文章编号　1004-924X(2023)20-3050-15

基于LL-GG-LG Net的CT和PET医学图像融合

周　涛[1,2]，张祥祥[1,2*]，陆惠玲[3]，李　琦[1,2]，程倩茹[1,2]

(1. 北方民族大学 计算机科学与工程学院，宁夏 银川 750021；
2. 北方民族大学图像图形智能处理国家民委重点实验室，宁夏 银川 750021；
3. 宁夏医科大学 医学信息工程学院，宁夏 银川 750004)

摘要：多模态医学图像融合在医学临床应用中起着至关重要的作用，为了解决现有方法大多数侧重于局部特征的提取，对全局依赖关系的探索不足，忽略了全局和局部信息交互，导致难以有效解决周围组织与病灶区域之间的模式复杂性和强度相似性问题。该文提出面向PET和CT医学图像融合的LL-GG-LG Net模型。首先，提出了局部-局部融合模块(Local-Local Fusion Module，LL Module)，该模块采用双层注意力机制更好地关注局部细节信息特征；其次，设计了全局-全局融合模块(Global-Global Fusion Module，GG Module)，该模块通过在Swin Transformer中加入残差连接机制将局部信息引入全局信息中，提高了Transformer对局部信息的关注程度；然后，提出一种基于可微分神经架构搜索自适应的密集融合网络的局部-全局融合模块(Local-Global Fusion Module，LG Module)，充分捕获全局关系并保留局部线索，有效解决背景和病灶区域相似度高问题；使用临床多模态肺部医学图像数据集验证模型的有效性，实验结果表明，该文方法在平均梯度，边缘强度，$Q^{AB/F}$，空间频率，标准差，信息熵等感知图像融合质量评价指标上与其他七种方法中最优的方法相比，分别平均提高了21.5%，11%，4%，13%，9%，3%。模型能够突出病变区域信息，融合图像结构清晰且纹理细节丰富。

关　键　词：医学图像融合；深度学习；注意力机制；可微分架构搜索；密集网

中图分类号：TP399　　**文献标识码**：A　　**doi**：10.37188/OPE.20233120.3050

CT and PET medical image fusion based on LL-GG-LG Net

ZHOU Tao[1,2], ZHANG Xiangxiang[1,2*], LU Huiling[3], LI Qi[1,2], CHENG Qianru[1,2]

(1. *School of computer science and engineering, North Minzu University, Yinchuan* 750021, *China*;
2. *Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan* 750021, *China*;
3. *School of Medical information engineering, Ningxia Medical University, Yinchuan* 750004, *China*)
* *Corresponding author, E-mail: zxx*19990503@163.*com*

Abstract: Multimodal medical image fusion plays a crucial role in clinical medical applications. Most of the existing methods have focused on local feature extraction, whereas global dependencies have been insufficiently explored; furthermore, interactions between global and local information have not been considered. This has led to difficulties in effectively addressing the complexity of patterns and the similarity be-

收稿日期：2023-03-31；**修订日期**：2023-05-08.

基金项目：国家自然科学基金资助项目(No. 62062003)；宁夏自然科学基金资助项目(No. 2022AAC03149)；宁夏自治区重点研发计划资助项目(引才专项)(No. 2020BEB04022)；北方民族大学2022年研究生创新项目资助(No. YCX22190)

tween the surrounding tissue (background) and the lesion area (foreground) in terms of intensity. To address such issues, this paper proposes an LL-GG-LG Net model for PET and CT medical image fusion. Firstly, a Local-Local fusion (LL) module is proposed, which uses a two-level attention mechanism to better focus on local detailed information features. Next, a Global-Global fusion (GG) module is designed, which introduces local information into the global information by adding a residual connection mechanism to the Swin Transformer, thereby improving the Transformer's attention to local information. Subsequently, a Local-Global fusion (LG) module is proposed based on a differentiable architecture search adaptive dense fusion network, which fully captures global relationships and retains local cues, thereby effectively solving the problem of high similarity between background and focus areas. The model's effectiveness is validated using a clinical multimodal lung medical image dataset. The experimental results show that, compared to seven other methods, the proposed method objectively improves the perceptual image fusion quality evaluation indexes such as the average gradient (AG), edge intensity (EI), $Q^{AB/F}$, spatial frequency (SF), standard deviation (SD) and information entropy (IE) edge retention by 21.5%, 11%, 4%, 13%, 9%, and 3%, respectively, on average. The model can highlight the information of the lesion areas. Moreover, the fused image structure is clear, and detailed texture information can be obtained.

Key words: medical image fusion; deep learning; attention mechanism; differentiable architecture search; dense network

1　引　言

医学图像融合是指将来自不同技术的图像融合成一幅融合图像，从而最大限度地利用有用信息，减少冗余，与单一模态的医学图像相比，融合图像所包含的纹理结构信息更加丰富，病灶更加明显，减少图像的不确定性和冗余信息，提高临床适用性[1]，从而能够帮助医生在许多临床应用中进行综合诊断、术前规划、术中指导和介入治疗[2]。由于成像方式不同，不同模态医学图像反应的器官结构信息也不同，如 CT 成像利用 X 射线检测骨骼和致密结构的信息，对骨骼的显示很清晰[3]，对病变的定位良好，但对病变本身的显示相对较差，软组织对比度有限。PET 图像对软组织、器官、血管等显示清晰，有利于确定病灶范围，但空间分辨率不如 CT，对刚性的骨组织显示差，并有一定的几何失真[4]。多模态医学图像融合技术通过综合不同模态医学图像之间的互补与冗余信息，为临床疾病诊断与科学研究提供丰富的信息，可以有效辅助医生对病灶进行诊断。

现有的图像融合方法可分为传统的融合方法和基于深度学习的融合方法，其中传统融合方法大致可以分为：基于多尺度变换的方法[5]、基于稀疏表示（Sparse Representation，SR）的方法[6]、混合方法[7]和其他方法[8]，这些方法通常需要手动设计特征提取机制和融合策略，如 Li 等人[9]提出潜在低秩表示（Latent Low-Rank Representation，LatLRR）图像融合方法，将源图像分为低秩部分和显著部分，有效保留边缘轮廓信息。Li 等人[10]提出拉普拉斯再分解（Laplacian Redecomposition，LRD）医学图像融合方法，有效解决颜色失真、模糊和噪声问题。这些方法仍然存在鲁棒性不足、泛化能力弱、优化困难、需要更多计算资源和细节丢失等缺点。

基于深度学习的融合方法可进一步划分为卷积神经网络、编解码网络和生成对抗网络的图像融合方法：基于卷积神经网络的图像融合方法通过利用卷积运算强大的特征提取和重建能力来获得更好的融合性能。Liu 等人[11]提出了一种用于医学图像融合的深度卷积神经网络（Convolutional Neural Networks，CNN），使用连体卷积网络生成权重图，对源图像的像素活动信息进行整合，并通过图像金字塔以多尺度方式进行融合。Tang 等人[12]提出基于残差编解码细节保留交叉网络（Detail Preserving Cross Network，DPCN），该网络采用结构引导的功能特征提取

分支、功能引导的结构特征提取分支，双分支提取架构提取源图像的功能信息和结构信息。但由于只有最后一层的结果被用作图像特征，因此容易丢失中间层信息。基于编解码网络的图像融合方法通过设计和训练由卷积神经网络构成的编码器和解码器得到融合图像，有效地避免神经网络深度对性能的影响，如DenseFuse[13]，在编码器中引入密集连接机制有效解决中间层信息丢失问题，Res2Net[14]将Resnet模块用于编码器中，提高网络的多尺度特征提取能力。DIFNet(Deep Image Fusion Net)[15]采用双编码器生成与高维输入图像具有相同对比度的输出图像。EMFusion(Enhanced Medical Image Fusion Network)[16]通过表层和深层约束以增强信息保存，其中表层水平约束基于显着性和丰富测量，深层约束是通过训练编码器定义的，有效解决了信息失真的问题。这些方法有效解决中间层信息丢失问题，但训练过程中专注于图像重建任务，无法正确的提取出融合所需要的显著特征，忽略全局特征的提取。基于生成对抗网络的图像融合方法在生成器和鉴别器之间建立对抗博弈，可以无监督地估计目标的概率分布，从而以隐式方式实现特征提取、特征融合和图像重建[17]。Ma等人将GAN(Generative Adversarial Network)[18]引入图像融合领域，提出一种名为FusionGAN的融合方法。DDcGAN(Dual-Discriminator Conditional Generative Adversarial Network)[19]包含两个鉴别器来驱动生成器融合源图像特征信息，以保持与两个输入图像之间的最大相似性。Zhang等人[20]提出了一种具有全尺度跳跃连接和双马尔可夫鉴别器(GAN with Full-scale skip connection and dual Markovian discriminators, GAN-FM)的生成对抗融合网络，以充分保留源图像中的有效信息，保留显着对比度和丰富纹理。然而这些方法模型复杂，使得训练过程不稳定，生成器和鉴别器之间的对抗不充分导致融合图像失真。

