

基于条件可逆神经网络的多模态医学图像融合

张倩楠, 王蒙

(昆明理工大学 信息工程与自动化学院, 昆明 650500)

摘要: 为减少图像融合在深层网络前馈过程中的空间细节损耗, 提出一种基于条件可逆神经网络 (CINN) 的医学图像融合方法; 通过应用可逆的分析-综合架构实现空间细节与关键语义互补的多模态融合; 在前向分析阶段, 将多分辨率特征嵌入 CINN 作为条件向量实现多模态表示学习; 在反向综合阶段, 使用一个基于小波的条件融合 (WCF) 网络引导 CINN 完成反向重构; 在特征融合中应用相关激活模板 (RAM), 聚焦多模态医学图像中的关键结构区域与纹理细节一致性信息; 构造前向分析-反向重构联合损失高效地优化网络参数, 以获得高质量的融合图像; 实验测试了 CT-MRI 及 MRI-PET 场景, 与现有融合基线相比, 提出方法在 SCD 以及 VIFF 等客观融合指标上性能分别提升了 15.16% 和 46.53%, 并且在主观视觉质量上均取得了优越的结果。

关键词: 图像融合; 多模态医学图像; 条件可逆神经网络; 深度生成网络; 多分辨率分析

Multi-modal Medical Image Fusion Based on Conditional Reversible Neural Networks

ZHANG Qiannan, WANG Meng

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: In order to reduce the spatial detail loss of image fusion in the feedforward process of deep networks, a medical image fusion method based on conditional invertible neural network (CINN) is proposed. Through the application of reversible analysis-synthesis architecture, the multi-modal fusion of spatial details and key semantic complementarity is realized. In the forward analysis phase, multi-resolution features are embedded into CINN as conditional vectors to achieve multi-modal representation learning. In the reverse synthesis stage, a wavelet-based conditional fusion (WCF) network is used to guide CINN to complete the reverse reconstruction. In feature fusion, the related activation model (RAM) is applied to focus on the consistency information of key structural regions and texture details in multi-modal medical images. The combination loss of forward analysis and reverse reconstruction is constructed to efficiently optimize the parameters of the network and to obtain high quality fusion images. Experimental testing on the CT-MRI and MRI-PET scenes, The results show that compared with the existing fusion baseline, the proposed method improves the fusion indicators of spatial color distribution (SCD) and visual information fidelity (VIFF) by 15.16% and 46.53%, respectively, achieving superior performance in subjective visual quality.

Keywords: image fusion; multi-modal medical image; conditional invertible neural network; deep generative networks; multi-resolution analysis

0 引言

近年来, 随着医学图像成像技术不断发展, 一些前沿成果有效改善了疾病诊断的质量。典型医学图像包

括: 磁共振成像 (MRI, magnetic resonance imaging)、计算机断层扫描 (CT, computed tomography) 和正电子发射计算机断层显像 (PET, positron emission tomography) 等。其中 CT 图像能够清晰地显示骨骼、钙化

收稿日期: 2023-11-10; 修回日期: 2023-12-20。

基金项目: 中国国家自然科学基金(62062048); 云南省科技计划(202201AT070113)。

作者简介: 张倩楠(1995-), 女, 硕士研究生。

通讯作者: 王蒙(1981-), 男, 博士, 硕导, 副教授。

引用格式: 张倩楠, 王蒙. 基于条件可逆神经网络的多模态医学图像融合[J]. 计算机测量与控制, 2025, 33(1): 147-154, 162.

灶等硬组织结构, MRI 图像具有良好的软组织对比度, PET 图像能够反映人体内部分子水平上的代谢活动和功能状态。通过融合不同模态医学图像可以综合利用各种图像的优势、弥补不足,从而获得更全面的解剖结构、病变信息和功能信息,帮助医生做出更准确的诊疗决策。目前,图像融合在多个计算机视觉应用领域发挥着重要的价值,如监测^[1]、遥感^[2]、医学辅助^[3]等。近年来,随着深度学习技术的成功应用,基于卷积神经网络^[4](CNN, convolutional neural network)、生成对抗网络^[5](GAN, generative adversarial network)等深度生成网络的图像融合架构也得到了充分研究,如 CNN 通过一系列卷积层和池化层来逐步学习图像的空间表示,GAN 通过对抗训练生成逼真的图像。然而,这些深层网络架构在前馈过程中通常存在图像空间细节损耗的问题,导致最终生成图像质量下降。

迄今为止,根据模型结构可将融合方案分为基于浅层解析模型、基于深层网络的卷积神经网络(CNN, convolutional neural network)、生成对抗网络(GAN, generative adversarial network)等几大类,这些类别之间对源图像进行分析—重构的处理明显不同,并且采用不同的训练及优化策略也会影响最终输出。在文献[6]中阐述了早期为将双筒望远镜图像通过拉普拉斯金字塔(LP, laplacian pyramid)方法进行融合,首次提出了基于多分辨率分解的图像融合方法,能够同时保留两幅图像中的高频细节信息和低频背景信息。随后文献[7]中作者等人提出了在图像融合中通过引入小波分析,将图像分解为不同尺度和方向的小波系数并对应系数进行融合,最后逆变换得到融合图像。通常浅层模型解析的分析—重构处理较为简单,对复杂的数据分布及非线性变换关系的拟合能力较弱,因此近年来引入了更复杂的深度学习模型。

