E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)05-1510-18

论文引用格式:Li F C, Gao S S, Liu Z, Zhang C M and Zhou Y F. 2025. Multimodal medical image fusion with progressive feature extraction and frequency domain information complementation. Journal of Image and Graphics, 30(5):1510-1527(李夫辰, 高珊珊, 刘峥, 张彩明, 周元峰. 2025. 渐进特征提取和频域信息补充的多模态医学图像融合. 中国图象图形学报, 30(5):1510-1527)[DOI:10.11834/jig.240509]

渐进特征提取和频域信息补充的多模态医学图像融合

李夫辰[1],高珊珊[1,2,3*],刘峥[1],张彩明[4,5],周元峰[4]

1. 山东财经大学计算机科学与技术学院,济南　250014; 2. 山东省数字经济轻量智算与可视化重点实验室,济南　250014;
3. 山东省中美数字媒体国际合作研究中心,济南　250014; 4. 山东大学软件学院,济南　250101;
5. 山东省未来智能计算协同创新中心,烟台　264025

摘　要:**目的**　如何充分保留各模态独有特征的细节以及有效整合模态间共有特征是多模态医学图像融合领域亟待突破的共性问题。目前常用的双分支图像编码方法存在对模态间相互依赖、相互关联的共有特征信息交互方式不够完善、过程不够充分的问题,影响了多模态成像间特征相关性的建立。对此,设计了基于渐进特征提取、频域信息补充以及Swin Transformer结合卷积神经网络(convolutional neural network,CNN)重建的多尺度医学图像融合网络。**方法**　首先,设计基于梯度信息引导的多尺度特征提取模块,渐进式提取图像的共有特征以及不同模态的独有特征;然后,结合交叉注意力设计渐进式融合模块,实现不同模态间的空域信息交互增强,以及频域高低频位置信息引导的针对性的多模态信息融合;最后,设计Swin-CNN重建模块,建立医学图像全局和局部区域相应特征之间的联系。**结果**　在3个公开融合数据集进行实验,在MRI-SPECT(magnetic resonance imaging-single photon emission computed tomography)和MRI-PET(magnetic resonance imaging-positron emission tomography)融合任务中,与8种最先进方法相比,视觉融合任务评价指标均取得最优,相比性能第2的模型,互信息(mutual information,MI)分别提高4.42%和17.30%,离散余弦特征互信息(discrete cosine transform feature mutual information,FMI_{dct})分别提高5.17%和11%。在GFP-PC(green fluorescent protein-phase contrast)融合任务中,取得6项最优和2项次优,相比性能第2的模型,MI和VIF(visual information fidelity)分别提高16.43%和16.87%。**结论**　提出的融合模型通过三分支架构的共同和独有特征提取能力,充分挖掘不同模态图像特征并渐进式整合多尺度信息,利用结合交叉注意力的渐进融合模块引导模型针对性地融合高低频特征,在重建过程中同时关注医学图像全局和局部的属性信息,有效提高了多模态医学图像融合的质量。

关键词:多模态医学图像融合;多尺度;渐进式提取融合;频域信息引导;全局—局部重建

Multimodal medical image fusion with progressive feature extraction and frequency domain information complementation

Li Fuchen[1], Gao Shanshan[1,2,3*], Liu Zheng[1], Zhang Caiming[4,5], Zhou Yuanfeng[4]

1. School of Computer Science and Technology, Shandong University of Finance and Economics, Ji'nan　250014, China;
2. Shandong Provincial Key Laboratory of Lightweight Intelligent Computing and Visualization for Digital Economy, Ji'nan　250014, China; 3. Shandong China-U. S. Digital Media International Cooperation Research Center, Ji'nan　250014, China;

收稿日期:2024-09-02;**修回日期**:2024-10-16;**预印本日期**:2024-10-23

*通信作者:高珊珊　gss_sdufe@sdufe.edu.cn

基金项目:国家自然科学基金项目(62172257);济南市科研带头人工作室项目(202228105, 2021GXRC092);教育部人文社会科学研究项目(20YJA870013);山东省高等学校"青创团队计划"团队项目(2022KJ185)

Supported by: National Natural Science Foundation of China(62172257); Humanities and Social Science Research of the Ministry of Education(20YJA870013); Shandong Provincial Colleges and Universities "Youth Innovation Team Plan" Team(2022KJ185)

4. School of Software, Shandong University, Ji'nan 250101, China;
5. Shandong Co-Innovation Center of Future Intelligent Computing, Yantai 264025, China

Abstract: Objective Single-modality medical imaging is often insufficient for providing a comprehensive review of lesion characteristics, including structure, metabolism, and other critical details. Medical images can generally be categorized into anatomical medical imaging and functional medical imaging. Anatomical medical imaging offers rich information on the structure of the body, but it lacks insight into metabolic processes. In contrast, functional medical imaging is the opposite. In clinical applications, doctors use medical imaging from multiple modalities to diagnose diseases, localize lesions, and plan surgeries. However, simultaneously observing multimodal medical images is not intuitive and may not fully capture all the relevant features of the lesion. Therefore, multimodal medical image fusion is commonly employed in practice to integrate and enhance the information from different imaging techniques. How to fully retain the unique features of each modality while effectively integrating the shared features between modalities is a common challenge in medical image fusion. The information interaction of shared modal features in currently used two-branch image coding methods is often underdeveloped, and the process is somewhat inadequate. This condition limits the establishment of feature correlations between multimodal images. A multiscale medical image fusion network is designed to address these issues. This network is based on progressive feature extraction, frequency domain information supplementation, and image reconstruction by Swin Transformer and convolutional neural network (CNN). **Method** First, a multiscale feature extraction module guided by gradient information was designed, which can be integrated into a three-branch feature extraction architecture. The left and right branches are responsible for extracting the unique features from each modality of the medical images, while the middle branch extracts the shared features between modalities. The extraction architecture comprises several multiscale feature extraction modules, each based on gradient information guidance. These submodule can simultaneously integrate features from all scale levels. The extraction architecture fully considers the information interaction between modalities and can progressively extract the common and unique features across different modalities. In addition, this extraction architecture effectively integrates multiscale features from multimodal medical images. A progressive fusion module that integrates cross-attention mechanisms was designed to fully utilize the frequency domain information and guide the fusion process at the modal level. This fusion module enhances the interaction of spatial domain information between different modalities and leverages high- and low-frequency positional information from the frequency domain, guiding the model for more targeted multimodal fusion. Finally, a Swin-CNN reconstruction module was designed to determine the relationship between global and local area features of medical images. The reconstruction module uses Swin Transformer to capture global information, such as the overall structure and shape of the image, while simultaneously employing CNN to extract regional features, such as local texture details. The reconstruction module can effectively improve the quality of fused images by integrating the global and local feature information of medical images simultaneously. **Result** The datasets used for the experiments include the MRI-SPECT and MRI-PET fusion datasets from the whole brain database at Harvard Medical School and the GFP-PC fusion dataset from the John Innes Center, respectively. Considering the visual effect of the fused images, the proposed fusion model effectively preserves the structural and functional features of different medical image modalities and improves the quality of the fused images. The advantages of the fused images generated by this model are as follows: 1) The fused image has richer texture details and sharper features such as edges and contours. These images effectively preserve the information-rich regions of each modal image. 2) The fused image also effectively preserves the visual features in all original medical images, which ensures no bias toward preserving information from only one modality of the medical image. 3) The fused image is rendered effectively, with no artifacts affecting the visual effect. In addition, in terms of comparison of quantitative indicators, the model achieves optimization for all eight image fusion evaluation metrics in MRI-SPECT and MRI-PET fusion tasks. Compared to the model with the second-best performance, the mutual information (MI) and discrete cosine transform feature mutual information (FMI_{dct}) are drastically improved. MI demonstrated an improvement of 4.42% and 17.30%, respectively, and FMI_{dct} showed improvements of 5.17% and 11%, respectively. In the GFP-PC fusion task, six optimal and two sub-optimal results are achieved. Compared to the model with the second-best performance, MI and visual information fidelity (VIF) are substantially improved by 16.43% and 16.87%, respectively. Abla-

tion experiments were also conducted for the network structure and loss function of the model to effectively analyze the experimental results and evaluate the effectiveness of each part of the model in this paper. Experimental results show that all model components and the loss function enhance the image fusion effect. **Conclusion** The proposed fusion model leverages the common and unique features of different medical image modalities and progressively integrates multiscale information using a three-branch architecture. The model also utilizes a progressive fusion module that incorporates cross-attention to fuse high- and low-frequency features in a highly targeted manner. Furthermore, the model focuses on the global and local attribute information of medical images in the reconstruction process, effectively enhancing the quality of multimodal medical image fusion. The proposed model in this paper performs well in three medical image fusion tasks with good generalization capability. This model can provide multimodal medical fusion images with clear contour structures and rich texture details, aiding doctors in clinical diagnosis and improving diagnostic efficiency and accuracy. Future studies will investigate the constraints or effects of downstream medical semantic segmentation and other tasks on image fusion. The network architecture will also be optimized for specific tasks, ensuring a close integration between tasks such as semantic segmentation and image fusion. This research aims to improve the quality of fused images while enhancing the performance of downstream tasks, thereby expanding the application possibilities of multimodal medical image fusion.

Key words: multimodal medical image fusion; multiscale; progressive extraction and fusion; frequency domain information guidance; global-local reconstruction

0 引 言

由于成像系统和传感器硬件的限制，单一模态医学图像特征信息量较少，无法充分显示病灶外观和内部结构。临床诊断中，医生需先观察一种模态的医学成像，在脑中形成记忆后再观察另一种模态的对应成像，这为医生的诊断带来不便。而多模态医学图像融合技术是将不同模态的医学图像进行融合，使融合图像同时包含多种模态信息，从而将患者病灶的结构、代谢等信息整合为一幅图像，为临床检查和诊断提供高质量的病灶成像。目前，多模态医学图像融合在疾病诊断、病灶分割、手术导航等领域得到了广泛的应用。

常见医学成像的种类有功能性成像和解剖性成像。功能性成像包括单光子发射计算机断层扫描(single-photon emission computed tomography，SPECT)、正电子发射断层扫描(positron emission tomography，PET)等，可反映功能、代谢等信息，但难以提供解剖结构和定位信息；解剖性成像包括计算机断层扫描(computed tomography，CT)、磁共振成像(magnetic resonance imaging，MRI)等，可提供清晰的解剖结构信息，但无法提供人体功能、代谢等信息。结构、纹理等形态特征通常由图像梯度信息所反映，而葡萄糖、氧气等人体代谢信息在功能性成像中表现为大片不同像素强度的颜色块和阴影区域。将两者融合是医学图像融合研究及临床应用中最为常见的任务，融合后的图像同时包含解剖形态及功能代谢信息，保留来自两种不同模态医学成像的病灶特征，可充分发挥两种图像的优势，弥补各自缺点并提升诊断效果，对病灶早期检测、精确定位和外科手术制定具有重要的临床价值。

如何让融合图像尽可能地保留解剖性成像和功能性成像中的形态和代谢特征是不同模态医学图像在融合过程中面临的重要挑战。合理地设计模型及损失函数以充分利用梯度信息和像素强度信息的约束，使融合图像病灶特征更丰富、颜色对比度更清晰是解决该问题的关键思路之一。