基于深度学习网络的医学图像融合是近几年的研究热点。但是根据上述文献报道以及医学图像的特点，现有融合模型还存在一些不足，从问题的角度来看，由于成像机制的不同，不同模态的医学图像侧重于不同类别的器官或组织信息，存在周围组织与病灶区域之间的模式复杂性和强度相似性问题；从方法的角度来看，由于卷积神经网络有限的感受野，特征提取过程中主要关注图像的局部信息，难以捕获全局上下文语义信息，忽略全局特征与局部特征的交互。为此本文从问题角度出发，充分考虑CNN网络的特点，加强全局特征与局部特征的交互，提出一种用于PET和CT图像的医学图像融合方法LL-GG-LG Net，其主要贡献是：(1)构造了用于提取局部-全局信息的三分支融合网络，有效提取源图像的局部信息和全局信息，增强局部-全局的信息交互能力；(2)设计了局部-局部融合模块(Local-Local Fusion Module，LL Module)，通过两次空间注意力获取PET/CT的局部融合信息，生成局部融合图像；(3)提出了全局-全局融合模块(Global-Global Fusion Module，GG Module)，通过在Swin Transformer中添加残差连接机制以提高全局信息的融合性能，生成全局融合图像；(4)为了提高局部-全局信息的交互能力，进一步增强融合图像质量，提出局部-全局融合模块(Local-Global Fusion Module，LG Module)，聚合局部融合图像特征和全局融合图像特征，使融合图像病变区域显著、细节丰富且鲁棒性高。

2　LL-GG-LG Net模型

本文提出用于PET和CT医学图像融合的LL-GG-LG Net融合方法，该方法框架由局部-局部融合模块(Local-Local Fusion Module，LL Module)、全局-全局融合模块(Global-Global Fusion Module，GG Module)和局部-全局融合模块(Local-Global Fusion Module，LG Module)三部分组成，首先将配准好的医学图像I_{PET}，I_{CT}分别经过LL Module和GG Module，得到局部融合图像F_L和全局融合图像F_G，然后将F_L和F_G输入到LG Module中重建得到最终融合图像F_M。该方法有效解决背景和病灶区域相似度高，提取全局信息特征能力有限，局部-全局信息交互能力弱，难以有效保留病变区域复杂信息的问题。

2.1　LL-GG-LG Net网络结构

LL-GG-LG Net网络结构如图1所示，首先将源图像转换为卷积层的特征表示，并输入LL-

GG-LG Net 网络的局部-局部融合模块分支中，通过空间注意力机制提取源图像边缘纹理信息，生成注意力图，并将源图像和生成的权重图进行相乘操作，最后进行累加生成局部融合图像；然后通过 1×1 卷积将源图像进行位置编码，生成序列向量，并将其输入到全局-全局融合模块中，提取全局特征，采用 L_1-norm 融合规则生成全局融合图像；最后，将双分支融合网络所生成的局部融合图像和全局融合图像采用局部-全局融合模块进行图像重构，提高局部-全局信息的交互，增强融合图像质量。

2.2　局部-局部融合模块（LL Module）

注意力机制通常根据源有特征图设计一个权重分布，通过其权重分布区分每个区域的重要性，再将该权重分布施加到源特征图上，抑制背景中的不同干扰，使得不同特征拥有不同权值，其中权值大的特征更加容易被注意到，因此在计算机视觉领域得到广泛应用[21]。空间注意力机制[22]通过平均池化和最大池化获得全局信息，但由于不同的信息表征，层次化特征的注意焦点有很大的不同，因此为了有效提取源图像局部特征，本文设计了双层注意力模块提取局部细节信息，该模块通过两次平均池化和最大池化提取源图像的边缘和背景信息，用来凸显像素层次上的重要空间位置特征，使得病变区域和骨骼纹理的边界特征明显。

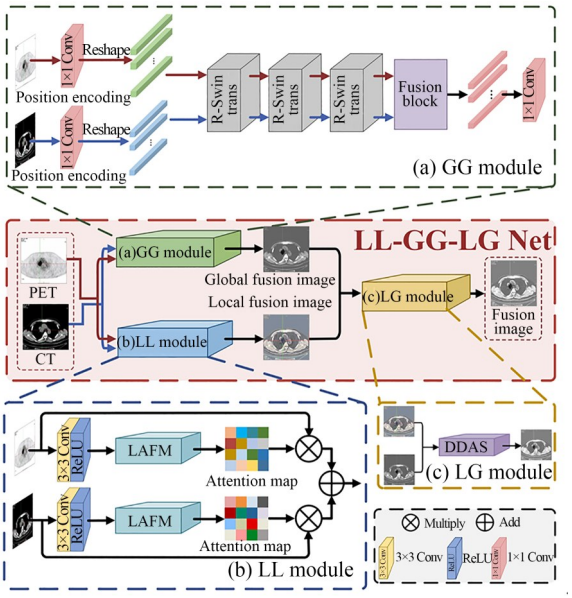

图 1　LL-GG-LG Net 医学图像融合网络

Fig. 1　LL-GG-LG Net Medical Image Fusion Network

双层注意力模块结构如图 2 所示。首先将源图像转换为卷积层的特征表示，并输入空间注意力模块中，空间注意力模块首先通过平均池化层和最大池化层分别对输入特征通道域上进行池化操作后再拼接在一起，得到 2 倍通道特征图，接着通过 1×1 卷积将其压缩为原通道，并再次通过平均池化层和最大池化层分别对输入特征通道域上进行池化操作后再拼接在一起，得到深层特征图，并通过 3×3 卷积压缩其通道，此外，为了补偿在最后卷积层中的上采样操作期间丢失的特征，本文使用卷积层的一个跳跃连接，在下/上采样操作后，局部-局部融合模块可以完全包含源图像局部特征信息，最后通过 sigmoid 激活函数归一化空间权重信息，得到空间注意力权重图 $\boldsymbol{M}_i^2$，最后将输入特征图 $\boldsymbol{I}_i$ 和权重图 $\boldsymbol{M}_i^2$ 对应元素相乘，得到最终的局部融合图像；局部-局部融合模块公式如式（1）所示：

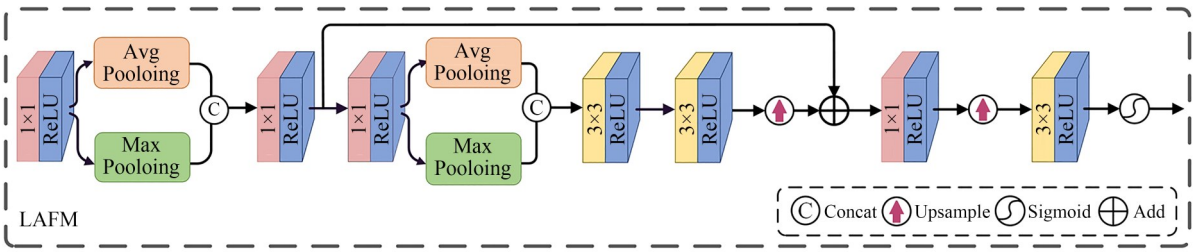

图 2　双层注意力模块

Fig. 2　Two-layer attention module

第一层局部特征信息表示：

$$\boldsymbol{M}_i^1 = \boldsymbol{C}^{1\times1}\left[\mathrm{Avgpool}(\boldsymbol{I}_i), \mathrm{Maxpool}(\boldsymbol{I}_i)\right], \quad (1)$$

其中：$\boldsymbol{M}_i^1$ 表示第一层局部特征信息，$\boldsymbol{I}_i$ 表示源图像，$i \in 1, 2$，源图像 PET 图像记为 $\boldsymbol{I}_1$，源图像 CT

记为 I_2，$C^{1\times1}$ 表示 1×1 大小卷积核的卷积层。

最终权重图表示公式如下：

$$M_i^2=\delta\left[C^{3\times3}\left(\text{Avgpool}\left(M_i^1\right),\text{Maxpool}\left(M_i^1\right)\right)\right], \tag{2}$$

其中：$\delta(\cdot)$ 表示 Sigmoid 激活函数，$C^{3\times3}$ 表示 3×3 大小卷积核的卷积层，生成最终的局部融合图，公式表达为：

$$I_A=\sum_{i=1}^{2}M_i^2\odot I_i, \tag{3}$$

其中：I_A 表示最终生成的局部融合图像，⊙ 表示点乘操作。

2.3　全局-全局融合模块(GG Module)

