DOI:10.3969/j.issn.1005-202X.2026.06.010

医学影像物理

基于改进U-Net模型的脑部MRI生成合成CT研究

杨智慧[1,2],李好[1,2],李丰森[1,2]
1.甘肃中医药大学医学信息工程学院,甘肃 兰州 730000;2.甘肃省人民医院(甘肃中医药大学第一临床医学院),甘肃 兰州 730000

【摘要】**目的:**提出一种基于改进U-Net的脑部MRI生成合成CT(sCT)的方法,实现MRI到CT图像的转换,并比较传统U-Net模型及改进模型在生成脑部sCT方面的性能差异。**方法:**基于用于放疗计划的CT图像生成2023挑战赛(SynthRAD2023)数据集,选取120例患者的脑部MRI、CT图像作为训练集,39例作为测试集。分别构建传统U-Net模型以及基于U-Net改进的模型。对于二者生成的sCT,分别从图像质量、强度值等角度进行评估。**结果:**对于测试集中的病例,使用传统U-Net模型和改进模型生成的sCT图像与真实CT间的平均绝对误差分别为(100.28±49.67) HU和(94.87±42.50) HU,峰值信噪比分别为(26.43±2.70) dB和(26.85±2.73) dB,结构相似性分别为0.808±0.092和0.820±0.085。**结论:**本研究基于U-Net改进的模型在脑部MRI生成sCT精度方面优于传统的U-Net模型。

【关键词】磁共振成像;脑;合成CT;U-Net
【中图分类号】R318;TP391.4 【文献标志码】A 【文章编号】1005-202X(2026)06-0774-08

Improved U-Net model for generating synthetic CT from brain MRI

YANG Zhihui[1,2], LI Hao[1,2], LI Fengsen[1,2]
1. School of Medical Information Engineering, Gansu University of Chinese Medicine, Lanzhou 730000, China; 2. Gansu Provincial Hospital (First Clinical Medical School, Gansu University of Chinese Medicine), Lanzhou 730000, China

Abstract: Objective To propose an improved U-Net approach for generating synthetic computed tomography (sCT) from brain magnetic resonance imaging (MRI), achieving MRI-to-sCT translation, and further compare the brain sCT generation performance of the improved model with that of the traditional U-Net model. **Methods** The SynthRAD2023 dataset for CT image generation in radiotherapy planning was adopted in the study. The brain MRI and CT images of 120 patients were selected as the training set, while those from 39 patients were used as the test set. A traditional U-Net model and an improved U-Net model were constructed, and their generated sCT were evaluated regarding image quality and intensity values. **Results** For the cases in the test set, the average absolute error relative to ground truth CT reached (100.28±49.67) HU for the traditional U-Net model and (94.87±42.50) HU for the improved model. The corresponding peak signal-to-noise ratios were (26.43±2.70) dB and (26.85±2.73) dB, and the structural similarities were 0.808±0.092 and 0.820±0.085, respectively. **Conclusion** The proposed improved U-Net model outperforms the traditional U-Net model in terms of accuracy for brain MRI-to-sCT translation.

Keywords: magnetic resonance imaging; brain; synthetic computed tomography; U-Net

前 言

在放疗计划中,磁共振成像(Magnetic Resonance Imaging, MRI)与计算机断层扫描(Computed Tomography, CT)的联合应用已成为临床常规影像学流程[1]。MRI具有优越的软组织对比度,在肿瘤靶区和关键解剖结构的精准勾画中发挥着关键作用;CT则能够提供电子密度信息,是实现精确剂量计算的必要条件。两者协同能够提供互补的信息。然而,现有方案通常需要患者分别接受MRI和CT扫描,并在治疗计划阶段将两种模态进行配准,以补偿不同扫描间可能产生的体位差异或运动伪影。这一流程不仅增加成像和操作的复杂性,还可能引入配准误差,从而影响后续剂量分布的准确性与治

【收稿日期】2026-01-15
【基金项目】甘肃省科技计划项目-软科学专项(25JRZA204)
【作者简介】杨智慧,硕士研究生,研究方向:医学图像处理、深度学习,E-mail: y2218784238@163.com
【通信作者】李丰森,正高级工程师,硕士生导师,研究方向:医学信息工程,E-mail: 1641673807@qq.com

疗效果。为克服上述不足，MRI-only放疗计划近年来得到广泛研究[2-5]。其核心思想是在治疗计划设计流程中仅使用MRI图像，无需CT图像及配准操作，避免患者额外的CT成像剂量。近年研究表明，将该流程与MRI引导自适应放疗技术相结合可以有效简化流程，提高效率[6]。然而，MRI缺乏精确剂量计算所需的电子密度信息，该策略必须依赖替代方法加以补偿。