过去几十年来，人们提出许多医学图像融合方法，这些方法大致可分为两类：传统医学图像融合方法和基于深度学习的医学图像融合方法。传统方法主要包括：基于空间域、基于变换域、基于稀疏表示和基于显著性的图像融合方法等。这些方法基于人类对医学成像特征的认知，针对不同模态的医学成像设计出对应的特征提取方法和特征融合策略。然而受限于人类的认知水平对图像的特征信息提取不太充分，且针对多种特征的内在深层规律设计的融合策略准确性不太充足。基于深度学习的方法主要包括基于卷积神经网络、基于生成对抗网络、基于Transformer模型和基于扩散模型的方法等。这些方法具有自主的深度特征提取能力，且简单的端到端框架避免了手工设计融合策略的限制，有效提高了

融合图像的质量。但基于深度学习的方法目前仍存在一些缺陷需要解决：

1）在特征编码部分，通常采用双分支对不同模态医学成像进行编码，但往往这两个分支之间的交互方式相对简单，导致共同特征提取过程中的多模态特征交互稍显不足，一定程度影响着多模态成像间特征相关性的建立。共同特征指模态间共享的特征，不同模态间的共同特征是一致的或相似的，如MRI和SPECT成像都具有组织边缘、轮廓等信息，而独有特征指模态内独特的属性信息，即与另一模态不相关的特征，如MRI中解剖结构的纹理细节和SPECT中的功能代谢信息。准确整合不同模态的独有特征和共同特征可使融合结果充分保留原始医学图像所包含的病灶信息。此外，多尺度特征信息在模态融合过程中发挥了重要作用，而一次性同时整合所有尺度层次的特征信息可有效建立不同尺度下的特征联系。同时，渐进式提取方式通过多个特征提取子模块对图像信息的充分挖掘，可更深入地捕捉图像的多尺度特征。

因此，为充分提取模态间共同特征和模态内独有特征，本文编码器采用3个分支，两侧分支只接受单模态输入，提取模态内的独有特征；中间分支整合不同模态的医学图像特征，学习来自模态间的共同特征信息。此外，设计的渐进式多尺度图像特征编码器可通过渐进式策略充分提取不同模态图像的多尺度信息，所设计的编码器子模块可同时整合所有尺度层次的特征。

2）多数深度学习图像融合方法对图像频域信息的关注程度不太充分。在频域中图像表现为多种幅度和相位的叠加态，不同频率的特征信息代表图像的多种属性。具体来说，高频信息代表图像灰度值快速变化的区域，如边缘、纹理等特征。低频信息的灰度值变化较为平滑，反映图像连续渐变的区域，如器官或组织的内部结构块。直接在空间域进行多模态融合仅是在图像的对应位置进行像素值融合，而通过将图像在空间域的特征转换为频域的表现形式，使图像在空间域中分散的同一频率特征在频域中聚合到一起，让两种模态的高频和低频信息在特征图中的位置相对应再进行融合，利用高低频的位置信息引导模型更具针对性地分别对图像的边缘、纹理等高频信息和阴影、团块等低频信息进行融合，可提高融合图像的质量。此外，梯度信息的提取和融合是保证融合图像充分包含原模态图像边缘和形状信息的关键，针对性地为其设计网络结构并结合梯度损失函数可更好地实现对梯度特征的提取和利用，提高融合图像对原模态结构的保留程度。

为学习图像在频域的表现形式，同时增强其在空间域中的特征交互，本文设计了结合交叉注意力的渐进融合模块，利用交叉注意力机制的特征交互能力增强不同模态的空域特征并在频域进行高低频位置信息引导融合，利用不同频率的图像特征对模型进行深度训练，从而使融合图像充分整合纹理、细节等多种频率信息。此外，空间域和频域的交替增强融合可提高模型融合能力。同时，本文采用梯度信息提取模块为模型进一步补充图像的边缘特征。

3）现存方法大多都采用卷积神经网络（convolutional neural network，CNN）重建图像，而普通卷积因其感受野受限，从而无法很好地捕获全局特征，即使通过堆叠卷积等方式，整合图像全局特征的能力与Transformer相比仍然有一定差距。

在深度学习中，医学图像的全局特征和局部特征分别指图像的整体或局部属性信息。全局特征一般指模型根据图像在空间域所有像素值计算而来的特征，而局部特征则指模型根据图像各子区域在空间域的像素值计算而来的特征。在实际应用场景中，医学图像的全局特征反映了人体器官或组织整体的结构、形状等属性，而局部特征表现了医学图像某些局部区域的纹理细节等信息。临床中医生根据这些特征信息结合医学知识判断人体的健康程度。

为有效利用全局和局部特征信息进行图像重建，本文设计基于Swin Transformer（Swin-T）和CNN的全局—局部特征信息重建模块，利用Swin Transformer的自注意力机制和移位窗口机制自适应地关注图像不同区域，建立图像的全局特征联系，深入挖掘图像的整体性特征信息，并通过跨窗口的交互灵活处理不同位置特征的相互依赖关系。同时结合CNN提取的区域性特征信息以有效建立图像的全局和局部间特征联系，从而提高重建图像质量。

综上，本文设计了基于渐进特征提取、频域信息补充以及Swin Transformer结合CNN重建的多尺度医学图像融合网络（progressive feature extraction and frequency domain information complementation network，PEFI），该模型可以有效提取图像的多尺度特

征，并利用频域高低频位置信息引导增强特征融合效果，同时提高融合图像的重建质量。

1 相关工作

本节对基于传统和基于深度学习的多模态医学图像融合进行简要回顾，并介绍本文模型中采用的Vision Transformer和傅里叶变换方法。

1.1 基于传统方法的多模态医学图像融合

传统多模态医学图像融合方法为充分保留图像的原始信息并提高计算效率，一般采用像素级或特征级图像融合策略，通常分为基于空间域、基于变换域和基于稀疏表示的方法（贾紫婷，2020）。

基于空间域的融合方法直接对图像像素值进行处理，即对经过预处理后的医学图像的像素在空间域进行各种运算，如：平均、加权平均、主成分分析等。此外，研究者还利用各种滤波器提取医学图像边缘、结构等信息并进行融合（Chen等，2021）。

基于变换域的融合方法通过数学变换将图像分解为不同频率特征，通过手工设计融合方案进行图像融合，最后采用逆变换操作得到最终融合图像。特征金字塔分解、小波变换（许蓉 等，2021）和多尺度几何变换（Faragallah 等，2022）是其中常用的方法。

基于稀疏表示的融合方法通过构造过完备的冗余字典将图像表示为基向量的稀疏线性组合，从而更好地保留图像信号特征（王兆滨 等，2022）。

1.2 基于深度学习的多模态医学图像融合

随着深度学习技术的飞速发展，在图像融合领域占据了主要地位。目前基于深度学习的融合方法主要分为基于卷积神经网络、基于生成对抗网络、基于Transformer模型和基于去噪扩散模型的方法等（黄渝萍和李伟生，2023）。

基于卷积神经网络的方法利用CNN提取图像局部特征并融合。Zhang等人（2020b）提出端到端全卷积图像融合模型，从原图像中提取突出特征进行融合重构，提升融合图像视觉效果。Xu等人（2022）提出一种基于卷积神经网络的统一无监督融合框架，将不同融合任务统一到同一框架，解决包括多模态、多曝光和多焦点等图像融合问题。

基于生成对抗网络的方法利用生成器和判别器的对抗性学习使融合图像与原图像分布趋于一致（王丽芳 等，2022）。Ma等人（2020）将生成对抗网络引入图像融合领域，利用双生成器和判别器融合不同分辨率的多模态医学图像，避免传统方法中显示模糊和可见光纹理丢失等问题。林予松等人（2025）结合多尺度空间注意力并利用双鉴别器进行训练，全面融合不同模态的细节特征。

基于Transformer的方法利用Transformer的全局特征提取能力，弥补传统卷积感受野受限的问题。Zhao等人（2023a）利用多种Transformer对图像的全局特征进行处理，有效保留每种模态的互补特征。Liu等人（2024）提出一种跨窗口和维度的特征混合变换器（mixing features across windows and dimensions，MixFormer），在多个不同尺度对不同模态的医学图像特征进行融合，有效提高了融合质量和泛化能力。

基于去噪扩散模型的方法利用前向加噪和逆向去噪过程学习图像特征的深层次关联。Zhao等人（2023b）提出一种基于去噪扩散概率模型的新型融合算法，将图像融合任务划分为无条件生成子问题和最大似然子问题，利用自然图像先验补充原图像的特征信息，提高了融合图像的生成质量。

此外，U-Net及其改进方法如deep supervision（Zhou等，2018）在多尺度特征提取方面给予了有价值的启发，deep supervision通过使用稠密连接将网络中的各层级互相联系起来，并对模型的不同尺度进行多重监督，通过多项损失函数的联合训练使模型学习到图像的多尺度特征信息。

1.3 Vision Transformer

Transformer最先在自然语言处理领域提出（Vaswani等，2017），利用自注意力机制捕捉序列中的长距离依赖关系，有效提高了模型对全局上下文信息的感知能力。受其启发，研究人员对Transformer进行改进并引入计算机视觉领域。Dosovitskiy等人（2021）设计ViT（vision Transformer），大幅提升图像分类任务的性能。Liu等人（2021）提出Swin Transformer，通过移位窗口机制将注意力计算限制在局部窗口并建立不同窗口之间的连接，从而减少了计算开销。为了重建出医学图像的全局特征信息，本文模型解码器中使用了Swin Transformer。

1.4 傅里叶变换

图像频率是表示图像灰度值变换剧烈程度的指标，利用不同频率分布的信息可对图像进行更直观

的分析。傅里叶变换可将图像从空间域转换到频域，是图像处理领域中的关键算法之一。Wen 等人(2024)利用傅里叶变换提取频域特征用于对医学图像的融合，在保留原图像特征的基础上进一步提取图像关键语义信息，有效提高了图像融合质量。为充分挖掘图像不同频率特征之间的联系，本文在融合模块采用傅里叶变换将医学图像从空间域变换到频域。

2 方 法

本节详细介绍所提方法的网络框架及其具体的结构，图 1 为融合网络的整体架构。

2.1 总体框架

由图 1 可以看出，融合网络由编码器、特征融合模块和解码器 3 部分组成。首先，设计的渐进式三分支多尺度图像特征编码器，可有效提取不同模态医学图像的多尺度特征信息；其次，结合交叉注意力设计的渐进融合模块，可充分利用多模态医学图像在频域的特征位置信息引导模态间的高低频特征更具针对性的融合；最后，设计的基于 Swin-CNN 的图像重建解码器，利用 Swin Transformer 的全局特征建模能力和 CNN 的局部特征提取能力实现图像全局范围内的特征信息重建。

