

# 一种用于低剂量CT的微小细节保护CNN与Transformer融合去噪方法

李晓增<sup>1</sup>, 王宝珠<sup>1</sup>, 郭志涛<sup>1,2</sup>, Shanaz Sharmin Jui<sup>1</sup>

1. 河北工业大学电子信息工程学院, 天津 300401; 2. 河北工业大学创新研究院(石家庄), 河北 石家庄 050299

**【摘要】**为解决低剂量CT图像因辐射剂量降低而引入大量噪声,导致图像质量下降,从而影响临床诊断准确性问题,构建一种结合卷积神经网络(CNN)与Transformer的网络模型,并在此模型中引入一种内部块特征提取模块,以更好地保护图像中的微小细节。此外,为了解决应用Swin Transformer去噪时出现恢复错误纹理细节的问题,在自注意力部分并入一个多输入通道注意力模块,进而构建一种双重注意力Transformer。本研究在AAPM数据集上进行测试,实验结果表明,与现有的去噪算法相比,本文提出的算法在去噪方面表现出色,可以更好地保护图像的微小细节。

**【关键词】**低剂量CT; 图像去噪; 深度学习; 微小细节保护

**【中图分类号】**R318

**【文献标志码】**A

**【文章编号】**1005-202X(2024)07-0842-09

## Low-dose CT denoising method with CNN and Transformer to preserve tiny details

LI Xiaozeng<sup>1</sup>, WANG Baozhu<sup>1</sup>, GUO Zhitao<sup>1,2</sup>, Shanaz Sharmin Jui<sup>1</sup>

1. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; 2. Innovation and Research Institute of Hebei University of Technology in Shijiazhuang, Shijiazhuang 050299, China

**Abstract:** Given that low-dose computed tomography significantly amplifies image noise due to the mitigation of radiation exposure, which degrades image quality and lowers the precision of clinical diagnoses, a novel model incorporating convolutional neural network and Transformer is established, in which an intra-patch feature extraction module is used to effectively preserve tiny details in the image. A double attention Transformer is constructed by incorporating a multiple-input channel attention module into the self-attention for tackling the problem of incorrect restoration of texture details during denoising using Swin Transformer. AAPM dataset is used for testing, and the results demonstrate that the proposed algorithm not only surpasses the existing algorithms in denoising performance, but also excels in preserving tiny details in the image.

**Keywords:** low-dose computed tomography; image denoising; deep learning; tiny detail preservation

## 前言

计算机断层扫描(Computed Tomography, CT)成像技术已经成为现代临床医学影像诊断的重要工具,但是常规的CT扫描放射性较高,对患者的身体健康会产生一定的影响。因此,亟需降低CT扫描的

辐射剂量。低剂量CT(Low-Dose Computed Tomography, LDCT)虽然成功降低辐射的危害,但图像存在显著的噪声、伪影以及低对比度等问题,对临床诊断造成干扰。如何改善LDCT图像的成像质量,已经成为医学影像处理领域的一个研究热点。为了提高LDCT图像质量,许多算法被开发出来。通常来说,这些算法分为3大类:投影域滤波<sup>[1]</sup>、迭代重建<sup>[2-3]</sup>、图像后处理降噪<sup>[4-5]</sup>。投影域滤波和迭代重建所依赖的原始投影数据往往难以获取,这使得相关的研究难以有效进行。另一方面,后处理方法直接处理重建后的LDCT图像,不需要依赖原始投影数据,因此对研究工作带来极大的便利。然而,传统的后处理方法不能完全解决处理后图像中存在的过平滑、引入新噪声等问题。

**【收稿日期】**2023-12-22

**【基金项目】**河北省高等学校科学技术研究项目(ZD2022115);河北工业大学创新研究院(石家庄)石家庄市科技合作专项基金(SJZZXB23005)

**【作者简介】**李晓增,硕士,研究方向:医学图像处理、计算机视觉, E-mail: lixiaozeng2023@163.com

**【通信作者】**郭志涛,博士,教授,研究方向:医学图像处理、深度学习, E-mail: mrnow@hebut.edu.cn

近年来,深度学习的迅速发展为医学成像领域中 LDCT 图像的增强提供全新的解决思路。Chen 等<sup>[6]</sup>对深度学习用于 LDCT 图像去噪的研究实现重要突破,首次提出用于降低 LDCT 图像噪声的卷积神经网络(Convolutional Neural Network, CNN)。此后,基于 CNN 的 LDCT 图像去噪方法应运而生。Gondara<sup>[7]</sup>提出的模型证明编码器-解码器模型在医学图像去噪方面具有优秀的性能。Chen 等<sup>[8]</sup>在先前研究成果<sup>[6]</sup>的基础上提出一种去噪模型 RED-CNN,巧妙地将快捷连接引入到编码器-解码器网络中。Gholizadeh-Ansari 等<sup>[9]</sup>设计一种采用扩展卷积代替标准卷积的网络。Liang 等<sup>[10]</sup>设计一种具有密集连接的去噪模型,能有效地通过后处理的方式实现 LDCT 图像去噪。此外,一些研究选择将 CNN 与解析重建或迭代重建相结合,进一步提升 LDCT 图像的质量<sup>[11-12]</sup>。近期,Transformer 的卓越性能鼓励研究人员将其引入到 LDCT 图像去噪的工作中,从而取得一系列显著的成果。CTformer 是首个将纯粹的 Transformer 架构应用于 LDCT 去噪的案例<sup>[13]</sup>。TED-Net 则运用无卷积 Transformer 结构,在编码器-解码器结构中运用令牌化和去令牌化的操作<sup>[14]</sup>。Eformer 结合 Transformer 的全局建模能力和残差学习方法,创新性地提出一种残差 Transformer,并利用边缘增强模块来生成高质量的去噪图像<sup>[15]</sup>。

尽管已经有许多模型和算法应用 LDCT 图像去噪,但这个任务并未完全解决,现有的模型也存在一些问题。例如,许多现行的算法基于纯卷积的 CNN 方法,这类方法主要依赖级联的卷积层来提取特征。然而,卷积操作的感受野有限,只能感知局部区域<sup>[16]</sup>,这使得基于 CNN 的方法虽在局部有优秀的特征提取能力,但难以有效地对 LDCT 图像的复杂组织结构进行全局建模<sup>[17]</sup>。基于自注意力机制的 Transformer 结构能有效克服这一限制<sup>[18]</sup>,但其计算复杂度与图像大小成二次方关系,这使得标准 Transformer 的计算量过于庞大<sup>[19]</sup>。一些研究成果表明,通过合理地将 CNN 与 Transformer 结合,整合二者的优势,能够显著提高模型的去噪性能<sup>[20-22]</sup>。此外,现有方法在保护 LDCT 图像中微小细节方面还存在不足,如边缘模糊、纹理细节丢失、纹理细节恢复错误等问题,而这些细节信息在临床诊断中至关重要。因此,如何确保 LDCT 图像中微小细节的保护,提升 LDCT 数据去噪质量,依然是研究人员需要着力解决的关键问题。