基于MRI的合成CT（Synthetic CT, sCT）生成技术是目前最直接，也是最常见的实现方式。该方法能够直接由MRI数据生成CT图像，从而避免多模态配准过程中潜在的误差与不确定性。同时，该方法显著减少因重复CT扫描所带来的额外辐射暴露，这对于儿童等敏感人群具有重要意义[7]。更为关键的是在自适应质子治疗等应用场景中，该技术不仅能够提高影像精度，还能够在保障治疗安全性的同时推动精准医学的发展[8-9]。鉴于MRI与CT在影像强度及结构特征上的本质差别，利用MRI直接生成sCT图像依旧是一项复杂的挑战。常见的传统方法包括体积密度赋值法、体素转换法、图谱法等，但往往基于人工设计的特征与经验规则[10]，因此不仅容易引入主观偏差，泛化能力还有限，难以适应多样化的临床数据。

借助深度学习方法生成sCT图像，有望突破传统方法的限制并支持MRI-only放疗。这类方法通常利用卷积神经网络、生成对抗网络或其变种来建立MRI与对应CT的HU值之间的映射关系。在众多深度学习架构中，U-Net是一种被广泛用于医学图像处理的网络模型。相较于传统的卷积神经网络，它引入跳跃连接的概念，使得网络的编码器和解码器路径得以连接，允许将早期层的高分辨率特征直接合并到后续层中。这减轻了下采样过程中空间信息的丢失，对于分割等像素精确任务至关重要。但是该架构最初是为分割而开发的，而不是为生成全新的图像而开发的，因此U-Net在应用于图像生成任务时存在缺乏对多模态医学数据有效支持的局限性。

为了解决合成图像的清晰度和结构保真度不足这一问题，本研究提出基于U-Net改进模型来实现MRI跨模态转换CT图像。在U-Net模型基础上引入协调坐标卷积（Coordinate Convolution, CoordConv）模块、ConvNeXt Block模块、Triplet Attention模块和Cross Attention模块。CoordConv模块引入坐标信息以提高模型对空间位置的敏感度，ConvNeXt Block模块增加网络深度以提高细节捕获能力，Triplet Attention模块用于增强跨通道与空间的交互，Cross Attention模块用于促进全局语义与局部细节的融合。通过改进U-Net模型，期望改善传统方法和U-Net模型方法在MRI生成sCT过程中存在的细节保真度不足的问题，从而使生成的sCT更全面地保留组织结构和纹理特征。

1 数据集与预处理

本研究使用的CT和MRI图像来自于荷兰3所大学医学中心的放射肿瘤科，即乌得勒支大学医学中心（Utrecht University Medical Center）、格罗宁根大学医学中心（Groningen University Medical Center）和拉德堡德大学医学中心（Radboud University Medical Center）组织的用于放疗计划的CT图像生成2023挑战赛（SynthRAD2023 grand challenge dataset: generating synthetic CT for radiotherapy, SynthRAD2023）中提供的免费公开使用的数据集[11]。所采用的数据集共计180对脑部MRI与CT成像数据，数据按照随机策略分配为3个子集：训练集120对、验证集21对、测试集39对。

在数据预处理部分，对MRI数据进行z-score标准化处理，对CT数据进行最小值-最大值归一化处理。所有图像均被填充到统一的大小，288×288，用0值填补边界。同时保存原始位置参数，方便后续恢复。每个病例的MRI图像均采用2.5D的方法进行预处理。通过处理来自多个平面的切片来集成3D上下文信息。传统的2D方法由于缺乏相邻的上下文信息而面临局限性，存在跨切片的结构之间断开连接的问题。但是，全3D方法面临重大的内存挑战，处理整个卷的能力受限，而且它们往往需要更多数据才能实现稳健的性能。于是，为了克服这些挑战，2.5D方法已成为一种有前途的中间立场。2.5D方法通过保持具有较少参数的轻量级模型来实现平衡，从而实现更快的训练和推理，同时仍然受益于3D上下文提供的增强的空间连贯性，兼具2D模型的效率和3D模型的空间一致性，为全3D模型提供一种轻量而有效的替代方案。本研究中网络的输入为5通道的2D MRI图像（尺寸为288×288×5），其中通道维度对应当前切片及其前后各两张相邻切片。

2 网络结构

2.1 主体结构

U-Net是一种对称的编码器解码器结构，由代表编码器分支的堆叠卷积层和最大池化层的收缩路径和表示解码器的反卷积层组成的相应扩展路径组成[12]。U-Net能够通过跳跃连接从相应对称编码器

和解码器层的不同级别连接各自的特征图，再从编码器卷积中提取上下文特征和从解码器卷积中提取语义特征之间进行权衡。通过这种方式，U-Net将低级细节和上下文信息与高级语义和位置属性相结合，从而实现两者之间的权衡。

本研究所提出的网络模型基于传统的U-Net架构，如图1所示，采用自定义的残差块和编码解码结构，并在此基础上进行扩展，旨在通过深度学习方法实现医学图像的跨模态转换。网络的输入为MRI图像，目标输出为合成的CT图像。该网络由编码器、解码器和跳跃连接组成，在每个编码层之后进行下采样，并在解码层进行上采样。在此基础上，针对医学图像特有的复杂性，本研究引入多个创新模块以提升模型性能。