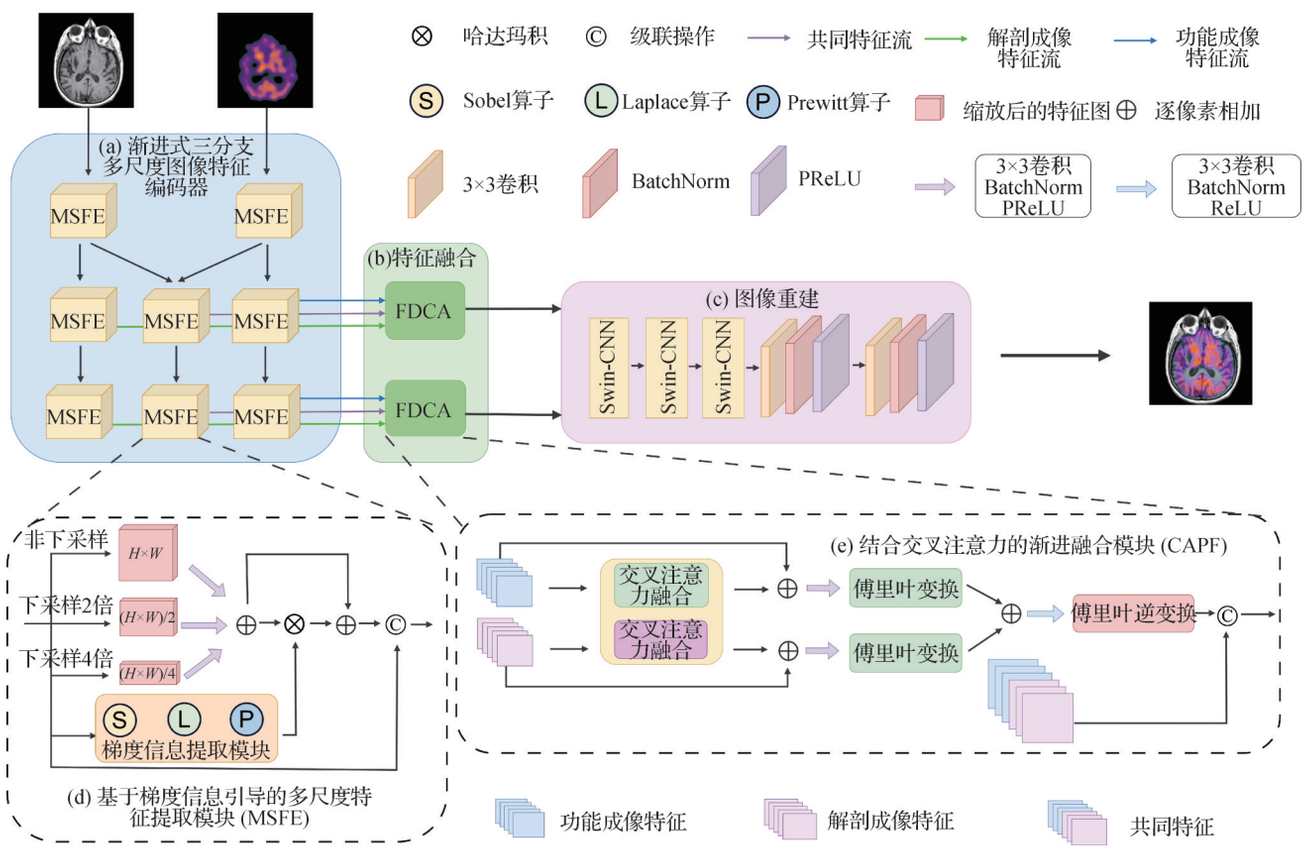


图 1 渐进特征提取和频域信息补充网络框架图

Fig. 1 Progressive feature extraction and frequency domain information complementation network (PEFI) frame diagram

具体地，编码器分为 3 个分支：左右两个分支提取不同模态图像的独特信息，中间分支提取它们的共同信息。首先用 2 个基于梯度信息引导的多尺度特征提取器(multi-scale feature extractor guided by gradient information, MSFE)提取 2 幅原图像特征，并将特征图进行级联操作输入到中间分支中，利用 MSFE 重复渐进式地提取原图像的共同特征，同时左右两个分支继续渐进式地提取独有特征。

将编码器每一层提取的模态间共同特征和 2 种不同模态的独有特征输入结合交叉注意力的渐进融合模块(progressive fusion module with cross- attention, CAPF)中进行模态间融合。融合后的特征进行级联操作后输入解码器，利用基于 Swin-CNN 的全局—局部特征信息重建模块进行图像重建，最后输入到

2个卷积块中进行通道压缩，得到最终的融合图像。

此外，为解决功能图像和解剖图像通道不匹配的问题，在图像预处理和后处理阶段，本文采用RGB-YCrCb颜色空间变换，将RGB功能图像颜色空间由RGB转换为YCrCb，取Y通道与解剖图像进行融合。融合后的图像作为Y通道与功能图像的色度分量Cr和Cb进行YCrCb-RGB颜色空间的逆变换，从而得到最终的融合图像。

2.2 渐进式三分支多尺度图像特征编码器

图像融合网络的特征提取部分通常设计2个独有编码器分别提取各模态特征信息，共同特征提取过程中特征交互不太充分。因此，本文编码器设计了3个分支分别提取模态间的共同信息和各模态内的独有信息，如图1(a)所示。编码器运算可表述为

$$\begin{cases} Encoder_{\text{unique}A_i} = MSFE(MSFE_{i-1}(\boldsymbol{I}_A)) \\ Encoder_{\text{unique}B_i} = MSFE(MSFE_{i-1}(\boldsymbol{I}_B)) \\ Encoder_{\text{joint}_i} = \\ \quad MSFE(MSFE_{i-1}(Concatenate(Encoder_{\text{unique}A_1}, \\ \quad Encoder_{\text{unique}B_1}))) \end{cases} \tag{1}$$

式中，$MSFE(\cdot)$为基于梯度信息引导的多尺度特征提取器，$\boldsymbol{I}_A$和$\boldsymbol{I}_B$分别是功能和解剖图像，下标uniqueA_i和uniqueB_i是指第i个$MSFE$模块的输出，$Concatenate(\cdot)$表示通道级联。

梯度信息在图像融合过程中发挥了重要的作用，表示图像灰度值变化快慢程度及方向，可为模型提供各医学图像的边缘特征。同时，高质量的融合图像必须充分保留原图像的多尺度信息。因此，受HRNet(high-resolution network)启发(Wang等，2021)，本文将多尺度特征提取和融合的过程进行简化并作为一个单独的特征提取模块，在特征提取过程中重复使用MSFE模块对图像多尺度特征进行渐进式提取和融合，借助所有深度层次的低分辨率特征来增强高分辨率特征。如图1(d)所示，首先将输入特征图分别进行2倍和4倍下采样，然后将下采样后的特征图与原特征图分别输入到3个3×3卷积块中进行多尺度特征提取，并执行逐元素相加操作。这个过程可表示为

$$\begin{cases} Conv_{\text{out}} = PReLU(BN(Conv_{3\times3}(\boldsymbol{S}))) \\ \boldsymbol{D}_1 = Down(\boldsymbol{Input}, 2) \\ \boldsymbol{D}_2 = Down(\boldsymbol{Input}, 4) \\ \boldsymbol{F}_{\text{conv}} = Conv_{\text{out}}(\boldsymbol{Input}) + \\ \quad Conv_{\text{out}}(\boldsymbol{D}_1) + Conv_{\text{out}}(\boldsymbol{D}_2) \end{cases} \tag{2}$$

式中，$Conv_{\text{out}}$是卷积块运算，$Conv_{3\times3}(\cdot)$是核大小为3的卷积，$BN(\cdot)$和$PReLU(\cdot)$分别表示批归一化和参数化线性激活函数。$Down(\boldsymbol{Input}, N)$表示对输入进行$N$倍下采样。

同时，利用梯度信息提取模块提取输入特征图的梯度信息。具体来说，通过Sobel、Laplace和Prewitt算子分别提取图像的梯度信息，经过通道级联后利用空间注意力机制进行空间位置权重的调整并进行融合。这个过程可表示为

$$\begin{aligned} \boldsymbol{F}_{\text{grad}} = Spa(&Concatenate(Sobel(\boldsymbol{Input}), \\ &Laplace(\boldsymbol{Input}), Prewitt(\boldsymbol{Input}))) \end{aligned} \tag{3}$$

式中，$Spa(\cdot)$表示空间注意力机制，$Sobel(\cdot)$，$Laplace(\cdot)$和$Prewitt(\cdot)$分别表示Sobel、Laplace和Prewitt算子。最后将多尺度特征图与梯度信息进行相乘并相加，并与模块输入级联后得到融合结果。最终结果可表示为

$$\boldsymbol{F}_{\text{out}} = Concatenate((\boldsymbol{F}_{\text{conv}} + (\boldsymbol{F}_{\text{conv}} \times \boldsymbol{F}_{\text{grad}}), \boldsymbol{Input}) \tag{4}$$

2.3 结合交叉注意力的渐进融合模块

交叉注意力机制可对空间域中多模态信息进行相互增强，通过查询、键和值的运算使两种不同模态的医学图像特征序列可以互相关注，促进功能性成像和结构性成像中的结构位置和功能代谢等特征进行交互完善，以进一步丰富图像的特征表示，为重建模块提供图像信息保留程度较高的特征，进一步提高多模态医学融合图像的质量。

为使单模态特征从全局角度补充另一模态的特征，本文在特征融合部分对来自两个不同模态的独有信息在空间域进行交叉注意力融合。交叉注意力模块(即图1(e)中交叉注意力融合部分)如图2所示。

具体来说，给定功能医学图像特征$\boldsymbol{F}_A \in \mathbf{R}^{H\times W}$和解剖医学图像特征$\boldsymbol{F}_B \in \mathbf{R}^{H\times W}$，将特征图由$\mathbf{R}^{H\times W}$缩放到$\mathbf{R}^{\frac{1}{8}H\times\frac{1}{8}W}$，以减少计算量。然后分别将下采样后的功能医学图像特征和解剖医学图像特征通过线性变换投影到矩阵$\boldsymbol{q}_1, \boldsymbol{k}_1, \boldsymbol{v}_1$和$\boldsymbol{q}_2, \boldsymbol{k}_2, \boldsymbol{v}_2$，具体为

$$\begin{cases} \boldsymbol{q}_1 = \boldsymbol{F}_A\boldsymbol{W}^{q_1}, \boldsymbol{k}_1 = \boldsymbol{F}_A\boldsymbol{W}^{k_1}, \boldsymbol{v}_1 = \boldsymbol{F}_A\boldsymbol{W}^{v_1} \\ \boldsymbol{q}_2 = \boldsymbol{F}_B\boldsymbol{W}^{q_2}, \boldsymbol{k}_2 = \boldsymbol{F}_B\boldsymbol{W}^{k_2}, \boldsymbol{v}_2 = \boldsymbol{F}_B\boldsymbol{W}^{v_2} \end{cases} \tag{5}$$

式中，$\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{v}$分别代表查询、键和值。

其次，通过点积运算建立模态间的相关性矩阵。以向特征$\boldsymbol{F}_A$中补充特征$\boldsymbol{F}_B$的信息为例：用特征$\boldsymbol{F}_A$投影得到的$\boldsymbol{q}_1$矩阵和特征$\boldsymbol{F}_B$投影得到的$\boldsymbol{k}_2$矩阵做

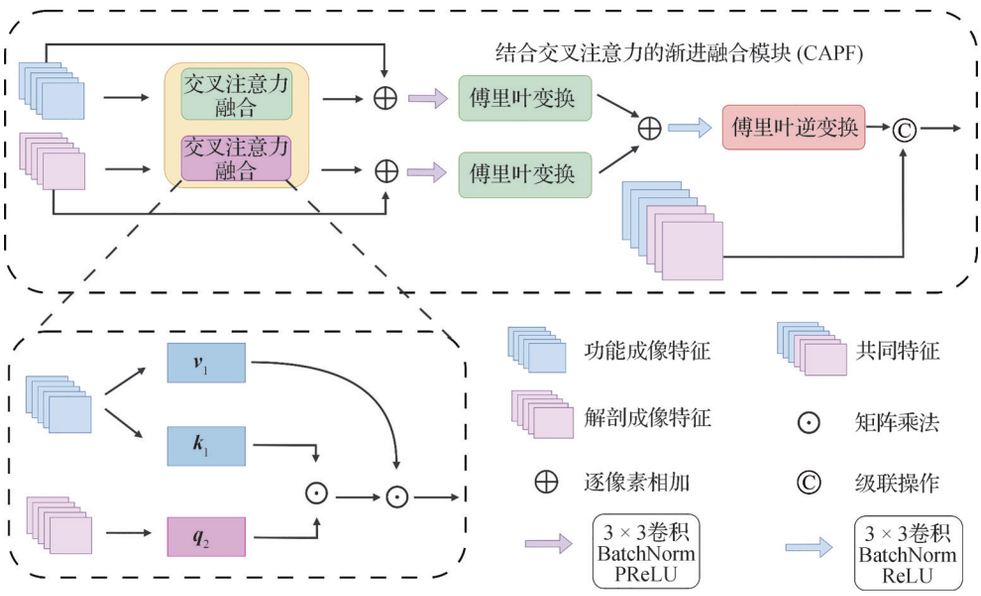


图2　交叉注意力模块

Fig. 2　Cross-attention module

点积运算并使用softmax函数对结果做归一化处理，得到表示功能图像和解剖图像之间的相关性权重。最后用特征$\boldsymbol{F}_B$投影得到的$\boldsymbol{v}_2$矩阵与相关性权重做点积得到最终的结果。d是输入特征维数，则上述过程可以表示为

$$\boldsymbol{CrossAttention}_1 = softmax\left(\frac{\boldsymbol{q}_1\boldsymbol{k}_2}{\sqrt{d}}\right)\boldsymbol{v}_2 \tag{6}$$

