

文章编号 1004-924X(2024)02-0221-16

基于 DCIF-GAN 的肺部肿瘤 PET/CT 跨模态 医学图像融合

周涛^{1,3}, 程倩茹^{1,3*}, 张祥祥^{1,3}, 李琦^{1,3}, 陆惠玲²

(1. 北方民族大学 计算机科学与工程学院, 宁夏 银川 750021;

2. 宁夏医科大学 医学信息工程学院, 宁夏 银川 750004;

3. 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)

摘要: 基于生成对抗网络 (Generative Adversarial Network, GAN) 的医学图像融合是计算机辅助诊断领域的研究热点之一, 但是现有基于 GAN 的融合方法存在训练不稳定, 提取图像的局部和全局上下文语义信息能力不足, 交互融合程度不够等问题。针对上述问题, 本文提出了双耦合交互式融合 GAN (Dual-Coupled Interactive Fusion GAN, DCIF-GAN)。首先, 设计了双生成器双鉴别器 GAN, 通过权值共享机制实现生成器之间和鉴别器之间的耦合, 通过全局自注意力机制实现交互式融合; 第二, 设计耦合 CNN-Transformer 的特征提取模块 (Coupled CNN-Transformer Feature Extraction Module, CC-TFEM) 和特征重构模块 (CNN-Transformer Feature Reconstruction Module, C-TFRM), 提升了对同一模态图像内部的局部和全局特征信息提取能力; 第三, 设计跨模态交互式融合模块 (Cross Model Intermodal Fusion Module, CMIFM), 通过跨模态自注意力机制, 进一步整合不同模态间的全局交互信息。为了验证本文模型的有效性, 在肺部肿瘤 PET/CT 医学图像数据集上进行实验, 该文方法在平均梯度, 空间频率, 结构相似度, 标准差, 峰值信噪比, 信息熵等上与其他四种方法中最优方法相比, 分别提高了 1.38%, 0.39%, 29.05%, 30.23%, 0.18%, 4.63%。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; 图像融合; PET/CT; 耦合生成对抗网络; Swin Transformer

中图分类号: TP399 **文献标识码:** A **doi:** 10.37188/OPE.20243202.0221

PET/CT Cross-modal medical image fusion of lung tumors based on DCIF-GAN

ZHOU Tao^{1,3}, CHENG Qianru^{1,3*}, ZHANG Xiangxiang^{1,3}, LI Qi^{1,3}, LU Huiling²

(1. School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China;

2. School of Medical Information and Engineering, Ningxia Medical University,
Yinchuan 750004, China;

3. Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission,
North Minzu University, Yinchuan 750021, China)

* Corresponding author, E-mail: chengqianru5@163.com

Abstract: Medical image fusion based on Generative Adversarial Network (GAN) is one of the research

收稿日期: 2023-08-02; 修订日期: 2023-09-14.

基金项目: 宁夏自然科学基金资助项目 (No. 2022AAC03149); 北方民族大学 2022 年研究生创新项目资助 (No. YCX22190)

hotspots in the field of computer-aided diagnosis. However, the problems of GAN-based image fusion methods such as unstable training, insufficient ability to extract local and global contextual semantic information of the images, and insufficient interactive fusion. To solve these problems, this paper proposed a dual-coupled interactive fusion GAN (DCIF-GAN). Firstly, a dual generator and dual discriminator GAN was designed, the coupling between generators and the coupling between discriminators was realized through the weight sharing mechanism, and the interactive fusion was realized through the global self-attention mechanism; secondly, coupled CNN-Transformer feature extraction module and feature reconstruction module were designed, which improved the ability to extract local and global feature information inside the same modal image; thirdly, a cross modal interactive fusion module (CMIFM) was designed, which interactively fuse image feature information of different modalities. In order to verify the effectiveness of the proposed model, the experiment was carried out on the lung tumor PET/CT medical image dataset. Compared with the best method of the other four methods, the proposed method in the average gradient, spatial frequency, structural similarity, standard deviation, peak signal-to-noise ratio, and information entropy are improved by 1.38%, 0.39%, 29.05%, 30.23%, 0.18%, 4.63% respectively. The model can highlight the information of the lesion areas, and the fused image has clear structure and rich texture details.

Key words: medical image; image fusion; PET/CT; coupled generative adversarial network; swin transformer

1 引言

不同模态的医学图像都有其优点和局限性,例如,计算机断层扫描(Computed Tomography, CT)图像具有很高的空间分辨率和几何特性^[1],可以清晰地显示密集的结构信息,但软组织的对比度相对较低^[2],尤其是无法清楚表示软组织和浸润性肿瘤的边界^[3];PET (Positron Emission Tomography)图像对软组织、器官、血管等显示清晰,提供了功能信息,有利于确定病灶范围,但其空间分辨率不如CT图像,对刚性的骨组织显示差,并有一定的几何失真^[4]。在临床实践中,单模态的医学图像不能完全准确反映病理组织的信息,医生需要检查同一部位的不同模态图像来诊断病人的病情^[5]。医学图像融合将多模态图像中的典型信息和互补信息合并到一个图像中,帮助医生理解图像内容^[6],融合的结果更有利于人类的视觉感知或机器的自动检测^[7],为医生诊断和治疗疾病开辟了新的前景^[8]。

自2014年生成对抗网络(Generative Adversarial Network, GAN)^[9]提出以来,由于其灵活性和优异的性能,在成像领域得到了广泛应用, GAN在图像融合任务中有网络灵活性强、对外部条件依赖性弱、能够自主优化融合图像质量等

优势。按照网络结构,基于GAN的图像融合方法可以分为三种:传统GAN的图像融合方法,双鉴别器GAN的图像融合方法,多生成器多鉴别器GAN的图像融合方法。Fusion GAN^[10], FL-GC-Fusion GAN^[11]和文献[12]都是基于传统GAN的图像融合方法,将拼接后的源图像作为生成器的输入,并生成同时保留两幅源图像信息的融合图像,但是这些方法将融合图像与单个模态的源图像作为鉴别器的输入,导致融合结果更倾向于保留某一个模态的特征信息;D2WGAN^[13], DDcGAN^[14]和DFPGAN^[15]是基于双鉴别器GAN的图像融合方法,采用两个独立的鉴别器分别计算融合图像与两幅源图像的结构差异,这样能够使融合图像更多地保留多源特征信息,但这些方法都以通道拼接的源图像作为生成器网络的输入,以相同的方式对不同模态的图像进行处理,然而简单的拼接不能有效地利用多模态图像的信息,网络只能学习图像语义层次之间的一种关系,从而限制了模型的性能;MGMDcGAN^[16]和RCGAN^[17]是基于多生成器多鉴别器GAN的图像融合方法,这些模型中包含多组GAN,可以充分获取不同模态源图像中的信息,提高融合图像的质量,但由于多个GAN的网络结构参数量大,导致在训练过程中非常耗时。

多生成器多鉴别器GAN充分地考虑了不同模态图像的差异性与互补性,在图像融合任务中取得了良好的效果,但还存在以下问题:多生成器或多鉴别器之间的关联性较弱,导致训练过程不稳定;基于GAN的图像融合方法大多都以CNN为基本结构,能够有效提取图像的局部细节特征,但关注图像全局信息的能力不足;基于多生成器多鉴别器GAN的图像融合方法模态间特征的交互程度不够,导致图像融合不充分。针对上述问题,本文提出了双耦合交互式融合GAN (Dual-Coupled Interactive Fusion GAN, DCIF-GAN),用于多模态CT和PET图像融合,主要贡献如下:

(1) 提出跨模态耦合生成器,处理PET图像中的病灶目标和CT图像中丰富的纹理特征,学习跨模态图像之间的联合分布;提出跨模态耦合鉴别器分别用于计算预融合图像与CT和PET图像间的结构差异,并使训练过程更加稳定。

(2) 设计耦合CNN-Transformer特征提取模块(Coupled CNN-Transformer Feature Extraction Module, CC-TFEM)和CNN-Transformer特征重构模块(CNN-Transformer Feature Reconstruction Module, C-TFRM),结合了Transformer和CNN的优势,在挖掘源图像中局部信息的同时也能学习特征之间的全局交互信息,实现更好的跨模态互补语义信息集成。

(3) 提出基于Swin Transformer的跨模态交互式融合模块(Cross Model Intermodal Fusion Module, CMIFM),通过跨模态自注意力机制,可以进一步整合不同模态图像之间的全局交互信息。

2 双耦合交互式融合DCIF-GAN

2.1 整体网络结构

CT图像和PET图像是一组多模态图像,CT图像缺乏组织信息,不能很好地表现病灶,而PET图像缺乏纹理结构细节,不能体现病灶所在组织结构中的位置,为了充分利用CT图像的解剖结构和PET图像中的功能信息,本文提出了双耦合交互式融合GAN,模型结构如图1所示,该模型以双生成器双鉴别器的结构来进行一个两团队的博弈游戏,每个团队分别包含一个生成器和一个鉴别器,分别为 G_1, D_1 和 G_2, D_2 。其中,

