

文章编号: 1671-7449(2024)06-0686-09

基于改进U-Net和跨模态自蒸馏的医学图像融合

刘势杰, 王丽芳*, 郁晓庆

(中北大学 计算机科学与技术学院, 山西 太原 030051)

摘要: 针对医学图像融合方法中存在融合图像信息完整性丢失和跨模态特征提取不足的问题, 提出了一种基于改进U-Net和跨模态自蒸馏的医学图像融合方法。该方法改进了U-Net的编码部分, 设计了一个双分支编码器, 它结合了CNN和Transformer的优势, 能够更有效地捕捉和保留医学图像的局部特征和全局特征, 解决了信息完整性丢失的问题。采用跨模态自蒸馏技术, 在两幅医学图像的CNN分支之间、Transformer分支之间进行信息传递, 加强不同模态特征之间的交互, 最大程度地获取跨模态特征。在解码阶段, 提出注意力门机制代替U-Net中的跳跃连接, 保证网络能够有效关注关键特征, 进一步增强了融合图像的信息完整性。实验结果表明, 相较于其他方法, 该方法得到的融合图像不仅保留了更完整的纹理细节和边缘信息, 而且有效地解决了跨模态特征提取不足的问题。

关键词: 医学图像融合; U-Net; 跨模态自蒸馏; 跨模态特征; 注意力门

中图分类号: TP391.4

文献标识码: A

doi: 10.3969/j.issn.1671-7449.2024086

引用格式: 刘势杰, 王丽芳, 郁晓庆. 基于改进U-Net和跨模态自蒸馏的医学图像融合[J]. 测试技术学报, 2024, 38(6): 686-694.

LIU Shijie, WANG Lifang, YU Xiaoqing. Medical image fusion based on improved U-Net and cross-modal self-distillation[J]. Journal of Test and Measurement Technology, 2024, 38(6): 686-694.

Medical Image Fusion Based on Improved U-Net and Cross-Modal Self-Distillation

LIU Shijie, WANG Lifang*, YU Xiaoqing

(College of Computer Science and Technology, North University of China, Taiyuan 030051, China)

Abstract: To address the issues of information integrity loss and insufficient cross-modal feature extraction in medical image fusion methods, a novel medical image fusion approach based on an improved U-Net and cross-modal self-distillation is proposed. This method enhances the encoding part of U-Net by designing a dual-branch encoder that combines the strengths of CNN and Transformer, enabling more effective capture and retention of local and global features in medical images, thus solving the problem of information integrity loss. Cross-modal self-distillation technology is employed to facilitate information exchange between the CNN branches and the Transformer branches of two medical images, enhancing the interaction between different modal features and maximizing the acquisition of cross-modal features. In the decoding stage, an Attention Gate mechanism is proposed to replace the skip connections in U-Net, ensuring

收稿日期: 2024-04-01

基金项目: 山西省重点研发计划资助项目(202102010101011)

作者简介: 刘势杰(1998—), 男, 硕士生, 主要从事图像融合研究。E-mail: 654995535@qq.com。

* **通信作者:** 王丽芳(1977—), 女, 教授, 博士, 主要从事计算机视觉、图像重建、智能与网络化控制、人工智能及其在医学影像方面的应用研究。E-mail: 727690392@qq.com。

ing that the network can effectively focus on key features and further enhance the integrity of the fused images. Experimental results show that, compared to other methods, the fusion images obtained by this method not only preserve more complete texture details and edge information but also effectively solve the problem of insufficient cross-modal feature extraction.

Key words: medical image fusion; U-Net; cross-modal self-distillation; cross-modal features; attention gate

0 引言

医学图像是临床诊断的重要依据,适用于临床诊断策略的制订和外科手术的指引等重要领域。伴随现代医学影像技术的持续进步,涌现出诸多不同的成像模态^[1],如 X 射线、磁共振成像(MRI)、计算机断层扫描(CT)等。这些多样化的影像模态可以互补地反映病患信息,为了提升临床诊断的精确性,需要整合不同模态的医学图像信息,将其融合为一幅图像,辅助医生对患者的健康状况有更全面和确切的掌握^[2]。

现有图像融合方法可分为传统融合方法和基于深度学习的融合方法。传统融合方法主要包括基于多尺度变换的融合方法^[3-4]和基于稀疏表示的融合方法^[5-6]。前者能捕捉到源图像在不同尺度上的特征,然而在图像分解和重构的过程中,容易造成源图像细节纹理和结构的丢失,导致融合后的图像细节特征未能得到充分保留;后者尽管能有效避免信息丢失等问题,但在处理尺寸庞大或结构复杂的图像时,常会遇到融合图像出现伪影、扭曲、失真等缺陷,在一定程度上限制了该方法的适用性^[7]。

近年来,随着深度学习技术的迅猛发展,研究者们为了突破传统图像融合方法的限制,提出了许多基于深度学习的图像融合方法。Liu 等^[8]利用卷积神经网络(Convolutional Neural Network, CNN)得到的权重图对多源图像的像素活动数据进行组合,实现了基于 CNN 的医学图像融合,但存在特征提取不充分、融合图像信息丢失等问题。Jian 等^[9]提出了 SEDRFuse 网络,该网络通过在编码器中引入残差块,并在编码器与解码器之间采用跳跃连接,有效降低了源图像在浅层纹理细节的丢失,避免了融合图像中的伪影,但由于其对浅层特征关注较多,导致融合图像丧失图像信息完整性。Zhao 等^[10]提出两个模态的低频特征是相关的,而高频特征是不相关的,通过共享编码器在浅层提取低频特征,使用可逆神经网络^[11]防止高频特征的丢失,并将 Transformer 提取全局特征