向特征$\boldsymbol{F}_B$补充特征$\boldsymbol{F}_A$信息的过程为

$$\boldsymbol{CrossAttention}_2 = softmax\left(\frac{\boldsymbol{q}_2\boldsymbol{k}_1}{\sqrt{d}}\right)\boldsymbol{v}_1 \tag{7}$$

多数深度学习模型在特征融合时通常注重图像特征在空间域的处理，对图像在频域的表征信息关注度不太充足。因此，本文将经过交叉注意力融合后的2个特征图与原特征图相加，并用3 × 3卷积块进一步提取特征，通过傅里叶变换将特征图转换到频域中。在频域对特征图进行逐元素相加融合，并经过3 × 3卷积块建立局部特征间的联系，将融合结果通过傅里叶逆变换到空间域。最后与编码器提取的模态间共同特征进行级联。上述过程可表示为

$$\begin{cases} PBC(\boldsymbol{x}) = PReLU(BN(Conv_{3\times3}(\boldsymbol{x}))) \\ \boldsymbol{f}_1 = Fourier(PBC(\boldsymbol{CrossAttention}_1 + \boldsymbol{F}_A)) \\ \boldsymbol{f}_2 = Fourier(PBC(\boldsymbol{CrossAttention}_2 + \boldsymbol{F}_B)) \\ \boldsymbol{f}_{\text{out}} = \\ Concatenate(Inverse(ReLU(BN(Conv_{3\times3}(\boldsymbol{f}_1 + \boldsymbol{f}_2)))), \boldsymbol{F}_j) \end{cases} \tag{8}$$

式中，$Fourier(\cdot)$表示二维离散傅里叶变换，$Inverse(\cdot)$表示傅里叶逆变换，$\boldsymbol{F}_j$表示编码器提取的模态间共同特征。二维离散傅里叶变换和逆变换的定义为

$$\begin{cases} \boldsymbol{F}(u,v) = \dfrac{1}{HW}\displaystyle\sum_{x=0}^{H-1}\sum_{y=0}^{W-1} f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux/H + vy/W)} \\ f(u,v) = \displaystyle\sum_{u=0}^{H-1}\sum_{v=0}^{W-1} \boldsymbol{F}(u,v)\mathrm{e}^{\mathrm{j}2\pi(ux/H + vy/W)} \end{cases} \tag{9}$$

模型通过傅里叶变换将图像在空间域中位置分散、联系较少的相同频率特征聚合到一起，使高低频信息在两种不同模态的特征图中处于相同位置，利用频域高低频的位置信息引导模态特征融合，从而更具针对性地对人体器官边缘、纹理等高频成分和组织代谢等灰度值变化平缓的低频成分进行融合，丰富模型学习到的知识，进一步提高融合性能。

2.4　基于Swin-CNN的全局—局部特征信息重建模块

Swin Transformer的核心思路是窗口注意力和移位窗口机制，通过将图像划分为多个不重叠的窗口并进行窗口内自注意力运算可有效建立医学图像的局部依赖关系，使医学图像的局部特征充分保留；而利用移位窗口机制进行不同窗口间的特征交互可减少计算量并充分建立各局部窗口特征间的全局联系，从而提高模型对全局医学图像特征序列的关注能力，有效建立不同位置特征间的关联。

受Swin Transformer启发，为改善CNN因感受野受限导致全局依赖建模能力较差的问题，从而充分提取多模态医学图像的全局特征，建立跨窗口的远

程依赖关系，本文将卷积层和Swin Transformer进行并行组合，利用3×3卷积层、批归一化层和PReLU激活函数（parametric rectified linear unit）来提取特征图的局部信息，同时利用Swin Transformer提取特征图的全局信息，然后将局部和全局信息进行级联。该过程可表示为

$$Swin\text{-}CNN_{\text{out}} = Concatenate(Swin(\boldsymbol{Input}), PReLU(BN(Conv_{3\times3}(\boldsymbol{Input})))) \tag{10}$$

将经过3个Swin-CNN块得到的特征图依次输入到2个由卷积、批归一化层和激活函数层组成的模块后输出最终融合图像。基于Swin-CNN的全局—局部特征信息重建模块（即图1（c）中Swin-CNN部分）如图3所示。为进一步减少计算量，本文中的Swin Transformer只使用了原Swin Transformer框架中的1个基础层。

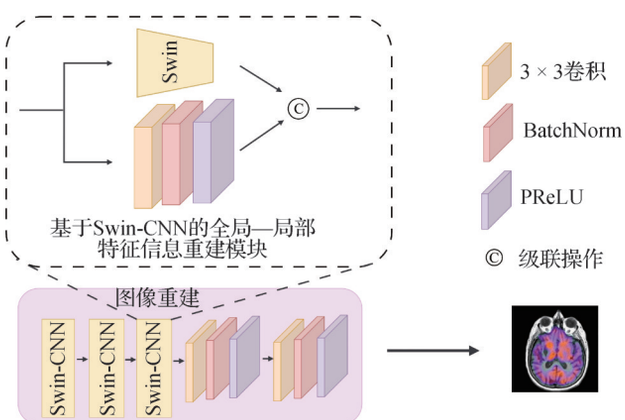


图3　基于Swin-CNN的全局—局部特征信息重建模块

Fig. 3　Global-local feature information reconstruction module based on Swin-CNN

2.5　损失函数

为了充分保留原图像的特征，本文采用了由结构相似度损失、像素级损失、梯度损失和区域互信息损失组成的损失函数。

2.5.1　基于显著性和信息丰富度测量的损失权重

为更好地确定损失权重，使融合结果偏向信息量更丰富、灰度值更显著的原图像，本文采用基于显著性和信息熵测量的损失权重（Xu和Ma，2021），具体为

$$\begin{cases} s_a = \lambda m_A^{\text{saliency}} + m_A^{\text{abundance}} \\ s_b = \lambda m_B^{\text{saliency}} + m_B^{\text{abundance}} \\ \omega_a = \dfrac{e^{s_a}}{e^{s_a} + e^{s_b}} \\ \omega_b = \dfrac{e^{s_b}}{e^{s_a} + e^{s_b}} \end{cases} \tag{11}$$

式中，m^{saliency}和$m^{\text{abundance}}$分别是显著性测量和信息丰富度测量，下标A和B分别代表功能图像和解剖图像，λ是超参数，这里设置为3。显著度测量为

$$\begin{cases} Sign(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases} \\ m_A^{\text{saliency}} = \dfrac{1}{HW}\displaystyle\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{Sign(I_{A_{i,j}} - \tau) + 1}{2} \times I_{A_{i,j}} \\ m_B^{\text{saliency}} = \dfrac{1}{HW}\displaystyle\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{Sign(I_{B_{i,j}} - \tau) + 1}{2} \times I_{B_{i,j}} \end{cases} \tag{12}$$

式中，τ是筛选突出区域的阈值，$I_{A_{i,j}}$和$I_{B_{i,j}}$分别是功能图像和解剖图像在(i,j)处的像素值。

信息丰富度的定义为

$$m_A^{\text{abundance}} = EN(\boldsymbol{I}_A),\ m_B^{\text{abundance}} = EN(\boldsymbol{I}_B) \tag{13}$$

式中，$\boldsymbol{I}_A$和$\boldsymbol{I}_B$分别代表功能图像和解剖图像，$EN(\cdot)$代表信息熵函数。

2.5.2　结构相似度损失

为了使融合图像保留来自不同模态原图像的结构特征，本文引入了结构相似度损失，具体为

$$L_{\text{S}} = \omega_a(1 - SSIM(\boldsymbol{I}_{\text{F}}, \boldsymbol{I}_A)) + \omega_b(1 - SSIM(\boldsymbol{I}_{\text{F}}, \boldsymbol{I}_B)) \tag{14}$$

式中，$\boldsymbol{I}_{\text{F}}$是融合图像，$SSIM(\cdot)$是结构相似性函数。

2.5.3　像素级损失

为了使融合后的图像尽可能保留原图像的视觉显著性区域，本文引入像素强度损失L_{int}来促进模型生成具有适当强度的融合图像，具体为

$$L_{\text{int}} = \frac{1}{HW}\left\| \boldsymbol{I}_{\text{F}} - \max(\boldsymbol{I}_A, \boldsymbol{I}_B) \right\|_1 \tag{15}$$

式中，$\|\cdot\|_1$是L1准则，$\max(\cdot)$表示取元素最大值。

此外，如果融合模型只有像素强度损失，那么模型的融合结果的最优解已经人为固定，即功能图像和解剖图像在各空间位置的像素最大值矩阵。为了让模型能够突破人为限制的最优解，得到更好的结果，本文引入均方误差损失函数L_{MSE}，具体为

$$L_{\text{MSE}} = \omega_a \times \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(I_{\text{F}_{i,j}} - I_{A_{i,j}})^2 + \omega_b \times \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(I_{\text{F}_{i,j}} - I_{B_{i,j}})^2 \tag{16}$$

式中，$I_{\text{F}_{i,j}}$是融合图像在(i,j)处的像素值。

2.5.4　梯度损失

为了充分保留不同模态图像的梯度信息，使融

合图像同时具有两幅原图像的纹理细节，本文引入梯度损失L_{grad}，具体为

$$L_{\text{grad}}=\frac{1}{HW}\left\|\nabla \boldsymbol{I}_{\text{F}}-\max\left(\nabla \boldsymbol{I}_{A},\nabla \boldsymbol{I}_{B}\right)\right\|_{1} \tag{17}$$

式中，∇表示Sobel梯度算子。

2.5.5 区域互信息损失

优秀图像融合模型可充分保留各模态图像子区域信息。因此，本文引入区域互信息损失L_{RMI}(Tang等，2022；Zhao等，2019)，它从区域到区域的角度促使融合结果有效聚合两幅原图像的信息，具体为

$$L_{\text{RMI}}=\omega_{a}\times RMI\left(\boldsymbol{I}_{\text{F}},\boldsymbol{I}_{A}\right)+\omega_{b}\times RMI\left(\boldsymbol{I}_{\text{F}},\boldsymbol{I}_{B}\right) \tag{18}$$

$$RMI\left(\boldsymbol{I}_{\text{F}},\boldsymbol{S}\right)=\eta\mathcal{L}_{\text{ce}}\left(\boldsymbol{I}_{\text{F}},\boldsymbol{S}\right)-(1-\eta)\frac{1}{B}\sum_{b=1}^{B}I_{l}^{b}\left(\boldsymbol{I}_{\text{F}};\boldsymbol{S}\right) \tag{19}$$