CNN 通过利用卷积运算强大的特征提取和重建能力来获取图像特征并重建融合图像，然而根据局部处理的原理，CNN 对远程依赖建模的能力有限。Transformer[23] 模型通过自注意力机制来捕获上下文之间的全局交互信息，突出病变组织结构特征，但忽略了局部相关性对于病灶特征融合的重要性。因此，为了有效提高网络全局感知能力，保留病变区域的有效特征，本文使用全局-全局融合模块，该模块在 Swin Transformer 中添加残差连接机制(Residual-Swin Transformer Module，RSTM)有效聚合不同层次的特征，弥补了 Transformer 对病变区域特征提取弱的问题，提高 Transformer 在全局特征提取过程中关注病变局部特征的相关性，并采用 L_1-$norm$ 融合规则生成全局融合图像。

全局-全局融合模块内部结构如图 1(a)所示，本文首先将输入图像 $I_i\in R^{H\times W\times3}$(其中 H、W 和 3 分别表示其高度、宽度和通道大小，$i=1$ 表示源图像 PET 图像，$i=2$ 表示源图像 CT 图像)，首先采用 1×1 卷积操作对源图像进行位置编码，并将特征维度映射到维度 C，C 设置成 96，生成序列向量 $F_{SV}^i\in R^{MN\times C}$($i=1$ 表示源图像 PET 图像，$i=2$ 表示源图像 CT 图像)，然后应用三个 R-Swin Transformer 块提取全局特征 F_G^i，其表达公式如下：

$$F_G^i=H_{RSTB_m}\left(F_{SV}^i\right), \tag{4}$$

其中：$RSTB_m$ 表示第 m 个 R-Swin Transformer 块，通过以上操作，提取 PET 和 CT 图像的全局特征，然后，采用基于行向量维数和列向量维数的 L_1-norm 融合规则进行特征融合，得到融合的全局特征 F_G，其公式表示为：

$$F_G=H_{\text{Norm}}\left(F_G^i\right), \tag{5}$$

其中：H_{Norm} 表示 L_1-norm 融合操作，最后使用卷积层重构融合图像 I_F，其公式为：

$$I_F=H_{\text{Conv}}\left(F_G\right), \tag{6}$$

其中：H_{Conv} 表示特征重构卷积操作，该卷积为 1×1 卷积，padding 设置为 0。

2.3.1　R-Swin Transformer 模块

图 3(a)展示了 R-Swin Transformer 模块(Residual-Swin Transformer Module，RSTM)的体系结构，它包括一系列 STL 以及残差连接，给定输入序列向量 $F_{SV,0}^i$，本文应用 n 个 STL 来提取中间全局特征 $F_{SV,n-1}^i$($i=1$ 表示源图像 PET 图像，$i=2$ 表示源图像 CT 图像)，RSTM 的最终输出由式(7)计算：

$$F_{SV}^i=H_{RSTB_n}\left(F_{SV,n-1}^i\right)+F_{SV,0}^i, \tag{7}$$

其中：H_{RSTB_n} 表示第 n 个 STL。类似于 CNN 的架构，多层 Swin Transformer 可以有效地对全局特征进行建模，残差连接可以对不同层次的特征进行聚合。

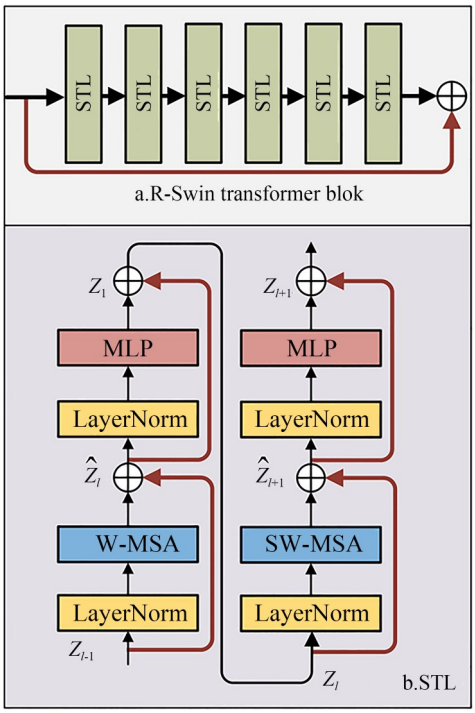

图 3　R-Swin Transformer 模块结构

Fig. 3　R-Swin Transformer module structure

如图 3(b)所示，STL 由多头自注意(Multi-headed Self-attention，MSA)和多层感知器(Multiple perceptron，MLP)组成。此外在每个 MSA 模块和每个 MLP 之前都会应用 LN(LayerNorm)层，并且在每个模块之后采用残差连接，因此可以将 Swin Transformer 中的第 l 层的输出表示为：

$$\widehat{Z}_l = \mathrm{MSA}\left(LN\left(Z_{l-1}\right)\right) + Z_{l+1}, \tag{8}$$

$$Z_l = MLP\left(LN\left(\widehat{Z}_l\right)\right) + \widehat{Z}_l. \tag{9}$$

由于 W-MSA 的窗口之间相互作用较弱，为了在不增加计算量的情况下引入跨窗口交互，Swin Transformer 则在 W-MSA 结构之后添加了 SW-MSA 模块，该模块的窗口配置与 W-MSA 不同，其通过向左上方向循环移动来开发高效的批处理方法，在此移动之后，批处理窗口可以由特征图中的多个不相邻窗口组成，同时在 W-MAS 中保持相同数量的批处理窗口作为规则划分。W-MSA 和 SW-MAS 中保持相同数量的批处理窗口作为规则划分。W-MSA 和 SW-MSA 在局部窗口内进行自注意计算时，在计算相似度时都考虑了相对位置偏差。

利用这种移位窗口划分机制，SW-MSA 和 MLP 模块的输出可以写为：

$$\hat{Z}_{l+1} = \mathrm{SW\text{-}MSA}\left(\mathrm{LN}\left(Z_l\right)\right) + Z_l, \tag{10}$$

$$Z_{l+1} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{Z}_{l+1}\right)\right) + \hat{Z}_{l+1}, \tag{11}$$

其中：$\hat{Z}_{l+1}$ 和 Z_{l+1} 分别表示第 $l+1$ 层中的 SW-MSA 和 MLP 的输出，如图 2(b)所示。在 W-MSA 和 SW-MSA 中计算的自注意可以写成：

$$\mathbf{Q} = Z_l W_Q, K = Z_l W_K, V = Z_l W_V, \tag{12}$$

$$\mathrm{Attention}\left(Z_l\right) = \mathrm{SoftMax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d}} + B\right)V, \tag{13}$$

其中：W_Q，W_K 和 $W_V \in R^{D\times d}$ 是跨不同窗口共享的三个线性投影层的可学习参数，Q，K. $V \in R^{L\times d}$ 为查询矩阵、键矩阵、值矩阵，d 表示查询或的维度，$B \in R^{L\times L}$ 表示相对位置偏差。

2.3.2　L_1-norm 融合规则

基于 L_1-norm 的 PET 和 CT 图像序列矩阵的融合策略，并测量了它们的行和列向量维的活跃度。本文将 PET 和 CT 的全局特征分别定义为 $F_G^{\mathrm{PET}}(i,j)$ 和 $F_G^{\mathrm{CT}}(i,j)$，首先使用 L_1-norm 计算它们的行向量权重，通过 Softmax 函数获得它们的活跃度，即 $H_r^{\mathrm{PET}}(i)$ 和 $H_r^{\mathrm{CT}}(i)$，它们由以下方程表示：

$$H_r^{\mathrm{PET}}(i) = \frac{\exp\left(\left\| F_G^{\mathrm{PET}}(i)\right\|_1\right)}{\exp\left(\left\| F_G^{\mathrm{PET}}(i)\right\|_1\right) + \exp\left(\left\| F_G^{\mathrm{CT}}(i)\right\|_1\right)}, \tag{14}$$

$$H_r^{\mathrm{CT}}(i) = \frac{\exp\left(\left\| F_G^{\mathrm{CT}}(i)\right\|_1\right)}{\exp\left(\left\| F_G^{\mathrm{PET}}(i)\right\|_1\right) + \exp\left(\left\| F_G^{\mathrm{CT}}(i)\right\|_1\right)}, \tag{15}$$

其中：$\|\cdot\|_1$ 表示 L_1-norm，然后直接将它们的活跃度与相应的全局特征相乘，从行向量维度获得融合的全局特征，称为 $F_{\mathrm{row}}(i,j)$，其公式如下：

$$F_{\mathrm{row}}(i,j) = F_G^{\mathrm{PET}}(i,j) \times H_r^{\mathrm{PET}}(i) + F_G^{\mathrm{CT}}(i,j) \times H_r^{\mathrm{CT}}(i). \tag{16}$$

