基于 GAN 和多尺度空间注意力的多模态医学图像融合

林予松[1,2,3], 李孟娅[1,2], 李英豪[1,2], 赵　哲[1,2]

(1.郑州大学 网络空间安全学院,河南 郑州 450002;2.郑州大学 互联网医疗与健康服务河南省协同创新中心,河南 郑州 450052;3.郑州大学 汉威物联网研究院,河南 郑州 450002)

摘　要:针对多模态医学图像融合过程中多尺度特征和纹理细节信息丢失的问题,提出一种基于生成对抗网络和多尺度空间注意力的图像融合算法。首先,生成器采用自编码器结构,分别利用编码器和解码器对输入图像进行特征提取、融合和重建,生成融合图像。其次,整个对抗网络框架采用双鉴别器结构,使得生成器生成的融合图像同时保留多个模态图像的显著特征。最后,构建一种多尺度空间注意力作为编码器进行特征提取的基本模块,利用多尺度结构充分捕获并保留源图像的多尺度特征,并且引入空间注意力更好地保留源图像的结构和细节信息。在全脑图谱数据库上的实验结果表明:所提算法生成的融合图像不仅纹理细节更为丰富,有助于人类视觉观察,而且在 3 种不同类型的医学图像融合任务上平均梯度、峰值信噪比、互信息、视觉信息保真度等客观评价指标的平均值分别达到 0.302 3、20.720 7、1.441 4、0.649 8,与其他先进的算法相比具有一定的优势。

关键词:图像融合; 多模态医学图像; 生成对抗网络; 特征金字塔; 注意力机制

中图分类号:TP391　　**文献标志码**:A　　**doi**:10.13705/j.issn.1671-6833.2025.01.001

随着医学成像技术的不断发展,医学图像在现代医学诊断和治疗中发挥着不可或缺的作用。常见的医学图像包括磁共振图像(MRI)、计算机断层扫描(CT)图像、正电子发射断层(PET)图像和单光子发射计算机断层(SPECT)图像等。然而,由于成像原理的差异,不同模态的医学图像显示的疾病信息不同[1],例如 CT 图像主要显示骨骼、血管和软组织钙化等,MRI 显示高分辨率的软组织结构信息,PET 和 SPECT 图像主要显示组织、器官或病变的功能性信息。每种成像模态都有其局限性和特点,仅凭单模态的图像往往无法提供全面、准确的医学信息。医生通常需要整合来自同一位置的多种模态的图像来进行诊断。如果仅凭医生的空间想象和推测对多种模态的医学图像进行分析,将耗费大量时间和精力,分析精度也会受到主观影响。因此多模态医学图像融合技术应运而生。将不同模态的医学图像整合成一个融合图像,可以弥补各模态的不足,提高图像的对比度、分辨率和空间定位能力,提供更全面、准确和可靠的医学信息,进而提高疾病诊断的效率和准确性[2]。

为了实现多模态医学图像融合,现有研究已经提出了多种融合算法,主要分为传统算法和深度学习算法两大类。传统算法中,常用的方法包括空间域和变换域两种。空间域方法直接对图像像素进行运算,容易造成空间失真。变换域方法对图像多尺度分解后的系数进行处理。因为该过程与人类视觉系统以多分辨率方式处理信息的过程类似[3],所以采用变换域方法产生的融合图像更有助于人眼视觉观察[4],因此受到了学者的广泛研究。Yin 等[5]使用加权局部能量和拉普拉斯算子作为活动水平对分解后得到的低频图像进行融合。Zhu 等[6]结合局部相位一致性、局部突变度量、局部能量信息和局部拉普拉斯能量作为活动水平对分解后得到的多尺度图像进行融合。Dogra 等[7]利用引导图像滤波器和图像统计规则对分解后得到的基层分量进行融合。尽管上述方法通过多尺度分解提高了融合图像的视觉效果,但复杂的特征提取和融合规则都需要人工设计。

基于深度学习的方法可以从大量的数据中自主学习特征表示和融合规则,具有更强的泛化能力和

收稿日期:2024-04-20;**修订日期**:2024-05-16

基金项目:国家自然科学基金资助项目(62206252);郑州市协同创新重大专项(20XTZX06013,20XTZX05015)

作者简介:林予松(1973—),男,四川西昌人,教授,博士,博士生导师,主要从事医学影像与人工智能、互联网新技术研究,E-mail:yslin@ha.edu.cn。

通信作者:赵哲(1983—),女,河南周口人,讲师,主要从事医学影像与人工智能研究,E-mail:sevenzz@zzu.edu.cn。

适应性,同时也能解决因人为主观判断错误导致的特征提取不准确和不完整问题,被广泛应用于图像融合领域。Zhang 等[8]提出了一种端到端的模型 IFCNN,不需要任何后期处理。Li 等[9]利用卷积层和密集块组成的编码器对源图像的显著特征进行提取。Fu 等[10]结合残差注意力和特征金字塔注意力的优势,提出了 MSRPAN 网络,能够更好地捕获源图像的深层特征。然而,由于医学图像融合缺乏金标准,因此无法对融合结果进行有效约束。