的能力与 CNN 局部特征提取和计算效率的优势相结合,提升网络的特征提取能力,但此结构仅在浅层考虑共享特征,对跨模态特征提取不足,且 Transformer 的计算量较大,影响整体网络的计算效率。

跨模态知识蒸馏(Cross-Modal Knowledge Distillation, CMKD)将传统的知识蒸馏扩展到多模态学习领域,并在各种应用中取得了成功^[12]。其通过采用一种模态的预训练网络作为教师,向另一个模态的学生网络提供监督信号。Zhou 等^[13]采用知识蒸馏技术,实现了将多模态中的互补信息迁移至单模态检测器上,从而增强其获得多模态知识的能力并提升了网络效果。田加林等^[14]将图像的知识通过知识蒸馏迁移至草图检索网络,并利用跨模态自蒸馏进一步将知识从图像模态迁移到草图模态,增强了图像与草图之间的关联度,有助于消除模态上的差异和语义不一致性。

综上,为了解决多模态医学图像融合中存在融合图像信息完整性丢失和跨模态特征提取不足的问题,本文提出了一种基于改进 U-Net 和跨模态自蒸馏的医学图像融合方法。该方法以 U-Net 为基础网络结构,在编码器中结合了具有 Inception 结构^[15]的卷积块和 Biformer 块^[16]作为特征提取模块,用于捕获局部特征和全局特征,增强特征的丰富度和表征能力。此外,引入跨模态自蒸馏技术,在特征提取过程中促进模态间的知识传递,提高不同模态特征的相互影响,从而实现不同模态之间的特征信息交互。由于 U-Net 中传统的跳跃连接通过直接的特征传递会限制网络的表达能力^[17],本文设计了注意力门 AD 代替跳跃连接后的拼接操作。

1 基于改进 U-Net 和跨模态自蒸馏的医学图像融合

本文提出的多模态医学图像融合方法的整体结构如图 1 所示。网络基于 U-Net 结构对编码器和解码器部分进行了改进。在编码器部分,设计了一个由基于 InceptionNeXt 的 CNN 分支编码器

和基于Biformer的Transformer分支编码器共同组成的CNN-Transformer双分支编码器。在解码器部分,设计了注意力门AD代替U-Net中的跳跃连接,构成新的解码器 $D_k(k=1, 2, 3, 4, 5, 6)$ 。在图像处理过程中,首先将配准好的医学图像 $I_{CT}/I_{MRI} \in \mathbb{R}^{H \times W \times 1}$ ($H, W, 1$ 分别为图像的高度、宽度和通道大小)分别经过相同的双分支编码器,提取出各层的特征 $f_{C,i}^{CT}, f_{B,i}^{CT}, f_{C,i}^{MRI}, f_{B,i}^{MRI}$ ($i=1, 2,$

$3, 4$);接着,利用跨模态自蒸馏在 $f_{C,i}^{CT}$ 和 $f_{C,i}^{MRI}, f_{B,i}^{CT}$ 和 $f_{B,i}^{MRI}$ 之间进行信息传递,以促进模态间的特征交互;再将通过跨模态自蒸馏后的特征进行级联得到融合特征 f_i ($i=1, 2, 3, 4$);然后,将 f_i 输入改进后的解码器中进行特征重建,每个解码器的输出特征表示为 f_{Dk} ($k=1, 2, 3, 4, 5, 6$);最后将特征 f_{D6} 进行上采样得到融合图像 I_f 。