深度学习模型能够通过多层神经网络学习数据的深层次表示,从而有效地提高复杂模型的泛化能力及预测性能。其应用涵盖了图像识别到最近的图像生成与重构等,例如文献[8]提出用于图像融合的统一密集连接网络(FusionDNN),通过保持训练密度和强度比例实现快速统一的图像融合网络。文献[9]提出了基于卷积神经网络的图像融合模型(IFCNN),融合多个输入图像的卷积特征以重构融合图像。文献[10]介绍了卷积稀疏表示在图像融合中的应用,将反卷积网络设计成结构化的层次。生成对抗网络(GAN, generative adversarial network)则是通过引入生成器和判别器的对抗训练机制,在保证细节融合的同时,对全局语义的生成提供有效的约束与指导。文献[11]中作者将GAN应用于医学图像融合,通过使用GAN模型来生成融合图像实现了高质量的医学图像融合,并提高了图像的清晰度和细节保留。

文献[12]中介绍了一种基于编码—解码器结构的生成对抗网络(GAN)的医学图像融合方法。其中编码器用于提取输入图像的特征表示,而解码器则用于将这些特征表示转换为融合图像。然而,以上基于CNN和GAN的融合方案,源图像在网络的前馈过程中空间细节会受到一定损耗,还可能受到噪声的干扰,导致融合结果在空间细节和高层语义上容易出现一定程度的退化。因此,解决这些固有问题成为深度图像融合研究的重要目标之一。

本文针对上述基于深度神经网络的多模态医学图像融合存在的问题,如在深层网络前馈过程中的空间细节损耗过多等问题,尝试提出一种新的融合架构。文献[13]中提出了可逆卷积生成流(Glow),其中使用了可逆的 1×1 卷积层和池化层。文献[14]提出了一种基于INN的残差网络结构,其中使用了可逆的残差块,实现了可逆的前向和反向传播。随后文献[15]中作者等人提出了条件可逆神经网络(CINN, conditional invertible neural network),在可逆网络的基础上引入了条件机制,将输入的条件信息与可逆层的计算相结合。这使得CINN能够根据输入的条件信息生成具有个性化特征的输出。受此启发,本文尝试引入条件可逆神经网络(CINN)作为多模态医学图像融合的基础模块,利用可逆的分析—综合网络架构来减少前馈网络的空间特征退化问题。提出方法(FusionCINN)的主要特点包括:

- 1) 利用条件可逆神经网络(CINN)实现空间细节损耗最小化,通过提取源图像的互补特征,处理多模态数据,并通过构造前向分析—反向重构联合损失函数,高效地优化模型参数,从而有效提升融合图像质量。

- 2) 在前向分析过程中,将多分辨率特征提取作为嵌入条件向量来实现条件可逆神经网络(CINN)的前向表示学习。在反向重构过程中,使用一个基于小波的条件融合(WCF)网络嵌入并引导CINN完成反向重构。

- 3) 在特征融合中,应用相关激活模板(RAM),通过引入注意力机制使特征融合能更准确地聚焦于多模态医学图像中的关键结构区域以及纹理细节一致性信息。

1 条件可逆神经网络(CINN)

可逆神经网络(INN, invertible neural networks)是一种基于流的生成模型,起源于NICE(非线性独立成分估计)^[16]和RNVP(实值非体积保持)^[17]的结合。INN利用可逆特性,通过最大化生成逆过程来提高生成能力和与真实数据的匹配度。它能在正向反向传播中保持输入输出对应,还原输入而不丢失信息。INN的

设计基于可逆函数概念, 通过堆叠多个可逆层实现, 每层包含正向反向传播子层, 确保输入和输出完全对应。其全局映射由一系列变换函数组成, 这样 I 和 z 之间的关系可以表示为:

$$I \leftrightarrow h_1 \xrightarrow{f_1} h_2 \xrightarrow{f_2} \dots \xrightarrow{f_n} z \quad (1)$$

之后, 为增强对处理过程的控制, 条件可逆神经网络 (CINN, conditional invertible neural networks) 在可逆网络的基础上引入了条件机制, 将输入条件信息与可逆层的计算相结合^[15]。这使得网络能够根据输入的条件信息生成具有个性化特征的输出。条件可逆神经网络的正向生成是将输入图像 I 和条件 C 输入到函数 f 中得到潜在变量 z 。由于函数是可逆的, 其反向生成是从标准正态分布中采样潜在变量 z , 该分布的均值为 0, 方差为 1。然后, 通过将潜在变量 z 和条件 c 输入到函数 f 的逆函数中, 计算得到 I , 其公式可以写为:

$$z = f_\theta(I; C) = g_\theta^{-1}(I; C) \quad (2)$$

$$I = g(\theta)(z; C) \quad (3)$$

为了确定模型参数, 通常使用最大似然估计方法。目标是选择参数的值 θ , 使得给定观测数据的条件 C , 潜在变量 z 的概率密度函数能够最大化。通过最大化似然函数, 获得最佳的参数估计结果 θ^* , 从而更好地描述潜在变量的分布情况。 θ^* 的定义如下公式 (4) 所示:

$$\theta^* = \arg \max_{\theta} p(\theta) \prod_{i=1}^N p_{\theta}(I_i; C_i) \quad (4)$$

上述 CINN 前向特征分析的变换过程是可逆的, 这意味着可以根据融合结果指导还原原始的特征潜在变量分布。这种可逆性也允许在融合过程中通过前向或后向网络传播在不同的特征表示域之间进行变换, 从而灵活地处理不同模态医学图像中的显著空间特征。给定数据点 x 的模型的概率密度函数可写为如下公式 (5) 所示:

$$\log p_{\theta}(I_i; C_i) = \sum_{i=1}^N \log p(f_{\theta}(I_i; C_i)) + \log \left| \det \left(\frac{\partial (f_{\theta}(I_i; C_i))}{\partial I_i} \right) \right| \quad (5)$$