生成对抗网络(GAN)通过生成器和鉴别器之间的对抗博弈最小化融合图像与真实图像的概率分布差异[11],从而对融合过程进行约束。近些年,该方法在图像融合领域取得了巨大进步。Ma[12]等提出了 FusionGAN 模型,尽可能地保留具有高分辨率的源图像的信息。Zhao[13]等结合密集块和编码器提出了 DCGAN 模型,进一步加强生成器网络的特征提取能力。上述模型均使用 1 个生成器和 1 个鉴别器进行对抗训练,导致在融合过程中损失了另一源图像的信息。为了解决该问题,Ma 等[14]提出了 DDcGAN 模型,引入 2 个鉴别器分别和 1 个生成器进行对抗,同时保留了 2 个源图像的显著信息。

虽然上述模型通过生成器和鉴别器之间的对抗博弈取得了较好的融合效果,但在图像特征提取过程中没有考虑到人类视觉系统的特点,而是采用单一尺度卷积层提取图像的特征,导致在融合过程中丢失了源图像的多尺度信息,不利于人类视觉观察。为了解决上述问题,本文提出一种基于生成对抗网络和多尺度空间注意力的多模态医学图像融合方法(multiscale spatial attention GAN, MSAGAN)。主要贡献和创新点如下。

(1)为了确保在训练过程中不丢失任何一种模态图像的信息,整个算法框架采用 2 个鉴别器和 1 个生成器进行对抗训练,使生成器生成的融合图像同时保留 2 种源图像的显著特征。

(2)为了捕获源图像的多尺度特征和细节信息,本文构建一个多尺度空间注意力(multiscale spatial attention, MSA)模块作为编码器进行特征提取的基本模块。该模块采用多尺度结构捕获医学图像的多尺度特征,并利用注意力机制更全面地保留图像的细节信息。

1 理论基础

1.1 生成对抗网络

GAN 是一种无监督的图像生成模型,由生成器(G)和鉴别器(D)两部分组成[15]。通过它们之间的对抗训练,GAN 可以估计样本的概率分布并生成新的数据。生成器的目标是根据输入的噪声 z 生成与训练集特征相似的数据 $G(z)$。鉴别器的目标是尽可能区分 $G(z)$ 和真实数据。在训练过程中,生成器和鉴别器交替进行训练,生成器根据鉴别器的判别结果不断更新参数,提高生成能力。鉴别器则跟随生成器的优化不断更新网络参数,以提高自身的判别能力。生成器和鉴别器通过不断地对抗博弈,最终可以达到纳什均衡状态[16]。即无论生成器和鉴别器的网络参数如何调整,鉴别器都无法区分真实数据和生成数据。G 和 D 的对抗过程为

$$\min_G \max_D V_{\mathrm{GAN}}(G,D) = E_{x\sim P_{\mathrm{data}}(x)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z)))]。\quad (1)$$

式中:E 为数据的期望;x 为输入样本;$P_{\mathrm{data}}(x)$ 和 $P_z(z)$ 分别为真实数据和噪声的分布;$D(x)$ 为鉴别器判断真实数据是否真实的概率;$D(G(z))$ 为鉴别器判断生成数据是否真实的概率。

1.2 特征金字塔注意力

Li 等[17]提出了特征金字塔注意力(feature pyramid attention, FPA)模块用于图像分割,该模块结合注意力机制和金字塔结构的优势,提取图像不同尺度的特征并增强获取图像语义上下文信息的能力,原理如图 1 所示。FPA 模块利用金字塔结构融合不同尺度的信息,从而更精确地整合相邻尺度上下文特征。将卷积神经网络提取的原始特征经过 1×1 卷积后与金字塔特征逐像素相乘,并且通过引入全局平均池化分支,进一步提升 FPA 模块的性能。该模块的计算过程可表示为