为应对上述问题,本文提出一种将残差卷积与 Swin Transformer 相结合的 LDCT 去噪网络(RSNet)。RSNet 融合了残差卷积在局部特征提取方面的优势以及 Swin Transformer 在全局特征提取方面的优势,

具备出色的去噪性能。同时,本文还提出一种强化微小细节保护的编码器结构,在编码器中加入一种内部块特征提取模块,以增强对图像内微小细节的特征提取能力。通过实验证明,本文提出的网络模型能够有效提升 LDCT 图像的质量。

## 1 网络结构设计

### 1.1 去噪模型

深度学习方法进行 LDCT 去噪的过程中,其目标是得到与正常剂量目标图像尽可能接近的结果。假设  $X \in \mathbf{R}^{m \times n}$  为 LDCT 图像,  $Y \in \mathbf{R}^{m \times n}$  是相应的常规剂量 CT (Normal Dose Computed Tomography, NDCT) 图像,它们的关系表示为:

$$X = \sigma(Y) \quad (1)$$

其中,  $\sigma$  表示图像的降质过程。由 LDCT 图像  $X$  恢复为 NDCT 图像  $Y$  的过程,可以理解为通过网络训练得到一个函数  $F$ ,使得  $F$  近似于  $\sigma^{-1}$ ,并且满足以下条件:

$$f = \arg \min_f \|F(X) - Y\|_2^2 \quad (2)$$

其中,  $f$  表示使  $\|F(X) - Y\|_2^2$  最小的自变量取值。

### 1.2 整体网络设计

本文提出的模型是一种 U 型的层次网络,编码器和解码器之间存在跳跃连接,如图 1 所示。对于输入的 LDCT 图像  $X \in \mathbf{R}^{C \times H \times W}$ ,首先经过一个  $3 \times 3$  卷积层提取底层特征,然后进入编码器阶段。在编码器中,每个阶段都包含若干个残差卷积与 Swin 融合模块(ResConv-Swin Block, RSB)和一个下采样层,并行地在每个编码器阶段中都存在若干个内部块特征提取模块(Intra-Patch Feature Extraction Module, IPFEM),RSB 和 IPFEM 的结构如图 1a 和图 1b 所示。在每个 RSB 中,都包含一个双重注意力 Transformer 模块(Double Attention Transformer, DAT)和一个残差卷积模块<sup>[23-24]</sup>,它们通过两个  $1 \times 1$  卷积、拆分、连接以及残差连接操作进行融合。在下采样部分,使用步长为 2 的  $4 \times 4$  卷积对特征图进行下采样,使得输出的特征尺寸减半,通道数增加一倍。之后,在编码器的末端添加一个带有 RSB 的瓶颈层,在此阶段 RSB 能够捕捉更长距离的依赖关系。

特征重建阶段由数量相当于编码器阶段的解码器组成,每一阶段都包含一个上采样层和若干个与编码器相同的 RSB。在上采样部分,选择使用步长为 2 的  $2 \times 2$  反卷积用于上采样使得上采样后的特征尺寸加倍,同时通道数减半。输入到 RSB 中的特征源自编码器输出特征的跳跃连接以及上采样特征。经过若干个解码器阶段处理后,使用一个  $3 \times 3$  的卷积生成残差图像  $R$ 。最后,将 LDCT 图像与残差图像相加,得到恢复图像  $Y$ ,即  $Y = X + R$ 。

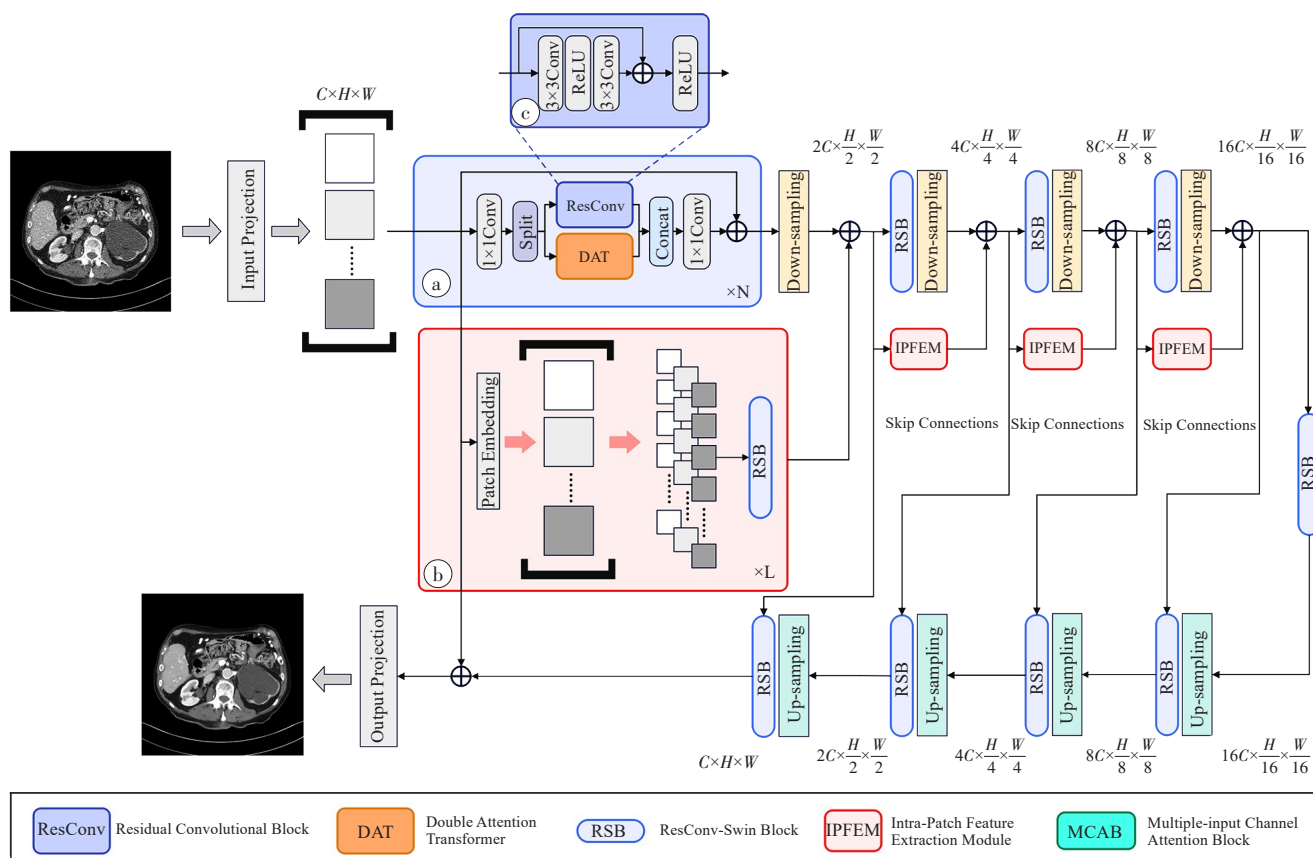


图1 RSNet网络结构图

Figure 1 RSNet structure

### 1.3 残差卷积与Swin融合模块

RSB的结构如图1a所示,RSB通过一个 $1 \times 1$ 卷积层和拆分操作,将输入特征 $X$ 平均分为两个特征,分别为 $X_1$ 和 $X_2$ ,这一过程表示为:

$$X_1, X_2 = \text{Split}(\text{Conv}1 \times 1(X)) \quad (3)$$

将 $X_1$ 和 $X_2$ 作为输入特征,分别送入一个残差卷积块和一个双注意力Transformer模块中,从而得到:

$$Y_1, Y_2 = \text{ResConv}(X_1), \text{DAT}(X_2) \quad (4)$$

其中,ResConv表示残差卷积块,如图1c所示。ResConv由数个卷积层组成,每个卷积层后都有一个激活函数,而且它的输入和输出之间存在残差连接。DAT是双注意力Transformer模块,在后面的部分中将会对其进行详细介绍。最后,将 $Y_1$ 和 $Y_2$ 输入到另一个 $1 \times 1$ 的卷积层中,该卷积层的输出与 $X$ 之间存在残差连接,因此RSB的最终输出为:

$$Z = \text{Conv}1 \times 1(\text{Concat}(Y_1, Y_2)) + X \quad (5)$$

本文提出的RSB融合残差卷积的局部建模能力和Swin Transformer的非局部建模能力,两次卷积操作可以有效促进残差卷积块和DAT之间的信息融合。

### 1.4 内部块特征提取模块

为了增强模型对输入特征中微小细节特征的提取能力,本文在RSB的基础上设计一个名为内部块

特征提取模块的结构,并将其融入到编码器中。如图1b所示,IPFEM模块首先会对输入特征进行图像块嵌入操作,将其切分为大小为原来高度和宽度一半的内部块,之后将这些内部块输入到多个RSB中。当原始输入特征较大时,RSB往往无法过度关注到特征中的细微部分。而在每个IPFEM里,RSB可以将更多的注意力集中到每一个输入特征的内部,从而能够更加精准地捕捉其中的微小细节信息。每个IPFEM都与一个编码器主块并行,除了第一阶段,IPFEM的输出特征都会与相应的编码器的输出特征相加,之后输入到模型的下一阶段。因此,编码器的每个阶段可以表示为:

$$Z_i = \text{RSB}_i(X_i) + \text{RSB}'_i(P(X_i)) \quad (6)$$

其中, $X_i$ 表示当前阶段编码器的输入特征,RSB<sub>i</sub>表示编码器主块,RSB'<sub>i</sub>表示与当前编码器主块并行IPFEM中的融合模块, $P$ 表示根据输入特征图创建内部特征图的过程。

### 1.5 双重注意力Transformer模块

本文提出一种双重注意力Transformer模块。传统的Transformer模型在提取图像特征时,强烈依赖于在图像全局范围内进行自注意力计算<sup>[18]</sup>。然而,这种计算方式由于与输入图像大小成二次关系,因



此计算复杂度较高。为了解决这个问题, Swin 在传统 Transformer 的基础上交替使用 Window Attention (W-MSA) 和 Shifted Window Attention (SW-MSA) 来实现注意力机制的运算, 这种滑动窗口机制能大幅度节省计算量<sup>[23]</sup>。但由于滑动窗口的机制所利用的像素范围有限, Swin Transformer 在构建去噪网络时可能会恢复出错误的纹理特征。因此, 如果模型能够利用更多的输入像素, 这一问题将得到改善<sup>[21,25]</sup>。

在使用通道注意力时,会利用全局信息计算每个特征通道的权重,这一操作能够激活输入特征更多的像素,并且卷积操作会提升Transformer的效果。因此,将通道注意力的卷积块加入到Swin Transformer结构中,可以增强网络的表达能力。但传统的通道注意力卷积块通常由多个堆叠在一起的卷积层和一个注意力层组成,而注意力层通常只接受最后一个卷积层的输出,导致生成的注意力权重与之前的卷积层输出缺乏关联<sup>[26]</sup>。因此,本文提出一种多输入通道注意力的卷积模块(Multiple-input Channel Attention Block, MCAB)。如图2所示,输入特征在MCAB中,首先会经过两个步长为1的 $3\times 3$ 卷积层和一个激活函数。为了节省计算成本,在第一个卷积层中将输出特征的通道数压缩为原来的 $1/3$ ,在第二个卷积层中再将其扩展到原有通道数。之后,将激活函数和第二个卷积层的输出分别输入到多输入通道注意力层(Multiple-input Channel Attention, MCA)中。在MCA中,首先通过全局平均池化得到两个输入特征的整体信息,之后分别经过一个 $1\times 1$ 卷积和一个BN层,将两者得到的结果相加后再经过激活函数、卷积层和Sigmoid非线性函数。最后,将得到的特征与卷积部分中最后一个卷积层的输出相乘,得到通道的特征向量。

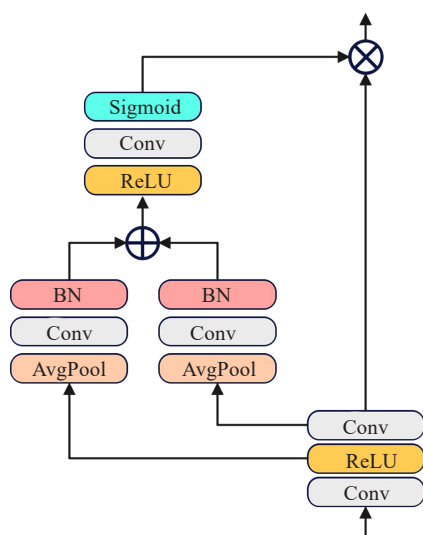


图2 MCAB 结构图

Figure 2 MCAB architecture

DAT的结构如图3所示,MCAB被插入到LN层之后,与W-MSA和SW-MSA并行,同时为了避免MCAB和W-MSA、SW-MSA之间发生优化上的冲突,在MCAB的输出中会乘一个小的常量。整个DAT的计算过程如下:

$$\mathbf{X}_{\text{LN}} = \text{LN}(\mathbf{X}_{\text{in}}) \quad (7)$$

$$\mathbf{X}_l = \text{MSA}(\mathbf{X}_{\text{LN}}) + \alpha \text{MCAB}(\mathbf{X}_{\text{LN}}) + \mathbf{X}_{\text{in}} \quad (8)$$

$$\mathbf{X}_{\text{out}} = \mathbf{X}_1 + \text{MLP}(\text{LN}(\mathbf{X}_1)) \quad (9)$$