具体来说, 给定一个概率密度函数, CINN 的损失函数可以表示为:

$$L(\theta) = - \sum_i \log p_{\theta}(I_i | C_i) - \log p_{\theta}(I_i) \quad (6)$$

其中: θ 表示 CINN 的模型参数, I_i 表示输入图像样本, C_i 表示对应的条件信息。

在训练过程中, CINN 会通过反向传播更新模型参数, 使得生成的图像样本与真实图像样本之间的似然度最大化。通过这种方式, CINN 可以学习到输入条件信息与输出图像之间的复杂映射关系, 生成逼真的图像样本。综上, 尝试将条件可逆神经网络引入多模态医学图像融合, 可生成保留更多原始细节信息与空间语义的融合图像。

2 基于条件可逆网络的多模态医学图像融合

2.1 提出网络结构

本文旨在设计一种基于 CINN 的多模态医学图像融合算法, 以获得具备丰富的高层语义信息和清晰的空间细节的优质融合图像。如图 1 所示, 本文提出的 FusionCINN 由 3 个共享权值的 CINN 模块组成, 分别是基于小波的条件特征提取模块、基于 RAM 的融合网络和基于小波融合的反向综合网络。通过设计多尺度的特征提取网络, 利用基础小波来提取不同模态医学图像中的空间细节、纹理等部分, 并将融合指导特征图以多种尺度的方式作为 CINN 的条件输入。另外, 在特征融合模块中嵌入相关激活模块 (RAM), 以便聚焦多模态医学图像中的关键结构区域和纹理细节的一致性信息。

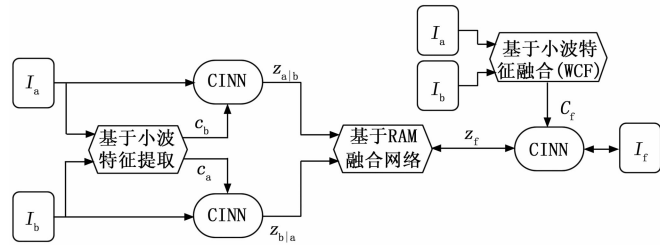


图 1 条件可逆融合网络 (FusionCINN) 模型结构图

2.1.1 基于小波的条件特征提取的前向分析网络

为了更好地融合不同模态医学图像, FusionCINN 使用具有基于小波的条件特征提取对条件输入进行预处理。将条件设定为通过卷积层进行预处理, 使用 Haar 基作为卷积核, 步长为 2, 以分别提取不同模态医学图像的小波特征。通过这种方式, 可以将输入图像分解为低频分量和 3 个高频分量 (H 、 V 和 D), 从而更好地捕捉图像的细节信息。

在小波融合中, 将待融合的两幅图像分别进行 l 层小波分解, 得到每个子频带的小波系数。在每一层小波分解中, 将低频分量和高频分量分别进行下一层的分解, 直到达到 l 层为止。最终, 得到 l 个尺度的子频带, 每个子频带对应不同的频率范围, 其中低频子带包含信号的大体结构信息, 而高频子带则包含信号的细节信息。具体的二维小波分解公式如下:

$$c_k^j = \sum_n h(n-2k)c_n^{j-1}; d_k^j = \sum_n g(n-2k)d_n^{j-1} \quad (7)$$

其中: 波变换中的系数计算过程, 其中 c 和 d 表示不同的小波系数, j 表示分解的层数, k 和 n 表示索引值。

$$\begin{cases} c_a^{(l)} = h_{\text{haar}}(I_a^{l-1}); c_b^{(l)} = \{I_a^{(l)}, I_a^{H(l)}, I_a^{V(l)}, I_a^{D(l)}\} \\ c_b^{(l)} = h_{\text{haar}}(I_b^{l-1}); c_b^{(l)} = \{I_b^{(l)}, I_b^{H(l)}, I_b^{V(l)}, I_b^{D(l)}\} \end{cases} \quad (8)$$

其中: I_a, I_b 是输入图像, h_{haar} 表示使用 Haar 基进

行分解以这样做可以实现融合调节。那么，所提取的多分辨率小波特征可表示为：

$$\begin{cases} C_a = \{c_a^{(l)}\}; l = \{1, 2, \dots, L\} \\ C_b = \{c_b^{(l)}\}; l = \{1, 2, \dots, L\} \end{cases} \quad (9)$$

之后，两个前向 CINN 模块嵌入多个尺度的小波特征，并通过将其作为条件输入来指导多模态源图像的前向分析过程。通过不同模态的源图像与条件进行组合，得到两个条件输出，分别为 $z_{a|b}$ 和 $z_{b|a}$ 。

$$\begin{cases} z_{a|b} = f_{\text{cinn}}(I_a, C_b), z_{a|b} \in R^{w \times h \times c} \\ z_{b|a} = f_{\text{cinn}}(I_b, C_a), z_{b|a} \in R^{w \times h \times c} \end{cases} \quad (10)$$

其中：多分辨率条件特征提取网络设计较为紧凑，对不同模态输入可共享运算资源且不包含训练参数。这意味着在训练过程中不会引入过多额外的计算开销。因此，前向分析网络的整体运行效率较高，能够为后续处理提供高效的多分辨率特征提取及条件网络嵌入。

2.1.2 基于 RAM 的融合网络

融合规则是融合任务中的关键部分，它对融合速度和质量有重要的影响。为了在保留医学图像的结构信息的同时，最大程度地保留其特征信息，FusionCINN 采用简单的平均融合法则。平均融合法则能有效指导条件可逆网络高维潜在变量空间的均匀采样，并用于重构高层语义。公式如下：

$$z_{\text{mean}} = f_{\text{mean}}(z_{a|b}, z_{b|a}) = \frac{z_{a|b} + z_{b|a}}{z_{\text{mean}} \in R^{w \times h \times c}} \quad (11)$$