$$H(x) = (1 + P_1(P_2(P_3(x)))) \times V(x)。\quad (2)$$

式中:$H(x)$ 为输出特征;$V(\cdot)$ 为卷积操作;P_1、P_2 和 P_3 分别为金字塔网络的 3 层卷积操作。

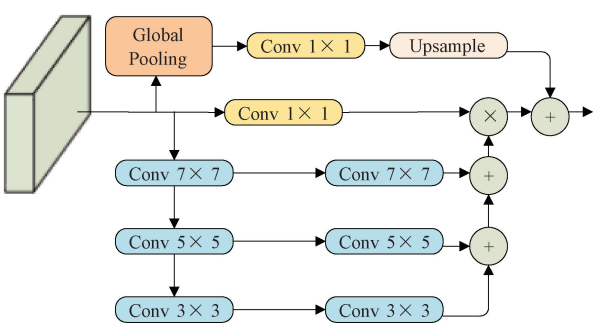

图 1 特征金字塔注意力模块

Figure 1 Feature pyramid attention module

2 MSAGAN 算法

2.1 算法整体框架

本文将多模态医学图像融合问题转化为一个条

件生成对抗模型训练问题,构建一个拥有双鉴别器的生成对抗网络 MSAGAN,实现多模态医学图像融合。整个模型的框架如图 2 所示。

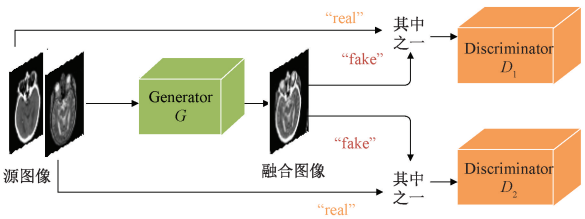

图 2　MSAGAN 整体框架图

Figure 2　Overall framework of MSAGAN

在给定多模态医学图像的情况下,该模型的最终目标是训练生成器 G,使其能够生成一个融合图像,该图像同时包含输入的两种模态图像的特征。生成器的主要任务是对输入的图像进行特征提取、融合和重建,生成融合图像。鉴别器由 D_1 和 D_2 两部分组成,每个鉴别器都用于判别生成器生成的融合图像和其中之一的真实参考图像,并将结果反馈给生成器以更新网络参数。通过两个鉴别器与生成器之间的对抗训练,确保生成器生成的融合图像能够同时保留输入的两种模态图像的显著特征。G 和 D 的对抗关系为

$$\min_{G}\max_{D_1,D_2}\{E[\log D_1(x_1)]+E[\log(1-D_1(G(x_1,x_2)))]+E[\log D_2(x_2)]+E[\log(1-D_2(G(x_1,x_2)))]\}。\tag{3}$$

生成器和鉴别器交替进行训练。通过生成器和 2 个鉴别器之间的对抗过程,生成器所生成的融合图像 $G(x_1,x_2)$ 的概率分布与两个真实图像 x_1、x_2 的概率分布越来越相似,从而使生成的融合图像能够同时保留输入的多模态医学图像的特征信息。模型整体的训练流程如下所示。

算法 1　MSAGAN 训练流程

参数描述:生成器 G、鉴别器 D_1 和 D_2 的训练次数分别为 I_G、I_{D_1} 和 I_{D_2};$I_{\max}$ 为网络迭代训练的最大次数;$L_{\max}$ 和 $L_{\min}$ 分别为生成器和鉴别器对抗损失的最大和最小值;$L_{G_{\max}}$ 为生成器的总损失。

Step1　初始化网络参数;

Step2　采样 m 对多模态医学图像 $\{I_1^1,I_1^2,\cdots,I_1^m\}$ 和 $\{I_2^1,I_2^2,\cdots,I_2^m\}$;

获得生成的融合图像 $\{G(I_1^1,I_1^1),G(I_1^2,I_2^2),\cdots,G(I_1^m,I_2^m)\}$;

使用优化器更新鉴别器 D_1 的网络参数,用最小化鉴别器对抗损失函数;

Step3　使用优化器更新鉴别器 D_2 的网络参数,用最小化鉴别器对抗损失函数;

若 $L_{D_1}>L_{\max}$ 且 $I_{D_1}<I_{\max}$,重复 Step2,$I_{D_1}+1\rightarrow I_{D_1}$;

若 $L_{D_2}>L_{\max}$ 且 $I_{D_2}<I_{\max}$,重复 Step3,$I_{D_2}+1\rightarrow I_{D_2}$;

Step4　采样 m 对多模态医学图像 $\{I_1^1,I_1^2,\cdots,I_1^m\}$ 和 $\{I_2^1,I_2^2,\cdots,I_2^m\}$;

获得生成的融合图像 $\{G(I_1^1,I_2^1),G(I_1^2,I_2^2),\cdots,G(I_1^m,I_2^m)\}$;

使用优化器更新生成器 G 的网络参数,以最小化生成器损失函数;

若 $L_{D_1}<L_{\min}$ 或 $L_{D_2}<L_{\min}$,并且 $I_G<I_{\max}$,更新生成器网络参数,用最小化生成器对抗损失函数,$I_G+1\rightarrow I_G$;

若 $L_G>L_{G_{\max}}$ 且 $I_G<I_{\max}$,重复 Step4,$I_G+1\rightarrow I_G$。

2.2　生成器和鉴别器

生成器 G 由一个编码器(Encoder)和一个解码器(Decoder)组成,其网络结构如图 3 所示。首先,将 2 种模态图像在通道维度上连接的结果作为编码器的输入;其次,通过编码器进行特征提取,输出融合特征图;最后,融合特征图输入解码器进行重建,生成同时包含两种模态图像特征的融合图像。编码器由 5 个 MSA 模块(如图 4 所示)组成,每个模块通过整合不同尺度的特征获得 48 个特征图。为了避免丢失中间模块的特征,实现特征重用,编码器采用 DenseNet,使每个模块与后续所有模块相连,尽可能保留图像的深层特征。解码器采用 5 层 CNN 结构。为了减少源图像信息损失,所有卷积层步长都设为 1,并且采用批归一化和 Relu 激活函数加快训练速度,最后 1 层使用 tanh 作为激活函数,输出融合图像。