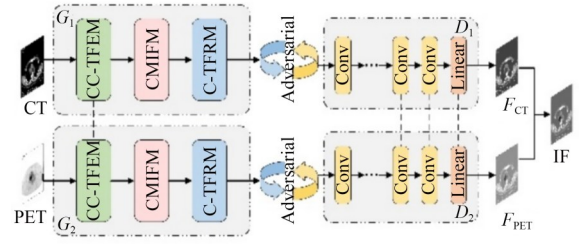


图1 双耦合交互融合GAN整体网络结构

Fig. 1 Dual-coupled interactive fusion GAN overall network architecture

生成器由基于耦合CNN-Transformer的特征提取模块(CC-TFEM)、跨模态与融合模块(CMIFM)和基于联合CNN-Transformer的特征重构模块(C-TFRM)构成,Liu等人^[18]提出的Co-GAN证明,通过在生成器中强制执行模块间权值共享,能够使网络学习多域图像间的联合分布,受此启发,本文通过共享特征提取模块的权值来实现生成器的“耦合”,从而使网络能够学习CT与PET图像浅层纹理特征中的联合分布,实现跨模态医学图像上下文语义信息的互补;鉴别器由四个卷积块和一个Linear层构成,鉴别器的“耦合”通过网络最后几层共享权值,此操作可以有效降低网络的参数量^[18]。

在双耦合交互式融合GAN中,第一生成器 G_1 的目的是生成具有CT图像纹理信息的预融合图像 F_{CT} ,对应的第一鉴别器 D_1 的目的是计算 F_{CT} 与源PET图像的相对偏移量并反馈,以增强 F_{CT} 中的功能信息;第二生成器 G_2 用于生成具有PET图像功能信息的预融合图像 F_{PET} ,第二鉴别器 D_2 计算 F_{PET} 与源CT图像的相对偏移量并反馈,以增强 F_{PET} 中的纹理信息,随着迭代次数的增加,两个生成器都可以生成足以欺骗鉴别器的预融合图像,生成的图像分别会相对偏向于其中一幅源图像,故将生成的两幅预融合图像进行加权融合,得到最终的融合图像 IF 。网络的极大极小博弈可以表示为:

$$\min_{G_1, G_2} \max_{D_1, D_2} L(G_1, G_2, D_1, D_2) = E_{I_{PET}} [\log D_1(I_{PET})] + E_{I_{CT}} [\log (1 - D_1(G_1(I_{CT})))] + E_{I_{CT}} [\log D_2(I_{CT})] + E_{I_{PET}} [\log (1 - D_2(G_2(I_{PET})))] \quad (1)$$

其中: G_1 和 G_2 是生成网络模型, D_1 和 D_2 是鉴别网络模型, I_{CT} 和 I_{PET} 是源CT和PET图像, $G_1(I_{CT})$ 和

$G_2(I_{PET})$ 分别是两个生成器生成的预融合图像, E 是期望输出值。 G_1 和 G_2 的目的是最小化目标函数(1), D_1 和 D_2 的目的是最大化目标函数(1),通过生成器和鉴别器的对抗,生成样本的概率分布和两个真实图像分布之间的分歧变得更小。

2.2 耦合生成器结构

由于CNN良好的性能和强大的特征提取能力,已被广泛应用于基于GAN的图像融合任务中,CNN能够简单而有效地提取图像的特征信息,但由于CNN的感受野受卷积核大小的影响,主要关注图像的局部特征信息,捕获多模态医学图像的全局背景信息的能力较弱;Transformer将整个图像转换为一维向量组作为输入,有效解决了CNN感受野有限的弱点,并通过使用自注

意力来捕获全局特征信息,弥补CNN只能提取局部特征的缺陷,但由于全局特征信息的秩往往较低,降低了前景与背景之间的可辨别性,导致融合图像病灶区域不明显。因此,为了有效利用多模态医学图像中的局部和全局互补特征,提高模型的融合性能,本文以联合CNN-Transformer为基本架构,提出了双耦合交互式融合GAN。生成器网络的结构如图2所示,其中包含三个模块:耦合CNN-Transformer特征提取模块(CC-TFEM)、跨模态预融合模块(CMIFM)和CNN-Transformer特征提取模块(C-TFRM)。跨模态耦合生成器的作用是生成融合图像,但是每个生成的图像都偏向于一个特定的源图像,这种有偏倚性的问题将在随后的耦合鉴别器中得到解决。

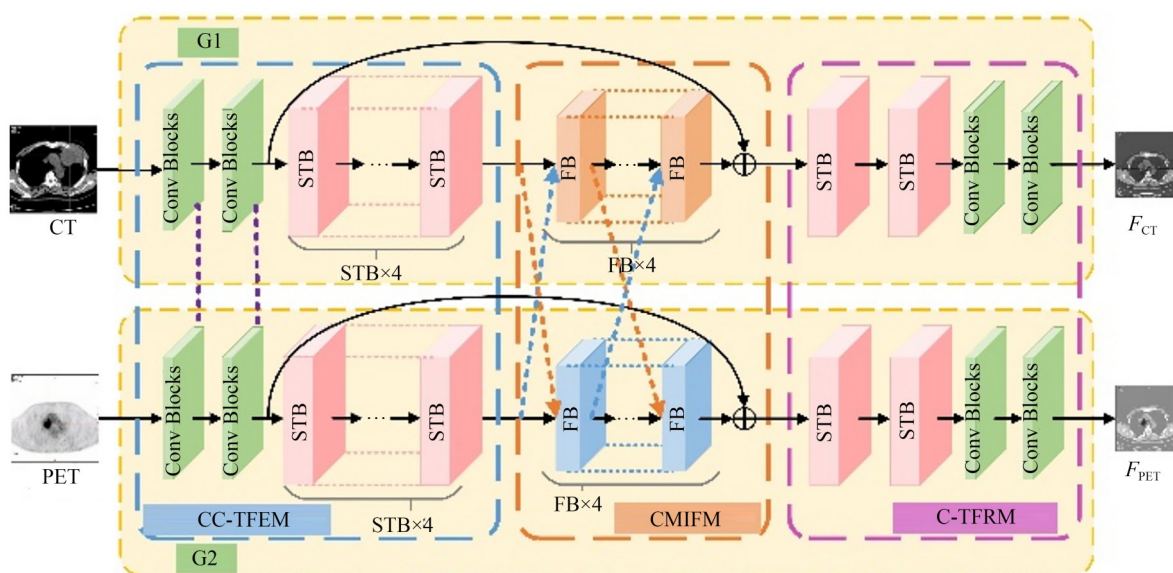


图2 生成器网络结构图

Fig. 2 The generator network structure

2.2.1 耦合CNN-Transformer特征提取模块(CC-TFEM)

充分提取源图像的特征信息是实现有效融合的关键,CNN只能通过加深网络,逐渐扩大接受域,从而捕捉到从细节到整体的特征信息,但与原始图像特征相比,CNN提取的特征更多地保留了浅层纹理信息,更有利于Transformer的训练,故本文结合CNN与Transformer的结构,设计了耦合CNN-Transformer特征提取模块(CC-TFEM)。如图3所示,采用基于CNN的浅层特征提取单元挖掘源图像中的局部特征,以保

留浅层纹理信息,基于Swin-Transformer的深度特征提取单元挖掘浅层特征之间的全局交互信息,生成包含高级语义信息的深度特征,两者的互补组合提高了特征提取的能力,并保留了更有效的信息。每个特征提取模块中包含2个卷积块和4个Swin Transformer块(Swin Transformer Block, STB),其中每个卷积块包含一个卷积层和一个Leaky ReLU层,卷积核的大小设置为 3×3 ,步长设置为1。为了将两生成器耦合在一起,需要共享特征提取模块的权值,通过在生成器之间的权值共享,一方面有助于学习多模态图