该融合法则不包含学习参数，可以无偏地融合不同分辨率的多模态医学图像。

此外，受注意力模型的启发，FusionCINN 将相关激活模板 (RAM) 嵌入特征融合中，以准确地聚焦多模态医学图像中的关键结构区域和纹理细节一致性信息。最近，文献 [18] 中作者等人将自关注引入到 GAN 网络中，并提出了用于图像生成任务的自关注 GAN 框架。文献 [19] 中提出了一种多尺度双分支残差注意力 (MSDRA) 网络，用于融合解剖结构和功能性医学图像。FusionCINN 在 RAM 中引入基于相关度量的注意力机制，通过在路径中加入一个卷积网络，使得该特征融合可以更加聚焦于通道的输入图像 (本文实验中对应 MRI 以及 PET 图像)。这类图像比较适合人眼观测，从而更有利于后续医学诊断，同时该方案也能更有效地融合 CINN 输出特征的显著关联信息。基于相关激活模板 (RAM) 的融合网络结构图如图 2 所示，过程可表示为：

$$z_{\text{atten}} = f_{\text{conv}}(z_{a|b}) \times z_{b|a}, z_{\text{atten}} \in R^{w \times h \times c} \quad (12)$$

$$z_f = z_{\text{mean}} + \lambda z_{\text{atten}} \quad (13)$$

在特征融合中嵌入相关激活模板 (RAM) 使融合网络能够动态地关注输入数据的关键部分，提高了模型的效率和泛化能力，使其能够更灵活地处理多模态图像

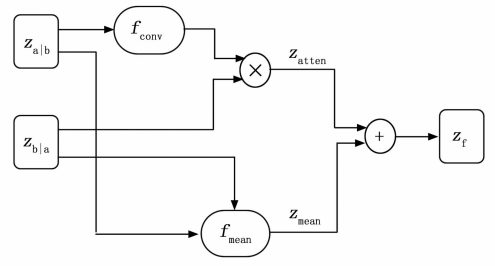


图 2 基于 RAM 的融合网络结构图

的输入。总的来说，引入注意力机制在融合网络中为模型提供了更强的适应能力和处理能力。

2.1.3 基于小波的条件 (WCF) 融合反向综合网络

在 CINN 逆向重构模块中，根据可逆仿射变换操作将融合后的潜在变量 z_f 以融合条件特征 c_f 输入生成预测融合图像 x_f 。通过前向分析中的小波分解公式 (7) 得到的小波特征用绝对值最大化策略进行融合，最后得到融合指导特征图 C_f 。

$$C_f = \max(|C_a|, |C_b|) \quad (14)$$

根据小波特征融合公式，获得融合指导特征图 $C_f = \{x(l), x_{(l)}^H, x_{(l), x_{(l)}^V}\}_{l=1}^L$ 。WCF 将融合指导特征图以多种尺度输出，同时作为 INN 反向综合网络的条件输入。这个模块中的双射函数 $g_{\theta}(\cdot) = f_{\theta}^{-1}(\cdot)$ 是可逆的，可以实现重构图像。最后，可逆体系结构将融合后的潜在变量 z_f 转换为以 C_f 条件输入生成预测融合图像 I_f ，如下式：

$$I_f = g_{\theta}(z_f; C_f) \quad (15)$$

多尺度模块处理数据，融合不同信息层次，低尺度获取细节，高尺度得上下文，提供全面数据表示。条件可逆多尺度方法用于处理高维数据，实现精确重构和融合。WCF 引导特征图融合为多尺度输出，作为 INN 的条件输入。Haar 和 Max-abs 融合策略无需可训练参数，减少了融合中的空间细节损耗。此节介绍了将潜在变量 z 嵌入图像语义空间的紧凑编码方法，降低了计算成本，传递高分辨率特征，并紧凑地编码图像语义信息，实现高质量的多模态医学图像融合。

2.2 融合网络训练策略

本文基于可逆神经网络实现的任务属于回归问题，通过综合多个输入图像得到一幅融合图像。若假设输入数据可映射到连续的输出空间，采用均方误差损失函数 (MSE) 可衡量模型的预测图像与理想参考图像之间的差异。对于条件可逆神经网络 (CINN) 使用均方误差损失函数，一对高度为 H ，宽度为 W ，通道数为 C 的源图像，其前向传播损失、逆向传播损失和总损失的表示如下所示。

1) 前向传播损失：假设地面真实图像为，通过条件可逆神经网络的前向传播操作得到预测图像，前向传

播损失为地面真实图像与预测图像之间的均方误差, 公式 (16) 表示为:

$$L_f = \frac{1}{H \times W \times C} \sum \| I_{ref} - I_f \|^2 \quad (16)$$

其中: \sum 表示对所有像素点进行求和操作, $\|\dots\|$ 表示范数 (通常是 L_2 范数)。

2) 逆向传播损失: 逆向传播损失用于保证逆向传播操作的正确性, 确保输入图像能够通过逆向传播操作恢复到原始状态。逆向传播损失为地面真实潜在变量 z_{ref} 与预测潜在变量 z_f 之间的均方误差表示为:

$$L_b = \frac{1}{H \times W \times C} \sum \| z_{ref} - z_f \|^2 \quad (17)$$

3) 总损失: 总损失是前向传播损失和逆向传播损失以及最大似然估计损失的和, 用于在训练过程中综合考虑两个方向上的误差。根据公式 (4) 中 $\log p(f_\theta(x^i; C_j))$ 的将简化为 $\frac{\|f_\theta(x^i; C_j)\|_2^2}{2}$ 得到最大似然估计损失的定义如公式 (18) 所示, 且总损失如公式 (19) 所示:

$$L_{ML} = \sum_{i=1}^N \left[\frac{\|f_\theta(x^i; C_j)\|_2^2}{2} - \log |J_i| \right] - \log p(\theta) \quad (18)$$

$$L = L_f + L_b + L_{ML} \quad (19)$$

本节中 CINN 的总损失 L 包括预测融合图像和地面真实图像的前向和反向均方误差以及最大似然估计作为基本损失, 可以更细致地调整模型, 进一步提高图像质量, 使生成图像更具结构相似性。并且通过减少在编码过程中的信息损耗, 高效地实现图像空间细节的分析与重构。