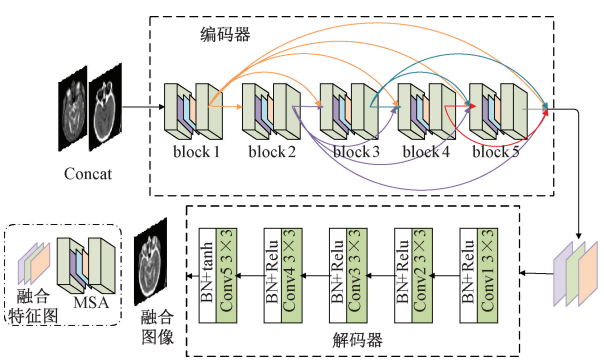

图 3　生成器结构图

Figure 3　Generator structure diagram

受 FPA 模块在图像分割应用中的启发,为解决医学图像融合过程中单一尺度卷积层特征提取不充

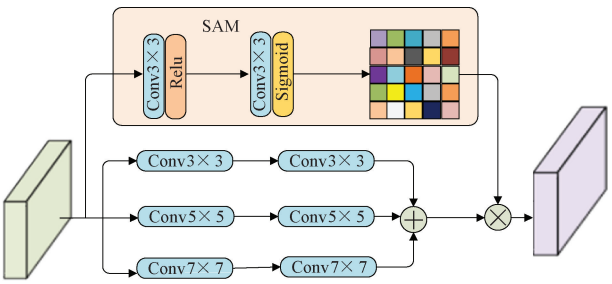

图 4　MSA 结构图

Figure 4　MSA structure diagram

分的问题,本文构建了多尺度空间注意力(MSA)模块,并把该模块作为编码器进行特征提取的基本模块,更好地保留图像的多尺度特征和细节信息。该模块由多尺度结构和空间注意力模块(spatial attention module, SAM)[18]两部分组成。

(1)多尺度结构。本文采用 3 种卷积尺度获取图像的多尺度特征。模块输入的特征图分别经过 3×3、5×5 和 7×7 的滤波器,这种多尺度结构可以帮助网络从不同的尺度提取特征,以获得更全面的信息表示。

(2)SAM。本文利用空间注意力获取图像全局信息的相关性和依赖关系,提高网络对关键信息的表达能力,全面捕获医学图像的细节信息。SAM 由 2 个 3×3 卷积层组成,输出 1 个单通道掩膜,用于强调空间中信息量更大的特征。

最后,多尺度结构整合后的多尺度特征和 SAM 得到的掩膜进行逐像素相乘,以获得 MSA 模块最终输出的特征。

2 个鉴别器 D_1 和 D_2 网络结构相同,如图 5 所示。每个鉴别器包含 4 层网络模型,前 3 层网络由卷积核大小为 3×3 的网络组成。为了和生成器对应,第 1 层网络仅使用 Relu 激活函数,其他 2 层包括批归一化和 Relu 激活函数,最后 1 层为全连接层,使用 tanh 激活函数生成 1 个标量,用于估计输入图像来自真实图像而不是生成器生成图像的概率。

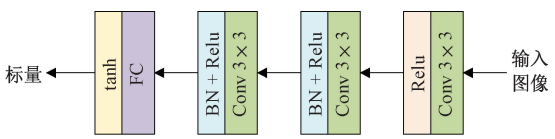

图 5　鉴别器结构图

Figure 5　Discriminator structure diagram

2.3　损失函数设计

本文所提 MSAGAN 网络的损失函数由生成器损失函数和鉴别器损失函数组成:

$$L = L_G + L_{D_1} + L_{D_2}。\tag{4}$$

式中:L_G 为生成器损失函数;L_{D_1} 和 L_{D_2} 分别为 2 个鉴别器的损失函数。

对生成器 G 而言,除了根据鉴别器的结果对其进行对抗训练,本文还引入内容损失约束生成图像和真实图像之间的相似性,解决 GAN 训练不稳定的问题。因此 G 的损失函数由对抗损失和内容损失两部分组成:

$$L_G = L_G^{adv} + \lambda L_{con};\tag{5}$$

$$L_G^{adv} = E[\log(1 - D_1(G(x_1, x_2)))] + E[\log(1 - D_2(G(x_1, x_2)))];\tag{6}$$

$$L_{con} = E[\ \| G(x_1, x_2) - x_1 \|_F^2 + \varphi \| G(x_1, x_2) - x_2 \|_{TV}]。\tag{7}$$

式中:L_G^{adv} 为生成器的对抗损失函数;L_{con} 为内容损失函数;λ 和 φ 用于平衡不同的损失函数,分别取 0.6 和 1.2;x_1 表示 PET、CT 或 SPECT 图像,这些影像中病变的亮度信息主要通过像素强度进行表示;x_2 表示 MRI 图像,主要显示病灶的纹理细节等信息。本文根据不同模态图像的特点,分别采用 Frobenius 范数和 TV 范数表示内容损失,使生成器生成的融合图像同时保留不同模态医学图像的显著特征。