其中,  $X_{LN}$  和  $X_I$  表示中间特征, LN 表示层归一化,  $X_{out}$  为 DAT 的输出, MLP 表示多层感知器。

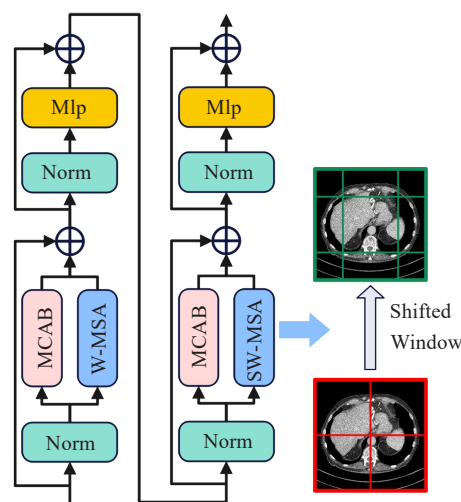


Figure 3 DAT architecture

在 W-MSA 和 SW-MSA 中, 输入的特征图  $\mathbf{X} \in \mathbf{R}^{C \times H \times W}$  被划分为  $HW/M^2$  个不重叠的局部窗口, 每个窗口大小为  $M \times M$ , 然后在每个窗口内计算自注意力。对于局部窗口特征  $\mathbf{X}_w \in \mathbf{R}^{M^2 \times C}$ ,  $\text{query}(\mathbf{Q})$ 、 $\text{key}(\mathbf{K})$  和  $\text{value}(\mathbf{V})$  矩阵计算为:

$$\mathbf{Q} = \mathbf{X}_w \times \mathbf{P}_o \quad (10)$$

$$\mathbf{K} = \mathbf{X}_W \times \mathbf{P}_K \quad (11)$$

$$\mathbf{V} = \mathbf{X}_w \times \mathbf{P}_V \quad (12)$$

其中,  $P_Q$ 、 $P_K$ 、 $P_V$  是投影矩阵, 并且  $Q$ 、 $K$ 、 $V \in R^{M^2 \times C}$ , 注意力矩阵在窗口中的自注意力机制计算为:

$$\text{Attention} = \text{SoftMax}\left(\frac{\underline{Q}\mathbf{K}^T}{\sqrt{d}} + B\right)\mathbf{V} \quad (13)$$

其中,  $d$  为  $\mathbf{QK}$  的维度,  $B$  表示可学习的相对位置编码。此外, 为了相邻非重叠窗口之间能够进行有效的信息交互, 将 W-MSA 和 SW-MSA 交替使用, 并且将位移大小设置为窗口大小的一半。



## 1.6 复合损失函数

作为优化过程的一部分,实验中使用 Charbonnier 损失来优化本文提出的网络结构,Charbonnier 损失测量模型输出与 NDCT 图像之间的差异,其定义为:

$$L_{\text{char}} = \sqrt{\|F(X_i) - Y\|^2 + \varepsilon^2} \quad (14)$$

其中,  $X_i$  为输入图像,  $F$  表示去噪过程,  $Y$  是常规剂量图像,  $\varepsilon$  是一个极小的常量,通常为  $10^{-3}$ 。

通过实验和已有研究表明,仅依赖单一的损失函数会导致输出图像过于平滑,使图像更加模糊<sup>[14,19]</sup>。因此,本文在 Charbonnier 损失上融入一种多尺度感知损失,形成一种复合损失函数。由于 ResNet-50 能够提取图像的深层次信息,并且引入残差学习机制,使得在计算感知损失时,ResNet-50 能够更稳定地学习图像的特征<sup>[27]</sup>。因此,本文选用

ResNet-50 作为特征提取器来获取多尺度感知损失。具体操作就是删除 ResNet-50 末尾的池化层和全连接层,仅保留模型前面的卷积层。在计算感知损失时,如图 4 所示,将恢复图像和 NDCT 图像输入到特征提取器中进行前向传播,然后使用 4 个阶段后的特征计算 Charbonnier 损失,最后对这些损失值求平均,从而得到多尺度感知损失。其表达式如下:

$$L_{\text{per}} = \frac{1}{M} \sum_{j=1}^M \sqrt{\|\Phi_j(F(X_i)) - \Phi_j(Y)\|^2 + \varepsilon^2} \quad (15)$$

其中,  $X_i$  为输入图像,  $F$  表示去噪过程,  $Y$  是常规剂量图像,  $\Phi$  表示带有固定权重预训练的 ResNet-50 模型。将 Charbonnier 损失与多尺度感知损失相结合得到复合损失函数,如下所示:

$$L_{\text{compound}} = L_{\text{char}} + \lambda L_{\text{per}} \quad (16)$$

其中,通过调整超参数  $\lambda$  调整多尺度感知损失在复合损失中的权重。

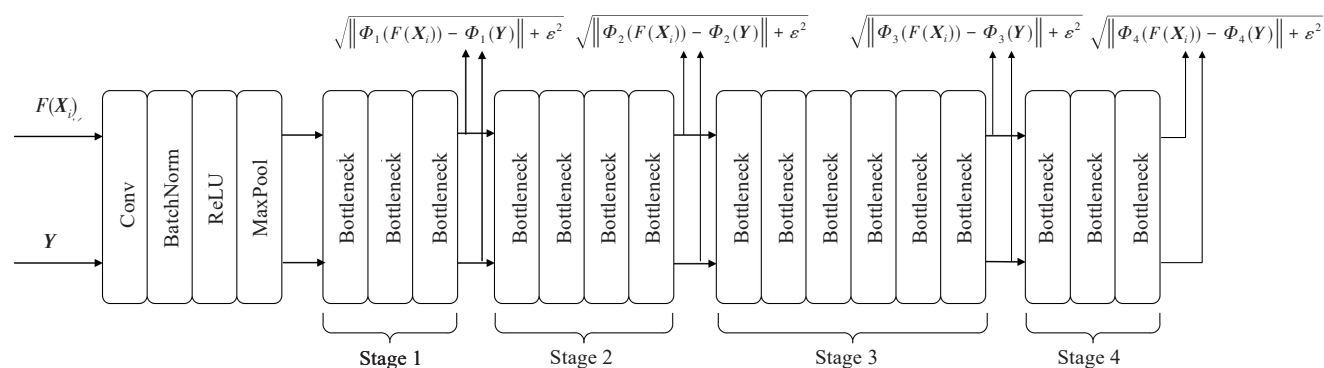


图4 基于 ResNet-50 的多尺度感知损失

Figure 4 Multi-scale perceptual loss based on ResNet-50

## 2 实验设计与结果

### 2.1 实验环境设置和评价指标

**2.1.1 数据集的构建** 实验中使用的数据集来自 Mayo Clinic 授权的“2016 年 NIH-AAPM-Mayo 诊所 LDCT 大挑战”的临床数据集<sup>[28]</sup>,其中包含来自 10 例患者的 2378 张正常剂量 CT 图像及其对应的低剂量(四分之一剂量)CT 图像,每张图像的大小为 512×512 像素。在实验中,选择来自 8 例患者的 2039 张图像作为训练集,1 例患者的 128 张图像作为验证集,以及另一例患者的 211 张图像作为测试集。

