保留的损失函数的消融实验足以说明,同时关注结构与细节信息对于取得优秀的融合结果是十分有效的。

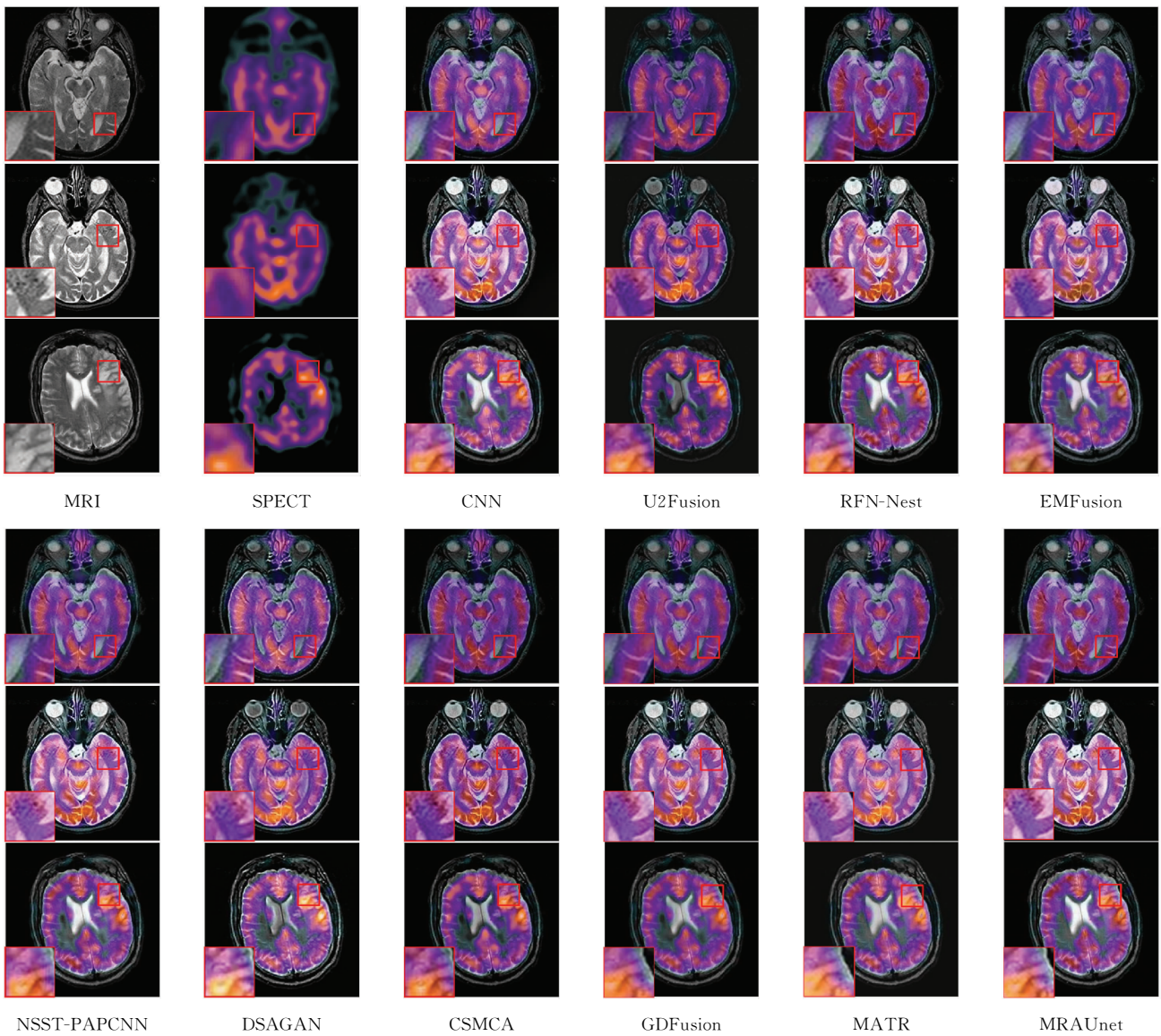


图 7 MRI-SPECT 融合任务上本文方法与其它 9 种方法的融合结果对比

Fig. 7 Comparisons with nine methods on MRI-SPECT fusion

表 4 MRI-PET 融合任务的客观评价指标对比

Tab. 4 Quantitative comparison on MRI-PET fusion

方法	SD	MI	VIF	AG	EN	QAB/F	SF
CNN	9.395 6	2.957 9	0.742 3	6.302 9	5.783 2	0.619 4	0.077 3
U2Fusion	8.531 5	2.701 6	0.085 2	4.118 3	4.848 8	0.397 2	0.051 6
RFN-Nest	9.505 5	3.483 9	0.890 5	6.094 5	5.298 0	0.633 7	0.075 1
EMFusion	9.618 2	3.074 3	0.641 1	5.631 6	5.347 1	0.622 2	0.067 8
NSST-PAPCNN	9.526 6	3.023 7	0.683 4	6.371 0	5.645 0	0.607 7	0.077 9
DSAGAN	9.637 7	2.634 3	0.542 6	7.567 2	5.678 5	0.506 5	0.091 4
CSMCA	9.359 1	2.901 3	0.727 0	5.980 0	5.336 8	0.618 5	0.074 5
MATR	9.281 1	2.912 8	0.590 1	5.601 3	5.273 0	0.580 4	0.067 7
GDFusion	9.534 5	3.079 0	0.748 3	5.504 7	5.482 3	0.607 6	0.067 9
MRAUnet	9.560 4	3.555 6	0.919 7	6.375 5	5.755 2	0.651 4	0.077 2