两个鉴别器 D_1 和 D_2 的损失函数由具有代表性的对抗损失表示,分别为

$$L_{D_1} = E[-\log D_1(x_1)] + E[-\log(1 - D_1(G(x_1, x_2)))];\tag{8}$$

$$L_{D_2} = E[-\log D_2(x_2)] + E[-\log(1 - D_2(G(x_1, x_2)))]。\tag{9}$$

3　实验结果与分析

3.1　数据集及实验设置

本文所用实验数据来自哈佛大学全脑图谱数据库[19]。为了验证本文所提算法 MSAGAN 对不同类型融合任务的有效性,从该数据库中分别选取了 124 对 CT-MRI、145 对 PET-MRI 和 117 对 SPECT-MRI 这 3 种类型的多模态医学图像融合任务进行实验,分别将其中的 104 对、125 对和 97 对图像作为训练集,剩余的 20 对图像作为测试集。所有源图像大小均为 256×256 像素,且每对图像都经过了精确配准。为了扩充训练集,将训练集图像裁剪为 84×84 像素大小的图像块。在训练过程中,网络使用 Adam 优化器,初始学习率为 0.000 2,衰减系数为 0.9,批量大小为 24。网络模型基于 TensorFlow1.14.0,编程语言为 Python3.7,实验环境为 64 位 Windows10 操作系统的台式电脑,硬件处理器为 Intel

Core i5-8500 CPU@3.00 GHz,RAM 为 16 GB。

3.2 实验结果分析

为了验证 MSAGAN 的融合效果,从主观和客观两个方面对实验结果进行评价。将 MSAGAN 与医学图像融合领域较为经典和先进的 7 种算法进行对比,这些算法包括 PA-PCNN[5]、LRD[20]、U2Fusion[21]、MSRPAN[10]、EMFusion[22]、DDcGAN[14]、GeSeNet[23]。

3.2.1 主观评价

从每种类型的医学图像融合任务中选取两对具有代表性的融合结果进行展示。图 6 为 CT 和 MRI 图像融合结果,可以明显看出,基于 U2Fusion 和 DDcGAN 的融合图像损失了 CT 图像部分的能量信息,导致图像对比度下降,不利于人眼观察。基于 LRD 和 MSRPAN 的融合图像虽然视觉对比度较好,但都在一定程度上损失了源图像的边缘纹理等细节信息,并且存在伪影。进一步观察 PA-PCNN 融合图像的骨骼区域,可以看到有边缘信息丢失的现象。与 EMFusion 相比,基于 GeSeNet 和 MSAGAN 的融合图像不仅具有更高的全局对比度和亮度,而且在保留 MRI 图像结构纹理等特征的基础上更好地保留了 CT 图像的能量信息。

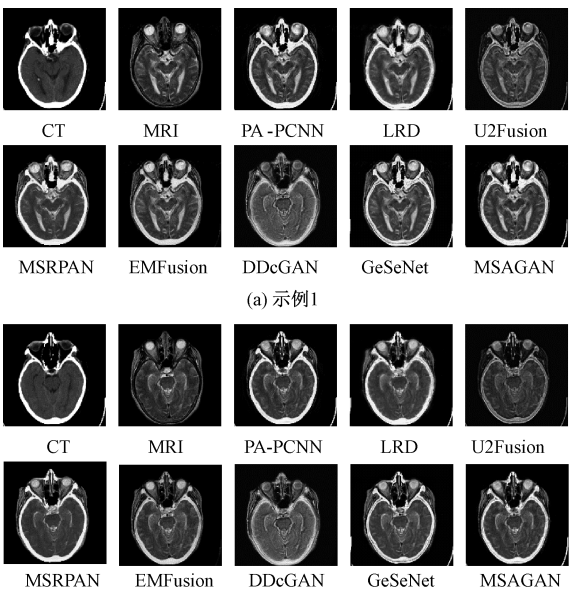

图 6 CT-MRI 融合结果

Figure 6 CT-MRI fusion results

图 7 和图 8 分别为 PET-MRI 和 SPECT-MRI 的融合结果,两者均表示功能图像和结构图像的融合。功能信息主要包含在 PET 或 SPECT 图像中。可以看出,除了 LRD 方法,其他方法都较好地保留了功能图像的颜色信息,主要区别在于 MRI 纹理细节信息的保留情况。观察发现,基于 MSRPAN 方法的融合图像较为模糊,基于 PA-PCNN 和 GeSeNet 的融合图像均有不同程度的细节损失。U2Fusion 和 EMFusion 方法引入了噪声,导致图像原始结构信息被破坏,基于 DDcGAN 方法获得的融合图像亮度较低。相比之下,基于 MSAGAN 获得的融合图像在纹理等梯度特征上更加明显,更利于人眼观察。