式中，$\eta\in[0,1]$，$\mathcal{L}_{\text{ce}}\left(\boldsymbol{I}_{\text{F}},\boldsymbol{S}\right)$是$\boldsymbol{I}_{\text{F}}$和$\boldsymbol{S}$之间的交叉熵损失，$B$是batchsize的大小，$I_{l}\left(\boldsymbol{I}_{\text{F}};\boldsymbol{S}\right)$代表互信息的下限。

综上，最终的目标函数为

$$L=L_{\text{RMI}}+\omega_{1}L_{\text{S}}+\omega_{2}L_{\text{int}}+\omega_{3}L_{\text{MSE}}+\omega_{4}L_{\text{grad}} \tag{20}$$

式中，ω_{1}，ω_{2}，ω_{3}和ω_{4}均是超参数，这里统一设置为$\omega_{1}=7$，$\omega_{2}=2$，$\omega_{3}=10$，$\omega_{4}=10$。

3 实 验

本节介绍数据集、实验细节、对比实验和消融实验，以证明模型的有效性。

3.1 数据集

本文的数据集来自美国哈佛医学院的全脑数据库(http://www.med.harvard.edu/AANLIB/home.html)，本文从中下载473对SPECT和MRI图像，图像的大小均为256×256像素且经过配准。其中，随机分配426对图像为训练集，47对图像为测试集。此外，从中随机下载27对SPECT和PET图像作为模型的泛化能力测试集，并从John Innes中心(http://data.jic.ac.uk/Gfp)随机下载27对绿色荧光蛋白(green fluorescent protein，GFP)和相位衬度(phase contrast，PC)成像作为另一个泛化能力测试集。

3.2 实验细节

在训练期间，输入图像先被放缩到256×256像素大小，并通过随机旋转、翻转和平移操作进行数据扩充。所提出的模型是基于NVIDIA GeForce GTX3090GPU上的PyTorch框架实现的。在训练过程中采用了Adam优化器，学习率设置为0.001，训练20个epoch，batchsize设置为1。

3.3 实验结果及分析

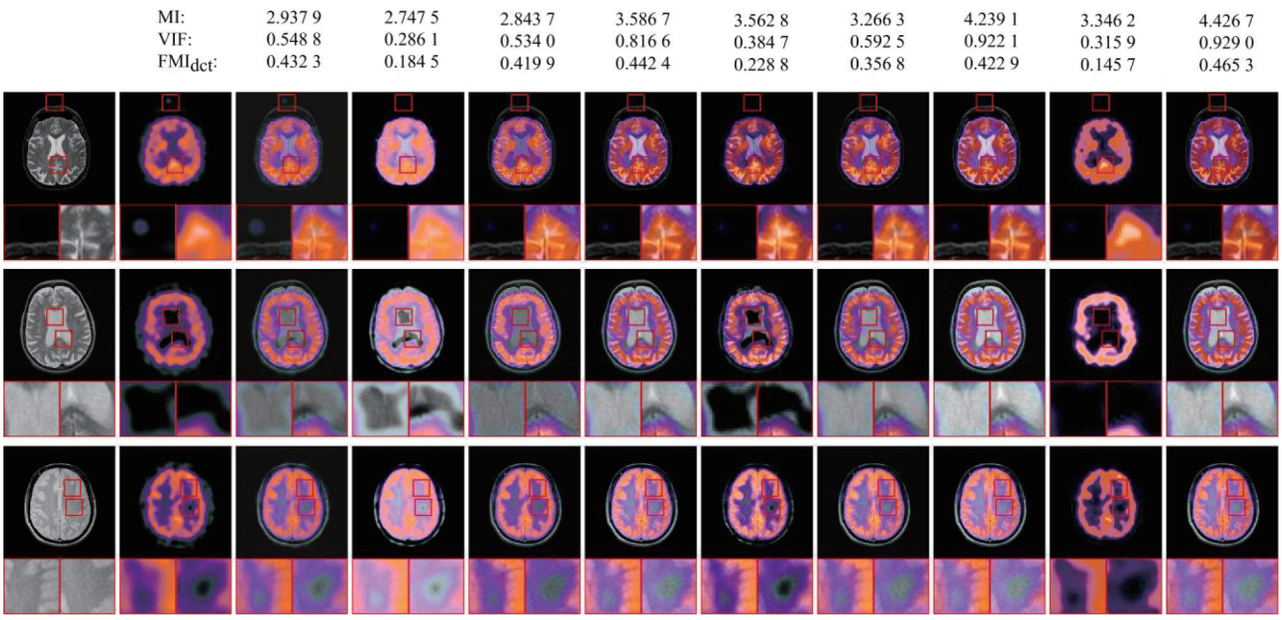
本文从定性和定量的角度对比了8种先进的医学图像融合方法，包括PMGI(proportional maintenance of gradient and intensity)(Zhang等，2020a)、U2Fusion(unified unsupervised fusion network)(Xu等，2022)、SDNet(squeeze and decomposition network)(Zhang和Ma，2021)、SwinFusion(Ma等，2022)、SwinFuse(Wang等，2022)、MATR(multiscale adaptive Transformer network)(Tang等，2022)、CDDFuse(correlation-driven feature decomposition fusion network)(Zhao等，2023a)和MsgFusion(medical semantic guided two-branch fusion network)(Wen等，2024)。其中，PMGI(Zhang等，2020a)、U2Fusion(Xu等，2022)、SDNet(Zhang和Ma，2021)和MsgFusion(Wen等，2024)是基于CNN的方法；SwinFusion(Ma等，2022)、SwinFuse(Wang等，2022)、MATR(Tang等，2022)和CDDFuse(Zhao等，2023a)方法为基于CNN和Transformer的方法。

3.3.1 定性分析

本文选择了8种具有代表性的先进的方法分别在3个数据集进行定性比较，每幅图像放大2个框选区域用于更好地视觉对比。其中图4—图6分别是MRI-SPECT模态融合、MRI-PET模态融合和PC-GFP模态融合。其中，第1、2列是2种不同模态医学图像，其余是各方法的融合结果。在每个方法的正上方增加了3个具有代表性的图像融合评价指标：互信息(mutual information，MI)、视觉信息保真度(visual information fidelity，VIF)和离散余弦特征互信息(discrete cosine transform feature mutual information，FMI_{dct})，以增加可读性。

在图4的MRI和SPECT图像融合任务中，MsgFusion方法倾向于保留SPECT功能图像的代谢信息，而无法提取到MRI图像的组织结构细节。U2Fusion和SwinFuse方法虽然比MsgFusion的效果有所改善，但未充分保留MRI图像的边缘信息。同时，U2Fusion、SwinFuse、PMGI和SDNet方法错误保留了SPECT图像中信息量较少的阴影部分信息(见第2和第3个例子)，且提取的MRI模态细节存在一定的模糊问题。此外，PMGI方法无法识别模态内的无用信息，融合了SPECT模态的噪声区域(见第1个例子)。SwinFusion和MATR方法虽然充分

保留了不同模态特征，但放大来看，它们的融合图像出现了一些棋盘状伪影（见第1个例子）。CDDFuse方法的视觉效果令人满意，但融合图像对比度方面仍存在一定缺陷（见第1个例子）。而本文方法可有效保留不同模态功能及细节信息，提高融合图像质量。



(a) MRI (b) SPECT (c) PMGI (d) U2Fusion (e) SDNet (f) SwinFusion (g) SwinFuse (h) MATR (i) CDDFuse (j) MsgFusion (k) 本文

图4 本文方法与其他8种方法在MRI和SPECT图像上的比较

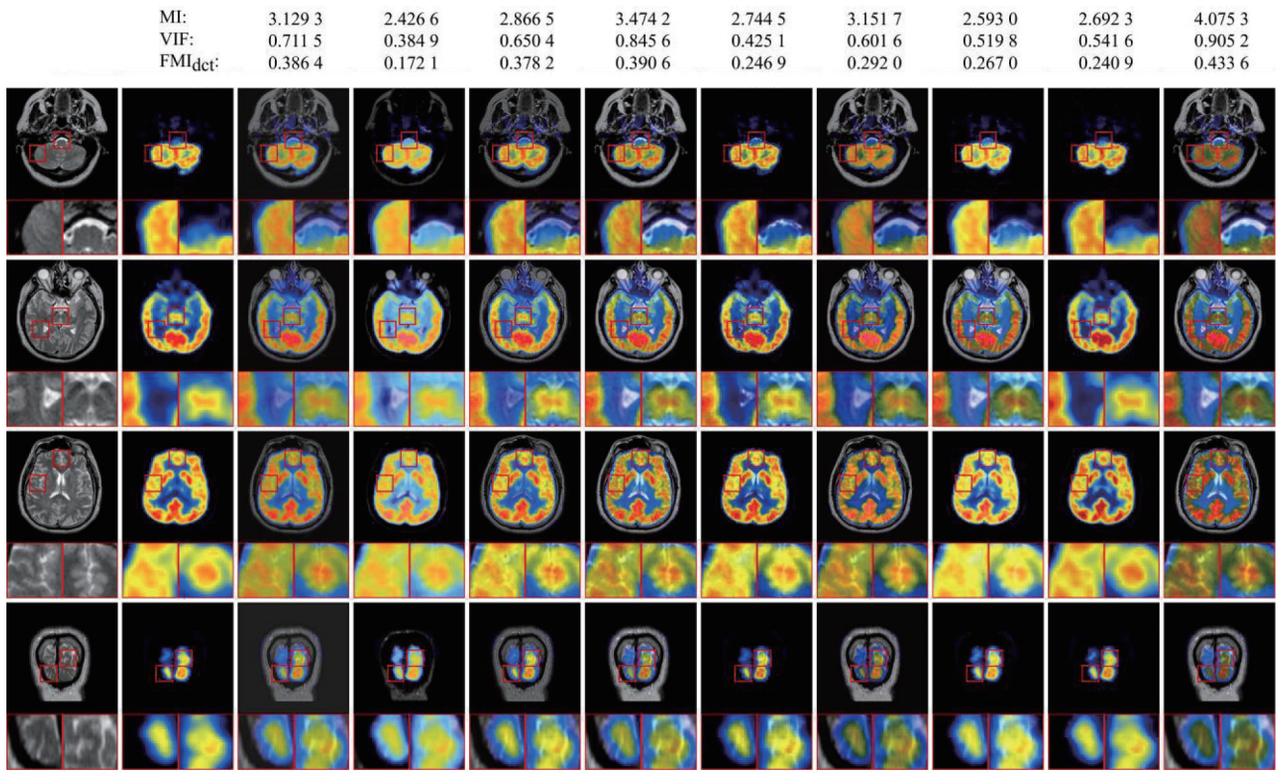
Fig. 4 Comparison of ours with eight other methods on MRI and SPECT images ((a) MRI; (b) SPECT; (c) PMGI; (d) U2Fusion; (e) SDNet; (f) SwinFusion; (g) SwinFuse; (h) MATR; (i) CDDFuse; (j) MsgFusion; (k) ours)

在图5的MRI和PET融合任务中，MsgFusion方法无法充分提取MRI图像的结构细节和边缘、纹理等特征信息，U2Fusion、SwinFuse方法提取的MRI细节虽然较MsgFusion有所提高但是仍然不尽如人意，没有深入融合来自MRI图像的边缘特征（见第2—4个例子）。此外，除本文方法外，其余方法的融合图像出现了明显的棋盘效应（见第4个例子）。CDDFuse方法在MRI和PET图像融合任务中的效果不如在MRI和SPECT图像融合任务中的一样好，偏向于融合SPECT模态，无法充分提取到MRI图像中的结构细节信息（见第1、3、4个例子）。SwinFusion和MATR方法取得了较为良好的视觉效果，但是在融合图像对比度方面存在一定的缺陷（见第2、3个例子）。本文方法的融合图像很好地保留了2种模态的共同信息和独有信息。