3 实验结果与分析

3.1 实验设置

为了充分验证 FusionCINN 在不同模态医学图像融合中的有效性, 实验部分的主要训练数据是多个不同模态组合的医学图像数据集由哈佛医学院网 (www.med.harvard.edu/AANLIB/home.html) 上获得的全脑图像数据集进行实验, 包括 CT、MRI 和 PET 等医学图像。所有源图像在预处理过程中都被调整到相同的尺寸, 并且所有图像对都经过配准处理, 使它们在空间上对齐。条件可逆融合网络 (CINN) 结构是由 4 个卷积可逆块和 1 个全连接可逆块组成, 同时在每个耦合块后加入下采样操作。本文将 CINN 的输入尺寸设为 128×128 , WCF 的输入尺寸设为 512×512 , 测试图像尺寸设为 128×128 。在实验中, 使用 Adam 优化器训练网络, 权重衰减设为 1×10^{-5} , 学习率为 5×10^{-5} , 批次大小为 8。

3.2 主观视觉评价

本节是将 FusionCINN 得到的融合结果通过主观视觉呈现的方式与 DTCWT^[20]、PAPCNN^[21]、DIDF^[22]、CNN^[23] 方法输出的融合结果进行比较。如图 3~6 所示, 其中包括两组 CT-MRI 图像对、两组 MRI-PET 图像以及其他几类不同融合场景下最终融合图像的视觉对比结果。

如图 3 和图 4 中的细节放大部分, 可以发现不同参与对比的方法在保留 CT 图像中的骨骼信息和 MRI 中的软组织信息方面表现各异。具体来说, FusionCINN 方法与 CNN 方法相比, 在图像中能够更为清晰地呈现骨骼结构, 而 DTCWT 和 DIDF 方法生成的图像亮度显著降低。然而, PAPCNN 方法生成的融合图像在对比度和清晰度方面存在一定不足, 即像素亮度层次相对较为模糊, 如图 3 和图 4 所示, 整体呈现较为昏暗的色调。相比之下, CNN 方法虽然在保留亮度信息方面较为出色, 但由于保留了过多的 CT 图像特征, 导致软组织的对比度相对较模糊。正如图 3 中细节放大图左下角所示, 软组织信息并不明显。另一方面, DTCWT 方法在亮度方面没能有明显的降低, 然而, 它得到的融合图像对比 MRI 输入模态不同程度地损失了软组织细节信息。从图 3 和图 4 中可以看出, FusionCINN 方法的融合图像能够有效地保留 CT 图像中显著特征的强度信息, 以及较多的 MRI 图像的细节纹理信息。原因在于 FusionCINN 方法采用的条件可逆神经网络 (CINN) 具有的可逆性, 这使得最后的融合图像最大程度地保留了不同模态的纹理空间细节, 生成图像细节较为清晰准确。并且, 多尺度的特征提取操作使得 CT 图像的骨骼亮度信息和 MRI 的软组织信息均能较为完整地获取, 并作为条件向量指导可逆网络的生成过程。

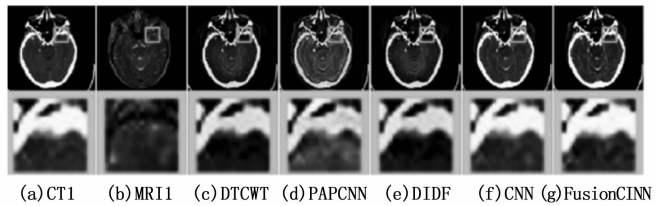


图 3 CT1-MRI1 图像融合结果比较

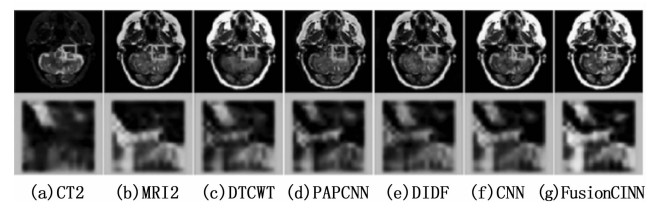


图 4 CT2-MRI2 图像融合结果比较

如图 5 和图 6 可以明显观察到 CNN 和 FusionCINN 方法在保留 MRI 和 PET 图像信息方面均表现出了卓越的性能。上述方法成功结合了两种不同光学模型的多模态图像,使显著空间信息得到完整而清晰的呈现。相比之下,DTCWT 方法在亮度和色彩保真度方面表现卓越,然而在细节处理方面存在一些问题。以图 6 的细节放大图为例,可以明显看到中间区域的信息混淆较为明显,导致 MRI 图像中的白色软组织细节被遮挡。这也导致了融合图像中整体轮廓信息的丧失,可能会对后续的一些特定诊疗判断产生不利影响。而图 6 中 DIDF 方法虽然明显降低了 PET 图像的色彩强度,但重要组织信息被严重丢失,如图 5 中间区域较为模糊。相比之下,提出的方法通过在融合网络中引入了 RAM 注意力机制,使得融合结果更加关注 PET 图像显著信息。PET 图像通常包含丰富的色彩及结构信息,如组织结构和相关代谢活动,由此看来 FusionCINN 方法的融合结果色彩层次更丰富,更有利于医生的直接观测,也为后续诊疗判断提供了更完善的支持。在最终融合图像对于不同模态显著特征的主观视觉感受上取得较好的平衡的同时,FusionCINN 方法还通过保持源图像的空间细节特征及对比度,呈现出最佳的主观融合视觉效果。