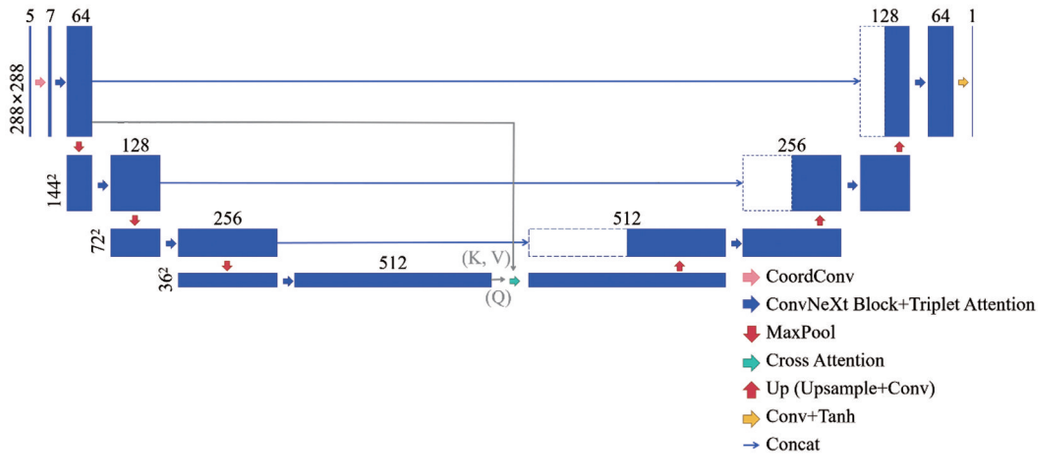

图1 基于U-Net改进的模型整体结构

Figure 1 Overall structure of the improved U-Net model

2.2 CoordConv模块

CoordConv层是标准卷积层的扩展，允许卷积滤波器考虑像素的空间坐标，表示卷积前通道维度上的串联坐标值。其目标是学习笛卡尔空间中的坐标与单热像素空间中的坐标之间的映射，将坐标信息引入卷积网络，有利于提高模型对空间位置的敏感度，特别是在医学图像中，不同位置的语义信息具有显著意义。传统假设认为卷积神经网络具有平移不变性，该特性有助于提升模型性能[13]。但研究表明，通过CoordConv在输入层向图像注入绝对位置信息，即使破坏平移不变性，仍能提高模型精度[12]。

CoordConv的实现是通过将两个额外的x和y通道连接到输入通道来完成的[14]，如式(1)所示：

$$Y = \mathrm{Conv}\left(\mathrm{Concat}\left(X, \left\{\left(X_{ij}, Y_{ij}\right)\right\}\right)\right) \tag{1}$$

其中，X代表输入特征，$\left(X_{ij}, Y_{ij}\right)$代表每个像素的坐标，Concat代表沿着维度方向的串联，Conv代表卷积层的处理，Y代表输出特征。

在本研究网络编码器的第一层中，笔者使用CoordConv取代标准卷积，使模型能够在特征提取开始时获取空间信息。CoordConv的结构如图2所示。在实验中，CoordConv是通过PyTorch库实现的，应用线性缩放以坐标层的值限制在-1~1。坐标信息的引入有助于模型在处理不同位置的特征时，能够考虑到其空间位置的特征，利用空间信息，模型可以构建更强的空间建模能力，增强对局部和全球位置的理解，提高基于位置的推理能力[15]。

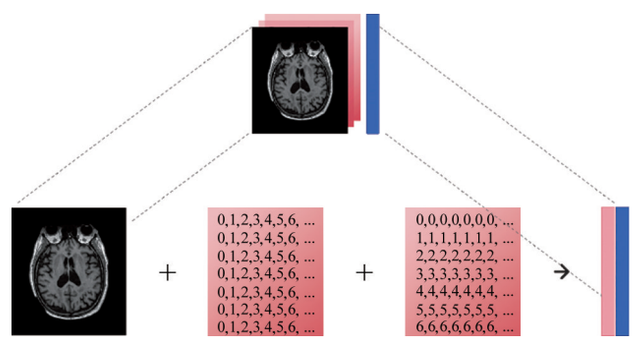

图2 CoordConv模块结构

Figure 2 CoordConv module structure

2.3 ConvNeXt Block模块

ConvNeXt由Liu等[16]在2022年提出，它集成ResNet和Swin Transformer的成功设计，以实现更平滑的网络梯度，从而实现更快的收敛和更高的模

型性能。ConvNeXt在图像分类任务中表现出优异的效果。在ConvNeXt块中，初始卷积层输出的数据进入深度卷积层进行进一步的卷积运算，这种卷积方法有利于增加网络的深度，通过逐个通道处理来增强模型捕获细节的能力[17]。此外，在ConvNeXt块中，常用的ReLU激活函数被更平滑的GeLU激活函数所取代，激活和归一化层的数量减少，并且使用层归一化LayerNorm代替批量归一化BatchNorm[18]。这些改进使ConvNeXt结合纯卷积网络和注意力网络的优点，保证效率和简单性，实现更高的精度[19]。