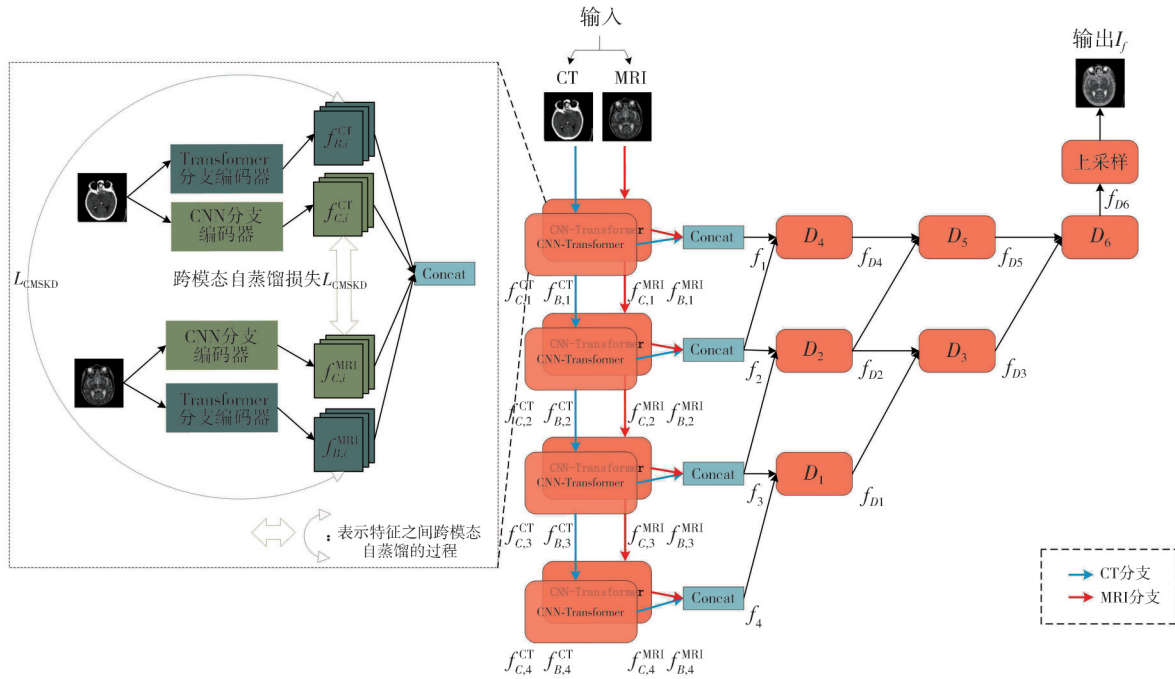


图1 整体网络结构

Fig. 1 Overall network structure

1.1 改进U-Net的编码结构

为了解决源图像特征提取不充分导致融合图像信息完整性丢失的问题,本文对U-Net网络的编码器部分进行了改进,提出了一种双分支编码器,其结构如图2所示。该网络中共包含4层双分支编码器,第1层、第2层和第4层每层由3个InceptionNeXt块和3个BiFormer块组成,第3层则由9个InceptionNeXt块和9个BiFormer块堆叠而成。CT和MRI图像经过每层双分支编码器得到的特征图大小分别为: $H/2 \times W/2 \times C_1, H/4 \times W/4 \times C_2, H/8 \times W/8 \times C_3$ 和 $H/16 \times W/16 \times C_4$,其中 $C_j(j=1, 2, 3, 4)$ 表示每层特征图的通道数。InceptionNeXt块中共包含4个分支,其中,3个分支代表3种不同尺寸的卷积块($3 \times 3, 1 \times 7, 7 \times 1$),并行地对下采样得到的原始特征进行卷积,1个分支表示Identity分支,通过保留原始特征,防止重要

细节特征的丢失。接下来,将4个分支得到的特征进行级联,通过归一化层得到局部特征,最后传递给下一层作为输入。在Transformer分支中,利用Biformer块中的双层路由注意力机制来捕捉特征中的上下文信息,从而增强对复杂图像特征的理解,有助于获得更完整的全局特征表示,通过这种方式使全局特征和局部特征之间达到平衡。

与原U-Net中的编码器相比,该编码器能够同时捕捉到图像的层间和层内多尺度特征,并实现对局部和全局特征的综合表征。

1.2 跨模态自蒸馏

本文在特征提取过程中引入了跨模态自蒸馏(Cross-Modal Self-Knowledge Distillation, CMSKD),蒸馏过程如图1左侧所示。将对CT和MRI图像的特征提取视为两个相互关联的任

务,利用一种模态的特征指导另一模态的特征提取,通过相互的教学关系来强化跨模态特征的学习。以第一层网络为例,对 CNN 分支提取到的特征 $f_{C,1}^{CT}$ 和 $f_{C,1}^{MRI}$,利用跨模态自蒸馏损失 L_{CMSKD} 衡量两种模态之间的差异,在网络反向传播过程中,使用梯度下降算法最小化 L_{CMSKD} 来促进两种模态间的信息交流和知识传递。同时,对 Transformer 分支提取到的特征 $f_{B,1}^{CT}$ 和 $f_{B,1}^{MRI}$ 也执行相同的操作。这种跨模态自蒸馏过程不仅限于网络的第一层,而是被应用于每一层的特征提取,对每一层提取

到的特征 $f_{C,i}^{CT}$ 和 $f_{C,i}^{MRI}$ 、 $f_{B,i}^{CT}$ 和 $f_{B,i}^{MRI}$ 都分别进行同样的操作。通过这种方式,利用不同模态之间的互补性和相关性,提高模型对多模态特征的深度理解,有效解决跨模态特征提取不足的问题。跨模态自蒸馏损失由特征提取损失(Feature Extraction Loss, FEL)和互信息损失(Mutual Information Loss, MIL)组成,如式(1)所示。