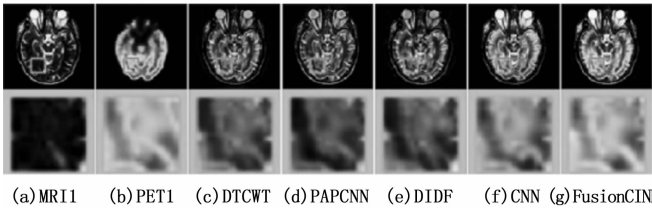


图 5 MRI1-PET1 图像融合结果比较

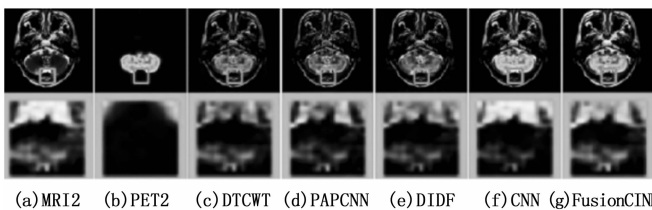


图 6 MRI2-PET2 图像融合结果比较

3.3 客观指标评价

在本节中,将对上述不同模态医学图像进行客观指标评价。为了确保定量评价的准确性和全面性,随机在两种模态融合图像对中各选取了 8 组测试图像进行了评估,并且使用了 5 项客观指标来综合评估融合结果,这些指标可以更好地分析不同算法的融合性能。采用的评价指标包括互信息 (MI)^[24]、空间颜色分布 (SCD)^[25]、结构相似性指数 (SSIM)^[26]、视觉信息保真度 (VIFF)^[27]、边缘强度 (ED)^[28]。为了更加清晰地展示定量评价结果,将 CT-MRI、MRI-PET 两种医学图像对通过与 5 种算法融合后得到的

指标平均值汇总在表 1 和表 2 中。在表格中,使用粗体标记最优值。

表 1 CT-MRI 的定量评价结果

模型/标准	MI	SCD	SSIM	VIFF	EI
DTCWT	0.683 4	1.318 7	0.628 9	0.268 4	0.147 8
PAPAN	0.713 9	1.232 9	0.689 7	0.380 9	0.130 7
DIDF	0.682 2	1.220 7	0.515 8	0.388 1	0.151 6
CNN	0.637 5	1.240 3	0.661 3	0.374 6	0.150 3
Fusion CINN	0.791 6	1.518 7	0.702 6	0.568 7	0.158 6

表 2 MRI-PET 的定量评价结果

模型/标准	MI	SCD	SSIM	VIFF	EI
DTCWT	0.637 5	1.338 7	0.626 3	0.346 3	0.133 4
PAPAN	0.791 5	1.307 9	0.602 6	0.582 4	0.139 7
DIDF	0.631 4	1.318 6	0.528 9	0.367 6	0.122 1
CNN	0.687 1	1.634 4	0.670 5	0.563 9	0.135 9
Fusion CINN	0.810 6	1.740 3	0.636 8	0.605 7	0.149 1

根据表 1 和表 2 的数据,FusionCINN 方法在许多客观评价指标上表现出色。特别是在 MI 和 EI 性能指标方面,提出方法表现出卓越的效果,取得了较好的评价数值。这意味着 FusionCINN 的融合结果在信息增强和边缘保持方面表现出色,成功地捕捉了更多源图像的特征信息。这与主观评价的结果基本一致,证明了 FusionCINN 方法的可靠性。这一优越性归因于 FusionCINN 引入了条件可逆神经网络,其可逆性该网络能够更好地提取源图像的语义信息和对比度信息。此外,FusionCINN 在视觉信息保真度 (VIFF) 方面也取得了令人满意的高分,与主观评价结果相符。提出方法在 RAM 融合网络中引入了注意力机制,使得融合结果巧妙地平衡了 PET 图像的色彩信息,这符合人眼观察的特点,并且对于医学诊断非常有帮助。FusionCINN 的定量评价结果虽然不在各个指标是最优值,如在 MRI-PET 图像融合中的 SSIM 指标为 0.636 8,仅次于 CNN 方法的 0.670 5,但融合后的图像细节内容信息比较丰富。FusionCINN 方法首先能够清晰地展示 CT 图像中的骨骼密集结构;其次,有效地保留了 MRI 图像中更多的纹理和细节。此外,该方法还能够有效地呈现 PET 图像的色彩饱和度和功能信息。

为了更全面地评估 FusionCINN 图像融合算法的性能,从包含 CT-MRI 和 MRI-PET 两组医学图像的样本中,随机选取了 8 张测试图像做进一步的性能评估。使用 5 个不同的指标对各种融合方法进行了定量分析,这些分析结果如图 7 和图 8 所示。DTCWT 方法在多尺度分解时未考虑边缘信息,且低频分量融合规则过于简单,导致了图像细节丧失和低对比度,客观指标