在保留ConvNeXt主干结构的基础上，引入可变形卷积以及后置的注意力模块作为精细特征重加权，如图3所示。首先，在卷积算子选择上，出于计算开销的考虑，仅在编码器的最后一个卷积层将深度卷积替换为可变形卷积。传统卷积在处理目标形变或解剖结构差异时表现受限，而可变形卷积通过引入可学习的采样偏移，实现对不规则结构的自适应建模[20]，从而更好地处理自然图像或医学影像中存在的形态不规则性。因为器官形态与病灶结构往往存在较大差异，可变形卷积能够提供更鲁棒的特征提取能力。其次，在残差连接之后引入注意力模块作为局部再校准器，可进一步对通道或空间维度的特征进行加权，从而抑制噪声信息并突出任务相关区域。

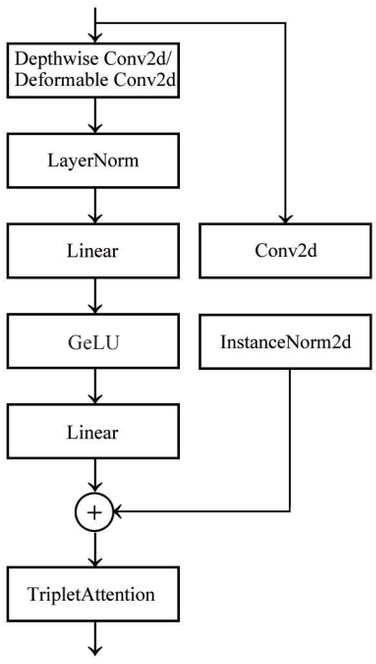

图3 ConvNeXt Block模块结构

Figure 3 ConvNeXt Block structure

2.4 Triplet Attention模块

Triplet Attention模块最早由Misra等[21]提出，其核心思想是通过旋转特征张量，从3个互补的二维视角建模注意力关系，从而在保持计算效率的同时增强特征表达能力。传统的通道注意力或空间注意力往往只关注单一维度的依赖关系，难以同时建模跨通道与跨空间的交互。为此，Triplet Attention将输入特征分别在3个维度组合下进行排列：通道-宽度（C-W）、高度-通道（H-C）、高度-宽度（H-W），并在每个分支上通过轻量的Z-Pool、卷积、sigmoid的操作来学习加权映射[22]。Z-Pool层旨在将输入张量的通道维度压缩为2个特征图，其实现方式是在通道维度上分别进行平均池化与最大池化操作，并将二者结果进行拼接。通过这种设计，Z-Pool在有效降低特征图深度、减轻后续计算负担的同时，仍能保留输入特征的显著响应与整体分布信息，从而为后续的注意力建模提供更为全面的特征表示[23]。最终，3个分支的输出特征被融合，形成同时包含空间与通道交互信息的增强特征[24]。Triplet Attention的结构如图4所示。

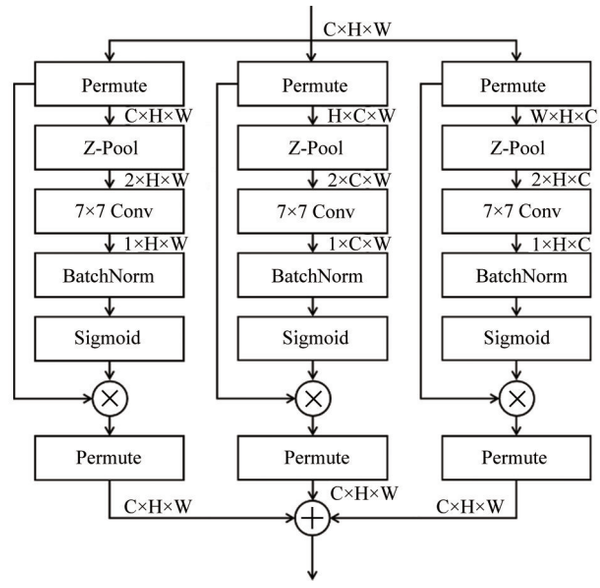

图4 Triplet Attention模块结构

Figure 4 Triplet Attention module structure

该设计的目标在于捕获输入特征在不同维度上的相互依赖，避免单一注意力机制对信息的忽略。与传统注意力模块相比，Triplet Attention结构简洁，无需显著增加计算开销，却能在多种视觉任务（分类、检测、分割等）中显著提升性能。对于医学图像