在图6的PC和GFP融合任务中，U2Fusion和MsgFusion方法提取的结构特征非常模糊，PMGI、SDNet、SwinFuse和CDDFuse的结构特征提取能力相较于PMGI有了明显的改进，但仍然有部分区域的组织细节未充分显示。此外，U2Fusion、SwinFuse、CDDFuse和MsgFusion方法的融合图像存在亮度较低的问题。SwinFusion和MATR的融合图像质量相对较好，但是仍然存在纹理细节模糊和颜色对比度的问题（见第2、3个例子）。本文的融合方法有效保留了结构和功能信息，且有着良好的颜色对比度。

3.3.2 定量分析

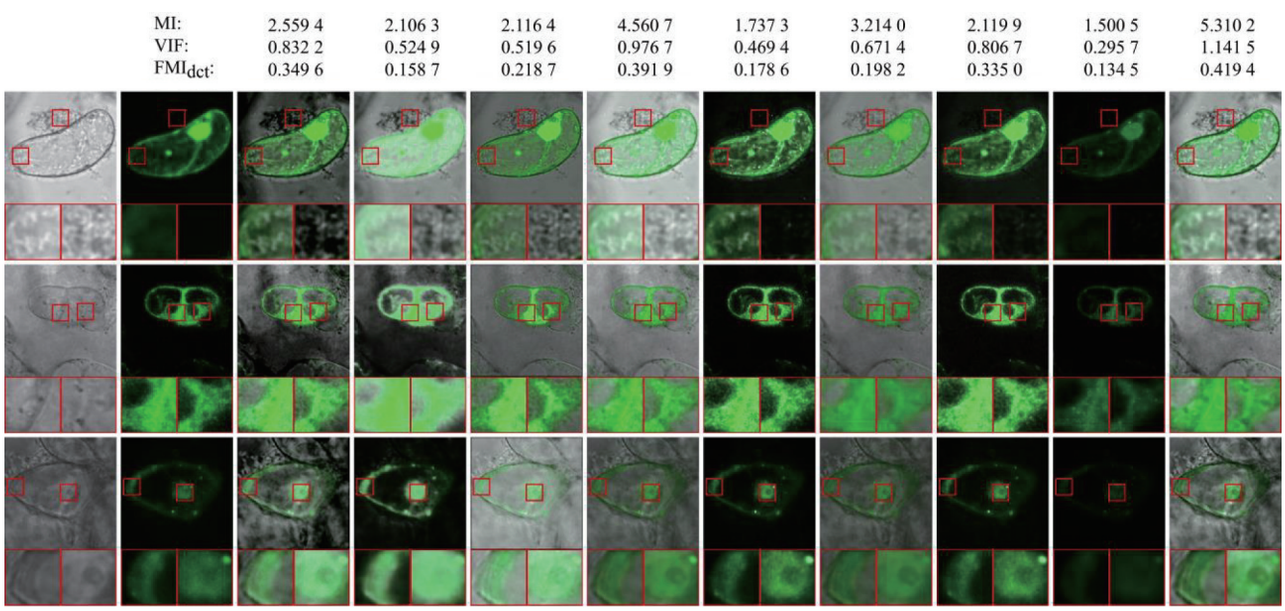
本文采用8个图像融合任务的客观评价指标，除MI（Qu等，2002）、VIF（Han等，2013）和FMI_{dct}（Haghighat和Razian，2014）外，还使用归一化互信息（normalized mutual information，Q_{MI}）（Hossny等，2008）、非线性相关信息熵（nonlinear correlation information entropy，Q_{NCIE}）（Wang等，2005）、基于梯度的度量（gradient-based metric，Q_G）（Xydeas和Petrović，2000）、像素特征互信息（pixel feature mutual information，FMI_{pixel}）（Peng等，2005）和小波特征互信息（wavelet feature mutual information，FMI_w）（Haghighat和Razian，2014）进行定量评价。表1—表3分别展示了MRI-SPECT、MRI-PET和GFP-PC图像融合结果，所有指标均分数越高越好。在MRI-SPECT和MRI-PET融合中本文方法获得最



(a) MRI　(b) PET　(c) PMGI　(d) U2Fusion　(e) SDNet　(f) SwinFusion　(g) SwinFuse　(h) MATR　(i) CDDFuse　(j) MsgFusion　(k) 本文

图5　本文方法与其他8种方法在MRI和PET图像上的比较

Fig. 5　Comparison of ours with eight other methods on MRI and PET images ((a) MRI; (b) PET; (c) PMGI; (d) U2Fusion; (e) SDNet; (f) SwinFusion; (g) SwinFuse; (h) MATR; (i) CDDFuse; (j) MsgFusion; (k) ours)



(a) PC　(b) GFP　(c) PMGI　(d) U2Fusion　(e) SDNet　(f) SwinFusion　(g) SwinFuse　(h) MATR　(i) CDDFuse　(j) MsgFusion　(k) 本文

图6　本文方法与其他8种方法在PC和GFP图像上的比较

Fig. 6　Comparison of ours with eight other methods on PC and GFP images ((a) PC; (b) GFP; (c) PMGI; (d) U2Fusion; (e) SDNet; (f) SwinFusion; (g) SwinFuse; (h) MATR; (i) CDDFuse; (j) MsgFusion; (k) ours)

表1　不同融合方法在MRI-SPECT融合任务上的定量比较

Table 1　Quantitative comparison of different fusion methods on the MRI-SPECT fusion task

方法	MI	VIF	Q_G	Q_{NCIE}	Q_{MI}	FMI_{pixel}	FMI_{dct}	FMI_w
PMGI (2020)	2.937 9	0.548 8	0.409 1	0.806 5	0.705 2	0.870 6	0.432 3	0.334 9
SDNet (2021)	2.843 7	0.534 0	0.595 2	0.806 3	0.695 5	0.860 8	0.419 9	0.337 7
U2Fusion (2022)	2.747 5	0.286 1	0.219 2	0.805 9	0.780 1	0.849 8	0.184 5	0.279 1
SwinFusion (2022)	3.586 7	0.816 6	0.734 1	0.809 2	0.919 7	0.881 7	0.442 4	0.468 3
SwinFuse (2022)	3.562 8	0.384 7	0.342 6	0.808 4	0.972 1	0.830 4	0.228 8	0.310 4
MATR (2022)	3.266 3	0.592 5	0.654 3	0.807 7	0.817 1	0.866 4	0.356 8	0.278 2
CDDFuse (2023)	4.239 1	0.922 1	0.750 4	0.812 4	1.095 1	0.881 7	0.422 9	0.450 0
MsgFusion (2024)	3.346 2	0.315 9	0.093 9	0.808 8	0.985 2	0.844 1	0.145 7	0.277 8
本文	**4.426 7**	**0.929 0**	**0.752 3**	**0.813 6**	**1.171 9**	**0.882 4**	**0.465 3**	**0.512 7**

注：加粗、下划线字体表示各列最优、次优结果。

表2　不同融合方法在MRI-PET融合任务上的定量比较

Table 2　Quantitative comparison of different fusion methods on the MRI-PET fusion task

方法	MI	VIF	Q_G	Q_{NCIE}	Q_{MI}	FMI_{pixel}	FMI_{dct}	FMI_w
PMGI (2020)	3.129 3	0.711 5	0.438 7	0.807 4	0.695 6	0.865 5	0.386 4	0.316 2
SDNet (2021)	2.866 5	0.650 4	0.580 3	0.806 4	0.637 3	0.849 4	0.378 2	0.324 2
U2Fusion (2022)	2.426 6	0.384 9	0.141 4	0.805 0	0.721 7	0.855 4	0.172 1	0.256 3
SwinFusion (2022)	3.474 2	0.845 6	0.694 1	0.809 5	0.819 1	0.876 2	0.390 6	0.463 1
SwinFuse (2022)	2.744 5	0.425 1	0.202 8	0.805 9	0.892 0	0.837 1	0.246 9	0.301 8
MATR (2022)	3.151 7	0.601 6	0.632 6	0.808 0	0.723 4	0.848 9	0.292 0	0.257 4
CDDFuse (2023)	2.593 0	0.519 8	0.287 3	0.805 7	0.751 5	0.851 9	0.267 0	0.300 8
MsgFusion (2024)	2.692 3	0.541 6	0.242 1	0.806 6	0.925 1	0.844 3	0.240 9	0.269 2
本文	**4.075 3**	**0.905 2**	**0.715 3**	**0.813 3**	**0.959 3**	**0.882 3**	**0.433 6**	**0.510 1**

注：加粗、下划线字体表示各列最优、次优结果。

表3　不同融合方法在PC-GFP融合任务上的定量比较

Table 3　Quantitative comparison of different fusion methods on the PC-GFP fusion task

方法	MI	VIF	Q_G	Q_{NCIE}	Q_{MI}	FMI_{pixel}	FMI_{dct}	FMI_w
PMGI (2020)	2.559 4	0.832 2	0.492 9	0.806 8	0.390 8	0.850 0	0.349 6	0.358 0
SDNet (2021)	2.116 4	0.519 6	0.444 9	0.805 3	0.351 0	0.841 6	0.218 7	0.224 1
U2Fusion (2022)	2.106 3	0.524 9	0.272 0	0.804 5	0.346 6	0.839 7	0.158 7	0.194 4
SwinFusion (2022)	4.560 7	0.976 7	**0.761 7**	0.821 1	0.713 2	0.882 7	0.391 9	**0.419 1**
SwinFuse (2022)	1.737 3	0.469 4	0.136 2	0.803 9	0.397 1	0.859 7	0.178 6	0.185 8
MATR (2022)	3.214 0	0.671 4	0.561 3	0.810 9	0.521 7	0.859 0	0.198 2	0.226 1
CDDFuse (2023)	2.119 9	0.806 7	0.231 9	0.805 0	0.427 0	0.860 4	0.335 0	0.317 3
MsgFusion (2024)	1.500 5	0.295 7	0.031 0	0.803 8	0.420 4	0.859 8	0.134 5	0.143 2
本文	**5.310 2**	**1.141 5**	0.694 6	**0.828 3**	**0.804 9**	**0.889 0**	**0.419 4**	0.411 5

注：加粗、下划线字体表示各列最优、次优结果。

优结果；在 PC-GFP 融合中获得 6 项最优和 2 项次优。综合来看，本文方法具有明显优势。

3.3.3　复杂度和效率分析

在 MRI-SPECT 融合任务的测试集上对各方法进行参数量和运行时间计算来评价不同方法的运行效率，结果如表 4 所示。本文方法由于使用了 Swin Transformer 和交叉注意力机制以更充分地建立图像整体结构、形状等全局特征同时进行特征增强，所以参数量和运行时间稍差，但在疾病诊断应用中仍然是可行的。

表 4　不同融合方法的参数量及运行时间比较

Table 4　Comparison of the number of parameters and runtime of different fusion methods

方法	参数量/M	运行时间/s
PMGI	0.042 0	0.049 1
U2Fusion	0.659 2	0.057 9
SDNet	0.067 0	0.042 6
SwinFusion	0.973 6	0.367 8
SwinFuse	8.085 1	0.111 9
MATR	**0.013 4**	**0.031 0**
CDDFuse	1.188 2	0.057 7
MsgFusion	2.971 1	0.041 6
本文	1.898 8	0.075 9