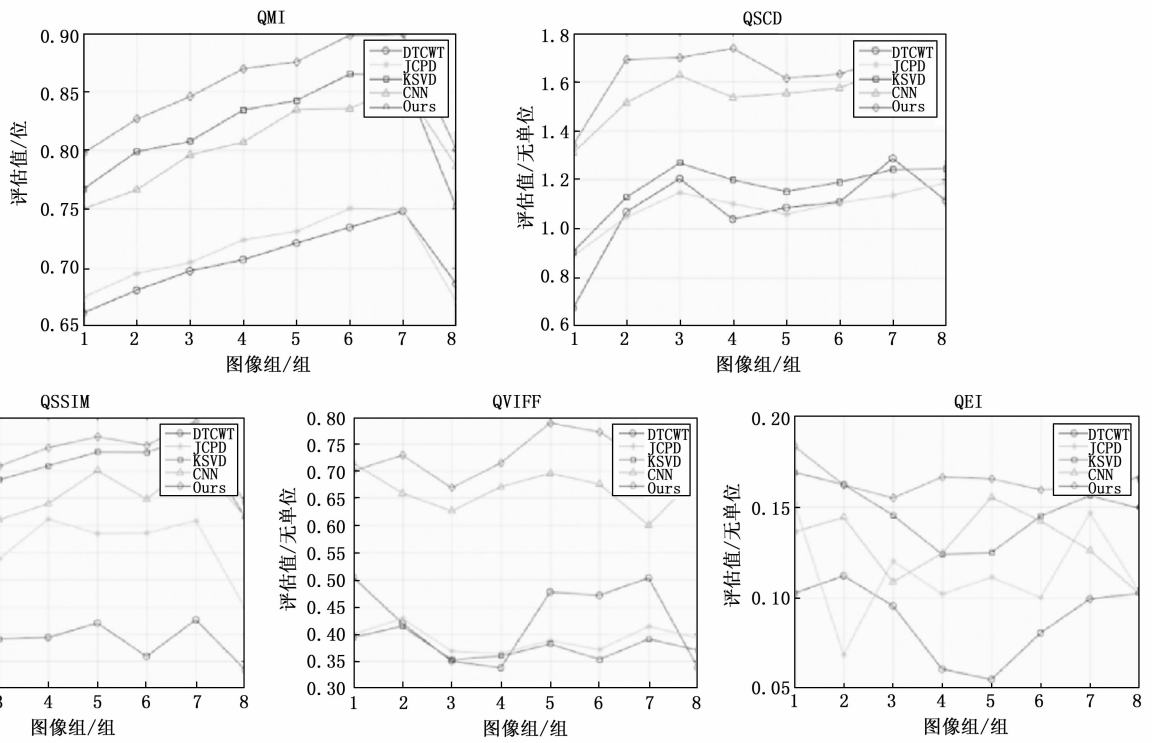


图 7 CT-MRI 的 8 个测试图像对融合结果定量比较

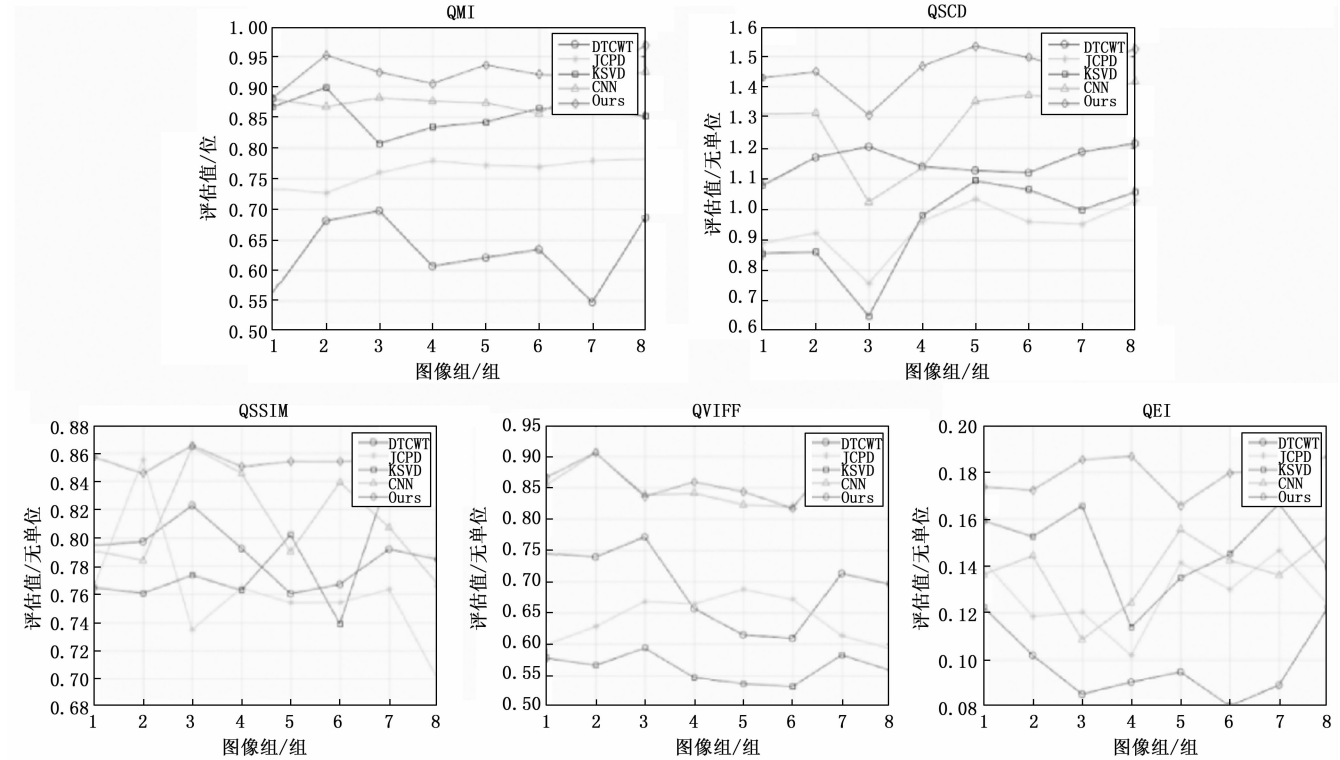


图 8 MRI-PET 的 8 个测试图像对融合结果定量比较

得分偏低。虽然 CNN 在特征提取方面有优势,但对不同模态图像的特征提取容易造成混乱,导致在 5 个客观指标上表现次于 FusionCINN 方法。FusionCINN 在