$$L_{CMSKD} = L_{FEL} + \lambda_1 L_{MIL}, \quad (1)$$

式中: L_{FEL} 为特征提取损失; L_{MIL} 为互信息损失; λ_1 为平衡两者的权重系数。

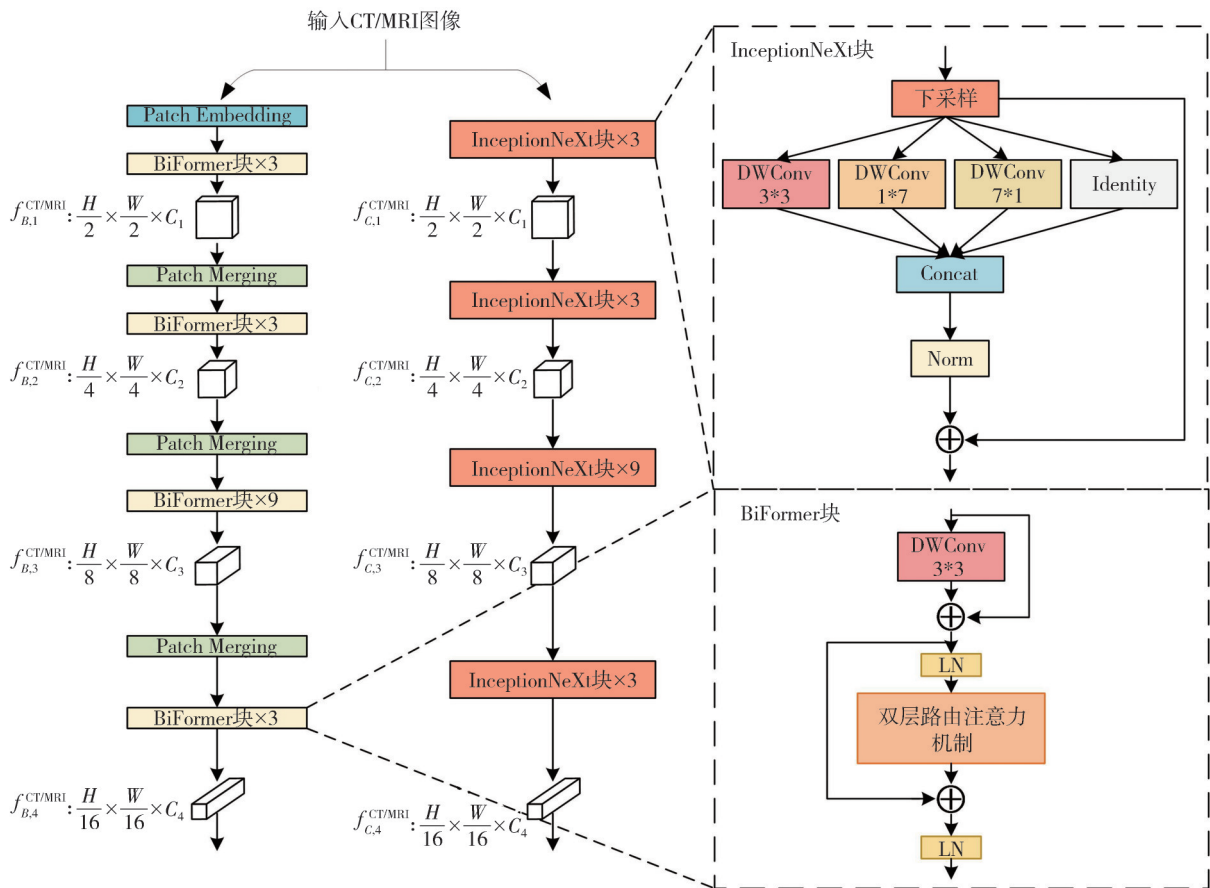


图 2 双分支编码器

Fig. 2 Double branch encode

1.2.1 特征提取损失

特征提取损失是衡量来自不同模态图像特征在特征空间中差异性的指标,本文采用 L2 范数来量化两个模态特征间的差异,进而强化编码器在识别模态间相似性和一致性方面的关注,通过这种方式,模型能够更好地理解和整合来自不同模态的信息,以提升整体的学习效果和性能。FEL 的定义如下

$$L_{CF} = \frac{1}{N} \sum_{n=1}^N \|f_{C,n}^{CT} - f_{C,n}^{MRI}\|_2^2, \quad (2)$$

$$L_{BF} = \frac{1}{N} \sum_{n=1}^N \|f_{B,n}^{CT} - f_{B,n}^{MRI}\|_2^2, \quad (3)$$

$$L_{FEL} = \frac{L_{CF} + L_{BF}}{2}, \quad (4)$$

式中: L_{CF} 为 CT 和 MRI 图像分别使用 CNN 分支提取到特征之间的特征提取损失; L_{BF} 为 Transformer 分支之间的特征提取损失; L_{FEL} 为总特征提取损失,取 L_{CF} 与 L_{BF} 的平均值; $f_{C,n}^{CT}$ 和 $f_{C,n}^{MRI}$ 分别为 CT 和 MRI 图像过双分支编码器中 CNN 分支得到的第 n 个通道的特征; $f_{B,n}^{CT}$ 和 $f_{B,n}^{MRI}$ 分别为 CT

和MRI图像经过双分支编码器中Transformer分支得到的第 n 个通道的特征； N 为特征总数。

1.2.2 互信息损失

在特征提取阶段，不同模态信息由于处理过程可能损失其相关性，易导致有益的跨模态信息被遗忘。为了解决这一问题，本文引入互信息损失作为优化目标，通过最大化不同模态特征提取结果之间的互信息，可以鼓励模态之间信息共享，促进模态间的信息交流，从而确保在特征提取过程中各模态间关联性的保留和增强。为了在训练过程中，能够使网络使用梯度下降算法最小化损失函数的同时，最大化互信息，本文在计算互信息损失时取负值。MIL的定义如下

$$L_{CM} = -\frac{1}{N} \sum_{n=1}^N T(f_{C,n}^{CT}, f_{C,n}^{MRI}), \quad (5)$$

$$T(f_{C,n}^{CT}, f_{C,n}^{MRI}) = H(f_{C,n}^{CT}) + H(f_{C,n}^{MRI}) - H(f_{C,n}^{CT}, f_{C,n}^{MRI}), \quad (6)$$

$$L_{BM} = -\frac{1}{N} \sum_{n=1}^N T(f_{B,n}^{CT}, f_{B,n}^{MRI}), \quad (7)$$

$$T(f_{B,n}^{CT}, f_{B,n}^{MRI}) = H(f_{B,n}^{CT}) + H(f_{B,n}^{MRI}) - H(f_{B,n}^{CT}, f_{B,n}^{MRI}), \quad (8)$$

$$L_{MIL} = \frac{L_{CE} + L_{BM}}{2}, \quad (9)$$

式中： L_{CM} 和 L_{BM} 分别为CT和MRI图像使用CNN分支和Transformer分支提取到特征之间的互信息损失； L_{MIL} 为总互信息损失，即 L_{CM} 和 L_{BM} 的平均值； $H(\cdot)$ 为特征图的熵； $H(\cdot, \cdot)$ 为两幅特征图的联合熵。

1.3 改进U-Net的解码结构

双分支编码器结合了CNN和Transformer的优势，能够提取出丰富的特征。而U-Net中的跳跃连接对特征的利用率较低，为了使模型能够更加关注关键特征，并进一步保证融合图像的信息完整性，本文在解码器中提出了注意力门AD代替U-Net中直接的跳跃连接。解码器内部结构如图3所示。