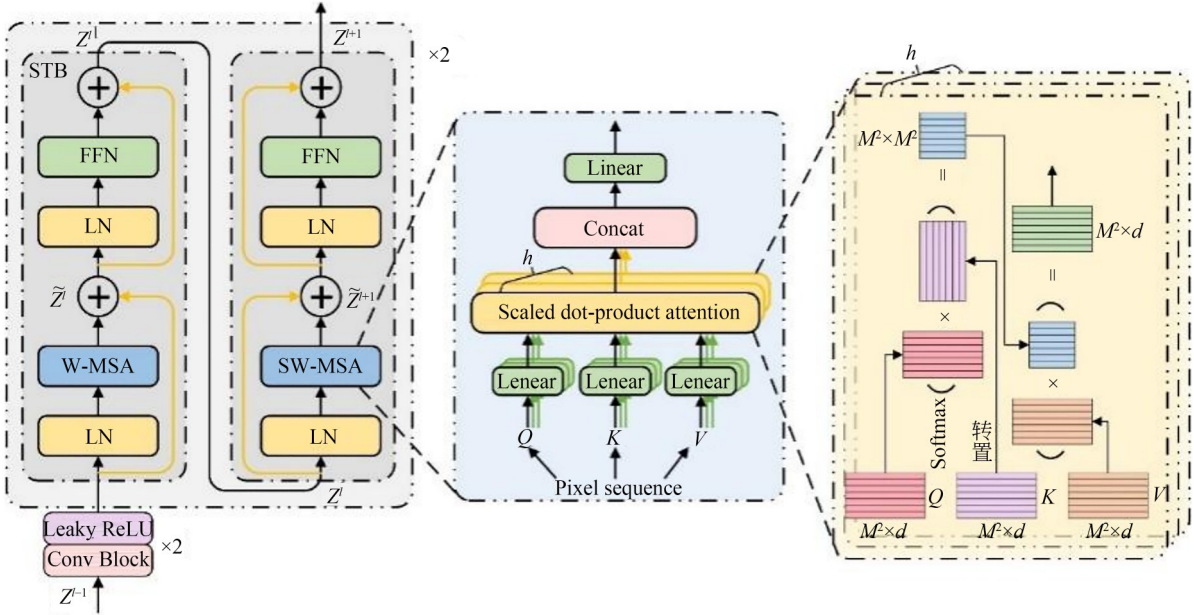


图3 特征提取模块

Fig. 3 Feature extraction module

像的联合分布,另一方面可以减少参数量。特征提取模块的步骤为:首先,通过浅层特征提取模块 $H_{SE}(\cdot)$ 提取源图像 I_{CT} 和 I_{PET} 的浅层特征 F_{SF}^{CT} 和 F_{SF}^{PET} ,然后,通过深度特征提取模块 $H_{DE}(\cdot)$ 从 F_{SF}^{CT} 和 F_{SF}^{PET} 中提取深度特征,最后,将 F_{DF}^{CT} , F_{DF}^{PET} 输入到跨模态预融合模块(CMIFM)中进行融合,这个过程可以表述为:

$$\begin{aligned} [F_{SF}^{CT}, F_{SF}^{PET}] &= [H_{SE}(I_{CT}), H_{SE}(I_{PET})] \\ [F_{DF}^{CT}, F_{DF}^{PET}] &= [H_{DE}(F_{SF}^{CT}), H_{DE}(F_{SF}^{PET})]. \end{aligned} \quad (2)$$

Vision Transformer (ViT)^[19] 首次将 Transformer^[20] 用于视觉任务中, Swin Transformer^[21] 是基于 ViT 的标准多头自注意力来实现的,与 ViT 不同, Swin Transformer 的局部注意力和窗口机制有效地降低了计算量。如图 3 右侧所示,给定大小为 $H \times W \times C$ 的特征 F , W-MSA 首先通过将输入划分为不重叠的 $M \times M$ 局部窗口,将输入特征 $F^{H \times W \times C}$ 重构为 $\frac{HW}{M^2} \times M^2 \times C$, 其中 $\frac{HW}{M^2}$ 为窗口总数。接下来,它对每个窗口分别执行标准的自注意力操作。对于局部窗口特征 $X \in R^{M^2 \times C}$, 使用三个线性变换矩阵 $W^Q \in R^{M^2 \times d}$, $W^K \in R^{M^2 \times d}$ 和 $W^V \in R^{M^2 \times d}$ 将其投影到 Q , K 和 V 中:

$$[Q, K, V] = [XW^Q, XW^K, XW^V]. \quad (3)$$

利用点积运算对矩阵 Q 和矩阵 K 中每个元素进行相似度匹配,然后用 softmax 进行归一化,得到注意力权重。过程可以表述为:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V, \quad (4)$$

其中: d_k 是键的维数, B 是可学习的相对位置编码,为了使注意力机制能够考虑不同的注意力分布,并使模型从不同的角度捕获信息,多头自注意力并行执行 h 次注意函数,并将每个注意力头的结果连接起来,在本文中, h 设置为 6。然后,通过由两个多层感知器 (Multi Layer Perceptron, MLP) 层组成的前馈网络 (Feed Forward Network, FFN) 来细化 W-MSA 产生的特征向量,过程可以表述为:

$$\begin{aligned} \tilde{Z}^l &= \text{MSA}(\text{LN}(Z^{l-1})) + Z^{l-1}, \\ Z^l &= \text{FFN}(\text{LN}(\tilde{Z}^l)) + \tilde{Z}^l, \end{aligned} \quad (5)$$

其中: Z^{l-1} 和 Z^l 分别表示 Swin Transformer 的输入和输出向量, \tilde{Z}^l 表示中间输出向量。前馈网络 $\text{FFN}(\cdot)$ 可以表述为:

$$\text{FFN}(X) = \text{GELU}(W_1 + b_1)W_2 + b_2, \quad (6)$$

其中, GELU 为高斯误差线性单元,使网络更快

地收敛,提高了模型的泛化能力。

图 4 显示了 Swin Transformer 层计算注意力的滑动窗口机制的示意图。在第 L 层中,采用 W-MSA 对图像进行计算,即一个规则的窗口划分方案,并在每个窗口内计算注意力,W-MSA 的弊端在于窗口之间的相互作用较弱,为了在不需要额外计算的情况下引入跨窗口交互,在下一层(第 $L+1$ 层)中,引入 SW-MSA 模块,SW-MSA 的窗口配置不同于 W-MSA 层,通过向左上方向循环移动,产生新的批窗口,在移动之后,窗口可以由特征图中的多个不相邻的子窗口组成,同时保持新窗口作为 W-MSA 中的常规分区进行计算,因此,新窗口中的注意力计算跨越了第 L 层中窗口的边界,提供了它们之间的连接。在 W-MSA 和 SW-MSA 中,当在局部窗口内进行自注意力计算时,在计算相似性中都包含了相对位置偏差。

2.2.2 跨模态交互式融合模块(CMIFM)

CC-TFEM 可以提取同一模态内的全局交互信息,为了进一步整合跨模态图像之间的全局

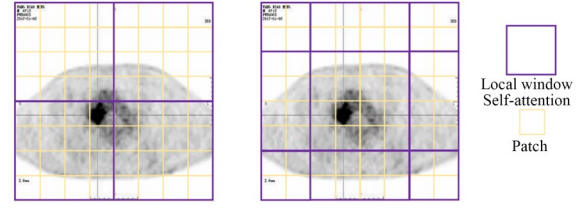


图 4 Swin Transformer 中的滑动窗口机制示意图

Fig. 4 Diagram of sliding window mechanism in Swin Transformer

交互信息,本文提出了跨模态交互式融合模块(CMIFM),其结构如图 5 所示,CMIFM 的每个分支由 4 个融合块(FB)构成,融合块通过自注意力,将特征图映射到查询(Query)、键(Key)和值(Value),并通过跨模态自注意力机制交换来自不同模态的键(Key)和值(Value)来实现全局特征信息融合。每个 CMIFM 由基于滑动窗口机制的多头自注意力(W-MSA)、基于滑动的 W-MSA (SW-MSA)、由两个多层感知器(MLP)层组成的前馈网络(FFN)和层归一化(LN)组成。

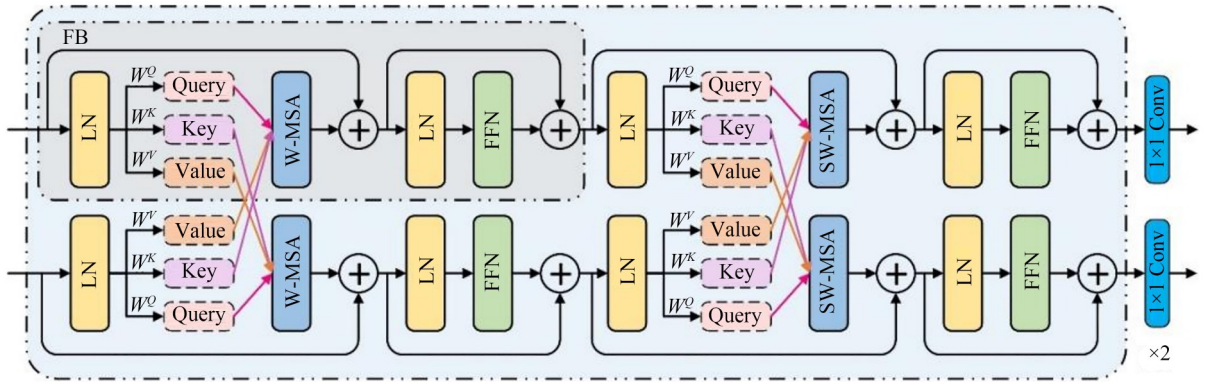


图 5 跨模态预融合模块(CMIFM)