同理，从列向量维测量其活跃度，并将 PET 和 CT 图像的列向量活跃度分别记为 $H_l^{\mathrm{PET}}(j)$ 和 $H_l^{\mathrm{CT}}(j)$，它们由以下公式表示：

$$H_l^{\mathrm{PET}}(j) = \frac{\exp\left(\left\| F_G^{\mathrm{PET}}(j)\right\|_1\right)}{\exp\left(\left\| F_G^{\mathrm{PET}}(j)\right\|_1\right) + \exp\left(\left\| F_G^{\mathrm{CT}}(j)\right\|_1\right)}, \tag{17}$$

$$H_l^{\mathrm{CT}}(j) = \frac{\exp\left(\left\| F_G^{\mathrm{CT}}(j)\right\|_1\right)}{\exp\left(\left\| F_G^{\mathrm{PET}}(j)\right\|_1\right) + \exp\left(\left\| F_G^{\mathrm{CT}}(j)\right\|_1\right)}. \tag{18}$$

然后，得到列向量维的融合全局特征，并将其称为 $F_{\mathrm{col}}(i,j)$，它由以下方程表示：

$$F_{\mathrm{col}}(i,j) = F_G^{\mathrm{PET}}(i,j) \times H_l^{\mathrm{PET}}(j) + F_G^{\mathrm{CT}}(i,j) \times H_l^{\mathrm{CT}}(j). \tag{19}$$

对融合后的全局特征在行向量维和列向量维度上进行逐元加法运算，得到最终的融合全局特征，其计算公式如下：

$$F(i,j) = F_{\mathrm{row}}(i,j) + F_{\mathrm{col}}(i,j). \tag{20}$$

最后利用融合后的全局特征通过卷积层重建全局融合图像。

2.4　局部-全局融合模块(LG Module)

由于不同病变区域的形状和大小均不同，导致网络在提取边缘和纹理信息的同时，难以保留病变特征，使得模型在关注局部病变特征的同时

难以进行病灶空间定位，因此本文通过添加局部-全局融合模块进行全局信息和局部信息的交互融合。根据密集网[24]的特性，即每一层的输出将级联到下一个输入，使得网络能够更好地保留每一层提取的特征，增强了特征提取能力，有效保留局部信息特征和全局信息，本文引入密集网络进行最后一步融合。此外受神经网络架构[25]的启发，考虑到输入图像特征不同，所需提取的信息也有所差异，相同的特征提取操作，难以兼顾不同源图像的特征，因此本文设计基于神经网络架构搜索策略训练的密集网进行局部-全局信息融合，为了降低计算成本，提高网络的性能和效率，本文采用可微分神经网络架构[26]自适应的构建和搜索以学习最佳密集网络架构（Darts: Differentiable Architecture Search，DDAS），保留局部融合图像和全局融合图像的差异信息，解决融合图像噪声和背景与病灶区域相似度高问题。

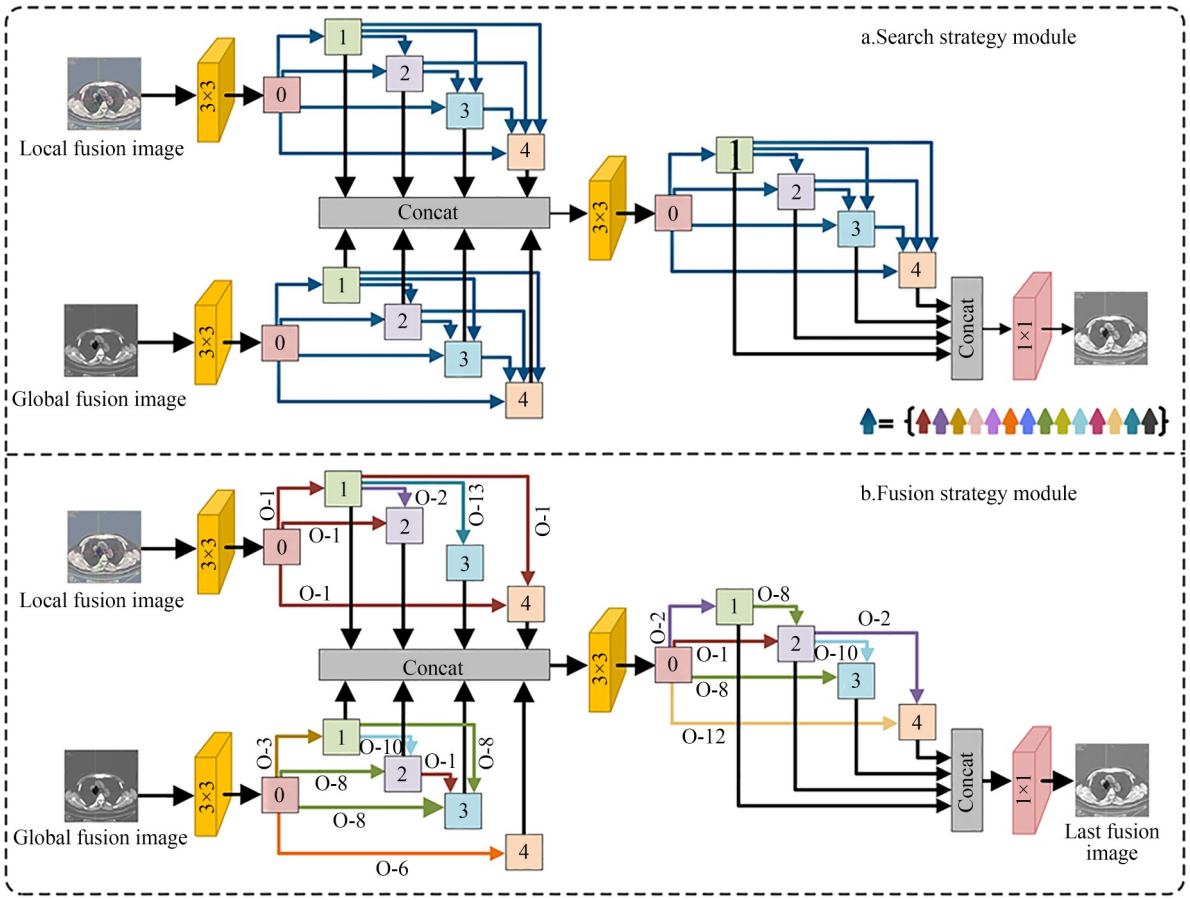

图 4　DDAS 模块结构

Fig. 4　DDAS module structure

DDAS 模块结构如图 4 所示。首先，采用搜索策略模块训练密集网络，预定义了三个相同的密集连接网络，每个网络都包含五个节点的非循环图，即 0，1，2，3，4，其中每个节点表示一组特征图，第一个节点由先前 3×3 卷积操作得到的特征图作为输入节点，每个节点的输出都级联到后续所有节点的输入，本文将局部融合图像和全局融合图像分别输入到密集网中，根据融合策略自适应的训练网络，经过训练选出最终的融合网络架构，如图 4 所示，将局部融合图像和全局融合图像分别输入到已经训练好的密集网中，进行特征提取和重构操作，生成最终的融合图像，实验表明该方法有效解决噪声问题以及背景和病灶区域相似度高问题。

2.4.1　搜索空间

本文搜索空间 0 选择如图 5 所示（彩图见期

刊电子版），其中运算集合 $O(\cdot)$ 主要包括 1×1 卷积、3×3 卷积、5×5 卷积、7×7 卷积、3×3 扩张卷积、5×5 卷积、7×7 扩张卷积、1×1 残差卷积、3×3 残差卷积、5×5 残差卷积、7×7 残差卷积、3×3 扩张卷积、5×5 残差扩张卷积、7×7 残差扩张卷积等 14 种操作，并以不同颜色箭头进行标注。

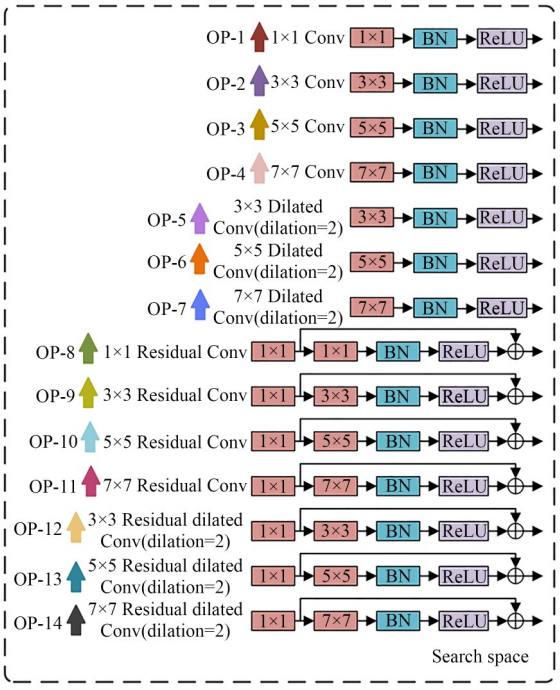

图 5　搜索空间模块结构

Fig. 5　Search space module structure

2.4.2　搜索策略

本文采用的可微搜索策略，具体结构如图 4 所示，首先预定义了三个相同的单元组成网络，每个单元包含五个节点的非循环图，即 0，1，2，3，4，其中每个节点表示一组特征图，第一个节点由先前 3×3 卷积操作得到的特征图作为输入节点，中间节点 j 与前身节点 i 之间的信息流由边 $E_{(i,j)}$ 连接，中间节点是其先前边的输出总和，其表示为 N_j，公式表达如下：