以网络第3层的解码器 D_1 为例，设第4层concat后的大小为 $16 \times 16 \times 2048$ 的输出特征 f_4 为 q ，第3层concat后的大小为 $32 \times 32 \times 1024$ 的输出 f_3 为 p 。首先将 q 进行上采样操作，将特征图的尺寸扩大一倍，然后将上采样后的特征平均分为 a_1 和 a_2 两部分，如式(10)所示。

$$a_1, a_2 = \text{split}(\text{Upsampling}(q)). \quad (10)$$

接下来， a_1 与 p 进行元素相加，以强化重要特

征表示并抑制图像噪声，将相加后的特征经过sigmoid函数激活后，进而与 p 进行元素相乘得到特征 C_1 ，以增强共同的特征并削弱不重要的特征。

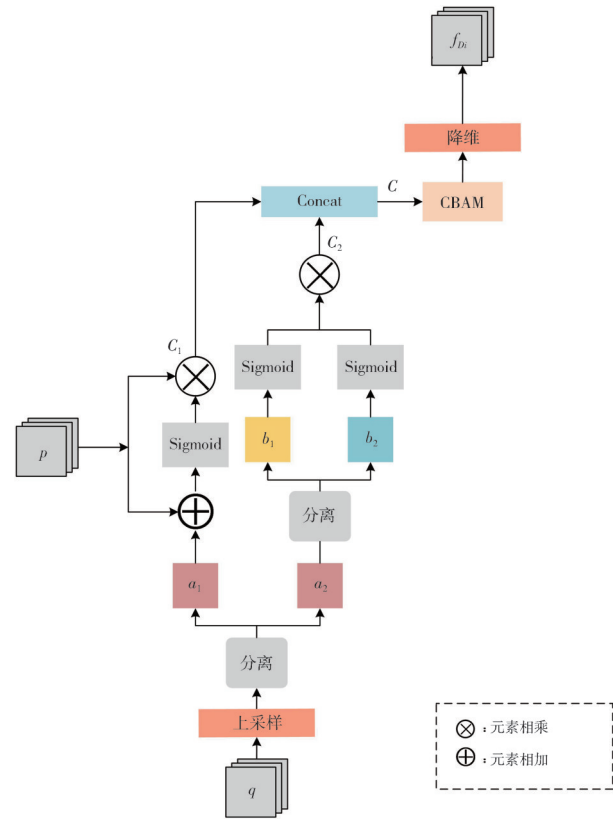


图3 解码器内部结构

Fig. 3 Internal structure of decoder

$$C_1 = \text{sigmoid}(p \oplus a_1) \otimes p, \quad (11)$$

式中： \oplus 为元素相加； \otimes 为元素相乘。如式(12)、式(13)所示，将 a_2 平均分为 b_1 和 b_2 两部分，再将 b_1 和 b_2 分别进行sigmoid函数激活后，进行元素相乘得到特征 C_2 。

$$b_1, b_2 = \text{split}(a_2), \quad (12)$$

$$C_2 = \text{sigmoid}(b_1) \otimes \text{sigmoid}(b_2). \quad (13)$$

最后，如式(14)所示，将 C_1 和 C_2 进行级联得到的特征输入卷积注意力模块^[18](CBAM)进行处理，并通过降维操作，输出大小为 $32 \times 32 \times 1024$ 的特征 f_{D1} 。

$$f_{D1} = \text{Reduction}(\text{CBAM}(\text{Concat}(C_1, C_2))). \quad (14)$$

1.4 损失函数

本文在整体网络的训练过程中，采用了L1损失和结构相似性损失作为编码器—解码器网络的训练目标，损失函数如式(15)所示

$$L = L_1 + \lambda_2 L_{SSIM}, \quad (15)$$

式中： L_1 和 L_{SSIM} 分别为L1损失和结构相似性损

失。鉴于 L1 损失与 SSIM 损失之间存在显著的数值差异,通过引入权重系数 λ_2 来平衡这两者之间的影响力,以实现综合的优化目标。

L1 损失通过分别计算两幅源图像和融合图像在同一像素点的绝对差值来实现,计算公式为

$$f_{\max} = \max(I_{CT}, I_{MRI}), \tag{16}$$

$$L_1 = \frac{1}{n} \sum |f_{\max} - I_f|, \tag{17}$$

式中: f_{\max} 为提取两幅源图像在同一像素点处的最大值; I_f 为融合图像; n 为像素点总数。

结构相似性损失通过比较源图像与融合图像之间的结构相似度来进行梯度优化,计算公式为

$$L_{SSIM} = 1 - SSIM(f_{\max}, I_f), \tag{18}$$

式中: SSIM 函数表示输入与输出两幅图像的结构相似度。

2 实验结果与分析

2.1 数据集与实验环境

本实验使用的的数据集来自美国哈佛医学院、山西省生物医学成像与影像大数据重点实验

室。选取 240 对经过配准后的 CT 和 MRI 图像,并将其中 180 对图像进行平移、旋转等操作扩充至 1 138 张作为训练数据集;剩余 60 对图像作为测试数据集,用于评估模型性能。

实验设置: GPU RTX3060Ti 8 GB GDR6; 环境框架为 Pytorch; Python 版本为 3.10.4。训练时使用 AdamW 优化器, batch_size 设置为 1, epoch 设置为 200, 初始学习率 1e-4, 学习率每 10 个 epoch 衰减 0.3。

2.2 对比实验

为验证本文方法的融合性能,选取基于 CNN 的 IFCNN^[19]、基于 GAN 的 U2Fusion^[20]、基于 Transformer 的 MATR^[21]、基于多尺度残差金字塔的 MSPRN^[22] 和基于双分支网络的 MsgFusion^[23] 以上 5 种先进的多模态图像融合方法进行对比,从定性和定量两个方面评价本文方法的有效性。