Fig. 5 Cross Modal Interactive Fusion Module

如果给定来自不同模态的两个局部窗口特征 X_1 和 X_2 ,则将跨模态融合单元的过程定义为:

$$[Q_1, K_1, V_1] = [X_1 W_1^Q, X_1 W_1^K, X_1 W_1^V]$$

$$[Q_2, K_2, V_2] = [X_2 W_2^Q, X_2 W_2^K, X_2 W_2^V]$$

$$\text{Attention}_1(Q_1, K_2, V_2) = \text{softmax}\left(\frac{Q_1 K_2^T}{\sqrt{d_k}} + B\right) V_2$$

$$\text{Attention}_2(Q_2, K_1, V_1) = \text{softmax}\left(\frac{Q_2 K_1^T}{\sqrt{d_k}} + B\right) V_1$$

$$\tilde{Z}_1^i = W - \text{MSA}\left(\text{LN}\left(Z_1^{i-1}\right)\right) + Z_1^{i-1}$$

$$\tilde{Z}_2^i = W - \text{MSA}\left(\text{LN}\left(Z_2^{i-1}\right)\right) + Z_2^{i-1}$$

$$Z_1^i = \text{FFN}\left(\text{LN}\left(\tilde{Z}_1^i\right)\right) + \tilde{Z}_1^i$$

$$Z_2^i = \text{FFN}\left(\text{LN}\left(\tilde{Z}_2^i\right)\right) + \tilde{Z}_2^i, \quad (7)$$

其中: Z_1^{i-1} 和 Z_1^i 分别表示 CT 域中融合单元的输入和输出向量, \tilde{Z}_1^i 为中间输出向量, Z_2^{i-1} 和 Z_2^i 分别表示 PET 域中融合单元的输入和输出向量, \tilde{Z}_2^i 为中间输出向量。对于 CT 域中的 Q_1 , 它通过对 PET

域中的 K_2 和 V_2 进行注意力加权来整合跨模态信息,同时通过残差连接保留CT域中的信息,PET域中相同。本文由级联域内融合单元和域间融合单元组成,交替集成全局域间和跨模态相互作用。

对特征图进行模态内和跨模态特征融合后,利用一个卷积层来对融合后的特征图进行降维,其过程可以表述为:

$$\begin{aligned} F_{\text{FDF}}^{\text{CT}} &= H_{\text{conv1}}(F_{\text{AF}}^{\text{CT}}) \\ F_{\text{FDF}}^{\text{PET}} &= H_{\text{conv2}}(F_{\text{AF}}^{\text{PET}}), \end{aligned} \quad (8)$$

其中: $F_{\text{AF}}^{\text{CT}}$ 和 $F_{\text{AF}}^{\text{PET}}$ 分别表示CMIFM以 $F_{\text{DF}}^{\text{CT}}$ 和 $F_{\text{DF}}^{\text{PET}}$ 为输入而融合的输出特征。 $H_{\text{conv}}(\cdot)$ 表示具有空间不变滤波器的卷积层。 $F_{\text{FDF}}^{\text{CT}}$ 和 $F_{\text{FDF}}^{\text{PET}}$ 分别表示融合的CT图像和PET图像的深度特征,它们是特征重构模块的输入。

2.2.3 CNN-Transformer 特征重构模块 (C-TFRM)

在融合不同模态的互补信息后,本文采用了基于CNN-Transformer的特征重构模块(C-TFRM),将融合后的深度特征映射回图像空间。由于浅层特征包含低频信息,深层特征包含高频信息,在重构模块中,采用一个长跳跃连接,将低频信息传输到重构模块,这样CNN提取的浅层特征得到充分利用。特征重构的过程为:首先,采用两个Swin Transformer块的深度特征重构单元 $H_{\text{DR}}(\cdot)$,从全局的角度对融合的深度特征进行细化,对浅层特征进行恢复;然后,为了充分利用深层特征的全局上下文来恢复融合的浅层特征,采用基于CNN的图像重构模块 $H_{\text{SR}}(\cdot)$,减少通道数量,融合图像重构模块包含两个卷积层,卷积核大小设置为 3×3 ,步长设置为1,卷积层之后是Leaky ReLU激活函数;最后,生成预融合图像 F_{CT} 和 F_{PET} 。过程可以表述为:

$$\begin{aligned} F_{\text{FSF}}^{\text{CT}} &= H_{\text{DR}}(F_{\text{FDF}}^{\text{CT}} + F_{\text{SF}}^{\text{CT}}) \\ F_{\text{FSF}}^{\text{PET}} &= H_{\text{DR}}(F_{\text{FDF}}^{\text{PET}} + F_{\text{SF}}^{\text{PET}}) \\ F_{\text{CT}} &= H_{\text{SR}}(F_{\text{FSF}}^{\text{CT}}) \\ F_{\text{PET}} &= H_{\text{SR}}(F_{\text{FSF}}^{\text{PET}}). \end{aligned} \quad (9)$$

2.2.4 损失函数

本文将第一生成器 G_1 生成的预融合图像更偏向于学习源CT图像的梯度信息。其损失函数可表述为:

$$L_{G_1} = \Phi(G_1) + \alpha L_{\text{content1}}, \quad (10)$$

其中: L_{G_1} 表示生成器总损失, $\Phi(G_1)$ 表示对抗损失, L_{content1} 表示第一生成器从源图像到预融合图像的内容损失, α 是控制源PET图像信息含量比例的元素。借助该损失函数,第一个生成的预融合图像可以在 F_{CT} 的基础上保留更多源PET图像的信息,对抗损失 $\Phi(G_1)$ 和内容损失 L_{content1} 可表述为:

$$\Phi(G_1) = \frac{1}{N} \sum_{n=1}^N (D_1(F_{\text{CT}}^n, I_{\text{PET}}^n) - I_{\text{PET}}^n)^2, \quad (11)$$

$$L_{\text{content1}} = L_{\text{int}(\text{CT})} + \mu L_{\text{ssim}(\text{CT})}, \quad (12)$$

其中: $D_1(I_{\text{PET}}, F_{\text{CT}})$ 是第一鉴别器的函数, I_{PET} 是真实的源图像, F_{CT} 表示 G_1 得到的第一个生成的预融合图像, L_{int} 和 L_{ssim} 表示强度损失函数和结构相似度损失函数, μ 表示正则化参数。

同样,第二生成器的损失函数设置为:

$$L_{G_2} = \Phi(G_2) + \alpha L_{\text{content2}}, \quad (13)$$

其中, $\Phi(G_2)$ 和 L_{content2} 可以表述为:

$$\Phi(G_2) = \frac{1}{N} \sum_{n=1}^N (D_2(F_{\text{PET}}^n, I_{\text{CT}}^n) - I_{\text{CT}}^n)^2, \quad (14)$$

$$L_{\text{content2}} = L_{\text{int}(\text{PET})} + \mu L_{\text{ssim}(\text{PET})}, \quad (15)$$

其中: $D_2(I_{\text{CT}}, F_{\text{PET}})$ 是第二个鉴别器的函数, I_{CT} 是真实的源图像, F_{PET} 是第二个生成器生成的图像,第二个生成的预融合图像可以在 F_{PET} 的基础上保留更多源CT图像信息。

2.3 耦合鉴别器结构

在双鉴别器模型中,不仅要考虑生成器和鉴别器之间的对抗关系,还要考虑两个鉴别器之间的平衡,否则随着训练的进行,一个鉴别器的优势最终会导致另一个鉴别器的效率低下,为了更好地平衡两个鉴别器,本文设计了跨模态耦合鉴别器,其结构如图6所示。每个鉴别器由四个卷积块和一个线性层组成,卷积块由一个卷积层、一个BN层和一个Leaky ReLU激活层组成,所有卷积核大小设置为 3×3 ,步幅设置为2,四个卷积核个数分别设置为32,64,128和256,最后一层线性层将特征图转换为一个输出,表示融合图像与相应源图像之间的相对距离。鉴别器中第三、第四卷积块和线性层的共享权值,可以有效地减少网络的参数,且能够使网络更快收敛、训练过程更稳定^[18]。

耦合鉴别器的作用是通过损失函数使生成

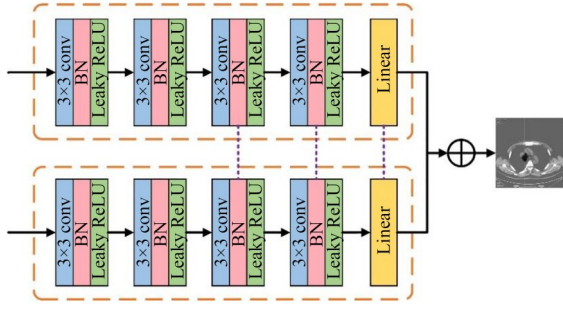


图6 鉴别器网络结构

Fig. 6 Discriminator network structure

的图像逼近另一幅源图像,通过反向传播,生成器与鉴别器反复博弈,生成的图像可以在包含当前源图像特征的同时,也一定程度上包含另一幅源图像的特征。其中,对于第一生成器 G_1 ,将源CT图像作为损失函数的一部分来优化结果,因此在第一鉴别器 D_1 中,计算第一个预融合图像 F_{CT} 与源PET图像 I_{PET} 的接近程度,以此来使 F_{CT} 更逼近 I_{PET} ,故本文将第一鉴别器 D_1 的损失函数设为:

$$L_1 = D_1(I_{PET}, F_{CT}). \quad (16)$$

相同,第二个鉴别器 D_2 的目的是通过损失函数使第二个预融合图像 F_{PET} 逼近源CT图像。因此,其损失函数可以表示为:

$$L_2 = D_2(I_{CT}, F_{PET}), \quad (17)$$

其中, L_1 和 L_2 分别表示第一鉴别器和第二鉴别器的损失函数。两个鉴别器的函数可以表示为:

$$\begin{aligned} D_1(I_{PET}, F_{CT}) &= C_1(I_{PET}) - E_{F_{CT}}(C_1(F_{CT})) \\ D_2(I_{CT}, F_{PET}) &= C_2(I_{CT}) - E_{F_{PET}}(C_2(F_{PET})), \end{aligned} \quad (18)$$