任务，它能够更好地强调解剖结构的关键区域，并在多模态特征融合或跨域合成中有效抑制冗余信息，提升模型的泛化与鲁棒性。

2.5 Cross Attention模块

Cross Attention机制旨在建立深层特征与浅层特征之间的关联，以促进全局语义与局部细节的融合[25]。其计算过程如图5所示。将深层特征定义为查询矩阵（***Q***），而浅层特征生成键（***K***）和值（***V***）矩阵。键矩阵与查询矩阵进行匹配，二者作内积运算计算相似度，经归一化后得到注意力权重。值矩阵则用于加权求和得到融合后的特征。

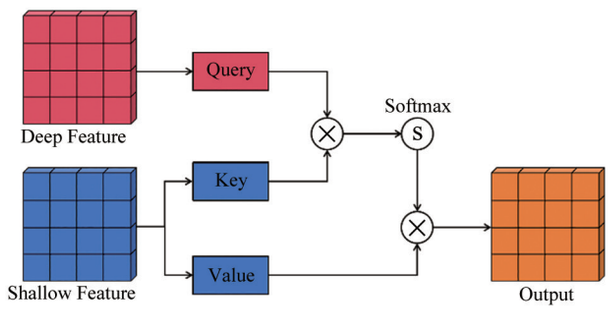

图5 Cross Attention模块特征融合示意图

Figure 5 Schematic diagram of Cross Attention module feature fusion

该机制的目的在于令深层特征作为查询，引导浅层分支的细粒度细节集成到深层分支中，从而实现浅层与深层特征的互补融合。借此，模型既能获取图像的全局上下文，也能保留局部结构特征，进而增强对整体图像构图与细节的理解。

出于计算开销的考虑，在具体实现时，笔者在网络瓶颈阶段引入Cross Attention，以深层特征为查询、浅层特征为键和值，从而实现跨层信息交互，并进行改进，旨在提升计算效率和特征融合能力。首先，动态下采样机制的引入，使得查询和键值特征图在计算注意力之前能够根据需要进行下采样，从而显著降低计算负担，尤其在处理高分辨率图像时，能够有效减少计算和内存开销。其次，采用多头注意力机制，通过将查询、键和值映射到多个子空间来增强模型对不同特征的学习能力[26]。此外，通道恢复机制通过卷积层恢复查询和键值特征图的原始通道数，确保输出特征图与输入特征图在通道维度上的一致性，从而避免可能出现的信息丢失或维度不匹配问题。在此基础上，缩放因子的引入对查询与键的内积结果进行归一化，缓解高维内积导致的数值不稳定问题。同时，针对下采样带来的分辨率变化，输出特征图通过双线性插值恢复至原始空间分辨率，确保输出在空间尺寸上的一致性。在网络模型中，不同于Triplet Attention聚焦于层内的特征，Cross Attention旨在跨层融合特征。

3 实验与结果分析

3.1 实验设置

本研究划分的120例训练集以2为间隔数切片，共11 518对图像，39例测试集以1为间隔数切片，共包含7 741对图像。本研究采用的所有模型均在操作系统Ubuntu 22.04（Canonical Ltd.，英国）上进行；建模基于深度学习架构Pytorch 2.3.0（Meta Platform Inc.，美国）和科学计算统一设备架构CUDA 12.1（Nvidia Corporation，美国）。实验的硬件环境为显卡（Nvidia GeFore GTX4090 GPU，NVIDIA Corporation，美国）。模型优化器使用自适应矩估计（Adaptive moment estimation，Adam）进行模型训练，将初始学习率设置为1×10^{-4}。批量大小设置为8，训练轮次为200。

损失函数根据掩膜将图像划分为感兴趣区域（ROI）与非ROI，并分别计算两个区域的损失。具体而言，掩膜为1的位置参与ROI损失计算；掩膜取反后得到的非ROI用于计算非ROI损失。两个区域的损失均由结构相似性（Structural Similarity，SSIM）损失和L1损失加权组成。SSIM损失用于衡量生成图像与真实图像之间的结构相似性，L1损失则评估两者在像素值上的绝对差异。最后，总损失由ROI区域的损失与非ROI区域的损失按比例加权求和。该设计旨在增强合成图像的结构保真度，同时优化像素级的精度，从而提高生成图像的整体质量。总损失如式（2）所示，其中两个区域的损失均采用相同的混合形式如式（3）所示：

$$L_{\text{total}} = 0.95 \times L_{\text{ROI}} + 0.05 \times L_{\text{nonROI}} \tag{2}$$

$$L = 0.5 \times \frac{1}{n}\sum_{i=1}^{n}\left(1 - \text{SSIM}\left(x_i, y_i\right)\right) + 0.5 \times \frac{1}{n}\sum_{i=1}^{n}\left|y_i - x_i\right| \tag{3}$$

其中，x为模型输出即sCT，y为真实CT，n为对应区域中的像素对数量，L_{ROI}为ROI区域中非零像素的数量，L_{nonROI}为非ROI区域的像素总数。SSIM(x,y)见式（6），其中K_1、K_2、L的取值为0.01、0.03、2。

3.2 评价指标

用于图像定量比较的评估指标，包括平均绝对误差（Mean Absolute Error，MAE），峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）和SSIM，这些指标用于评估生成性能。