**2.1.2 网络训练环境和参数设定** 实验通过 PyTorch 深度学习框架实现,所使用的 CPU 型号是 Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz,内存大小为 16 GB,使用的显卡型号是 NVIDIA GeForce GTX 1650 Ti,显存大小为 4 GB。在网络训练过程中使用

AdamW 优化器对损失函数进行优化,参数 betas 设为 0.9,初始学习率设为  $2 \times 10^{-4}$ ,随着迭代次数增加逐渐降为  $1 \times 10^{-6}$ ,调优方法采用 Cosine Warmup,训练次数为 100 个 epoch,训练过程中 batch size 设为 4。

**2.1.3 图像质量评价指标** 实验中使用峰值信噪比 (PSNR)、结构相似性 (SSIM) 和特征相似度 (FSIM) 3 种图像质量评价指标对实验中使用的模型进行定量比较。

PSNR 和 SSIM 表示为:

$$R_{\text{PSNR}} = 20 \log_{10} \frac{I_{\text{max}}}{E_{\text{RMSE}}} \quad (17)$$

$$M_{\text{SSIM}} = \frac{(2\mu_y + \mu_x + c_1)(2\delta_{yx} + c_2)}{(\mu_y^2 + \mu_x^2 + c_1)(\delta_y^2 + \delta_x^2 + c_2)} \quad (18)$$

其中,  $I_{\text{max}}$  表示最大像素值,  $E_{\text{RMSE}}$  表示均方根误差;  $x$  和  $y$  分别表示 LDCT 图像和恢复图像;  $\mu_x$ 、 $\mu_y$  分别表示

$x$ 和 $y$ 的均值; $\delta_x^2$ 和 $\delta_y^2$ 表示 $x$ 和 $y$ 的方差, $\delta_{yx}$ 表示 $x$ 和 $y$ 的协方差; $c_1$ 和 $c_2$ 分别是定义为 $(0.01 \times I_{\max})^2$ 和 $(0.03 \times I_{\max})^2$ 的两个默认SSIM参数。

FSIM表示为:

$$F_{\text{FSIM}} = \frac{\sum_{x \in \Omega} S_L PC_m}{\sum_{x \in \Omega} PC_m} \tag{19}$$

其中, $S_L$ 表示恢复图像和LDCT图像之间相位一致性(PC)和特征梯度(GM)融合的相似度, $PC_m$ 表示恢复图像的PC和LDCT图像的PC中的最大值。

2.2 对比实验及分析

实验选取EDCNN<sup>[10]</sup>、CTformer<sup>[13]</sup>、Restormer<sup>[29]</sup>、MPRNet<sup>[30]</sup>、Uformer<sup>[31]</sup>、NAFNet<sup>[32]</sup>、DDT<sup>[33]</sup>算法以及LDCT图像、NDCT图像与本文所提出的网络方法

(RSNet)进行比较。在实验中,所有参与比较的算法参数均按照各自推荐的最佳值进行设定。图5展示了将测试集中的一张具有复杂图像细节的CT图像作为测试样本1,通过不同算法进行去噪处理得到的对比结果,其中图5a是LDCT图像,图5f是NDCT图像,其余分别是其他算法和RSNet的结果。图6是图5中局部的感兴趣区域(ROI)的放大对比结果。对比LDCT图像与NDCT图像,可以发现LDCT图像中充斥着大量噪声和伪影,边缘模糊不清,纹理细节难以辨识,而相应的NDCT图像纹理清晰,病灶边缘更加清晰。与LDCT图像比较,RSNet能够有效去除噪声,显著提升图像质量,恢复图像与NDCT图像相似。通过与其他算法对比可知,RSNet的去噪效果最为显著,能够有效保护病灶的细节信息,图像轮廓也最为清晰。

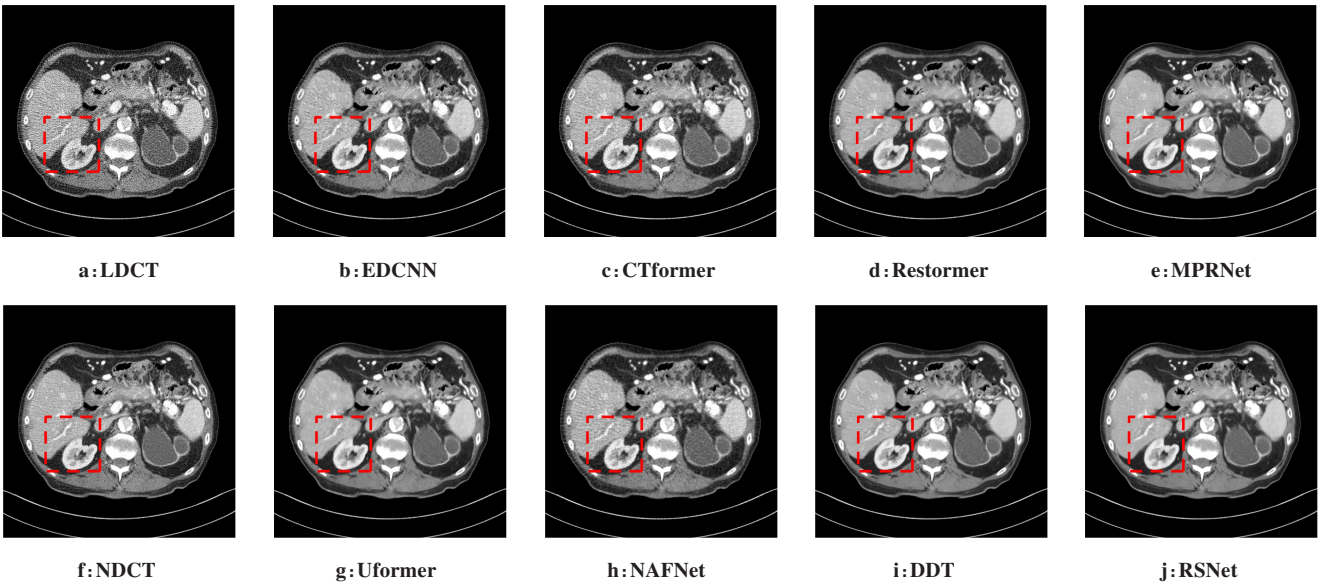


图5 测试样本1整体对比图  
Figure 5 Overall comparison of test sample 1

选取测试集中一张具有相似背景结构的图像作为测试样本2,图7展示了使用该样本进行降噪处理后的整体视觉对比效果,图8展示了图7中ROI区域放大的比较结果。通过图7的整体视觉对比,在背景结构相似的LDCT图像中,RSNet成功地消除噪声,使得图像质量得到显著提升。由图8的ROI放大图可以看出,RSNet面对具有相似背景结构的LDCT图像时,同样具有良好的去噪效果,并且能够较好地保护细节信息。