注:加粗的数字表示得分排名第 1,斜体的数字表示得分排名第 2。

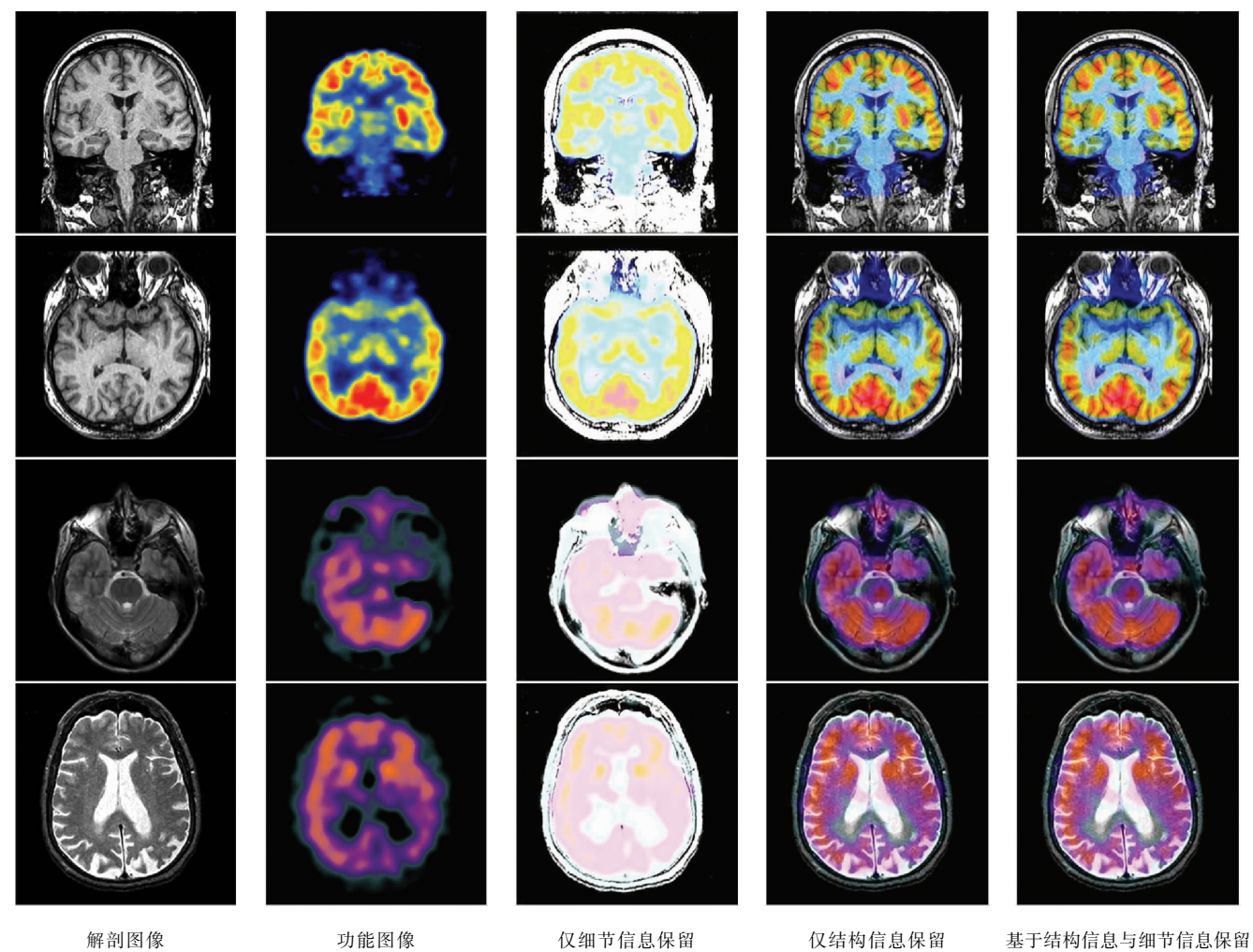


图 8 损失函数的消融实验结果

Fig. 8 The results of ablation experiments on loss function

表 5 消融实验不同组别的融合结果评价指标

Tab. 5 Quantitative comparison on MRI-PET fusion

评价指标	实验组一	实验组二	实验组三	实验组四
SD	10.351 7	10.360 7	9.551 1	9.556 2
MI	3.376 9	3.453 1	3.503 7	3.555 6
VIF	0.926 8	0.928 3	0.918 4	0.919 7
AG	11.103 0	11.356 3	6.148 2	6.347 1
EN	5.974 0	6.126 4	5.490 2	5.656 7
$Q_{AB/F}$	0.693 0	0.706 1	0.638 1	0.651 4
SF	0.131 0	0.137 8	0.073 9	0.077 2

注:加粗的数字分别表示实验组一、二和实验组三、四中较高的得分。

4.5 运行时间比较

表 6 给出了本文算法与其它 9 种融合算法在 MRI-PET 和 MRI-SPECT 融合任务上生成 1 幅 256×256 分辨率的图像所需要的平均时间。实验环境的软硬件配置为 Windows 10(64 位)系统,Intel Core i3-12100 CPU, NVIDIA GeForce GTX 1060 5G GPU 和 16 GB RAM,实验平台为 Matlab2016a 和 PyCharm。从表 6 可以看出,EMFusion 方法所需的时间是最短的,但是它的融合质量表现较差。而本文提出的方法虽然效率不是最高,但相比很多其它算法的用时较少,可以满足实时要求,因此是一种较为快速且高效的医学图像融合算法。

表 6 各方法的平均融合时间比较

Tab. 6 Fusion time comparison of each method

s

方法	CNN	U2Fusion	RFN-Nest	EMFusion	NSST-PAPCNN
平均时间	8.674 0	0.036 3	0.270 1	0.016 7	5.403 1
方法	DSAGAN	CSMAC	MATR	GDFusion	MRAUnet
平均时间	0.049 8	87.941 0	0.986 0	0.033 5	0.208 5

注:加黑的数字表示对应方法时间最短,斜体的数字表示对应方法用时第 2 短。

5 结束语

本文提出了一种多尺度残差注意力 Unet-Like 网络用于医学图像融合,解决了融合过程中融合策略难以设计的问题,同时丰富了融合图像中的模态信息。使用 3×3 卷积堆叠的方式来提取特征,与使用大卷积核相比在提高特征提取能力的同时减少了参数量。整体网络采用 Unet 的架构,将原始的多维特征转化成多尺度特征并利用 Unet 架构对多尺度特征的重用能力挖掘医学图像中的深层抽象特征。通过大量的对比、消融实验说明本文提出的算法能够有效的保留源图像中的纹理和结构等关键特征,可以高效的保留其中的模态信息。下一步将主要研究算法的复杂度与参数量。