注：加粗字体表示各列最优结果。

Swin Transformer 在计算机视觉领域应用广泛，可有效提高全局上下文表示能力，但相较于其他神经网络，Swin Transformer 运行速度稍慢，参数量较多。在医学图像融合领域，Swin Transformer 的窗口注意力机制实现了图像特征的跨窗口交互，可充分保留各模态独有信息，实现模态间的互补。因此，本文引入 Swin Transformer 架构作为图像重建模块。与使用 Swin Transformer 架构的模型 SwinFusion、SwinFuse 相比，本文方法的运行时间最短。更值得关注的是，本文方法在视觉效果和指标上提升显著。

医学图像融合的应用有医疗诊断、病灶定位、治疗规划和手术导航。除术中导航期间需高度实时图像融合技术外，其余临床应用中，相比于实时性而言更需要高质量融合图像，即在可接受的时间要求内，需要融合图像精确保留不同模态图像的病灶信息和解剖结构、位置信息，从而降低误诊可能性。因此，本文方法在临床中具备实际应用的可行性。

3.3.4　对比总结

为更好地分析实验结果，对本文方法和 8 种比较方法进行深入讨论，通过分析各自优缺点并展示不同方法的差异，找出本文方法的优势及不足。

PMGI 利用梯度分支和强度分支充分提取原图像中的相应信息，但对模态间共同特征的提取不太充分。U2Fusion 采用信息丰富度测量的方法使融合图像与包含重要信息的模态更接近，但对模态共同和独有信息的关注稍显不足，对模态间关联特征的挖掘有待加强。SDNet 设计挤压融合和分解一致性损失提高图像融合质量，但因卷积窗口大小的限制，对图像特征的提取不太全面。SwinFusion 和 SwinFuse 虽然利用 Swin Transformer 和 CNN 对不同模态图像的全局和局部特征进行充分挖掘，但未同时结合空间域和频域特征。MATR 和 CDDFuse 改进 CNN 和 Transformer 模块，使其更适应图像融合任务，但对频域特征及模态间和模态内的信息考虑不太充分。MsgFusion 虽然通过结合空间域和频域特征并采用分级融合策略有效保留和增强不同模态医学语义信息，但对模态共同和独有特征重视程度不太充足。

本文方法相较对比方法的优势在于：1）通过渐进式三分支多尺度图像特征编码器充分挖掘了模态间共同特征和模态内独有特征；2）通过结合交叉注意力的渐进融合模块对不同模态医学图像的高低频特征进行更具针对性的引导整合；3）通过基于 Swin-CNN 的全局—局部特征信息重建模块，充分结合 CNN 提取的局部特征信息和 Swin Transformer 提取的全局特征信息。

本文方法不足之处在于模型的参数量相对于 PMGI、U2Fusion 等网络较多，运行时间稍慢。此外，由于解剖性成像和功能性成像是多模态医学图像融合中最重要、应用最广泛的融合任务，因此本文方法仅专注于解剖成像和功能成像融合任务。

3.4　消融实验

3.4.1　网络结构分析

为了评估本文模型各部分的有效性，在 MRI-SPECT 数据集进行一系列消融实验。1）将编码器中基于梯度信息引导的多尺度特征提取模块（MSFE）改为普通卷积块进行训练；2）将编码器中模态间特

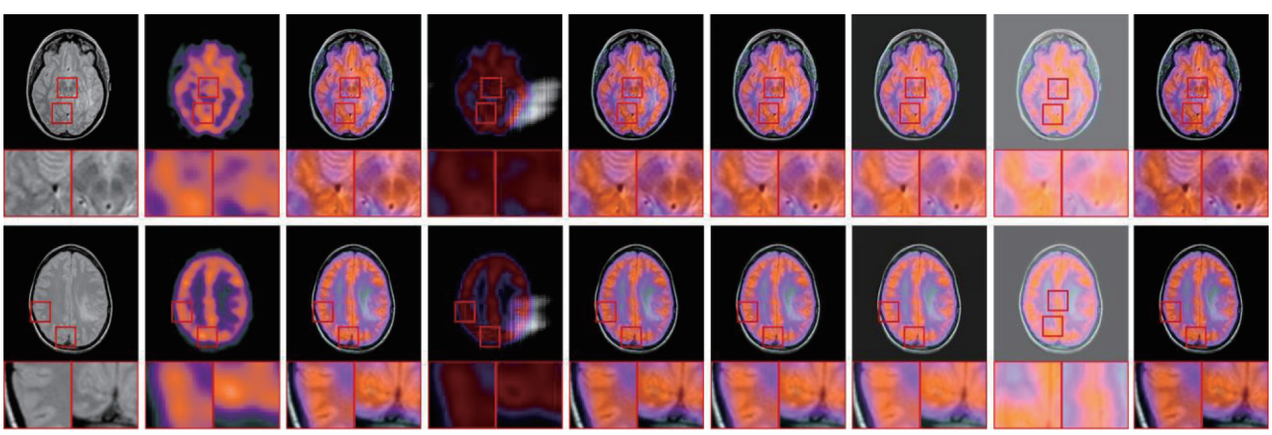
征提取分支(Encoder$_{joint}$)删除进行训练;3)将编码器和解码器框架改为U-Net架构进行训练;4)将结合交叉注意力的渐进融合模块(CAPF)删除进行训练;5)将基于Swin-CNN的全局—局部特征信息重建模块中的Swin Transformer替换为普通卷积块进行训练;6)舍弃频域处理操作。结果如表5和图7所示。结合可视化结果和客观指标比较可以看出,本文设计的各模块都增强了图像融合效果。

表5　MRI和SPECT医学图像融合任务数据集上不同结构的消融研究

Table 5　Ablation study of different structures on MRI and SPECT medical image fusion task datasets

设置	MI	VIF	Q_G	Q_{NCIE}	Q_{MI}	FMI_{pixel}	FMI_{dct}	FMI_w
w/o特征提取模块	3.835 1	0.876 2	0.739 4	0.810 3	0.989 7	0.881 2	0.455 6	0.460 9
w/o模态间特征提取分支	2.991 2	0.082 8	0.034 1	0.807 8	0.879 6	0.800 0	0.105 2	0.273 2
U-Net框架	4.109 8	0.894 8	0.745 0	0.811 8	1.066 8	0.881 7	0.431 1	0.473 2
w/o特征融合模块	4.120 4	0.881 7	0.749 7	0.812 4	0.994 5	0.881 6	**0.468 1**	0.346 1
w/o Swin Transformer	3.645 0	0.794 2	0.721 8	0.810 1	0.886 9	0.867 9	0.426 2	0.325 9
w/o 频域处理操作	3.209 2	0.613 5	0.425 9	0.808 1	0.799 8	0.865 0	0.346 5	0.276 3
本文	**4.426 7**	**0.929 0**	**0.752 3**	**0.813 6**	**1.171 9**	**0.882 4**	0.465 3	**0.512 7**

注:加粗字体表示各列最优结果。



(a) MRI　(b) SPECT　(c) w/o MSFE　(d)w/o Encoder$_{joint}$　(e) U-Net框架　(f) w/o CAPF　(g) w/o Swin-T　(h) w/o频域操作　(i) 本文

图7　MRI和SPECT医学图像融合任务数据集上不同模块组合的可视化结果

Fig. 7　Visualization results of different module combinations on the MRI and SPECT medical image fusion task datasets ((a) MRI; (b) SPECT; (c) w/o MSFE; (d) w/o Encoder$_{joint}$; (e) U-Net framework; (f) w/o CAPF; (g) w/o Swin-T; (h) w/o frequency domain operation; (i) ours)

3.4.2　损失函数分析

本文为平衡损失函数权重,首先将所有损失项平等看待,然后进行大量实验验证不同损失权重值对实验结果的影响,以学习到最优的损失函数权重。

为验证式(19)中损失函数的权重对实验结果的影响,在MRI-SPECT融合任务上进行大量实验,最终确定$\omega_1=7,\omega_2=2,\omega_3=10,\omega_4=10$。由于4个参数组合太多,因此只展示采用控制变量法的结果(即只改变1个参数,固定其余3个参数)。同时,为更好展示不同组合的性能差异,将所有MI指标与所有MI的最小值相减以放缩数据。如图8所示,当权重设置为7,2,10,10时融合性能最好。

此外,进行损失函数各项的消融实验。1)舍弃区域互信息损失L_{RMI};2)舍弃结构相似度损失L_S;3)舍弃像素强度损失L_{int};4)舍弃均方误差损失L_{MSE};5)舍弃梯度损失L_{grad}。结果如表6所示,舍弃L_{RMI}、L_{int}、L_{grad}会使模型融合效果有不同程度的下降。而舍弃L_S和L_{MSE}虽然使FMI_{pixel}和FMI_{dct}微弱提高,但其他指标都有不同程度下降。因此综合来看,使用全部损失项融合效果最好。

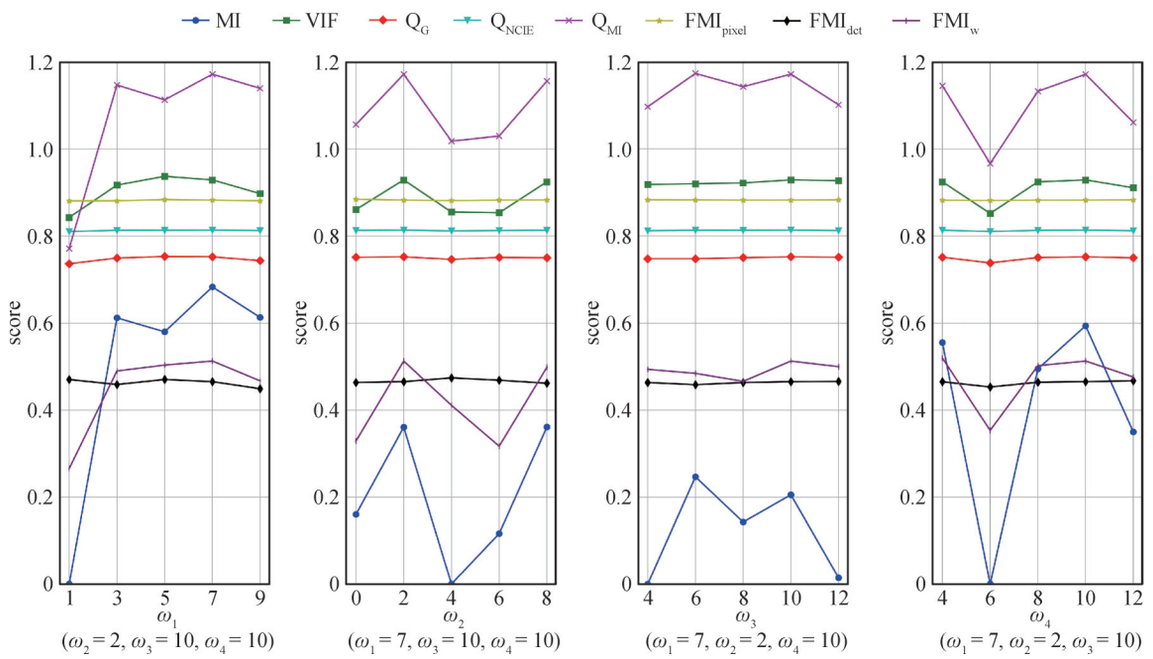


图 8　损失函数权重参数 $\omega_1, \omega_2, \omega_3, \omega_4$ 对 MRI-SPECT 融合任务的影响

Fig. 8　Effect of loss function weight parameters ω_1, ω_2, ω_3, ω_4 on MRI-SPECT fusion task