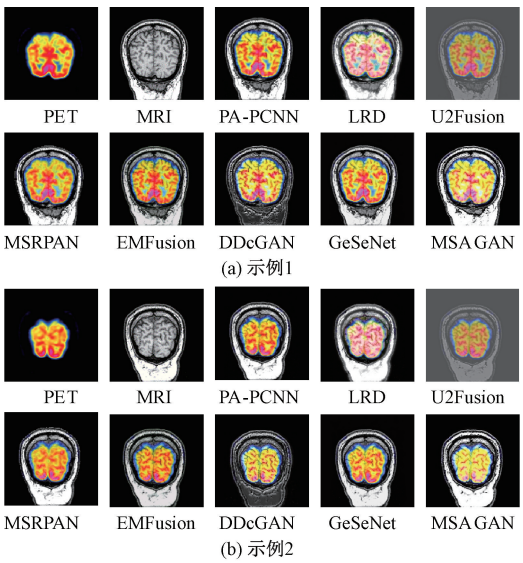

图 7 PET-MRI 融合结果

Figure 7 PET-MRI fusion results

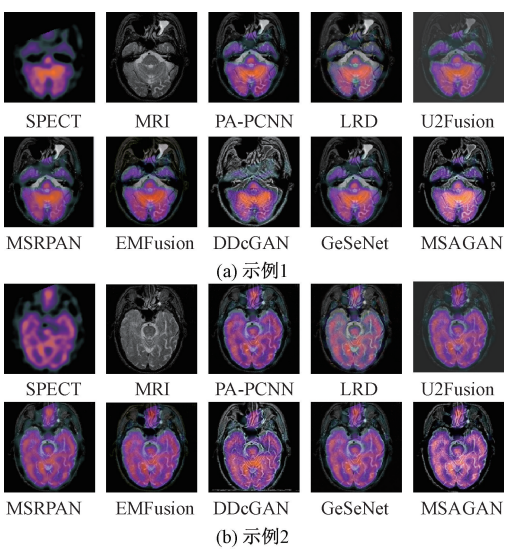

图 8 SPECT-MRI 融合结果

Figure 8 SPECT-MRI fusion results

3.2.2 客观评价

本文选取 6 种常见的图像融合评价指标[2,24],从多个角度对不同算法的融合效果进行客观评价。这些指标包括平均梯度(*AG*)、相关系数(*CC*)、信息熵(*EN*)、互信息(*MI*)、峰值信噪比(*PSNR*)、视觉信息保真度(*VIF*)。在 3 种类型医学图像融合任务中,信息使用每个方法在测试集上的平均值进行评估,指标的值越高,表示融合效果越好。此外,本文还引入运行时间评估不同算法的效率,时间单位为 s,运行时间

越少,表示算法性能越好。评价结果如表1所示。

由表1可知,本文算法MSAGAN在*AG*、*PSNR*、*VIF*和*MI*这4项指标整体上取得了最优。在3种融合任务上*AG*和*VIF*的平均值分别达到0.302 3和0.649 8,较高的*AG*和*VIF*表明融合图像不仅具有更强的细节表达能力,而且有助于人眼观察,这是由于本文所提算法采用多尺度结构进行图像特征的提取,符合人眼视觉感知的特点。在3种融合任务上*MI*和*PSNR*的平均值分别达到1.441 4和20.720 7,较高的*MI*和*PSNR*值表示在融合过程中引入的噪声较少,融合图像与原始图像之间的失真较小。这意味着融合图像在保持原始图像的整体结构和像素方面表现良好,这是由于本文采用空间注意力能够更全面地保留医学图像的结构和细节信息。整体而言,客观评价指标表明,MSAGAN算法能够提供更高质量的融合结果。

表1 融合图像客观评价指标平均值

Table 1 Average objective evaluation metrics for fused images

数据集	方法	*AG*	*CC*	*EN*	*MI*	*PSNR*	*VIF*	运行时间/s
CT-MRI	PA-PCNN	0.253 0	0.753 5	5.081 6	0.889 0	21.456 6	0.588 6	3.47
	LRD	0.224 6	0.757 6	4.895 4	0.897 2	21.194 2	0.585 1	200.98
	U2Fusion	0.194 2	0.778 1	4.991 5	0.924 7	15.714 0	0.259 5	0.57
	MSRPAN	0.208 9	0.746 0	4.164 7	1.171 5	22.097 8	0.572 8	0.30
	EMFusion	0.204 7	0.776 1	4.905 4	1.150 4	21.679 7	0.425 3	0.20
	DDcGAN	0.210 5	0.773 1	5.335 5	0.997 7	21.288 0	0.298 9	0.52
	GeSeNet	0.263 6	**0.795 1**	**5.386 3**	1.101 7	23.902 4	0.572 5	**0.02**
	MSAGAN	**0.274 5**	0.784 4	5.345 5	**1.192 8**	**24.066 1**	**0.593 6**	0.53
PET-MRI	PA-PCNN	0.264 4	0.738 7	4.323 5	1.663 9	15.878 4	**0.688 2**	3.50
	LRD	0.211 0	0.747 0	5.044 7	1.313 9	18.008 3	0.638 3	197.13
	U2Fusion	0.090 2	**0.831 5**	4.045 8	1.165 6	19.089 5	0.117 7	0.59
	MSRPAN	0.239 0	0.734 1	3.846 0	1.632 9	15.591 6	0.595 9	0.31
	EMFusion	0.242 4	0.747 8	4.555 7	1.491 1	13.729 2	0.530 5	0.17
	DDcGAN	0.284 1	0.714 7	4.991 7	1.375 1	15.848 6	0.296 8	0.51
	GeSeNet	0.264 3	0.754 0	**5.132 0**	1.433 9	16.181 5	0.679 5	**0.02**
	MSAGAN	**0.342 5**	0.794 1	4.927 9	**1.679 7**	**19.899 8**	0.635 3	0.52
SPECT-MRI	PA-PCNN	0.159 2	0.851 4	5.450 4	1.600 8	14.385 5	0.642 8	3.63
	LRD	0.142 3	0.852 0	**5.963 5**	1.390 2	14.733 1	0.616 8	201.69
	U2Fusion	0.080 7	**0.863 9**	4.678 5	1.141 5	14.283 7	0.242 3	0.57
	MSRPAN	0.124 1	0.845 2	4.850 9	**1.715 1**	14.133 1	0.481 4	0.31
	EMFusion	0.142 0	0.850 4	5.301 6	1.583 1	13.331 0	0.539 9	0.17
	DDcGAN	0.280 8	0.642 1	5.524 6	1.362 9	16.615 1	0.518 5	0.52
	GeSeNet	0.163 3	0.856 0	5.903 1	1.582 6	14.347 2	0.649 8	**0.02**
	MSAGAN	**0.289 9**	0.857 5	5.355 2	1.451 7	**18.196 2**	**0.720 4**	0.54