表1展示了测试样本1和测试样本2的客观评估结果。无论是处理拥有复杂结构的图像,还是处理背景结构相似的图像,本文提出的方法都能够表现出最佳的降噪效果。表2则呈现了不同算法对测试集数据进行去噪处理后的平均结果。相较于其他的

LDCT图像去噪模型,本文所提出的模型在3种评价指标上均实现不同程度的提升。无论是在整体图像,还是在局部的ROI上,RSNet都展示出了超越其他方法的去噪能力,并且具备优秀的细节保护能力。

2.3 消融实验及分析

2.3.1 不同网络结构对性能的影响 表3对比了RSNet中不同结构对于增强模型去噪效果的贡献。首先,从模型中移除IPFEM,在保持其他部分不变的情况下,选择使用等量的Swin Transformer层来替换网络结构中的RSB,从而形成一个U型的基准网络(Baseline)。紧接着,将Baseline中的Swin Transformer层替换为DAT,以验证DAT对模型性能提升的作用。在采用DAT后,3种评价指标均有所上升,这证实了DAT结构的引入能



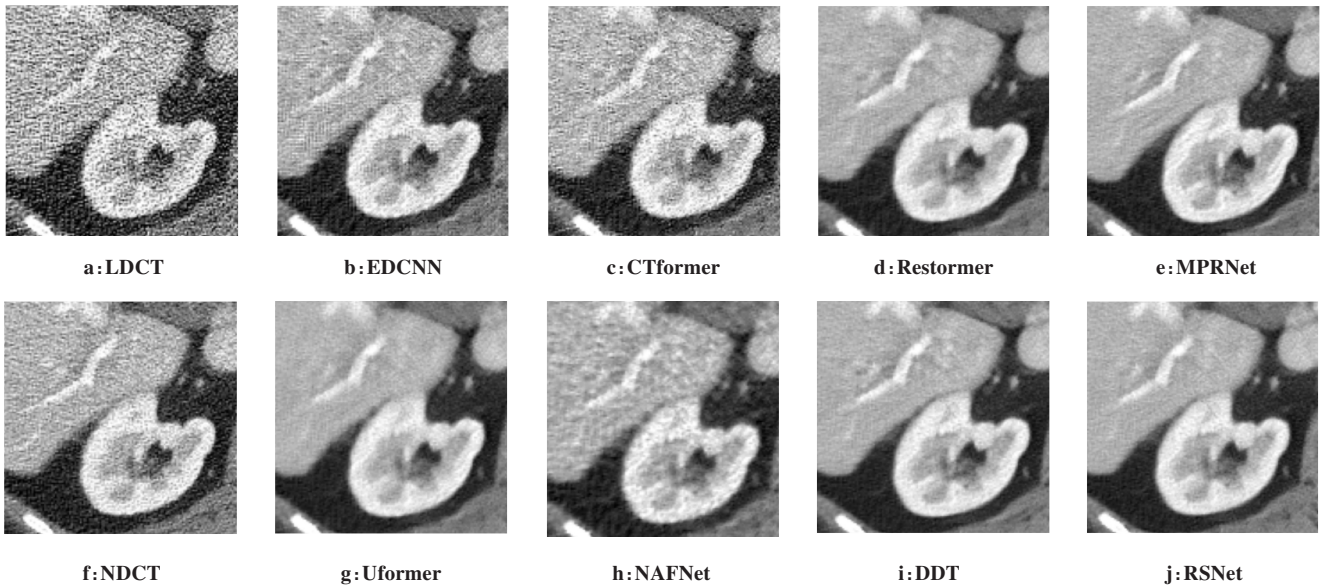


图 6 测试样本 1 的 ROI 区域放大对比图  
Figure 6 Comparison of enlarged ROI area of test sample 1

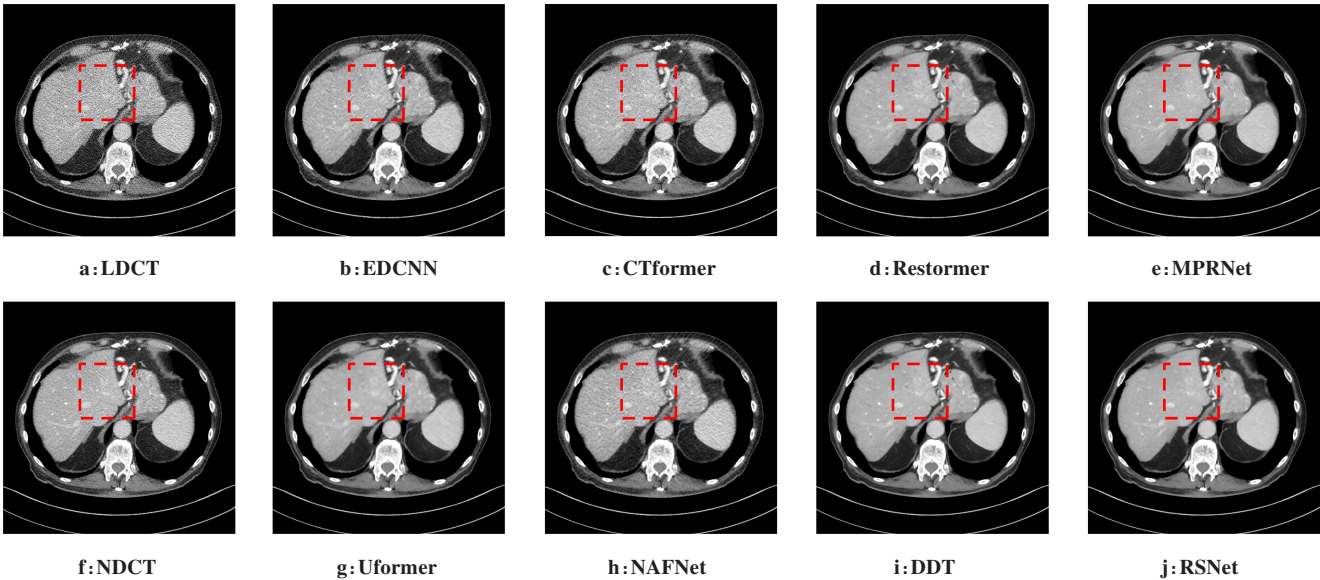


图 7 测试样本 2 整体对比图  
Figure 7 Overall comparison of test sample 2

有效提升模型的性能。随后,在 Baseline 基础上,使用 Swin Transformer 层与残差卷积块融合的结构,称之为“Use RSB without MCAB”。与 Baseline 相比,CNN 和 Transformer 的融合使得性能指标得到显著提升。进一步使用本文所提出的完整 RSB 结构替代 Baseline 中的 Swin Transformer 层进行实验,实验结果显示,与 Baseline 相比去噪效果有所提升。最后,将 IPFEM 模块加入模型中实验,评价指标均达到最佳水平。