$$\bar{o}^{(i,j)}(X_i)=\sum_{o\in 0_{i,j}}\frac{\exp\left(\alpha_o^{(i,j)}\right)}{\sum_{O'\in 0}\exp\left(\alpha_{O'}^{(i,j)}\right)}o(X_i),\quad(21)$$

$$N_j=\sum_{i<j}f_{i,j}(X_i).\quad(22)$$

中间节点是其先前边输出的总和，而输出节点则是通道维度中所有节点的串联，其表示为：

$$N_{\text{output}}=\text{Concat}(N_1,N_2,N_3,N_4),\quad(23)$$

其中：搜索空间 0 表示候选运算 $O(\cdot)$ 集合，$\alpha_0^{(i,j)}$ 表示归一化结构参数，X_i 表示第 i 个节点，体系结构搜索的任务归结为学习变量 $\alpha=\left\{\alpha^{(i,j)}\right\}$，在搜索结束时，可以通过用最有可能的操作替换每个混合操作 $\bar{o}^{(i,j)}$，进而获得离散结构，即 $o^{(i,j)}=argmax_{o\in 0}\alpha_o^{(i,j)}$（在集合 0 中选取一个子操作 $o^{(i,j)}$ 使得 $\alpha_o^{(i,j)}$ 最大）。

在搜索阶段，本文需要解决一个双层优化问题，本文使用 L_{Train} 和 L_{Val} 分别表示训练和验证损失，这两个损失由 α 和 w 确定（α 表示三个单元的体系结构，w 表示网络的权重，*表示最优权重），其中体系结构的目标是找到 α^*，使得 $L_{\text{Val}}(w,\alpha^*)$ 损失最小化；找出 w^* 使得 L_{Train} 训练的损失最小，即 $w^*=argmin_w L_{\text{Train}}(w,w^*)$。

$$\min_{\alpha} L_{\text{Val}}(w^*(\alpha),\alpha),\quad(24)$$

$$\text{st}\quad w^*(\alpha)=argmin_w L_{\text{Train}}(w,\alpha).\quad(25)$$

为了简化训练搜索任务，添加了特殊的空操作，及两个节点之间缺少连接；经训练选出一组最合适的融合架构，具体结构如图 5 所示，其中 0，1，2，3，4 五个节点之间的操作均采用不同颜色，其颜色与搜索空间颜色相对应，并对应其具体操作。

3　实验结果与分析

3.1　实验设置

数据集选用 2018 年 1 月至 2020 年 6 月在宁夏某三甲医院核医学进行 PET/CT 全身检查的肺部肿瘤临床患者，以图像质量符合分析要求（图像质量清晰、无伪影、病灶可见），患者未接收射频消融，肺切除治疗，且病理报告完整详细为实验纳入标准，有 95 名符合实验入选条件的患者纳入实验，身高不限，其中包括女性 46 例（占 48%），年龄 30～80 岁，平均年龄（54.32±4.21）岁。男性 49 例（占 52%），年龄 27～74 岁，平均年龄（50±5.11）岁。患者禁食 6 h，控制血糖 10 以下，显像前排尿，去除金属饰物。静脉注射氟[^{18}F]脱氧葡萄糖注射液（^{18}F-FDG）3.7 mbq/kg，注射完显像剂一小时后在安静、避光的房间平卧 45～60 min 进行肺部及躯干部 PET-CT 图像采集，扫描完成取横断面、矢状面与冠状面图像。

数据集图像标准化摄取值≥2.5为阳性，采用GE公司Discovery MI型PET/CT机进行扫描检查。所有CT扫描均固定电压在120 kV、电流在90～200 mA进行曝光，Thick为3.75 mm，Iterval为3 270，SFOV为Large，DFOV为50 cm，ReconType为std，操作者均为多年从事CT及核医学工作的资深技术人员，为确保对病变进行正确标注，为确保数据准确性，本次数据由三位专家医生结合临床综合诊断，进行评估，结果以多数人意见为准，三位专家医生包括一位具有8年临床经验的胸外科医生，一位具有5年临床经验的呼吸内科医生，一位影像科专业医生。数据集经旋转、镜像的数据增强与数据增广处理，三种模态图像数据集的最终样本数分别为2 430张，其中选取1 000张CT与PET图像作为训练集，200张作为验证集，200张作为测试集，图像标签由两位临床医师手动绘制。原始图像格式为DICOM格式，扫描层厚为7 mm，由于融合结果的效果还受到灰度不均匀性、伪影等因素的影响，且原图像直接输入网络会造成训练困难，因此有必要对图像进行预处理使网络实现更好的融合效果。本文用算法将数据读取之后转换为JPG格式，并进行Resize操作，将其变为356 pixel×356 pixel。

实验室硬件环境服务器Intel(R) Xeon(R) Gold6154 CPU，内存256 GB，显卡NVIDIA TITANV，实验环境框架采用pytorch，python版本为3.7.0，CUDA版本为11.1.106。训练时Batchsize被设置为4，训练150个epoch，选择学习速率为1e-4的Amda优化器。

3.2 对比实验

为了验证LL-GG-LG Net的有效性，选取两种基于分解变换的融合方法，五种基于深度学习的图像融合网络，对CT图像和PET图像的融合结果进行比较。方法一：LATLRR变换[9]。方法二：LRD[10]。方法三：DPCN[12]。方法四：Res2Net[14]。方法五：DIFNet[15]。方法六：EMFusion[16]。方法七：DDcGAN[19]。以上五种基于深度学习的图像融合网络的参数值设置为其作者指定的默认值。本文从定性与定量两个方面评价本文方法的有效性。

3.2.1 定性比较

图6展示了本文提出的方法与上述七种比较方法的融合结果图，针对CT纵膈窗和PET图像融合进行了主观对比实验，可以看出LATLRR结果展示图中融合结果比较模糊，病灶部位轮廓显示不清；LRD方法融合的图像中细节丢失严重，影响医生对疾病信息的识别；Res2Net增强了边缘信息，但融合图像未能保持适当的亮度，病灶骨骼信息模糊；DIFNet融合方法虽然增强了对比度，但图像亮度过高影响视觉效果，且具有一定的噪声；DPCN，DDcGAN和EMFusion方法可以突出病变区域信息，但融合后的图像对比度

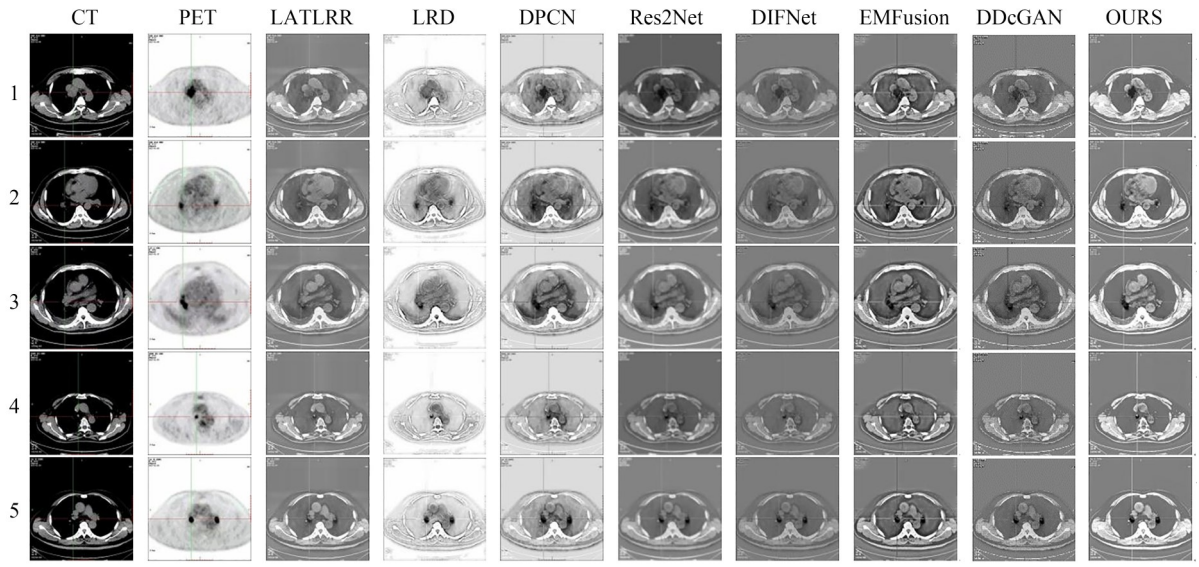

图6　不同方法主观对比融合图像

Fig. 6　Subjective comparison of fused images by different methods

较低，边缘信息模糊。相比之下，本文方法能够有效保留源图像的边缘细节及轮廓特征，病变区域的信息也更加丰富完整，有效解决背景与病灶区域之间的模式复杂性和强度相似性问题，方便医生的观察。