参考文献:

[1] ZHOU T,CHENG Q R,LU H L,et al. Deep learning methods for medical image fusion;a review[J]. Computers in Biology and Medicine,2023,160:106959.

[2] DU J,LI W,XIAO B,et al. Union Laplacian pyramid with multiple features for medical image fusion[J]. Neurocomputing,2016,194:326-339.

[3] WANG Z,CUI Z,ZHU Y. Multi-modal medical image fusion by Laplacian pyramid and adaptive sparse representation[J]. Computers in Biology and Medicine,2020,123:103823.

[4] XU H,WANG Y,WU Y,et al. Infrared and multi-type images fusion algorithm based on contrast pyramid transform[J]. Infrared Physics & Technology,2016,3:120-128.

[5] 杨立才,刘延梅,刘欣,等. 基于小波包变换的医学图像融合方法[J]. 中国生物医学工程学报,2009,28(1):12-16.

YANG L C,LIU Y M,LIU X,et al. A novel algorithm for medical image fusion based on wavelet packet transform[J]. Chinese Journal of Biomedical Engineering,2009,28(1):12-16.

[6] SINGH R,KHARE A. Fusion of multimodal medical images using Daubechies complex wavelet transform-A multiresolution approach[J]. Information Fusion,2014,19:49-60.

[7] IBRAHIM S I,MAKHLOUF M A,EL-TAWEL G S. Multimodal medical image fusion algorithm based on pulse coupled neural networks and nonsubsamped contourlet transform[J]. Medical & Biological Engineering & Computing,2023,61(1):155-177.