表 6　MRI 和 SPECT 医学图像融合任务数据集上不同损失函数的消融研究

Table 6　Ablation of different loss functions on MRI and SPECT medical image fusion task datasets

设置	MI	VIF	Q_G	Q_{NCIE}	Q_{MI}	FMI_{pixel}	FMI_{dct}	FMI_w
w/o L_{RMI}	3.088 4	0.674 6	0.704 2	0.807 0	0.788 2	0.880 6	0.417 2	0.458 8
w/o L_S	4.079 3	0.880 3	0.749 4	0.812 1	1.010 2	0.884 3	**0.475 6**	0.355 3
w/o L_{int}	4.226 2	0.860 9	0.751 3	0.812 9	1.056 3	0.884 5	0.463 4	0.328 9
w/o L_{MSE}	4.123 1	0.850 0	0.749 6	0.812 3	1.023 6	**0.885 1**	0.459 8	0.334 9
w/o L_{grad}	4.236 5	0.910 0	0.749 3	0.812 2	1.097 8	0.881 0	0.460 4	0.499 3
本文	**4.426 7**	**0.929 0**	**0.752 3**	**0.813 6**	**1.171 9**	0.882 4	0.465 3	**0.512 7**

注：加粗字体表示各列最优结果。

4　结　论

本文针对多种医学图像融合任务设计了渐进式特征提取、频域信息补充并结合 Swin Transformer 与 CNN 重建的融合框架，通过三分支多尺度编码器渐进式提取模态共同特征和独有特征并利用频域位置信息引导融合，最后通过 Swin-CNN 图像重建模块生成融合图像。在 3 个医学图像融合任务上的结果表明，本文方法融合效果提升显著，在多种医学图像融合任务中表现出良好的泛化能力。本文方法在临床中可为医生提供视觉效果良好、语义特征丰富的医学融合图像，辅助医生进行疾病诊断和手术制定，从而进一步提高诊断精度，优化治疗效果。

现有多模态医学图像融合方法的关注点大多集中于如何获取视觉效果更优质的融合图像，但与进一步的下游任务如医学图像语义分割等的联系不够充分。未来将研究下游任务目标对融合工作的约束或影响，如语义分割更关注如何充分提取和融合不同模态的语义特征。针对具体的任务进一步改进模型架构，从而提高融合图像的质量并提升下游任务的性能，拓宽模型实际应用场景。

参考文献（References）

Chen J Y, Zhang L, Lu L, Li Q L, Hu M F and Yang X M. 2021. A

novel medical image fusion method based on rolling guidance filtering. Internet of Things, 14: #100172 [DOI: 10.1016/j.iot.2020.100172]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16 × 16 words: Transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. Vienna, Austria: ICLR:#11929

Faragallah O S, El-Hoseny H, El-Shafai W, El-Rahman W A, El-Sayed H S, El-Rabaie E S, El-Samie F A, Mahmoud K R and Geweid G G N. 2022. Retracted article: optimized multimodal medical image fusion framework using multi-scale geometric and multi-resolution geometric analysis. Multimedia Tools and Applications, 81(10): 14379-14401 [DOI: 10.1007/s11042-024-20048-7]

Haghighat M and Razian M A. 2014. Fast-FMI: non-reference image fusion metric//Proceedings of the 8th IEEE International Conference on Application of Information and Communication Technologies (AICT). Astana, Kazakhstan: IEEE: 1-3 [DOI: 10.1109/ICAICT.2014.7036000]

Han Y, Cai Y Z, Cao Y and Xu X M. 2013. A new image fusion performance metric based on visual information fidelity. Information Fusion, 14(2): 127-135 [DOI: 10.1016/j.inffus.2011.08.002]

Hossny M, Nahavandi S and Creighton D. 2008. Comments on "information measure for performance of image fusion". Electronics Letters, 44(18): 1066-1067 [DOI: 10.1049/el:20081754]

Huang Y P and Li W S. 2023. A review of medical image fusion methods. Journal of Image and Graphics, 28(1): 118-143 (黄渝萍, 李伟生. 2023. 医学图像融合方法综述. 中国图象图形学报, 28(1): 118-143) [DOI: 10.11834/jig.220603]

Jia Z T. 2020. Overview of key technologies of multimodal medical image fusion. Computer Era, 7: 4-6, 11 (贾紫婷. 2020. 多模态医学图像融合的关键技术研究. 计算机时代, (7): 4-6, 11) [DOI: 10.16644/j.cnki.cn33-1094/tp.2020.07.002]

Lin Y S, Li M Y, Li Y H and Zhao Z. 2025. Multimodal medical image fusion based on GAN and multiscale spatial attention. Journal of Zhengzhou University (Engineering Science), 46(1): 1-8 (林予松, 李孟娅, 李英豪, 赵哲. 2025. 基于GAN和多尺度空间注意力的多模态医学图像融合. 郑州大学学报(工学版), 46(1): 1-8) [DOI: 10.13705/j.issn.1671-6833.2025.01.001]

Liu Y, Yu C, Cheng J, Jane Wang Z and Chen X. 2024. MM-Net: a mixformer-based multi-scale network for anatomical and functional image fusion. IEEE Transactions on Image Processing, 33: 2197-2212 [DOI: 10.1109/TIP.2024.3374072]

Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin Transformer: hierarchical vision Transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10012-10022 [DOI: 10.1109/ICCV48922.2021.00986]

Ma J Y, Tang L F, Fan F, Huang J, Mei X G and Ma Y. 2022. SwinFusion: cross-domain long-range learning for general image fusion via Swin Transformer. IEEE/CAA Journal of Automatica Sinica, 9(7): 1200-1217 [DOI: 10.1109/JAS.2022.105686]

Ma J Y, Xu H, Jiang J J, Mei X G and Zhang X P. 2020. DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing, 29: 4980-4995 [DOI: 10.1109/TIP.2020.2977573]

Peng H C, Long F H and Ding C. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8): 1226-1238 [DOI: 10.1109/TPAMI.2005.159]

Qu G H, Zhang D L and Yan P F. 2002. Information measure for performance of image fusion. Electronics Letters, 38(7): 313-315 [DOI: 10.1049/el:20020212]

Tang W, He F Z, Liu Y and Duan Y S. 2022. MATR: multimodal medical image fusion via multiscale adaptive Transformer. IEEE Transactions on Image Processing, 31: 5134-5149 [DOI: 10.1109/TIP.2022.3193288]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach, USA: MIT Press: 6000-6010

Wang J D, Sun K, Cheng T H, Jiang B R, Deng C R, Zhao Y, Liu D, Mu Y D, Tan M K, Wang X G, Liu W Y and Xiao B. 2021. Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10): 3349-3364 [DOI: 10.1109/TPAMI.2020.2983686]

Wang L F, Mi J, Qin P L, Lin S Z, Gao Y and Liu Y. 2022. Medical image fusion using improved U-Net3+ and cross-modal attention blocks. Journal of Image and Graphics, 27(12): 3622-3636 (王丽芳, 米嘉, 秦品乐, 蔺素珍, 高媛, 刘阳. 2022. 改进U-Net3+与跨模态注意力块的医学图像融合. 中国图象图形学报, 27(12): 3622-3636) [DOI: 10.11834/jig.211066]

Wang Q, Shen Y and Zhang J Q. 2005. A nonlinear correlation measure for multivariable data set. Physica D: Nonlinear Phenomena, 200(3/4): 287-295 [DOI: 10.1016/j.physd.2004.11.001]

Wang Z B, Ma Y K and Cui Z J. 2022. Medical image fusion based on guided filtering and sparse representation. Journal of University of Electronic Science and Technology of China, 51(2): 264-273 (王兆滨, 马一鲲, 崔子婧. 2022. 基于引导滤波与稀疏表示的医学图像融合. 电子科技大学学报, 51(2): 264-273) [DOI: 10.12178/1001-0548.2021165]

Wang Z S, Chen Y L, Shao W Y, Li H and Zhang L. 2022. SwinFuse: a residual swin Transformer fusion network for infrared and visible images. IEEE Transactions on Instrumentation and Measurement, 71: #5016412 [DOI: 10.1109/TIM.2022.3191664]

Wen J Y, Qin F W, Du J, Fang M E, Wei X H, Chen C L P and Li P.

2024. MsgFusion: medical semantic guided two-branch network for multimodal brain image fusion. IEEE Transactions on Multimedia, 26: 944-957 [DOI: 10.1109/TMM.2023.3273924]

Xu H and Ma J Y. 2021. EMFusion: an unsupervised enhanced medical image fusion network. Information Fusion, 76: 177-186 [DOI: 10.1016/j.inffus.2021.06.001]

Xu H, Ma J Y, Jiang J J, Guo X J and Ling H B. 2022. U2Fusion: a unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1): 502-518 [DOI: 10.1109/TPAMI.2020.3012548]

Xu R, Wang Z, Zong T, Lu R and Yang S S. 2021. Method of multi-focal medical image fusion based on improved wavelet transform. Computer and Digital Engineering, 49(8): 1571-1573, 1630 (许蓉, 王直, 宗涛, 陆蓉, 杨莎莎. 2021. 基于改进小波变换的多聚焦医学图像融合方法. 计算机与数字工程, 49(8): 1571-1573), 1630) [DOI: 10.3969/j.issn.1672-9722.2021.08.013]

Xydeas C S and Petrović V. 2000. Objective image fusion performance measure. Electronics Letters, 36(4): 308-309 [DOI: 10.1049/el: 20000267]

Zhang H and Ma J Y. 2021. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion. International Journal of Computer Vision, 129(10): 2761-2785 [DOI: 10.1007/s11263-021-01501-8]

Zhang H, Xu H, Xiao Y, Guo X J and Ma J Y. 2020a. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI: 12797-12804 [DOI: 10.1609/aaai.v34i07.6975]

Zhang Y, Liu Y, Sun P, Yan H, Zhao X L and Zhang L. 2020b. IFCNN: a general image fusion framework based on convolutional neural network. Information Fusion, 54: 99-118 [DOI: 10.1016/j.inffus.2019.07.011]

Zhao S, Wang Y, Yang Z and Cai D. 2019. Region mutual information loss for semantic segmentation//Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada: MIT Press: 11117-11127

Zhao Z X, Bai H W, Zhang J S, Zhang Y L, Xu S, Lin Z D, Timofte R and Van Gool L. 2023a. CDDFuse: correlation-driven dual-branch feature decomposition for multi-modality image fusion//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 5906-5916 [DOI: 10.1109/CVPR52729.2023.00572]

Zhao Z X, Bai H W, Zhu Y Z, Zhang J S, Xu S, Zhang Y L, Zhang K, Meng D Y, Timofte R and Van Gool L. 2023b. DDFM: denoising diffusion model for multi-modality image fusion//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 8082-8093 [DOI: 10.1109/ICCV51070.2023.00742]

Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada, Spain: Springer: 11045: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

李夫辰，男，硕士研究生，主要研究方向为图像处理和计算机视觉。E-mail: lifuchen163@163.com

高珊珊，通信作者，女，教授，主要研究方向为智能图形图像处理、数据挖掘与可视化、计算几何。

E-mail: gss_sdufe@sdufe.edu.cn

刘峥，男，教授，主要研究方向为图像处理、跨媒体信息检索和大数据分析挖掘。E-mail: liuzheng@sdufe.edu.cn

张彩明，男，教授，主要研究方向为计算机图形学、图像处理与计算机辅助几何设计。E-mail: czhang@sdu.edu.cn

周元峰，男，教授，主要研究方向为智能图形图像处理、计算几何和医工交叉。E-mail: yfzhou@sdu.edu.cn