3.2.2　定量比较

为了客观且全面的评价模型的融合性能，同时便于与其他算法进行比较，本文从融合图像细节信息丰富度，清晰度、边缘保留程度、边缘信息量、纹理信息多个角度评估融合性能，选取以下六种常见评价指标进行比较：平均梯度 AG、边缘强度 EI、边缘信息传递因子 $Q^{AB/F}$、空间频率 SF、标准差 SD、信息熵 IE。以上指标值越大，性能越好。

(1)平均梯度。

平均梯度(Average Gradient, AG)反映了融合图像的细节和纹理信息。该数值越大，融合图像信息越丰富，融合性能越好。公式如下：

$$AG=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2}{2}},\tag{26}$$

其中：M 和 N 分别代表融合图像的高度和宽度，$F(i,j)$ 表示图像的第 i 行第 j 列像素值。

(2)边缘强度。

边缘强度(Edge Intensity, EI)，边缘强度越大，融合图像质量越高。公式如下：

$$\mathrm{EI}=\frac{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}S_X(i,j)^2+S_Y(i,j)^2}}{M\times N},\tag{27}$$

$$S_X=F\times h_x,S_Y=F\times h_y,\tag{28}$$

其中：M,N 为图像的宽高；h_x,h_y 为 x 和 y 方向的 Sobel 算子；$\boldsymbol{S}_x$ 和 $\boldsymbol{S}_y$ 为 Sobel 算子卷积后的结果。

(3)基于边缘的相似性度量 $\boldsymbol{Q}^{AB/F}$。

$\boldsymbol{Q}^{AB/F}$ 衡量融合图像保留源图像的边缘信息数量，计算得到的 $\boldsymbol{Q}^{AB/F}$ 的取值范围为[0,1]，其值越接近 0，表示损失的边缘信息越多；该值越大，表示融合性能越好。设图像 A，B，大小为 $n\times m$，融合图像为 $\boldsymbol{F}$，边缘信息保持度 $\boldsymbol{Q}^{AF}(i,j)$ 和 $\boldsymbol{Q}^{BF}(i,j)$，分别用 $\boldsymbol{W}^A(i,j)$ 和 $\boldsymbol{W}^B(i,j)$ 进行加权，得到融合图像 $\boldsymbol{F}$ 相对于图像 A 和图像 B 的边缘保持度。公式如下：

$$\boldsymbol{Q}^{AB/F}(i,j)=\frac{\sum_{i=1}^{N}\sum_{m=1}^{M}\boldsymbol{Q}^{AF}(i,j)\boldsymbol{W}^A(i,j)+\boldsymbol{Q}^{BF}(i,j)\boldsymbol{W}^B(i,j)}{\sum_{i=1}^{N}\sum_{m=1}^{M}\left(\boldsymbol{W}^A(i,j)+\boldsymbol{W}^B(i,j)\right)},\tag{29}$$

其中：$\boldsymbol{W}$ 是固定大小的滑动窗口，$0\leqslant\boldsymbol{Q}^{AB/F}(i,j)\leqslant 1$。$\boldsymbol{Q}^{AB/F}$ 的值越高，融合图像保留源图像的边缘信息越丰富。

(4)空间频率。

空间频率(Spatial Frequency, SF)反映图像的整体清晰度，空间频率越大，融合图像包含的边缘和纹理信息越丰富，融合性能也就越好。公式如下：

$$\mathrm{SF}=\sqrt{RF^2+CF^2},\tag{30}$$

$$RF=\frac{1}{MN}\sqrt{\sum_{i=2}^{M}\sum_{j=1}^{N}\left(\left(I(i,j)-I(i-1,j)\right)\right)^2},\tag{31}$$

$$CF=\frac{1}{MN}\sqrt{\sum_{i=2}^{M}\sum_{j=1}^{N}\left(\left(I(i,j)-I(i,j-1)\right)\right)^2},\tag{32}$$

其中：RF,CF 分别表示空间行频率和空间列频率，M 和 N 分别表示融合图像的高度和宽度，$I(i,j)$ 表示图像的第 i 行第 j 列像素值。

(5)标准差。

标准差(Standard Deviation, SD)衡量信息的丰富程度，标准差越大，图像的灰度级分布越分散，图像的信息量越多。公式如下：

$$SD=\frac{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I(i,j)-\mu\right)^2}}{M\times N},\tag{33}$$

$$\mu=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}H(i,j),\tag{34}$$

其中，μ 表示均值，反映亮度信息。

(6)信息熵。

信息熵(Information Entropy, IE)衡量图像中所包含的信息数量。公式如下：

$$\mathrm{IE}=-\sum_{l=0}^{N-1}P_l\log_2 P_l,\tag{35}$$

其中：l 表示图像的灰度等级，P_l 表示融合图像中相应灰度级的归一化直方图。

本实验采用测试数据集的 200 对 CT 和 PET 图像分五组，根据对比的七种融合方法以及本文方法生成融合图像，表 1 展示了在不同指标上每组图像不同融合方法融合结果的平均值。

表 1　客观评价指标均值

Tab. 1　Mean value of objective evaluation metrics.

No.	Methods	AG	EI	$Q^{AB/F}$	SF	SD	IE
1	LATLRR	9.669	55.903 0	0.470 4	26.66	36.52	6.537 1
	LRD	10.009	67.610 9	0.378 6	27.85	39.11	6.007 8
	DPCN	8.568	65.025 6	0.372 7	24.31	41.52	6.057 2
	Res2Net	4.860	34.026 1	0.188 4	13.34	33.01	5.565 5
	DIFNet	5.676	44.391 2	0.271 6	17.92	23.78	5.834 2
	EMFusion	9.487	75.529 0	0.509 9	30.03	40.46	6.254 3
	DDcGAN	10.609	57.058 7	0.310 0	29.16	32.35	5.876 9
	OURS	**12.901**	**78.936 0**	**0.520 0**	**34.42**	**46.63**	**6.555 2**
2	LATLRR	9.494	54.709 2	0.469 8	25.63	36.33	6.445 1
	LRD	10.152	70.681 0	0.415 4	29.12	42.90	6.361 0
	DPCN	9.143	70.115 4	0.437 3	24.80	39.65	6.384 3
	Res2Net	3.568	34.822 2	0.188 1	13.74	33.13	5.657 4
	DIFNet	5.445	43.107 7	0.263 9	17.83	24.84	5.679 1
	EMFusion	9.225	70.558 6	0.512 3	29.14	40.73	6.080 6
	DDcGAN	10.383	56.541 6	0.302 5	27.57	32.41	6.146 7
	OURS	**12.577**	**75.507 8**	**0.532 8**	**32.91**	**44.72**	**6.586 3**
3	LATLRR	9.312	56.217 1	0.458 9	24.45	33.31	6.209 9
	LRD	10.568	75.193 2	0.418 6	29.38	43.38	6.067 0
	DPCN	9.234	70.159 1	0.416 1	24.76	42.98	6.656 2
	Res2Net	4.845	32.287 0	0.195 1	13.11	34.59	5.190 9
	DIFNet	6.593	45.018 0	0.263 7	19.95	26.04	5.224 6
	EMFusion	10.421	73.309 5	0.483 3	29.24	42.49	6.325 1
	DDcGAN	10.542	59.717 5	0.322 7	27.55	35.69	6.520 1
	OURS	**12.771**	**78.064 2**	**0.514 2**	**32.77**	**48.94**	**6.766 2**
4	LATLRR	7.344	52.859 9	0.463 6	24.14	40.98	6.777 6
	LRD	7.854	56.040 9	0.418 6	26.49	44.90	6.276 7
	DPCN	7.461	57.857 8	0.416 4	23.82	46.99	6.251 3
	Res2Net	4.735	33.297 2	0.186 9	13.26	34.48	5.921 4
	DIFNet	5.805	41.799 6	0.262 8	17.73	26.50	6.163 3
	EMFusion	9.236	65.856 7	0.503 3	29.90	44.49	**7.095 4**
	DDcGAN	8.390	47.059 3	0.323 7	27.55	35.69	6.521 0
	OURS	**10.257**	**74.071 5**	**0.518 1**	**33.77**	**48.65**	6.966 2
5	LATLRR	8.243	59.719 8	0.472 5	24.02	35.28	6.483 7
	LRD	8.715	63.709 9	0.393 2	27.34	39.86	5.926 2
	DPCN	8.040	62.557 3	0.407 1	23.81	41.66	5.923 7
	Res2Net	5.340	36.794 4	0.176 0	14.08	26.97	5.596 6
	DIFNet	6.960	48.410 1	0.280 5	19.41	22.82	5.754 2
	EMFusion	9.773	71.245 6	0.489 6	27.87	39.52	6.325 8
	DDcGAN	9.091	51.866 5	0.299 9	26.08	30.74	5.740 8
	OURS	**11.048**	**90.424 1**	**0.525 4**	**31.67**	**44.93**	**6.543 2**