[8] YIN M,LIU X,LIU Y,et al. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsamped shearlet transform domain[J]. IEEE Transactions on Instrumentation & Measurement,2018,68(1):49-64.

[9] CHENG J,LIU H,LIU T,et al. Remote sensing image fusion via wavelet transform and sparse representation[J]. Journal of Photogrammetry & Remote Sensing,2015,104:158-173.

[10] TANG D,XIONG Q,YIN H,et al. A novel sparse representation based fusion approach for multi-focus images[J]. Expert Systems with Applications,2022,197:116737.

[11] MAQSOOD S,JAVED U. Multi-modal medical image fusion based on two-scale image decomposition and sparse representation [J]. Biomedical Signal Processing and Control,2020,57:101810.

[12] PATEL V M,CHELLAPPA R. Sparse representations and compressive sensing for imaging and vision [M]. NewYork: Springer Science & Business Media,2013.

[13] BARBA-J L,VARGAS-QUINTERO L,CALDERÓN-AGUDELO J A. Bone SPECT/CT image fusion based on the discrete hermite transform and sparse representation[J]. Biomedical Signal Processing and Control,2022,71:103096.

[14] CHEN J,ZHANG L,LU L,et al. A novel medical image fusion method based on rolling guidance filtering[J]. Internet of Things,2020:100172.

- [15] 郭盼,何文超,梁龙凯,等. 基于导向滤波器的医学图像融合方法[J]. 液晶与显示,2019,34(6):605-612.
GUO P, HE W C, LIANG L K, et al. Medical image fusion method based on guided filter[J]. Chinese Journal of Liquid Crystals and Displays, 2019, 34(6): 605-612.
- [16] 赵贺,张金秀,张正刚. 基于 NSCT 与 DWT 的 PCNN 医学图像融合[J]. 激光与光电子学进展,2021,58(20):445-454.
ZHAO H, ZHANG J X, ZHANG Z G. PCNN medical image fusion based on NSCT and DWT[J]. Laser & Optoelectronics Progress, 2021, 58(20): 445-454.
- [17] HU Q, HU S, ZHANG F. Multi-modality medical image fusion based on separable dictionary learning and Gabor filtering[J]. Signal Processing Image Communication, 2020, 83: 115758.
- [18] PADMAVATHI K, ASHA C S, MAYA V K. A novel medical image fusion by combining TV-L1 decomposed textures based on adaptive weighting scheme[J]. Engineering Science and Technology, an International Journal, 2020, 23(1): 225-239.
- [19] DINH P H. A novel approach based on Three-scale image decomposition and Marine predators algorithm for multi-modal medical image fusion[J]. Biomedical Signal Processing and Control, 2021, 67(2): 1-14.
- [20] XU L, SI Y, JIANG S, et al. Medical image fusion using a modified shark smell optimization algorithm and hybrid wavelet-homomorphic filter[J]. Biomedical Signal Processing and Control, 2020, 59: 101885.
- [21] LIU Y, CHEN X, CHENG J, et al. A medical image fusion method based on convolutional neural networks[C]//2017 20th International Conference on Information Fusion, July 10-13, 2017, Xi'an China, Piscataway: IEEE, 2017: 1-7.
- [22] WANG L, ZHANG J, LIU Y, et al. Multimodal medical image fusion based on gabor representation combination of multi-CNN and fuzzy neural network[J]. IEEE Access, 2021, 9: 67634-67647.
- [23] ZHANG Y, LIU Y, SUN P, et al. IFCNN: a general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.
- [24] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale [EB/OL]. (2020-10-22)[2023-11-17]. <https://arxiv.org/abs/2010.11929>.
- [25] WANG Z, CHEN Y, SHAO W, et al. SwinFuse: a residual swin transformer fusion network for infrared and visible images [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-12.
- [26] FU Y, XU T, WU X, et al. Ppt fusion: pyramid patch transformer for a case study in image fusion [EB/OL]. (2021-07-29)[2023-11-17]. <https://arxiv.org/abs/2107.2107>.
- [27] TANG W, HE F, LIU Y, et al. MATR: multimodal medical image fusion via multiscale adaptive transformer[J]. IEEE Transactions on Image Processing, 2022, 31: 5134-5149.
- [28] LI J, ZHU J, LI C, et al. CGTF: convolution-guided transformer for infrared and visible image fusion [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.
- [29] ZHANG J, LIU A, WANG D, et al. Transformer-based end-to-end anatomical and functional image fusion [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 5019711.
- [30] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, Las Vegas, NV, USA. Piscataway: IEEE, 2016: 770-778.
- [31] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention[C]//GHAHRAMANI Z, WELLING M, CORTES C, et al. NIPS'14: Proceeding of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2204-2212.
- [32] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA, Piscataway: IEEE, 2017: 3156-3164.
- [33] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//NAVAB N, HORNEGGER J, WELLS W M, et al. Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241.
- [34] WANG X, WANG X, ZHAO K, et al. Fsl-unet: full-scale linked unet with spatial-spectral joint perceptual attention for hyperspectral and multispectral image fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.
- [35] WANG X, HUA Z, LI J. Cross-UNet: dual-branch infrared and visible image fusion framework based on cross-convolution and attention mechanism[J]. The Visual Computer, 2023, 39(10): 4801-4818.
- [36] XU H, MA J, JIANG J, et al. U2Fusion: a unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 502-518.
- [37] HUI L A, WU X J, KITTLER B. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J].