2.2.1 定性结果分析

本文对脑梗、脑血管、脑中风和脑肿瘤 4 种脑部疾病在不同对比方法以及本文方法上的融合结果进行了定性比较实验,融合结果如图 4 所示。

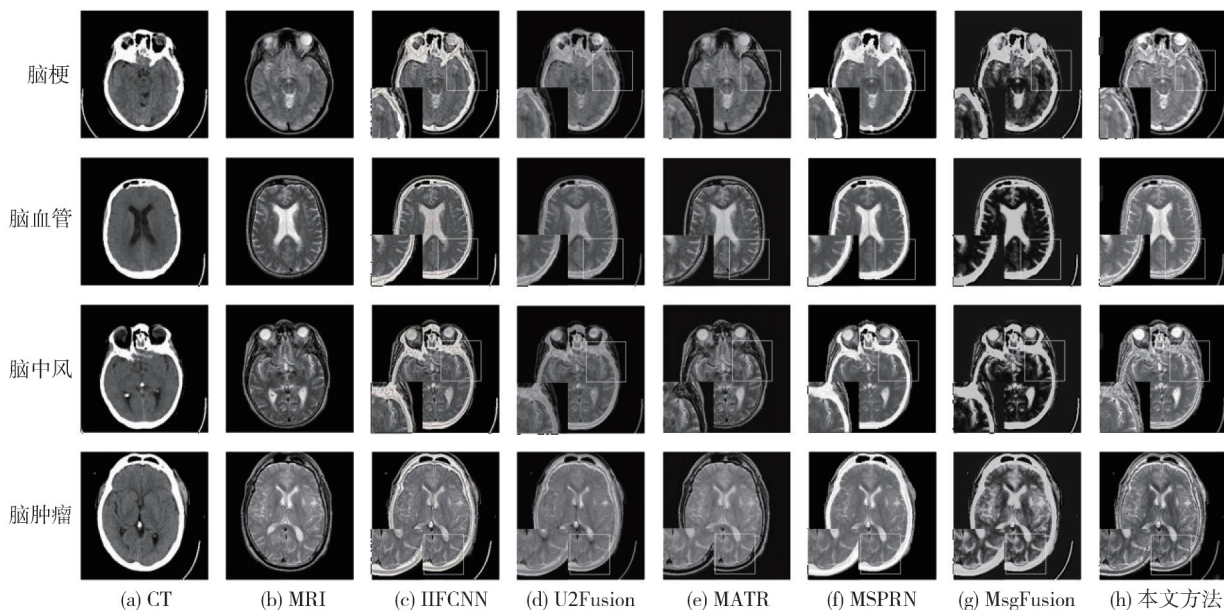


图 4 对比实验定性结果

Fig. 4 Comparative qualitative results of the experiment

在脑梗疾病中,IFCNN 方法生成的融合图像边缘细节丢失;U2Fusion 方法生成的融合图像亮度较低,无法显示梗塞区域的细微结构;MATR 结果中 CT 特征不明显,难以评估梗塞区域的密度变化;MSPRN 方法的融合图像特征明显,但细节过度锐化,导致脑梗区域边界模糊;MsgFusion 方法得到的

融合图像疾病区域边界清晰,但对比度过高,导致梗塞低密度区域被过度强化;而本文方法得到的融合图像脑部沟壑清晰,有助于更准确的识别脑梗区域。

在脑血管疾病中,IFCNN 方法生成的融合图像中血管与背景对比度不足,影响血管轮廓的识别;U2Fusion 结果灰度分布不均匀;MATR 方法

融合信息不充分; MSPRN方法得到的图像噪声增加; MsgFusion结果图像对比度过高, 血管内部的细节信息丢失; 而本文方法得到的融合图像成功保留了血管的纹理和细节信息, 有助于更好地理解病情的严重程度。

在脑中风疾病中, IFCNN方法得到的融合图像细节不清晰, 轮廓不平滑, 影响对中风程度的评估; U2Fusion结果亮度不平衡, 中风区域有微小结构丢失; MATR方法得到的融合图像CT信息欠佳, 灰度层次不清晰; MSPRN结果对比度不均, 影响了图像的纹理细节; MsgFusion结果灰度细节丢失, 影响对中风区域的准确评估; 相较之下, 本文方法得到的融合图像明显地展现了中风区域的变化, 有助于临床医生更好地理解病情的严重程度和影响范围。

在脑肿瘤疾病中, IFCNN方法的融合图像病灶区域与背景对比度不高, 影响对肿瘤的可观性; U2Fusion结果的图像细节不够清晰, 肿瘤结构信息丢失; MATR生成的融合图像难以区分肿瘤的不同成分; MSPRN结果锐化过度, 影响图像清晰性; MsgFusion中MRI图像细节丢失, 肿瘤区域的层次过于明显; 相比之下, 本文方法得到的融合图像有效地保留了脑髓质和骨骼的信息, 有助于提高肿瘤定位的准确性。

综上, 本文所提方法在脑梗、脑血管、脑中风和脑肿瘤4种疾病的融合图像中, 在细节清晰度和信息保留度上, 均优于其他对比方法。

2.2.2 客观评价指标及定量结果分析

为确保评价的客观性, 实验选取了熵(Entropy, EN)、标准差(Standard Deviation, SD)、差异相关和(Sum of the Correlations of Differences, SCD)、结构相似性指数(Structural Similarity Index, SSIM)、平均梯度(Average Gradient, AG)和相关系数(Correlation Coefficient, CC)以上6种客观评价指标评价本文方法在脑部医学图像数据集上的性能。其中, EN值越高, 表示图像包含的信息量越多; SD值越大, 表明图像像素值分布更为分散; SCD值越大, 融合图像与源图像之间的结构差异和内容差异越小; SSIM值越高, 表明融合图像与源图像的相似度更高, 且图像结构完整度越高; AG值越大, 说明图像的清晰度越好; CC值越接近于1, 融合图像与源图像的相关度越高, 融合质量越好。表1所示为不同融合方法在多个客观评价指标上的平均表现, 其中加粗字体标识为最优, 下划线标注为次优。