其中: E 是期望输出值, $C_1(\cdot)$ 和 $C_2(\cdot)$ 表示两个鉴别器的非线性变换。跨模态耦合鉴别器允许单个生成的图像具有相反图像的信息。但所得到的图像仍有一定程度的偏置,因此将生成的两幅图像进行平均,得到最终的融合结果 F 为:

$$F = 0.5 \times (F_{CT} + F_{PET}). \quad (19)$$

由于这两幅图像 F_{CT} 和 F_{PET} 都是基于源图像生成的有偏向性的与融合图像,因此平均操作可以使最终得到的融合图像既保留CT图像中的边缘、纹理等结构信息,又保留PET图像中的病灶等功能信息。

3 实验结果与分析

3.1 实验设置

数据集:选用2018年1月-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1 000组已配准的肺部肿瘤PET和CT影像,图像大小为:356 pixel×356 pixel。为了满足模型的输入条件,将原始RGB三通道图像转换为灰度图像。本文将处理好的数据集按照6:2:2比例划分为训练集、验证集和测试集,即训练集、验证集和测试集分别由600, 200, 200对CT-PET图像。

硬件与软件环境:Windows Server 2019 Datacenter的64位操作系统,计算机内存256 GB,显卡NVIDIA TITAN V,处理器Intel(R) Xeon(R) Gold 6154 CPU @ 3.00 GHz。Python 3.7编程语言,Pytorch1.7.0深度学习框架,CUDA版本为11.1.106。学习率为 1×10^{-4} ,为保障模型参数更新较快和在全局最优特点进行收敛,训练周期为1 000,批处理大小为4。

评价指标:选择8个评价指标来对融合图像进行定量评价,包括平均梯度(Average Gradient, AG)、空间频率(Spatial Frequency, SF)、结构相似度(Structural Similarity Index Measure, SSIM)、图像标准差(Standard Deviation, SD)、互信息(Mutual Information, MI)、峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)、信息熵(Information Entropy, IE)和基于梯度的融合性能(Edge based Similarity Measurement, $Q^{AB/F}$),这些评价指标都与融合质量正相关。

本节通过两组实验来证明本文模型的有效性,第一组实验是对比实验,将本文方法与医学图像领域的先进方法进行比较,本文选取了两类数据集进行对比实验,分别是PET/CT肺窗、PET/CT纵膈窗;第二组实验是消融实验,在基准模型的基础上,证明本文所提模块的有效性。

3.2 对比实验

为了验证该文模型的有效性,在两个数据集上进行实验,分别为CT肺窗和PET图像组与CT纵膈窗和PET图像组,在CT肺窗中,肺内部气管细节信息清晰,图像灰度值较大,CT纵膈窗中,纵膈信息清晰,图像灰度值偏小。将该文方

法与其他医学图像融合方法做比较,并从主客观上分析融合结果。

本节选择了4种图像融合方法在该文所提出的数据集上进行对比实验。分别是方法1:基于单生成器单鉴别器GAN的医学图像融合方法(Fusion GAN^[10]);方法2:基于单生成器双鉴别器GAN的医学图像融合方法(DDcGAN^[14]);方法3:基于多生成器多鉴别器GAN的医学图像融

合方法(MGMDcGAN^[16]);方法4:分解方法采用LatLRR和NSCT的嵌套分解,低频图像采用平均梯度自适应加权的融合规则,高频图像采用基于区域能量最大的融合规则(LatLRR+NSCT)。

3.2.1 实验一:PET/CT肺窗

定性实验:本节选取了5组融合图像来进行定性实验,图7提供了对比实验与本文模型在CT肺窗和PET图像组实验中的可视化结果,方

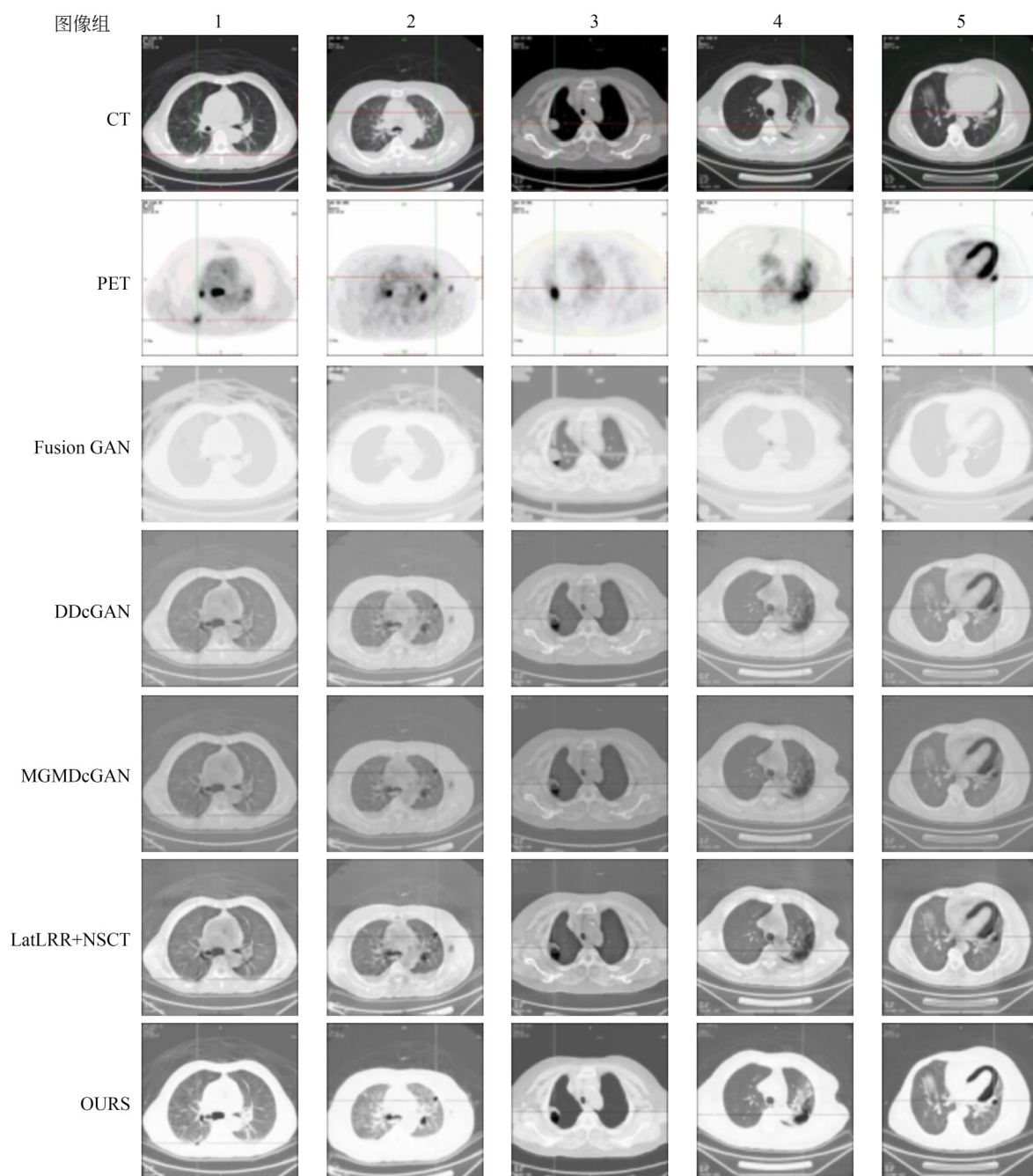


图7 对比实验一图像融合结果

Fig. 7 Contrast experiment 1 qualitative comparison

法 1、方法 2 和方法 3 对肺部支气管等细节信息不能较好地保留,虽然方法 3 能凸显病灶部位,但融合图像对比度低,边缘模糊,本文方法和方法 4 在显示肺部支气管等细节部分清晰,但本文方法的病灶区域更凸显。所以,本文方法较好得融合 PET 图像中的病灶信息和 CT 影像中的肺部支气管等信息。