MAE是最简单和最常用的指标之一，它计算合成图像和真实图像中像素强度之间的平均绝对差异，定

义为：

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-x_i\right| \tag{4}$$

其中，y_i和x_i分别是合成图像和真实图像中第i个像素的强度值，n是像素对的总数。MAE反映差异的幅度，较低的值代表较高的图像质量。

PSNR源自均方误差，是图像处理中广泛使用的指标。它将图像中最大可能的像素强度与合成图像和真实图像之间的误差进行比较：

$$\mathrm{PSNR}=10\log_{10}\left(\frac{\mathrm{MAX}^2}{\sum_{i}^{n}\left|y_i-x_i\right|^2/n}\right) \tag{5}$$

其中，MAX代表图像中的最大强度值。PSNR可有效检测全局强度差异，但在捕获结构差异方面存在局限性，而结构差异在医学图像合成中通常至关重要。PSNR值越高表示图像质量越好。

虽然MAE和PSNR等指标易于计算，但它们准确评估视觉质量的能力有限，尤其是在处理感知差异时。为了解决这个问题，Wang等[27]引入SSIM，利用人类视觉系统已知的特征，从而比较像素强度的局部差异，并对亮度和对比度进行归一化。此后，SSIM因其评估图像结构完整性的能力而在图像合成领域得到广泛采用。SSIM是通过评估合成图像和参考图像之间的亮度、对比度和结构进行计算的：

$$\mathrm{SSIM}(x,y)=\frac{\left(2\mu_x\mu_y+C_1\right)\left(2\sigma_{xy}+C_2\right)}{\left(\mu_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)} \tag{6}$$

其中，μ和σ分别是图像的平均值和标准差，σ_{xy}代表的是两图像的协方差。$C_1=\left(K_1L\right)^2$，$C_2=\left(K_2L\right)^2$，K_1和K_2是为了以避免分母接近零而引入的微小常数，通常取值为0.01和0.03，而L则表示图像的动态范围。SSIM值越高，合成图像的感知质量越好，特别是在结构一致性方面。

在本研究中，评估指标用于定量评估sCT与真实CT之间的差异。在测试阶段，每个样本的输出图像与真实CT图像均经过标准化与反标准化处理，以确保结果的一致性与准确性。首先，模型输出的sCT与真实CT图像会进行最小值-最大值归一化，将图像的像素值映射到指定的范围（通常为-1 024~3 000）。在计算评估指标之前，归一化后的图像会被恢复至原始的HU值范围。随后，为确保sCT与真实CT图像在空间上的一致性，所有图像会依据位置信息进行裁剪，只保留ROI。在裁剪后的图像上，评估指标计算仅考虑有效区域的像素，忽略无效区域。

3.3 对比分析

将本文改进模型与传统U-Net模型以及Pix2pix模型进行对比。其定量评估结果如表1所示，相应的可视化效果展示如图6所示。

表1 不同模型生成的sCT与真实CT间的图像质量指标比较

Table 1 Comparison of image quality metrics between ground truth CT and sCT generated by different models

模型	MAE/HU	PSNR/dB	SSIM
Pix2pix	111.83±47.96	25.69±2.53	0.780±0.095
U-Net	100.28±49.67	26.43±2.70	0.808±0.092
本文模型	94.87±42.50	26.85±2.73	0.820±0.085