表1 对比实验的客观评价指标

Tab.1 Objective evaluation index of the comparative experiment

融合方法	EN	SD	SCD	SSIM	AG	CC
IFCNN	4.24	<u>77.39</u>	0.90	1.33	<u>7.49</u>	0.78
U2Fusion	<u>4.79</u>	56.93	0.93	0.73	4.77	<u>0.81</u>
MATR	4.47	46.72	-0.42	0.52	4.50	0.72
MSPRN	3.86	85.92	<u>1.19</u>	<u>1.45</u>	5.41	0.76
MsgFusion	4.26	69.54	0.71	0.45	6.52	0.74
本文方法	4.92	76.83	1.46	1.48	7.55	0.83

可以看出, IFCNN方法的评价指标值均不突出, 融合图像综合质量相对较低; U2Fusion方法的SSIM和AG评价指标较低, 表示其融合图像与源图像的相似度较低, 且融合图像的亮度和清晰度较低; MATR的SCD值为负值, 且所有指标均为最低, 表明该方法得到的融合图像质量不佳, 在融合图像中没有有效融合和保留源图像的信息; MSPRN得到的融合图像SD值突出, 说明该方法得到的融合图像像素分布较为分散, 图像的对比较高, 但其EN与CC值较低, 表明融合图像提供的信息量不足, 且与源图像的相关度较低, 影响了融合图像的整体质量; MsgFusion方法在SCD和SSIM指标上表现较差, 这表明该方法在保持融合图像与源图像结构、亮度的完整性和相似度方面存在不足, 导致融合图像边缘细节丢失; 本文方法得到的融合图像在SD指标上相对较低, 表示融合图像的像素分布更为集中, 这是因为本文方法在特征提取阶段注重源图像间的互补性和相关性, 减少了源图像间的波动性和不一致性, 使得融合图像的像素值之间不存在明显差异, 最终导致SD指标降低, 但定性结果分析表明, 融合图像质量并未因此受损。

2.3 消融实验

为了验证跨模态自蒸馏与改进U-Net的编码器能够提升融合图像质量的效果, 进行了消融实验。其中, 实验一未采用跨模态自蒸馏技术, 实验二将Inception结构替换为3*3卷积进行特征提取。消融实验模型与本文方法效果对比如图5所示, 实验一图像细节信息模糊; 实验二图像边缘轮廓信息丢失。客观评价指标如表2所示, 实验一EN值较低, 表明融合图像中源图像的信息量少; 实验二AG值较低, 表明融合图像清晰度较差。实验结果证实了本文方法在医学图像融合任务中的优越性能。

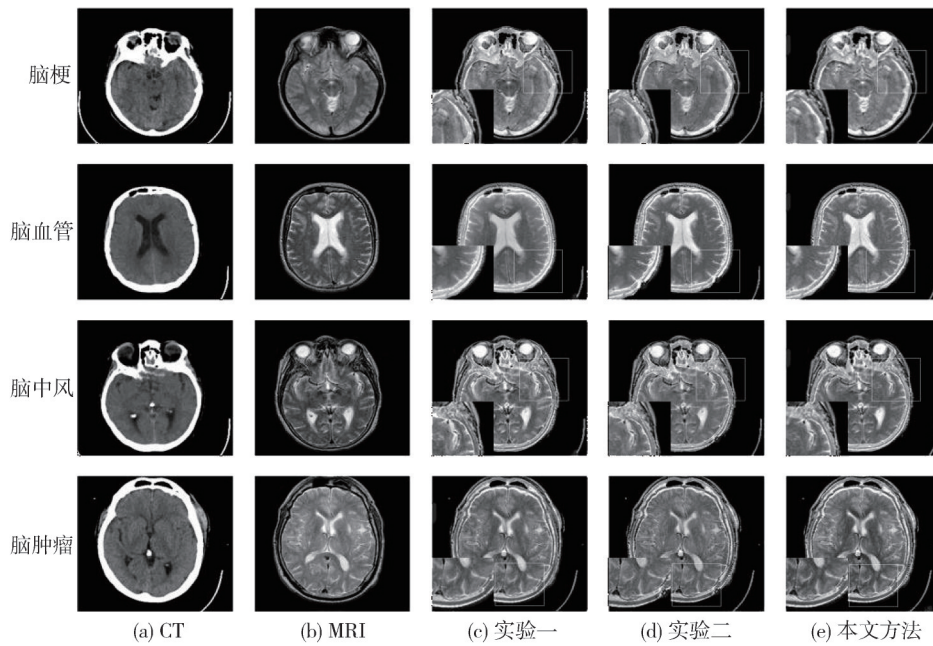


图 5 消融实验定性结果

Fig. 5 Qualitative results of ablation experiment

表 2 消融实验的客观评价指标

Tab. 2 Objective evaluation indexes of ablation experiments

融合方法	EN	SD	SCD	SSIM	AG	CC
实验1	4.13	78.32	1.43	1.36	8.01	0.79
实验2	4.61	75.56	1.45	1.38	6.69	0.82
本文方法	4.92	76.83	1.46	1.48	7.55	0.83