**2.3.2 MCAB 与通道注意力模块对比** 表 4 展示了将 DAT 中的 MCAB 替换为标准的 CAB 或不使用 CAB

的实验对比结果。其中,CAB 与 MCAB 在卷积层、BN 层、池化层以及激活函数的数量和参数设置上均保持一致。结果显示,采用 MCAB 的模型在去噪效果上表现最佳。这是因为在 MCAB 中,注意力层接收所有卷积层的输出,从而在最大程度上获取输入特征的信息,激活更多像素,更有效地解决 Swin Transformer 结构恢复错误纹理细节的问题。

**2.3.3 不同损失函数的比较** 表 5 展示了使用均方误差(MSE)损失、Charbonnier 损失和复合损失函数后模型的去噪性能对比。结果显示,相较于单独使用



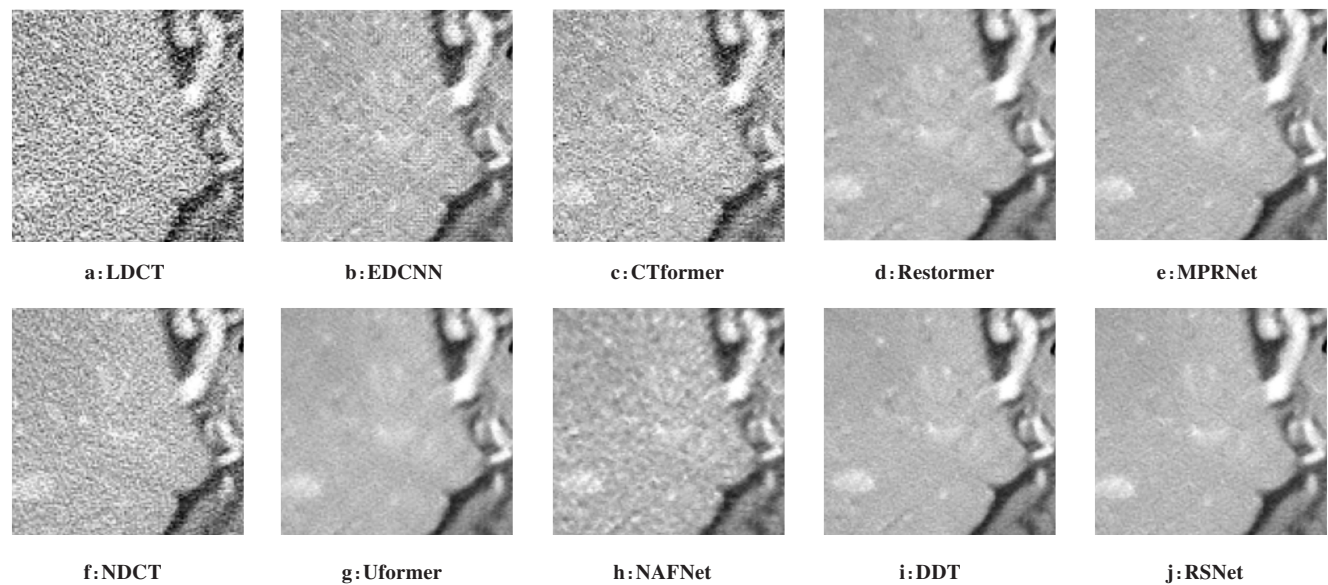


图8 测试样本2的ROI区域放大对比图

Figure 8 Comparison of enlarged ROI area of test sample 2

表1 测试样本1和测试样本2的客观评估结果

Table 1 Objective indicators of test sample 1 and test sample 2

方法	测试样本 1			测试样本 2		
	PSNR/dB	SSIM	FSIM	PSNR/dB	SSIM	FSIM
LDCT	18.014 5	0.714 6	0.603 7	18.562 8	0.741 0	0.626 7
EDCNN	22.426 5	0.766 1	0.638 5	22.869 3	0.782 3	0.645 4
CTformer	21.854 8	0.741 0	0.630 4	22.261 3	0.765 2	0.643 5
Restormer	23.844 3	0.766 7	0.645 3	24.302 8	0.779 0	0.655 2
MPRNet	23.877 5	0.763 8	0.652 7	24.331 6	0.784 4	0.666 4
Uformer	23.857 7	0.765 3	0.649 5	24.315 5	0.777 1	0.663 8
NAFNet	23.506 2	0.759 9	0.624 3	23.935 0	0.775 8	0.629 2
DDT	23.909 4	0.769 5	0.649 4	24.360 0	0.780 0	0.663 7
RSNet	23.914 0	0.770 8	0.652 9	24.363 3	0.792 6	0.668 0

表2 不同算法实验结果对比

Table 2 Comparison of experimental results obtained with different algorithms

方法	PSNR/dB	SSIM	FSIM
LDCT	21.604 8	0.801 7	0.648 1
EDCNN	25.871 3	0.838 3	0.667 8
CTformer	25.812 1	0.836 2	0.663 0
Restormer	27.168 0	0.841 4	0.676 4
MPRNet	27.220 2	0.844 3	0.680 7
Uformer	27.179 5	0.839 6	0.678 5
NAFNet	26.712 3	0.831 3	0.662 7
DDT	27.279 6	0.843 9	0.680 7
RSNet	27.294 3	0.845 5	0.683 5

表3 模型中不同结构对降噪性能的影响

Table 3 Effects of different structures in the model on denoising performance

方法	PSNR/dB	SSIM	FSIM
Baseline	27.157 1	0.839 5	0.680 8
Use DAT	27.258 8	0.841 3	0.681 3
Use RSB without MCAB	27.276 1	0.842 3	0.682 5
Use RSB	27.289 0	0.842 5	0.682 6
Use RSB+IPFEM (RSNet)	27.294 3	0.845 5	0.683 5

MSE 损失或 Charbonnier 损失, 本文采用的复合损失函数能够获得更优秀的去噪效果。

表4 不同通道注意力模块实验对比结果

Table 4 Comparison of experimental results using different channel attention modules