定量实验:表 1 展示了本文方法与对比方法的定量实验数据比较,本文方法在大部分指标中取得了最优结果。具体来说,最高的 SF 和 IE 代

表本文方法的融合图像从源图像中保留了更多的特征和边缘信息,五组实验中,本文方法与其他方法中最优的相比,SSIM 分别提高了 1.16%,10.79%,15.03%,10.26% 和 1.57%,数据显示了本文方法在维护结构信息方面的优势,PSNR 分别提高了 0.81%,6.62%,8.15%,0.86% 和 0.67%,代表本文方法在融合过程中的信息失真最小,本文方法通过充分集成源图像中的全局相互作用,更加关注 PET 图像中的显著病灶区域。

表 1 对比实验一融合图像指标评价结果
Tab. 1 Comparison experiment 1 fusion image index evaluation results

图像组	方法	AG	SF	SSIM	SD	MI	PSNR	IE	$Q^{AB/F}$
1	Fusion GAN	2.003	6.97	0.497	26.8	1.482	10.032	4.875 2	0.109 3
	DDcGAN	7.477	21.32	0.587	33.11	1.432	10.251	6.663 5	0.347 6
	MGMDcGAN	5.184	14.88	0.601	29.88	2.145	10.338	6.258 5	0.285 3
	LatLRR+NSCT	8.071	23.27	0.600	45.48	1.512	10.331	7.328 7	0.548 2
	OURS	8.706	23.37	0.608	59.04	1.897	10.422	7.395 0	0.497 3
2	Fusion GAN	1.78	6.2	0.535	28.24	1.354	10.133	5.120 0	0.105 9
	DDcGAN	6.866	20.08	0.622	35.08	1.441	10.622	6.625 3	0.338 0
	MGMDcGAN	4.748	14.34	0.630	31.53	1.95	10.657	6.163 2	0.276 4
	LatLRR+NSCT	7.320 0	22.020 0	0.531	46.03	1.386	10.644	7.306 2	0.531 8
	OURS	8.113	22.3	0.698	62.81	1.861	11.363	7.482 4	0.486 8
3	Fusion GAN	2.893	9.86	0.429	37.08	1.73	8.839	5.681 1	0.127 4
	DDcGAN	7.087	21.74	0.499	35.56	2.514	8.906	6.155 3	0.336 3
	MGMDcGAN	4.985	15.77	0.492	32.03	2.925	9.006	5.830 8	0.285 0
	LatLRR+NSCT	8.305	23.4	0.492	46.68	2.104	8.940	7.113 0	0.559 6
	OURS	8.388	23.99	0.574	66.75	2.029	9.740	7.502 8	0.526 4
4	Fusion GAN	1.893	6.29	0.517	25.13	1.428	10.286	5.120 8	0.096 6
	DDcGAN	7.84	22.12	0.590	36.61	1.711	10.322	6.886 5	0.338 3
	MGMDcGAN	5.565	15.63	0.601	32.84	2.071	10.419	6.529 6	0.291 1
	LatLRR+NSCT	8.668	23.43	0.604	50.14	1.548	10.425	7.474 5	0.563 2
	OURS	9.071	23.94	0.666	62.78	1.828	10.515	7.638 3	0.496 7
5	Fusion GAN	2.034	6.69	0.500	29.19	1.324	9.577	5.262 4	0.084 0
	DDcGAN	8.019	21.69	0.558	36.91	1.994	9.778	6.996 9	0.343 5
	MGMDcGAN	5.534	15.19	0.571	33.21	2.318	9.882	6.607 6	0.274 9
	LatLRR+NSCT	8.76	24.33	0.573	50.23	1.747	9.861	7.547 2	0.583 4
	OURS	9.329	24.61	0.580	63.85	1.923	9.949	7.567 9	0.506 4

3. 2. 2 实验二:PET/CT 纵膈窗

定性实验:图 8 提供了对比实验与本文模型在 CT 纵膈窗和 PET 图像组实验中的可视化结果,方法 1 和本文方法均能较好保留 CT 组织、骨

骼对比度等信息,但方法 1 在一定程度上削弱了功能信息,不能有效地表现病灶区域;方法 2、方法 3 和方法 4 对组织轮廓的表现均不如本文方法清晰,虽然方法 4 保留更多病灶信息,但是器官和

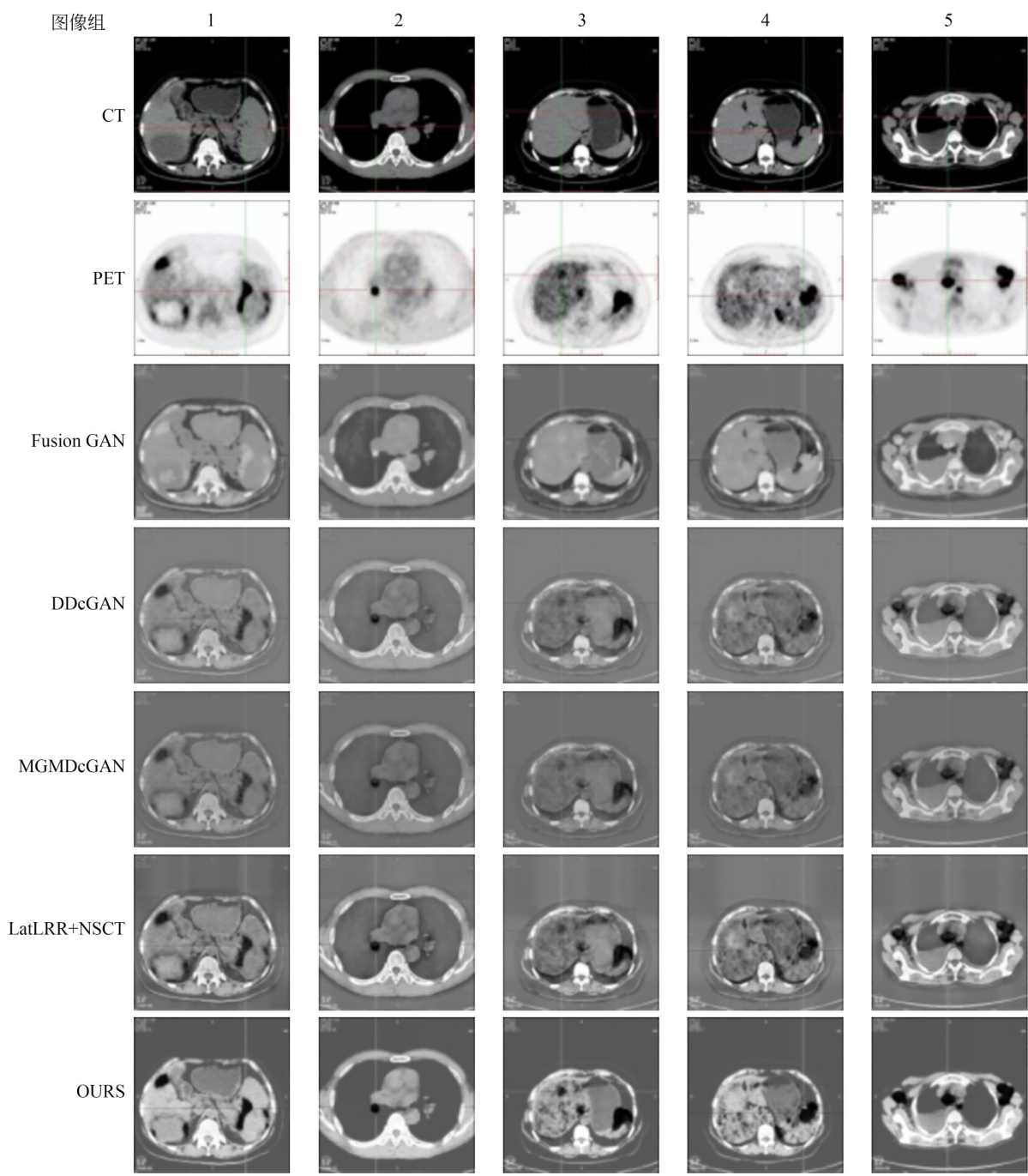


图 8 对比实验二图像融合结果

Fig. 8 Contrast experiment 2 qualitative comparison

骨骼等细节部分对比度低;本文方法不仅保留了 CT 纵膈窗图像的骨骼纹理信息,还更好地保留了 PET 图像的功能信息,证明本文方法可以通过跨模态交互自适应地关注功能信息和纹理信息。

定量实验:表 2 展示了本文模型与对比方法的定量实验数据比较,如数据所呈现的,本文方

法在大部分指标中取得了最优结果。具体来说,五组实验中,本文方法与其他方法中最优的相比,AG 分别提高了 0.27%,1.05%,3.14%,1.52%和 0.85%,代表本文方法对源图像梯度信息保留的更多;最高的 SSIM 显示了本文方法在维护结构信息方面的优势,最高的 PSNR 代表本文方法在融合过程中的信息失真最小,在本实验