3 结 论

本文提出了基于改进 U-Net 和跨模态自蒸馏的医学图像融合方法,该方法有效克服了现有多模态医学图像融合方法得到的融合图像信息完整性丢失和跨模态特征提取不足的问题。根据实验验证,本文方法在主观视觉评估和客观评价指标上均展现出优异的性能。该方法具有以下特点:1) Inception 结构的卷积能够提取局部层内多尺度信息,与传统的 CNN 相比可以获得更多的细节信息;2) Biformer 块比传统 Transformer 性能好,计算量少,参数量少;3) 跨模态自蒸馏技术可以最大程度地对两幅源图像特征进行跨模态信息传递,使网络能够获取跨模态特征;4) 注意力门比传统的跳跃连接更能有效关注关键特征,使不同尺度的特征进行高效融合,保留更丰富的特征信息。

参考文献:

[1] VESHKI F G, OUZIR N, VOROBYOV S A, et al. Coupled feature learning for multimodal medical image fusion [J/OL]. arXiv: 2102.08641, 2021vl. <https://www.semanticscholar.org/paper/Coupled-Feature-Learning-for-Multimodal-Medical-Veshki-Ouzir/4a75e2ccdea15cd37245d06443837626e2766f5e>.

[2] PALKAR B, MISHRA D. Fusion of multi-modal lumbar spine images using Kekre’s hybrid wavelet transform [J]. IET Image Processing, 2019, 13(12): 2271-2280.

[3] YANG Y. Multimodal medical image fusion through a new DWT based technique [C]//2010 4th International Conference on Bioinformatics and Biomedical Engineering, 2010: 1-4.

[4] DA CUNHA A L, ZHOU J, DO M N. The nonsubsampled contourlet transform: theory, design, and applications [J]. IEEE Transactions on Image Processing, 2006, 15(10): 3089-3101.

[5] 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.

[6] ZHANG Y, YANG M, LI N, et al. Analysis-synthesis dictionary pair learning and patch saliency measure for image fusion [J]. Signal Processing, 2020, 167: 107327.

[7] LI H, WU X J. Multi-focus image fusion using dictionary learning and low-rank representation [C]// International Conference on Image & Graphics. Cham: Springer, 2017: 675-686.

[8] 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, 2017: 1-7.
- [9] JIAN L, YANG X, LIU Z, et al. SEDRFuse: a symmetric encoder-decoder with residual block network for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-15.
- [10] ZHAO Z, BAI H, ZHANG J, et al. CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5906-5916.
- [11] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using Real NVP[J/OL]. [2024-02-18]. <http://arxiv.org/abs/1605.08803v3>.
- [12] XUE Z, GAO Z, REN S, et al. The modality focusing hypothesis: towards understanding crossmodal knowledge distillation[J/OL]. [2024-02-18]. <http://arxiv.org/abs/2206.06487v3>.
- [13] ZHOU S, LIU W, HU C, et al. UniDistill: a universal cross-modality knowledge distillation framework for 3D object detection in bird's-eye view[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5116-5125.
- [14] 田加林, 徐行, 沈复民, 等. 基于跨模态自蒸馏的零样本草图检索[J]. 软件学报, 2022, 33(9): 3152-3164.
TIAN Jialin, XU Xing, SHEN Fumin, et al. Cross-modal self-distillation for zero-shot sketch-based image retrieval[J]. Journal of Software, 2022, 33(9): 3152-3164. (in Chinese)
- [15] YU W, ZHOU P, YAN S, et al. Inceptionnext: when inception meets ConvNeXT[J/OL]. [2024-02-18]. <http://arxiv.org/abs/2303.16900v1>.
- [16] ZHU L, WANG X, KE Z, et al. Biformer: vision transformer with bi-level routing attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.
- [17] 谢宁, 陈梁, 裴自卿, 等. 基于改进 U-Net 的白车身焊点定位研究[J]. 测试技术学报, 2024, 38(1): 12-18.
XIE Ning, CHEN Liang, PEI Ziqing, et al. Weldingspot positioning method for body in white based on improved U-Net[J]. Journal of Test and Measurement Technology, 2024, 38(1): 12-18. (in Chinese)
- [18] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [19] 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.
- [20] XU H, MA J, JIANG J, et al. U2Fusion: a unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 502-518.
- [21] 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.
- [22] FU J, LI W, DU J, et al. A multiscale residual pyramid attention network for medical image fusion[J]. Biomedical Signal Processing and Control, 2021, 66: 102488.
- [23] WEN J, QIN F, DU J, et al. Msgfusion: medical semantic guided two-branch network for multimodal brain image fusion[J]. IEEE Transactions on Multimedia, 2023, 26: 944-957.

(上接第 685 页)

- [15] 王法臻, 崔少辉, 王成, 等. FPGA 可重构仪器 USB 通信接口的设计与实现[J]. 测试技术学报, 2020, 34(5): 451-456.
WANG Fazhen, CUI Shaohui, WANG Cheng, et al. Design and implementation of USB communication interface of FPGA reconstructed instrument[J]. Journal of Test and Measurement Technology, 2020, 34(5): 451-456. (in Chinese)
- [16] 王茂森, 李兴, 戴劲松. 基于现场可编程门阵列的高速多通道并行测量系统设计[J]. 测试技术学报, 2015, 29(4): 303-308.
WANG Maosen, LI Xing, DAI Jinsong. The design of high-speed multi-channel parallel measurement system based on field programmable gate arrays[J]. Journal of Test and Measurement Technology, 2015, 29(4): 303-308. (in Chinese)
- [17] 乔华. 基于 QT 的炮口冲击波测试系统软件设计[J]. 中北大学学报(自然科学版), 2021, 42(2): 140-145.
QIAO Hua. Software design based on QT for muzzle shock wave test system[J]. Journal of North University of China (Natural Science Edition), 2021, 42(2): 140-145. (in Chinese)