Bold font is the best result for each column

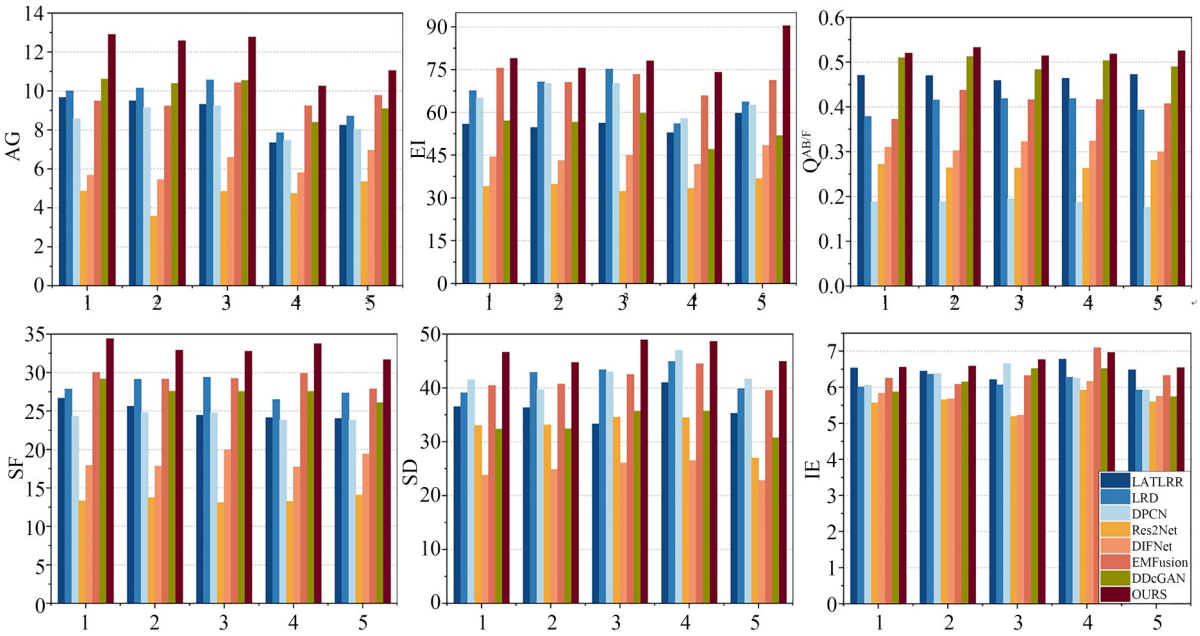

图 7　融合图像评价指标柱状图

Fig. 7　Histogram of fused image evaluation metrics

在表 1 和图 7 中可以看出，本文方法的空间频率（SF）和标准差（SD）明显高于其他七种方法，说明本文方法融合清晰度高，凸显了 PET 图像的病灶区域特征，AG，EI，$Q^{AB/F}$、IE 评价指标上，本文方法与其他七种方法相比有很大提升，说明本文方法保留边缘细节信息能力较好，融合图像病变区域信息丰富。

3.3　消融实验

为了验证 LL-GG-LG Net 模型中局部-局部融合模块（LL Module）、LL Module 中两层空间注意力模块、全局-全局融合模块（GG Module）、GG Module 中 Swin Transformer 添加的残差连接和局部-全局融合模块（LG Module）的有效性，本文设计了六组消融实验来进行对比，实验 1 是仅保留一层注意力机制的 LL Module，用来验证双层注意力机制对局部特征提取能力的影响。实验 2 在整体网络中去除 GG Module 和 LG Module，仅采用 LL Module，用来验证 LL Module 对图像融合中的影响。实验 3 是仅保留 Swin Transformer 中去除残差连接机制的 LL Module，用来验证残差连接机制对病变区域特征提取的影响。实验 4 在整体网络中去除 LL Module 和 LG Module，仅使用 GG Module，用来验证其对融合图像全局特征保留的影响。实验 5 在整体网络中去除 LG Mode，将 LL Module 生成的局部融合图像和 GG Module 生成的全局融合图像采用像素加权平均的方法进行融合，用来验证 LG Mode 对图像融合中细节保留以及对比度的影响。实验 6 是本文融合方法。具体如表 2 所示。表中√表示实验中添加此创新模块，×表示没有添加此创新模块。

图 8 展示了消融实验与本文方法在五组 PET 和 CT 医学图像融合后的定性对比效果，表 2 为六组图像在不同客观评价指标上的平均值，实验 1 的融合效果病灶明显且对比度高，但整体对比度较差，轮廓信息不明显，丢失了一定的边缘信息。实验 2 相比实验 1，AG 增加了 15%，SF 提升了 13%，融合得到的图像整体对比度较好，证明两层注意力网络能够有效保留细节信息和边缘信息。实验 3 的融合图像相比较实验 2，病灶信息更加清晰，但是忽略了边缘轮廓信息。评价指标 SD 增加了 10%，整体对比度明显提高，但从 AG，EI，SF 和 SD 指标值来看，实验 4 与实验 3 相比，评价指标 AG 提高了 21%，IE 提高了 16%，说明 R-Swin Transformer 有效保留了源图像细

节内容。实验5在视觉上突出了病灶区域信息，且边缘信息丰富，但在评价指标上略低。本文方法相比实验2在评价指标AG上增加了17%、EI提上了11%、IE提升了10%、SD增加了14%、相比实验5在$Q^{AB/F}$提升了4%，说明LG Module能够充分保留图像的边缘和纹理信息，保留丰富细节信息，并对图像的降噪以及区分背景和病灶区域相似度发挥了良好的作用。

表2　消融实验客观评价指标均值

Tab. 2　Mean value of objective evaluation metrics for ablation experiments

No.	LL Module		GG Module		LG Module		AG	EI	$Q^{AB/F}$	SF	SD	IE
	Number of attention layers		Residual connection		No (Average weighted)	Yes						
	Single	Double	No	Yes								
1	√	×	×	×	×	×	8.802 4	63.593 0	0.465 2	26.055 5	33.141 3	5.130 1
2	×	√	×	×	×	×	10.149 3	77.014 6	0.487 6	29.444 7	39.209 3	5.584 6
3	×	×	√	×	×	×	6.490 7	47.797 3	0.387 9	22.891 9	36.838 8	4.394 8
4	×	×	×	√	×	×	7.870 1	52.358 2	0.403 6	24.558 9	36.227 8	5.133 2
5	×	√	×	√	√	×	8.947 9	62.319 3	0.491 8	26.224 1	36.537 3	5.836 4
6	×	√	×	√	×	√	**11.906 6**	**85.917 3**	**0.516 0**	**32.293 6**	**44.694 9**	**6.232 5**

Bold font is the best result for each column

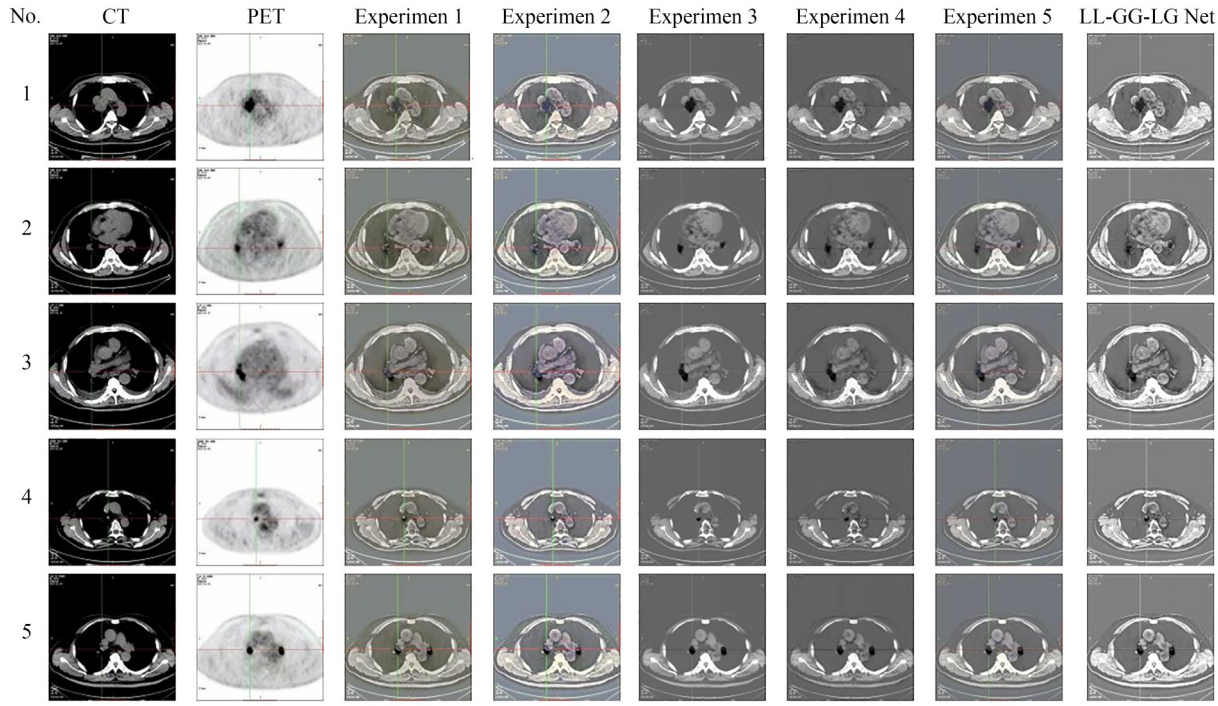

图8　消融实验定性比较

Fig. 8　Qualitative comparison of ablation experiments

表2、图8和图9，更能坚信本文所提出的局部-局部融合模块(LL Module)，全局-全局融合模块(GG Module)和局部-全局融合模块(LG Module)相结合方法有效的结合了全局特征和局