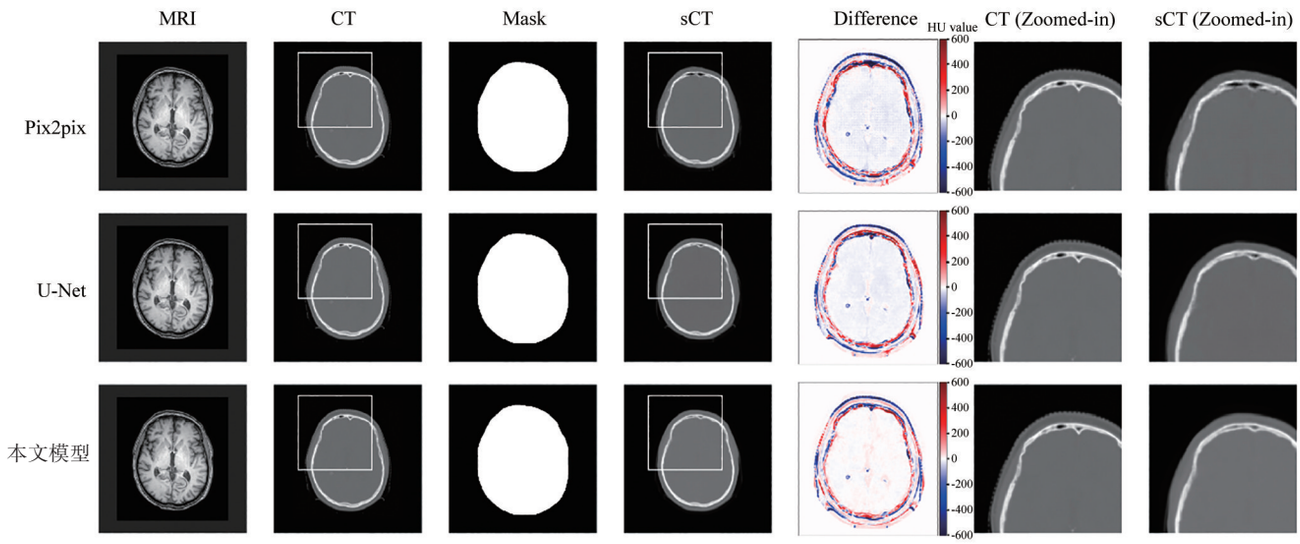

图6 U-Net模型与改进模型所生成的sCT与真实CT的差异比较

Figure 6 Comparison of sCT generated by different models and their differences with the ground truth CT

表1展示39例测试集数据在不同模型下生成的sCT图像质量评价指标的对比。从3个评价指标的平均值上来看，本研究提出的改进模型均优于U-Net与Pix2pix模型，相较于U-Net模型MAE降低5.39%，PSNR提高1.59%，SSIM提高1.49%。从标准差来看，本文模型除了PSNR较高、MAE和SSIM指标均优于U-Net与Pix2pix模型。这说明本研究生成的sCT具有较高的图像质量，噪声更低，与真实CT图像中的解剖结构也更为接近。

通过将脑部MRI生成的sCT结果进行可视化呈现，进一步展示改进模型优于U-Net模型与Pix2pix模型，结果如图6所示。第1列是MRI图像，第2列是真实CT图像，第3列是掩膜图像，第4列是sCT图像，第5列是差值图像，第6列是真实CT的局部放大图像，第7列是sCT的局部放大图像。差值图像中，红色为正差值即sCT相比真实CT偏高部分，白色为差值为零部分，蓝色为负差值即sCT相比真实CT偏低部分。视觉上，改进模型生成的sCT在颅骨与皮质骨边界处表现出更清晰、更锐利的骨质轮廓。差值图像显示改进模型在骨质边界处的高幅度误差斑块较U-Net模型更集中且总体强度略小，表明改进模型在高密度结构的HU恢复上更接近真实CT。

此外，本研究对39例病例在本文模型与U-Net模型的生成结果上，将切片按照病例分类，进行统计学分析。两组比较，MAE、PSNR、SSIM差异均有统计学意义（P<0.05）。MAE的效应大小为-0.119 3，且95%置信区间为（-5.738 8, -5.532 1），显示出本文模型在误差控制上具有明显优势。PSNR、SSIM的效应大小分别为0.153 8、0.137 5，95%置信区间分别为（0.415 5, 0.427 1）、（0.012 1, 0.012 5），显示出本文模型相较于U-Net模型在PSNR、SSIM上有一定效应的提升，表明改进模型的提升是可靠的。

4 结 论

本研究针对基于脑部MRI生成sCT进行探索。围绕MRI-only放疗计划中sCT生成的关键问题，提出一种基于U-Net改进的深度学习模型。实验结果表明，通过引入CoordConv、ConvNeXt Block、Triplet Attention和Cross Attention等模块，本文模型在39例测试数据上的综合性能均优于传统U-Net。具体表现为MAE降低5.39%，PSNR提高1.59%，SSIM提高1.49%，表明本文模型在保持结构细节与纹理特征的同时，有效提升sCT的整体质量与稳定性。这一结果揭示在通过引入空间坐标信息、多维注意力机制和更高效的卷积结构等深度学习方法，可以显著增强模型的特征表达能力、跨模态转换的鲁棒性，和从MRI到CT跨模态映射中的普遍优势，为实现MRI跨模态转换为CT图像提供新思路，为将来进一步开展MRI-Only流程的研究与验证提供方案，并有望在MRI引导的自适应放疗中提升效率和准确性。另外，本研究采用的样本数据是多中心异质性数据，有利于模型的泛化应用。未来研究计划探索其他模型的改进思路。

【参考文献】

[1] Simkó A, Bylund M, Jönsson G, et al. Towards MR contrast independent synthetic CT generation[J]. Z Med Phys, 2024, 34(2): 270-277.

[2] Chourak H, Barateau A, Tahri S, et al. Quality assurance for MRI-only radiation therapy: a voxel-wise population-based methodology for image and dose assessment of synthetic CT generation methods[J]. Front Oncol, 2022, 12: 968689.

[3] Lerner M, Medin J, Jamtheim Gustafsson C, et al. Prospective clinical feasibility study for MRI-only brain radiotherapy[J]. Front Oncol, 2022, 11: 812643.

[4] Yang X, Feng B, Yang H, et al. CNN-based multi-modal radiomics analysis of pseudo-CT utilization in MRI-only brain stereotactic radiotherapy: a feasibility study[J]. BMC Cancer, 2024, 24(1): 59.