- Information Fusion, 2021, 73: 72-86.
- [38] XU H, MA J. EMFusion: an unsupervised enhanced medical image fusion network[J]. Information Fusion, 2021, 76: 177-186.
- [39] FU J, LI W, DU J, et al. DSAGAN: a generative adversarial network based on dual-stream attention mechanism for anatomical and functional image fusion[J]. Information Sciences, 2021, 576: 484-506.
- [40] VANITHA K. Medical image fusion algorithm based on weighted local energy motivated PAPCNN in NSST domain[J]. Journal of Advanced Research in Dynamical and Control Systems, 2020, 12(SP3): 960-967.
- [41] LIU Y, CHEN X, WARD R K, et al. Medical image fusion via convolutional sparsity based morphological component analysis [J]. IEEE Signal Processing Letters, 2019, 26(3): 485-489.
- [42] KURBAN R. Gaussian of differences: a simple and efficient general image fusion method[J]. Entropy, 2023, 25(8): 1215.
- [43] LIU Y, LIU S, WANG Z. A general framework for image fusion based on multi-scale transform and sparse representation[J]. Information Fusion, 2015, 24: 147-164.
- [44] QU G H, ZHANG D L, YAN P F. Information measure for performance of image fusion[J]. Electronics Letters, 2002, 38(7): 313-315.
- [45] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444.
- [46] XYDEAS C S, PETROVIĆ V. Objective image fusion performance measure[J]. Electronics Letters, 2000, 36(4): 308-309.
- [47] ESKICIOGLU A M, FISHER P S. Image quality measures and their performance[J]. IEEE Transactions on Communications, 1995, 43(12): 2959-2965.
- [48] ROBERTS J W, Van AARDT J A, AHMED F B. Assessment of image fusion procedures using entropy, image quality, and multispectral classification[J]. Journal of Applied Remote Sensing, 2008, 2(1): 023522.

Multiscale Residual Attention Unet-Like Network for Medical Image Fusion

YANG Weiming, ZHANG Weihao, YU Lei

(College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China)

Abstract: Aiming at the difficulties of formulating fusion strategies and incomplete preservation of modal information in medical image fusion tasks, a multi-scale residual attention Unet-Like network is proposed for medical image fusion, namely MRAUnet (Multiscale Residual Attention Unet-like Network). Firstly, most traditional fusion algorithms cannot extract sufficient complementary information from multimodal medical images. To enhance the feature extraction ability and stability of the model, MRAUnet employs residual attention mechanism in the feature extraction processing. Secondly, the Unet structure has great advantages in the semantic perception of medical images, and its U-shaped structure enables Unet to have stronger perception and reuse capabilities for shallow simple features and deep abstract features. Therefore, MRAUnet adopts the Unet architecture to enhance the semantic perception ability and feature reuse of the model, while the overall network adopts a multi-scale fashion for feature processing to enhance the feature extraction ability of the Unet architecture. In addition, to capture more modal information, a loss function based on structural information and detail information retention is designed to train MRAUnet. Finally, MRAUnet is an end-to-end fusion network with no need to manually design fusion strategies. The proposed network can effectively capture the shallow structural features and deep abstract features from the source image, effectively balancing the anatomical and functional information in the fused image, while maintaining high clarity and displaying functional features such as element radiation and blood flow. The experimental results on the Harvard Medical School Whole Brain Atlas dataset indicate that the MRAUnet has significant improvements in both subjective and objective evaluation compared to some recent advanced algorithms, which has great advantages in terms of VIF based on human visual fidelity, and MI based on feature retention and other metrics.

Keywords: Unet; residual attention mechanism; medical image fusion; multiscale feature

(责任编辑 黄 颖)