表 2 对比实验二融合图像指标评价结果

Tab. 2 Comparison experiment 2 fusion image index evaluation results

图像组	方法	AG	SF	SSIM	SD	MI	PSNR	IE	$Q^{AB/F}$
1	Fusion GAN	4.877	14.56	0.303	32.42	1.736	8.234	5.982 8	0.120 7
	DDcGAN	10.897	25.58	0.381	27.48	1.611	8.376	6.471 3	0.289 5
	MGMDcGAN	7.440	17.34	0.384	23.63	0.757 1	8.489	6.232 1	0.253 1
	LatLRR+NSCT	13.582	30.49	0.384	38.87	1.524	8.469	7.491 3	0.551 9
	OURS	13.619	30.85	0.391	52.81	2.05	8.522	7.752 6	0.518 0
2	Fusion GAN	4.665	14.96	0.333	35.46	2.441	8.12	6.012 8	0.200 9
	DDcGAN	7.522	22.76	0.403	30.27	2.322	8.155	6.185 3	0.314 2
	MGMDcGAN	5.082	15.64	0.388	27.27	0.549 5	8.255	5.976 2	0.269 0
	LatLRR+NSCT	8.235	24.45	0.407	40.27	2.124	8.173	7.049 1	0.550 3
	OURS	8.322	24.55	0.407	52.68	2.914	8.266	7.387 5	0.488 8
3	Fusion GAN	4.215	13.08	0.279	31.19	1.811	7.714	5.469 3	0.120 2
	DDcGAN	9.933	26.19	0.367	27.53	1.842	7.755	5.681 7	0.290 6
	MGMDcGAN	6.909	18.32	0.36	22.9	0.703 4	7.859	5.457 9	0.245 5
	LatLRR+NSCT	11.858	31.09	0.382	35.890 0	1.247	7.772	6.901 1	0.510 3
	OURS	12.231	31.14	0.383	44.55	2.216	7.873	7.276 7	0.461 8
4	Fusion GAN	4.291	13.72	0.286	30.4	1.676	7.752	5.375 9	0.121 4
	DDcGAN	9.783	25.47	0.374	26.24	1.656	7.852	5.626 8	0.276
	MGMDcGAN	6.913	18.11	0.367	21.9	0.728 9	7.954	5.420 0	0.243 6
	LatLRR+NSCT	12.234	31.78	0.394	34.82	1.154	7.864 0	6.925 4	0.512 1
	OURS	12.42	31.25	0.396	46.6	2.006	7.961	7.291 5	0.468 8
5	Fusion GAN	5.015	16.28	0.317	36.79	2.217	7.495	5.473 6	0.164 7
	DDcGAN	8.253	24.19	0.375	33.53	2.417	7.637	5.52	0.298 1
	MGMDcGAN	5.372	16.15	0.366	29.67	0.610 87	7.737	5.322 2	0.239 7
	LatLRR+NSCT	9.324	27.06	0.38	43.610 0	1.858	7.671	6.736 8	0.533 6
	OURS	9.404	27.14	0.382	55.23	2.599	7.747	7.022 1	0.478 8

Bold font is the best result for each column

中,SSIM和 $Q^{AB/F}$ 值略落后于方法4。

3.3 消融实验

本节在基于生成对抗网络的图像融合方法的基础上,证明了本文提出的模块的有效性,本节基准网络的构建方法为:采用单生成器单鉴别器GAN(Network1)进行图像融合,生成器以CT和PET拼接图像作为输入,生成器内部采用五个卷积块作为主干,鉴别器对融合图像和CT图像进行偏移量计算并回馈,以此为基准网络,依次对本文提出的模块进行测试;跨模态耦合生成

器和跨模态耦合鉴别器(Network2)、耦合CNN-Transformer特征提取模块(CC-TFEM)(Network3)、跨模态特征融合模块(CMIFM)(Network4),证明本文模型的有效性。其中,Network2由两个生成器和两个鉴别器组成,每个生成器都和Network1中的生成器结构相同,并将前两个卷积块进行耦合,鉴别器与本文模型的鉴别器结构相同,并将第三、四个卷积块和最后一个Linear层进行耦合;Network3在Network2的基础上将前两个卷积块替换为CC-TFEM,后两

个卷积块替换为C-TFRM,并将CC-TFEM模块进行耦合,且在CC-TFEM与C-TFRM之间加入长跳跃连接;Network4在Network3的基础上,将中间卷积块替换为CMIFM。

定性实验:如图9所示,因单鉴别器仅对源CT图像和融合图像进行鉴别,所以Network1得到的结果整体比较模糊,对PET图像中病灶的表现能力较差;Network2对两幅源图像和融合图像进行鉴别,故病灶的表现能力有所提升,但整体依然比较模糊;Network3加入CC-TFEM模块

后,从指标值可以看出融合图像的亮度、对比度等有所提升,这说明了CC-TFEM可以充分提取图像的全局深度特征信息,且跳跃连接使得浅层特征提取模块中的纹理特征得到充分利用,该模块可以更好地保留源图像中重要的互补特征信息;CDIFM模块使得融合图像具有更多的梯度信息,提升了融合图像的质量,与Network3的融合结果相比,CDIFM使得融合结果更清晰地突出病灶的同时反映图像的纹理背景信息,Network4源图像的互补语义特征信息交互

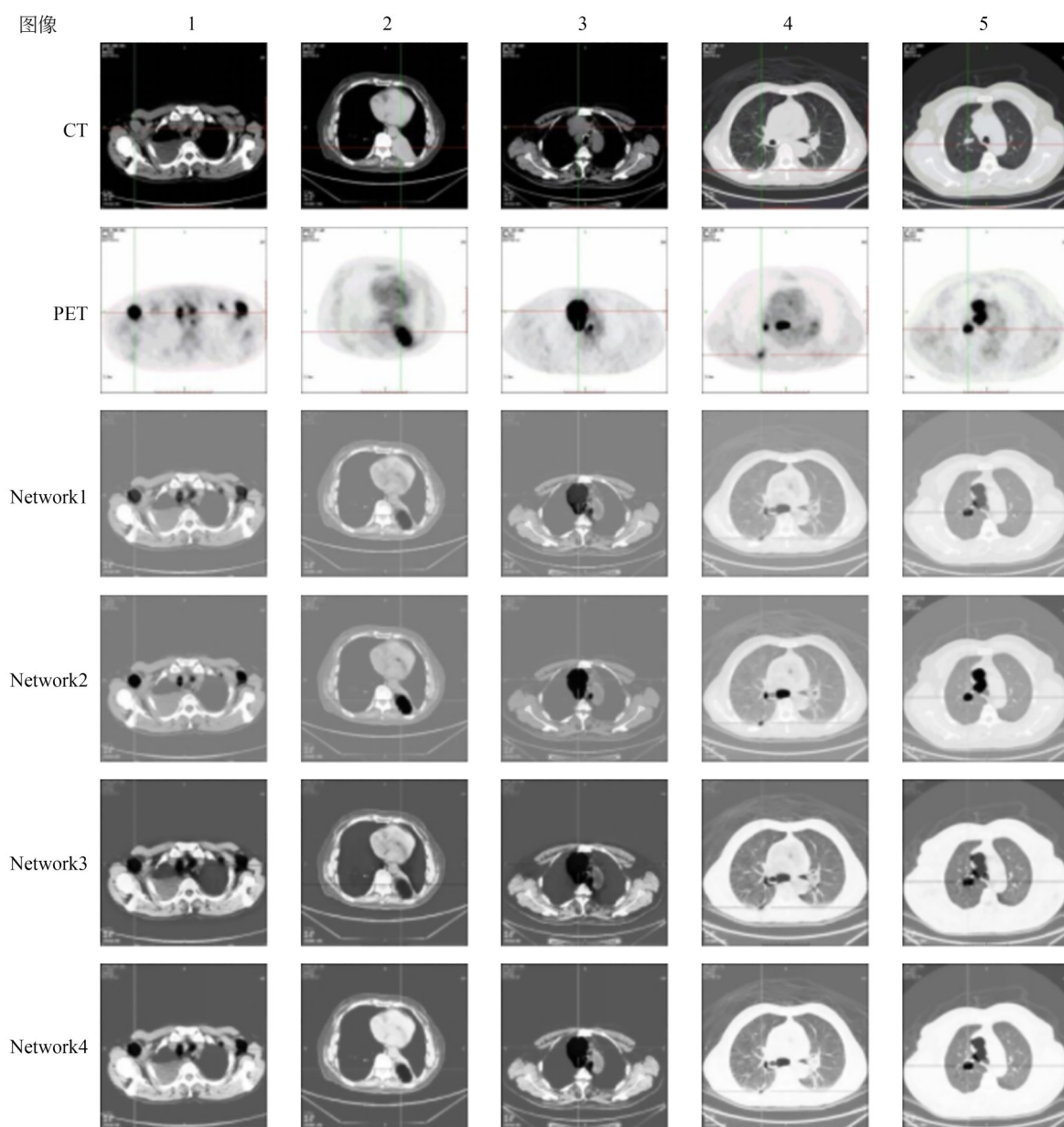


图9 消融实验定性比较

Fig. 9 Qualitative comparison of ablation experiments

更充分。

定量实验:表 3 展示了消融实验的评价指标数据,从数据结果中可以看出,本文所提出的模块都能够一定程度地提升融合的性能。具体来说,双耦合生成对抗网络相较于基准网络在除 MI 外都得到了一定程度的提升,Network4 与 Network3 相比,SD 分别提升了 4.66%, 11.45%,5.02%,3.87% 和 1.01%,SF 分别提升

了 8.76%,0.38%,10.49%,0.27% 和 2.26%,说明耦合生成对抗网络结构能够有效且更多地保留源图像特征信息;加入 CC-TFEM 和 C-TFRM 后,除 SSIM 和 SD 外的指标值都得到了提升,说明该模块可以有效地提取源图像的局部和全局上下文特征;加入 CMIFM 后,所有指标都得到了提升,说明 CMIFM 可以有效地交互跨模态图像之间的互补信息。

表 3 消融实验结果评价指标
Tab. 3 Evaluation index of ablation experimental results.