[5] Young T, Dowling J, Rai R, et al. Clinical validation of MR imaging time reduction for substitute/synthetic CT generation for prostate MRI-only treatment planning[J]. Phys Eng Sci Med, 2023, 46(3): 1015-1021.

[6] 李芳华, 丁寿亮, 李永宝, 等. 基于Transformer的生成对抗网络用于鼻咽癌MRI生成伪CT[J]. 中国医学物理学杂志, 2025, 42(6): 701-707.
Li FH, Ding SL, Li YB, et al. Synthetic CT generation from NPC MRI using Transformer-based generative adversarial network[J]. Chinese Journal of Medical Physics, 2025, 42(6): 701-707.

[7] Hoffmann A, Oborn B, Moteabbed M, et al. MR-guided proton therapy: a review and a preview[J]. Radiat Oncol, 2020, 15(1): 129.

[8] Albertini F, Matter M, Nenoff L, et al. Online daily adaptive proton therapy[J]. Br J Radiol, 2020, 93(1107): 20190594.

[9] Li X, Bellotti R, Meier G, et al. Uncertainty-aware MR-based CT synthesis for robust proton therapy planning of brain tumour[J]. Radiother Oncol, 2024, 191: 110056.

[10] 邵虹, 荆一烜, 崔文成. 基于扩散生成对抗网络的核磁共振图像与计算机断层扫描图像跨模态转换[J]. 生物医学工程学杂志, 2025, 42(3): 575-584.
Shao H, Jing YX, Cui WC. Cross modal translation of magnetic resonance imaging and computed tomography images based on diffusion generative adversarial networks[J]. Journal of Biomedical Engineering, 2025, 42(3): 575-584.

[11] Thummerer A, van der Bijl E, Galapon A Jr, et al. SynthRAD2023 grand challenge dataset: generating synthetic CT for radiotherapy[J]. Med Phys, 2023, 50(7): 4664-4674.

[12] El Jurdi R, Petitjean C, Honeine P, et al. CoordConv-Unet: investigating CoordConv for organ segmentation[J]. IRBM, 2021, 42(6): 415-423.

[13] Liu MK, Zhuo YT, Liu J, et al. Real-time estimation of lung deformation from body surface using a general CoordConv CNN[J]. Comput Methods Programs Biomed, 2024, 244: 107998.

[14] Yin AC, Ren C, Yue WT, et al. CDAU-Net: a novel CoordConv-integrated deep dual cross attention mechanism for enhanced road extraction in remote sensing imagery[J]. Remote Sens (Basel), 2023, 15(20): 4914.

[15] Fan XS, Ding WT, Qin WL, et al. Fusing self-attention and CoordConv to improve the YOLOv5s algorithm for infrared weak target detection[J]. Sensors (Basel), 2023, 23(15): 6755.

[16] Liu Z, Mao HZ, Wu CY, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 11966-11976.

[17] Song JH, Nie XB, Wu C, et al. A novel intelligent fault diagnosis method of rolling bearings based on the ConvNeXt network with improved DenseBlock[J]. Sensors (Basel), 2024, 24(24): 7909.
[18] Liu SK, Cao S, Lu X, et al. Lightweight deep learning model, ConvNeXt-U: an improved U-Net network for extracting cropland in complex landscapes from Gaofen-2 images[J]. Sensors (Basel), 2025, 25(1): 261.
[19] Wei FM, Wang X. Sar ship detection based on convnext with multi-pooling channel attention and feature intensification pyramid network [J]. Sensors (Basel), 2023, 23(17): 7641.
[20] Dai JF, Qi HZ, Xiong YW, et al. Deformable convolutional networks [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 764-773.
[21] Misra D, Nalamada T, Arasanipalai AU, et al. Rotate to attend: convolutional triplet attention module [C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2021: 3138-3147.
[22] Tang L, Yi JZ, Li XY. Improved multi-scale inverse bottleneck residual network based on triplet parallel attention for apple leaf disease identification[J]. J Integr Agric, 2024, 23(3): 901-922.
[23] Li Y, Yan BB, Hou JX, et al. UNet based on dynamic convolution decomposition and triplet attention[J]. Sci Rep, 2024, 14(1): 271.
[24] Li XX, Xue SY, Xie JY, et al. Interactive triplet attention for few-shot fine-grained image classification[J]. Neurocomputing, 2025, 655: 131377.
[25] Meng WJ, Shan LL, Ma SG, et al. DLNet: a dual-level network with self-and cross-attention for high-resolution remote sensing segmentation[J]. Remote Sens (Basel), 2025, 17(7): 1119.
[26] Li FD, Lu XY, Yuan JJ. MHA-CoroCapsule: multi-head attention routing-based capsule network for COVID-19 chest X-ray image classification[J]. IEEE Trans Med Imaging, 2022, 41(5): 1208-1218.
[27] Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans Image Process, 2004, 13(4): 600-612.

（编辑：薛泽玲）