所有 5 个指标上均取得最优值,特别是 VIFF 指标,显示 FusionCINN 更符合人眼视觉感知,而最大的 SSIM 值说明了 FusionCINN 有效地保留了源图像的边

缘细节纹理信息,减少了空间细节的损耗。在这些定量对比实验中,FusionCINN方法由于利用条件可逆神经网络(CINN)的可逆性,成功地保留了CT图像的亮度信息,捕捉了MRI图像的细节信息,并同时保留了PET图像的彩色信息。并且在多个关键指标上取得的令人满意的结果。

本文提出的融合方法FusionCINN在主观和客观评价上均体现了良好的性能,且融合后的图像保留了源图像中的显著空间细节特征,对源图像的语义信息进行了有效的视觉增强,实现了在不同场景下与多模态输入图像的有效交互,具有较强的通用性。然而,本文的方法只在特征融合模块中引入注意力机制来聚焦多模态输入图像的潜在空间特征,对更为复杂的不同模态通道独立的特征提取模块的深入研究将有望在本文基础上进一步提升最终融合图像的性能。

4 结束语

本文介绍了一种基于条件可逆神经网络(CINN)的多模态医学图像融合算法FusionCINN,旨在最小化深度网络前馈中的空间细节损失。通过将多分辨率特征嵌入CINN作为条件向量,成功实现了多模态表示的学习。采用基于小波的条件融合(WCF)网络,能够引导CINN反向重构融合图像。引入的相关激活模板(RAM)保持了多模态医学图像中关键结构区域与纹理细节的一致性信息。通过实验证明,相较于其他融合方法,本文方法在多个客观指标和主观效果上均取得了优越的性能,且紧凑的网络架构具有运算效率的优势,能适应多模态医学图像应用场景,显著地提高融合结果的对比度和细节保留,为医学诊疗进一步优化显著信息提供了稳定的支持。

参考文献:

- [1] 王达志,杨康,成果,等.多传感器采集和数据融合用于智能监控的目标检测[J].通信电源技术,2020,37(12):133-137.
- [2] 李俊杰,傅俏燕,姜涛.结合光谱响应函数和全局方差匹配的遥感图像融合[J].光子学报,2020,49(10):148-157.
- [3] SARMA M, UMER J. Multi-modal medical image fusion based on two-scale image decomposition and sparse representation [J]. Biomedical Signal Processing and Control, 2020, 57: 101-110.
- [4] LECUN Y, BOSER B, DENKER J S, et al. Back propagation Applied to Handwritten Zip Code Recognition [J]. Neural Computation, 1989, 1: 541-551.
- [5] GOODFELLOW I, MIRZA M, XU B, et al. Generative adversarial nets [J]. Computer Science, 2014, 24: 2672-2680.
- [6] BURT P J, ADELSON E H. The Laplacian pyramid as a compact image code [M]. Morgan Kaufmann: Readings in Computer Vision, 1987.
- [7] LI H, MANJUNATH B S, MITRA S K. Multisensor image fusion using the wavelet transform [J]. Graphical Models and Image Processing, 1995, 57(3): 235-245.
- [8] XU H, MA J, LE Z. FusionDNN: a unified densely connected network for image fusion [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12484-12491.
- [9] ZHANG Y, LIU Y, SUN P. IFCNN: A general image fusion framework based on convolutional neural network [J]. Information Fusion, 2020, 54: 99-118.
- [10] LIU Y, CHEN X, WARD R. Image fusion with convolutional sparse representation [J]. IEEE Signal Processing Letters, 2016, 99: 1882-1886.
- [11] JUN H, ZHU L L, YONG M. MGMDcGAN: medical image fusion using multi-generator multi-discriminator conditional generative adversarial network [J]. IEEE Access, 2020, 8: 55145-55157.
- [12] CHEN H. Medical image fusion using generative adversarial networks with encoder-decoder structure [J]. IEEE Access, 2019, 37: 10054-10063.
- [13] KINGMA D P, DHARIWAL P. Glow: Generative flow with invertible 1x1 convolutions [J]. Advances in Neural Information Processing Systems, 2018, 2: 331-337.
- [14] GOMEZ A N, REN M, URTASUN R. The reversible residual network: Backpropagation without storing activations [J]. Statistics, 2017, 10(3): 597-598.
- [15] ARDIZZONE L, LÜTH C, KRUSE J. Guided image generation with conditional invertible neural networks [J]. Computer Vision and Pattern Recognition, 2019, 4: 215-224.
- [16] DINH L, KRUEGER D, BENGIO Y. NICE: Non-linear independent components estimation [J]. Computer Science, 2014, 3: 114-127.
- [17] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using real NVP [J]. Statistics, 2016, 10(3): 143-148.
- [18] ZHANG H, GOODFELLOW I, METAXAS D. Self-attention generative adversarial networks [C]. International Conference on Machine Learning PMLR, 2019, 7354-7363.
- [19] LI W, PENG X, FU J. A multiscale double-branch residual attention network for anatomical-functional medical image fusion [J]. Computers in Biology and Medicine, 2020, 141: 100-105.

(下转第 162 页)