图像组	方法	AG	SF	SSIM	SD	MI	PSNR	IE	Q ^{AB/F}
1	Network1	8.159	23.34	0.352	39.14	2.525	7.613	4.943 4	0.294 7
	Network2	5.951	20.04	0.354	38.94	2.537	7.602	4.721 9	0.248 4
	Network3	7.899	24.99	0.346	54.4	2.609	7.674 0	5.518 4	0.432 4
	Network4	9.366	27.18	0.355	57.04	2.61	7.799	6.505 6	0.473 2
2	Network1	7.845	24.42	0.312	33.17	2.135	7.284	4.619 9	0.302 4
	Network2	5.644	19.73	0.315	32.56	2.219	7.278	4.528 0	0.238 6
	Network3	7.369	25.96	0.328	47.13	2.617	7.494	5.260 3	0.367 3
	Network4	8.615	26.00	0.333	52.53	2.971	7.523	6.454 7	0.460 3
3	Network1	8.974	26.4	0.309	33.11	2.647	7.278	4.802 8	0.320 3
	Network2	5.742	20.27	0.306	31.26	2.766	7.273	4.536 4	0.215 1
	Network3	8.233	27.25	0.315	45.13	2.904	7.444	5.357 2	0.452 8
	Network4	10.247	30.11	0.323	47.40	2.996	7.504	6.332 0	0.484 0
4	Network1	7.266	20.4	0.567	38.99	1.447	10.074	6.656 4	0.363 1
	Network2	6.052	19.26	0.564	40.67	1.617	10.059	6.443 9	0.339 9
	Network3	7.581	22.10	0.579	56.84	1.801	10.253	6.867 1	0.463 1
	Network4	8.024	22.16	0.568	59.04	1.897	10.302	6.867 3	0.485 0
5	Network1	6.623	19.42	0.587	43.87	1.826	10.397	6.382 4	0.338 6
	Network2	5.9	20.38	0.584	49.22	1.966	10.318	6.147 7	0.334 0
	Network3	6.606	21.21	0.603	64.77	2.08	10.435	6.232 5	0.393 7
	Network4	7.209	21.69	0.608	65.43	2.136	10.766	6.711 2	0.457 2

4 结 论

为解决跨模态医学图像融合不充分、病灶区域不明显等问题,本文提出了双耦合交互式融合 GAN,用于肺部肿瘤 PET/CT 跨模态医学图像融合。本文模型以耦合生成器耦合鉴别器的结构来进行一个两团队的博弈游戏,两个生

成器分别用于生成具有偏向性的预融合图像,鉴别器迫使融合图像包含更多源图像的特征信息。在生成器中,首先,提出耦合 CNN-Transformer 的特征提取模块(CC-TFEM),在挖掘源图像中的局部信息的同时也能学习特征之间的全局交互信息;然后,设计基于 Swin Transformer 的跨模态交互式融合模块(CMIFM),分别对

同一模态和跨模态图像进行融合,在整合同一模态内的全局交互信息的同时进一步整合不同模态之间的全局交互信息;最后,通过CNN-Transformer特征重构模块(C-TFRM)对融合后的图像进行重构,生成两幅“预融合”图像。在耦合鉴别器的最后几层共享权值,使训练过程更加稳定。为了验证本文模型的有效性,在肺部肿瘤PET/CT医学图像数据集上进行实验,通过与现有的医学图像融合模型的对比实验,在平均梯度,空间频率,结构相似度,标准差,峰

值信噪比,信息熵等上与其他四种方法中最优方法相比,分别提高了1.38%,0.39%,29.05%,30.23%,0.18%,4.63%,结果表明本文模型在绝大多数指标数据上取得了最优值,消融实验表明,本文所提模块都能在一定程度上提升模型的效果,通过定性实验证明,本文模型得到的融合图像符合人类视觉感知,能够较好地融合PET图像中的病灶信息和CT图像中的纹理信息,有助于医生更快速、更精准地定位肺部肿瘤在解剖结构中的位置。

参考文献:

- [1] LI Y, ZHAO J L, LV Z H, *et al.* Medical image fusion method by deep learning [J]. *International Journal of Cognitive Computing in Engineering*, 2021, 2: 21-29.
- [2] WANG L F, ZHANG J, LIU Y, *et al.* Multimodal medical image fusion based on Gabor representation combination of multi-CNN and fuzzy neural network[J]. *IEEE Access*, 2021, 9: 67634-67647.
- [3] ZHANG Y D, DONG Z C, WANG S H, *et al.* Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation [J]. *Information Fusion*, 2020, 64: 149-187.
- [4] POLINATI S, DHULI R. Multimodal medical image fusion using empirical wavelet decomposition and local energy maxima [J]. *Optik*, 2020, 205: 163947.
- [5] ZHANG J Z, ZHOU Z, TENG J H, *et al.* Fusion algorithm of functional images and anatomical images based on wavelet transform [C]. 2009 2nd International Conference on Biomedical Engineering and Informatics. Tianjin, China. IEEE, 2009: 1-5.
- [6] LIANG X C, HU P Y, ZHANG L G, *et al.* MCF-Net: multi-layer concatenation fusion network for medical images fusion [J]. *IEEE Sensors Journal*, 2019, 19(16): 7107-7119.
- [7] ZHAO W D, LU H M, WANG D. Multisensor image fusion and enhancement in spectral total variation domain [J]. *IEEE Transactions on Multimedia*, 2018, 20(4): 866-879.
- [8] LIANG M Y, REN Z Y, YANG J M, *et al.* Identification of colon cancer using multi-scale feature fusion convolutional neural network based on shearlet transform [J]. *IEEE Access*, 2020, 8: 208969-208977.
- [9] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, *et al.* Generative adversarial networks [J]. *Communications of the ACM*, 2014, 63: 139-144.
- [10] MA J Y, YU W, LIANG P W, *et al.* Fusion-GAN: a generative adversarial network for infrared and visible image fusion [J]. *Information Fusion*, 2019, 48: 11-26.
- [11] YUAN C, SUN C Q, TANG X Y, *et al.* FLGC-fusion GAN: an enhanced fusion GAN model by importing fully learnable group convolution [J]. *Mathematical Problems in Engineering*, 2020, 2020: 6384831.
- [12] MA J Y, LIANG P W, YU W, *et al.* Infrared and visible image fusion via detail preserving adversarial learning [J]. *Information Fusion*, 2020, 54: 85-98.
- [13] LI J, HUO H T, LIU K J, *et al.* Infrared and visible image fusion using dual discriminators generative adversarial networks with Wasserstein distance [J]. *Information Sciences*, 2020, 529: 28-41.
- [14] 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: 4980-4995.
- [15] YI S, LI J J, YUAN X S. DFPGAN: dual fusion path generative adversarial network for infrared and visible image fusion [J]. *Infrared Physics and Technology*, 2021, 119: 103947.
- [16] HUANG J, LE Z L, MA Y, *et al.* MGMDcGAN: medical image fusion using multi-generator multi-discriminator conditional generative adversari-

- al network [J]. *IEEE Access*, 2016, 8: 55145-55157.
- [17] LI Q L, LU L, LI Z, *et al.* Coupled GAN with relativistic discriminators for infrared and visible images fusion [J]. *IEEE Sensors Journal*, 2021, 21 (6): 7458-7467.
- [18] LIU M Y, TUZEL O. Coupled Generative Adversarial Networks [EB/OL]. 2016: *arXiv*: 1606.07536. <http://arxiv.org/abs/1606.07536.pdf>
- [19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [EB/OL]. 2020: *arXiv*: 2010.11929. <http://arxiv.org/abs/2010.11929.pdf>
- [20] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [C]. *Advances in neural information processing systems*, 2017, 30.
- [21] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021. *Montreal, QC, Canada*. IEEE, 2021: 10012-10022.

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: zhouta-onxmu@126.com

通讯作者:



程倩茹(1997—),女,山西晋中人,硕士研究生,2019年于中北大学获得学士学位,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: chengqianru5@163.com