在医学领域,图像的细节信息对于精确诊断和治疗至关重要。在这个问题上,MSAGAN算法表现出较好的融合效果。尽管运行时间与其他深度学习方法相比略有增加,但考虑到其能够提供更清晰、更细致的融合图像,为医生提供更准确和全面的诊断信息,这个额外的时间成本是完全可以接受的。

3.2.3 消融实验

为了验证所提出的MSAGAN方法的有效性,本文分别针对其中的多尺度结构和SAM开展消融实验。以CT-MRI数据集为例,消融实验结果如表2所示。

表2 CT-MRI数据集上的消融实验结果

Table 2 Results of ablation experiments on the CT-MRI dataset

多尺度	SAM	*AG*	*CC*	*EN*	*MI*	*PSNR*	*VIF*
×	×	0.210 5	0.773 1	5.335 5	0.997 7	21.288 0	0.298 9
√	×	0.242 6	0.782 1	5.338 9	1.094 8	23.146 1	0.497 2
×	√	0.247 3	0.779 4	5.342 3	1.128 4	23.154 0	0.485 6
√	√	**0.274 5**	**0.784 4**	**5.345 5**	**1.192 8**	**24.066 1**	**0.593 6**

表 2 结果显示,加入多尺度结构可以帮助网络更好地捕捉全局和局部特征,提高对图像的理解能力和表达能力。此外,引入 SAM 可以自适应地调整不同病变区域的权重,提高融合结果对关注区域的准确性和清晰度,进一步提升融合图像质量。本文提出的 MSAGAN 方法同时引入多尺度结构和 SAM,在 6 种评价指标上都表现出了最好的效果,验证了 MSAGAN 方法的有效性。

4 结论

本文提出了一种基于 GAN 和多尺度空间注意力的多模态医学图像融合算法 MSAGAN。整个网络框架包含 1 个生成器和 2 个鉴别器,通过它们之间的对抗博弈,融合图像能够同时保留 2 个输入图像的显著特征。为了更好地捕捉源图像的多尺度特征和细节信息,本文构建了 1 个 MSA 模块作为编码器的基本模块。该模块采用多尺度结构,能更有效地捕捉图像的多尺度特征;此外,该模块还引入空间注意力,能更好地保持原始图像的结构和细节信息。实验结果表明:基于 MSAGAN 方法生成的融合图像边缘纹理更加清晰,并且在 3 种不同类型的医学图像融合任务上 *AG*、*PSNR*、*MI*、*VIF* 这些客观评价指标的平均值分别达到 0.302 3、20.720 7、1.441 4、0.649 8,融合图像的质量更高。现有的多模态医学图像融合数据集较少,后续研究将考虑利用迁移学习的思想扩充数据集,更好地训练网络。

参考文献:

[1] HUANG B, YANG F, YIN M X, et al. A review of multimodal medical image fusion techniques[J]. Computational and Mathematical Methods in Medicine, 2020, 2020: 8279342.

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

[3] PIELLA G. A general framework for multiresolution image fusion: from pixels to regions[J]. Information Fusion, 2003, 4(4): 259-280.

[4] GUO P, XIE G Q, LI R F, et al. Multimodal medical image fusion with convolution sparse representation and mutual information correlation in NSST domain[J]. Complex & Intelligent Systems, 2023, 9(1): 317-328.

[5] YIN M, LIU X N, LIU Y, et al. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(1): 49-64.

[6] ZHU Z Q, ZHENG M Y, QI G Q, et al. A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain[J]. IEEE Access, 2019, 7: 20811-20824.