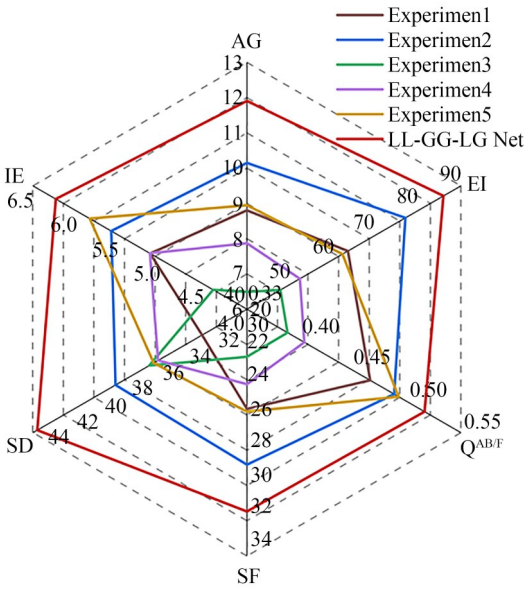

图 9　消融实验雷达图

Fig. 9　Radar map of ablation experiments

部特征，从各项数据的结果中更加体现了对源图像分别提取局部特征和全局特征并进行局部-全局信息交互融合的方法在多模态医学图像融合的优势。

4　结　论

针对多模态医学图像融合捕获全局特征能力有限，忽略了全局和局部特征关联性，周围组织与病变区域之间的模式复杂性和强度相似性的问题，本文提出面向 PET 和 CT 医学图像融合的 LL-GG-LG Net 模型。首先为了有效保留边缘和纹理等特征，局部-局部融合模块进行局部特征的提取融合。此外，设计了 R-Swin Transformer 模块保留病灶部位复杂信息。最后，采用局部-全局融合模块聚合全局特征和局部特征，有效保留纹理边缘等全局信息与局部病变区域。使用临床数据集对本文提出的方法进行验证，实验结果表明 LL-GG-LG Net 在 AG，EI，$Q^{AB/F}$，SF，SD，IE 6 种评价指标上分别平均提高了 21.5%，11%，4%，13%，9%，3%。7 组对比实验表明本文所提出的模型能够计算图像全局关系的同时关注病变区域局部特征，二者互为补充相互融合，使得融合图像能够突出病变区域信息，结构清晰且纹理细节丰富，为医生的辅助诊断，提高术前准备工作效率提供了有效帮助。

参 考 文 献：

[1] 周涛，霍兵强，陆惠玲，等. 融合多尺度图像的密集神经网络肺部肿瘤识别算法[J]. 光学 精密工程，2021，29(7)：1695-1708.
ZHOU T，HUO B Q，LU H L，*et al*. Lung tumor image recognition algorithm with densenet fusion multi-scale images[J]. *Opt. Precision Eng.*，2021，29(7)：1695-1708.(in Chinese)

[2] AZAM MA，KHAN KB，SALAHUDDIN S，*et al*. A review on multimodal medical image fusion：compendious analysis of medical modalities，multimodal databases，fusion techniques and quality metrics [J]. *Computers in Biology and Medicine*，2022，144：105253.

[3] LI Y，ZHAO J，LV Z，*et al*. Medical image fusion method by deep learning[J]. *International Journal of Cognitive Computing in Engineering*，2021，2：21-29.

[4] POLINATI S，DHULI R. Multimodal medical image fusion using empirical wavelet decomposition and local energy maxima[J]. *Optik*，2020，205：163947.

[5] HILL P，AL-MUALLA M E，BULL D. Perceptual image fusion using wavelets[J]. *IEEE Transactions on Image Processing*，2016，26(3)：1076-1088.

[6] LI S T，YIN H T，FANG L Y. Group-sparse representation with dictionary learning for medical image denoising and fusion[J]. *IEEE Transactions on Biomedical Engineering*，2012，59(12)：3450-3459.

[7] 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.

[8] SUN J，ZHU H，XU Z，*et al*. Poisson image fusion based on Markov random field fusion model [J]. *Information Fusion*，2013，14(3)：241-254.

[9] LI H，WU X. Infrared and Visible Image Fusion Using Latent Low-Rank Representation[EB/OL]. 2018：*arXiv*：1804.08992. https://arxiv.org/abs/1804.08992.pdf

[10] 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.

[11] LIU Y, CHEN X, CHENG J, *et al*. A medical image fusion method based on convolutional neural networks[C]. 2017 20*th International Conference on Information Fusion* (*Fusion*). 10-13, 2017, *Xi'an*, *China*. IEEE, 2017: 1-7.

[12] TANG W, LIU Y, CHENG J, *et al*. Green fluorescent protein and phase contrast image fusion via detail preserving cross network[J]. *IEEE Transactions on Computational Imaging*, 2021, 7: 584-597.

[13] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2019, 28(5): 2614-2623.

[14] SONG X, WU X, LI H, *et al*. Res2NetFuse: a Fusion Method for Infrared and Visible Images [EB/OL]. 2021: *arXiv*: 2112.14540. https://arxiv.org/abs/2112.14540.pdf

[15] JUNG H, KIM Y, JANG H, *et al*. Unsupervised deep image fusion with structure tensor representations [J]. *IEEE Transactions on Image Processing*, 2020, 29: 3845-3858.

[16] XU H, MA J. EMFusion: an unsupervised enhanced medical image fusion network[J]. *Information Fusion*, 2021, 76: 177-186.

[17] ZHOU T, LI Q, LU H, *et al*. GAN review: models and medical image fusion applications[J]. *Information Fusion*, 2023, 91: 134-148.

[18] GUO X P, NIE R C, CAO J D, *et al*. FuseGAN: learning to fuse multi-focus image via conditional generative adversarial network[J]. *IEEE Transactions on Multimedia*, 2019, 21(8): 1982-1996.

[19] MA J Y, XU H, JIANG J J, *et al*. DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4980-4995.

[20] ZHANG H, YUAN J T, TIAN X, *et al*. GAN-FM: infrared and visible image fusion using GAN with full-scale skip connection and dual Markovian discriminators[J]. *IEEE Transactions on Computational Imaging*, 2021, 7: 1134-1147.

[21] FANG P F, ZHOU J M, ROY S K, *et al*. Attention in attention networks for person retrieval[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(9): 4626-4641.

[22] ZHANG Z, LIN Z, XU J, *et al*. Bilateral attention network for RGB-D salient object detection [J]. *IEEE Transactions on Image Processing*, 2021, 30: 1949-1961.

[23] VASWANI A, SHAZEER N, NIKI P, *et al*. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30.

[24] ZHOU T, YE X Y, LU H L, *et al*. Dense convolutional network and its application in medical image analysis[J]. *BioMed Research International*, 2022, 2022: 1-22.

[25] SUGANUMA M, OZAY M, OKATANI T. Exploiting The Potential of Standard Convolutional Autoencoders for Image Restoration by Evolutionary Search[EB/OL]. 2018: *arXiv*: 1803.00370. https://arxiv.org/abs/1803.00370.pdf

[26] LIU H, SIMONYAN K, YANG Y. DARTS: Differentiable Architecture Search [EB/OL]. 2018: *arXiv*: 1806.09055. https://arxiv.org/abs/1806.09055.pdf

作者简介：

周　涛(1977－),男,宁夏同心人,博士,教授,2010年于西北工业大学获得博士学位,主要从事医学图像分析处理、计算机辅助诊断、深度学习、模式识别等方面的研究。

E-mail: zhoutaonxmu@126.com

通讯作者：

张祥祥(1999－),女,河南驻马店人,硕士研究生,2019年于商丘师范学院获得学士学位,主要从事医学图像处理、计算机辅助诊断、深度学习等方面的研究。

E-mail: zxx19990503@163.com