[7] DOGRA A, KUMAR S. Multi-modality medical image fusion based on guided filter and image statistics in multidirectional shearlet transform domain[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(9): 12191-12205.

[8] 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.

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

[10] FU J, LI W S, DU J, et al. A multiscale residual pyramid attention network for medical image fusion[J]. Biomedical Signal Processing and Control, 2021, 66: 102488.

[11] 许光宇, 陈浩宇, 张杰. 双路径双鉴别器生成对抗网络的红外与可见光图像融合[J/OL]. 计算机辅助设计与图形学学报, 1-14[2024-04-07]. http://kns.cnki.net/kcms/detail/11.2925.TP.20240204.1728.061.html.
XU G Y, CHEN H Y, ZHANG J. Infrared and visible image fusion based on dual-path and dual-discriminator generation adversarial network [J/OL]. Journal of Computer-Aided Design & Computer Graphics, 1-14[2024-04-07]. http://kns.cnki.net/kcms/detail/11.2925.TP.20240204.1728.061.html.

[12] MA J Y, YU W, LIANG P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11-26.

[13] ZHAO C, WANG T F, LEI B Y. Medical image fusion method based on dense block and deep convolutional generative adversarial network[J]. Neural Computing and Applications, 2021, 33(12): 6595-6610.

[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, 29: 4980-4995.

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

[16] 肖儿良, 林化溪, 简献忠. 基于生成对抗网络探索潜在空间的医学图像融合算法[J]. 信息与控制, 2021, 50(5): 538-549.
XIAO E L, LIN H X, JIAN X Z. Medical image fusion

algorithm adopting generative adversarial network to explore latent space[J]. Information and Control, 2021, 50(5): 538-549.

[17] LI H C, XIONG P F, AN J, et al. Pyramid attention network for semantic segmentation[EB/OL]. (2018-11-25)[2024-04-07]. https://arxiv.org/abs/1805.10180.

[18] LIU Y, WANG L, LI H F, et al. Multi-focus image fusion with deep residual learning and focus property detection[J]. Information Fusion, 2022, 86: 1-16.

[19] 尹海涛, 岳勇赢. 基于半监督学习和生成对抗网络的医学图像融合算法[J]. 激光与光电子学进展, 2022, 59(22): 245-254.
YIN H T, YUE Y Y. Medical image fusion based on semisupervised learning and generative adversarial network[J]. Laser & Optoelectronics Progress, 2022, 59(22): 245-254.

[20] LI X X, GUO X P, HAN P F, et al. Laplacian redecomposition for multimodal medical image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(9): 6880-6890.

[21] XU H, MA J Y, JIANG J J, et al. U2Fusion: a unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 502-518.

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

[23] LI J W, LIU J Y, ZHOU S H, et al. GeSeNet: a general semantic-guided network with couple mask ensemble for medical image fusion[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023,1: 14.

[24] 刘帅奇, 王洁, 安彦玲, 等. 基于CNN的非下采样剪切波域多聚焦图像融合[J]. 郑州大学学报(工学版), 2019, 40(4): 36-41.
LIU S Q, WANG J, AN Y L, et al. Multi-focus image fusion based on CNN in non-sampled shearlet domain[J]. Journal of Zhengzhou University (Engineering Science), 2019, 40(4): 36-41.

Multimodal Medical Image Fusion Based on GAN and Multiscale Spatial Attention

LIN Yusong[1,2,3], LI Mengya[1,2], LI Yinghao[1,2], ZHAO Zhe[1,2]

(1. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China; 2. Henan Provincial Collaborative Innovation Center for Internet Medical and Health Services, Zhengzhou University, Zhengzhou 450052, China; 3. Hanwei IoT Institute, Zhengzhou University, Zhengzhou 450002, China)

Abstract: Aiming to address the problem of multi-scale feature and texture detail information loss in the process of multimodal medical image fusion, a novel image fusion algorithm based on generative adversarial network (GAN) and multi-scale spatial attention mechanism was proposed. Firstly, the generator adopted an autoencoder structure to extract, fuse, and reconstructed the input images using an encoder and a decoder, generating the fused image. Secondly, the entire GAN framework employed a dual discriminator structure, enabling the generator to preserve salient features from multiple modal images in the fused image. Finally, a multi-scale spatial attention mechanism was constructed as a fundamental module for feature extraction in the encoder. It effectively captured and retained multi-scale features from the source images, and incorporated spatial attention mechanism to better preserve the structures and details of the source images. Experimental results on the Whole Brain Atlas database demonstrated that the fused images generated by the proposed algorithm exhibit richer texture details, enhancing human visual observation. Furthermore, the algorithm outperformed other advanced algorithms in such objective evaluation metrics as average gradient, peak signal-to-noise ratio, mutual information, and visual information fidelity for three different types of medical image fusion tasks, with average values of 0.302 3, 20.720 7, 1.441 4, and 0.649 8, respectively. Thus, the proposed algorithm demonstrated a certain advantage over other advanced algorithms.

Keywords: image fusion; multimodal medical images; generative adversarial network; feature pyramid; attention mechanism