方法	PSNR/dB	SSIM	FSIM
None	27.279 7	0.842 7	0.682 4
With CAB	27.284 3	0.842 8	0.682 5
With MCAB (RSNet)	27.294 3	0.845 5	0.683 5

表5 不同损失函数实验对比结果

Table 5 Comparison of experimental results using different loss functions

方法	PSNR/dB	SSIM	FSIM
MSE Loss	27.247 4	0.841 6	0.681 6
Charbonnier Loss	27.284 3	0.842 4	0.682 4
Compound Loss (RSNet)	27.294 3	0.845 5	0.683 5

### 3 结 语

为了提升LDCT图像的质量,同时更有效地保留图像的微小细节,本文提出一种融合残差卷积与Swin Transformer的网络(RSNet)。RSNet的核心部分是RSB,通过拆分、卷积、拼接等操作,有效地融合残差卷积的局部特征提取能力和Swin Transformer的全局特征提取能力。在编码器中,通过使用IPFEM来提取每一编码阶段输入特征的内部块信息,并将IPFEM的输出添加到编码器主块中,从而实现保护CT图像微小细节的目标。最后,为了解决Swin Transformer结构中移动窗口机制导致恢复错误纹理细节的问题,在Swin Transformer中引入多输入通道注意力机制,提出一种DAT模块。一系列的实验验证RSNet能有效去除LDCT图像中的噪声,同时较好地保护图像中的微小细节。与其他主流算法相比,RSNet能更有效地保护图像中的病灶和器官的边缘和纹理,降低图像噪声,获取更高质量的恢复图像,显示出优秀的性能。

### 【参考文献】

- [1] Liu Y, Gui Z, Zhang Q. Noise reduction for low-dose X-ray CT based on fuzzy logical in stationary wavelet domain[J]. Optik, 2013, 124 (18): 3348-3352.
- [2] Liu Y, Shangguan H, Zhang Q, et al. Median prior constrained TV algorithm for sparse view low-dose CT reconstruction[J]. Comput Biol Med, 2015, 60(3): 117-131.
- [3] Zhang R, Ye DH, Pal D, et al. A Gaussian mixture MRF for model-based iterative reconstruction with applications to low-dose X-ray CT[J]. IEEE Trans Comput Imaging, 2016, 2(3): 359-374.
- [4] Zhuang Z, Chen Y, Shu H, et al. Fast low-dose CT image processing using improved parallelized nonlocal means filtering[C]//International Conference on Medical Biometrics. IEEE, 2014: 147-150.
- [5] Chen Y, Yin X, Shi L, et al. Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing[J]. Phys Med Biol, 2013, 58(16): 5803.
- [6] Chen H, Zhang Y, Zhang W, et al. Low-dose CT via convolutional

- neural network[J]. Biomed Opt Express, 2017, 8(2): 679-694.
- [7] Gondara L. Medical image denoising using convolutional denoising autoencoders[C]//2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). IEEE, 2016: 241-246.
- [8] Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network[J]. IEEE Trans Med Imaging, 2017, 36(12): 2524-2535.
- [9] Gholizadeh-Ansari M, Alirezaie J, Babyn P. Deep learning for low-dose CT denoising using perceptual loss and edge detection layer[J]. J Digit Imaging, 2020, 33(2): 504-515.
- [10] Liang T, Jin Y, Li Y, et al. Edcnn: edge enhancement-based densely connected network with compound loss for low-dose ct denoising[C]//2020 15th IEEE International Conference on Signal Processing (ICSP). IEEE, 2020: 193-198.
- [11] Wu D, Kim K, El Fakhri G, et al. Iterative low-dose CT reconstruction with priors trained by artificial neural network[J]. IEEE Trans Med Imaging, 2017, 36(12): 2479-2486.
- [12] Kang E, Chang W, Yoo J, et al. Deep convolutional framelet denoising for low-dose CT via wavelet residual network[J]. IEEE Trans Med Imaging, 2018, 37(6): 1358-1369.
- [13] Wang D, Fan F, Wu Z, et al. CTformer: convolution-free Token2-Token dilated vision transformer for low-dose CT denoising[J]. Phys Med Biol, 2023, 68(6): 065012.
- [14] Wang D, Wu Z, Yu H. Ted-net: convolution-free t2t vision transformer-based encoder-decoder dilation network for low-dose ct denoising[C]//Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021. Springer International Publishing, 2021: 416-425.
- [15] Luthra A, Sulakhe H, Mittal T, et al. Eformer: edge enhancement based transformer for medical image denoising [J]. arXiv preprint arXiv: 2109.08044, 2021.
- [16] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7794-7803.
- [17] Li M, Hsu W, Xie X, et al. SACNN: self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network[J]. IEEE Trans Med Imaging, 2020, 39(7): 2289-2301.
- [18] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv preprint arXiv: 1706.03762, 2017.
- [19] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international Conference on Computer Vision. 2021: 10012-10022.
- [20] Zhang K, Li Y, Liang J, et al. Practical blind image denoising via Swin-Conv-UNet and data synthesis[J]. Mach Intell Res, 2023, 20(6): 822-836.
- [21] Chen X, Wang X, Zhou J, et al. Activating more pixels in image super-resolution transformer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 22367-22377.
- [22] Fan CM, Liu TJ, Liu KH. SUNet: Swin Transformer UNet for image denoising[C]//2022 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2022: 2333-2337.
- [23] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks [C]//Computer Vision-ECCV 2016: 14th European Conference. Springer International Publishing, 2016: 630-645.
- [24] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 136-144.
- [25] Liang J, Cao J, Sun G, et al. Swinir: image restoration using Swin Transformer [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1833-1844.
- [26] Zhao Y, Chen J, Zhang Z, et al. BA-Net: bridge attention for deep convolutional neural networks[C]//Computer Vision-ECCV 2022: 17th European Conference. Cham: Springer Nature Switzerland, 2022: 297-312.
- [27] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [28] AAPM. Low dose CT grand challenge[EB/OL]. <http://www.aapm.org/GrandChallenge/LowDoseCT/#>.
- [29] Zamir SW, Arora A, Khan S, et al. Restormer: efficient transformer for high-resolution image restoration[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5728-5739.
- [30] Zamir SW, Arora A, Khan S, et al. Multi-stage progressive image restoration [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14821-14831.
- [31] Wang Z, Cun X, Bao J, et al. Uformer: a general u-shaped transformer for image restoration[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17683-17693.
- [32] Chen L, Chu X, Zhang X, et al. Simple baselines for image restoration [C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 17-33.
- [33] Liu K, Du X, Liu S, et al. DDT: dual-branch deformable transformer for image denoising[J]. arXiv preprint arXiv: 2304.06346, 2023.

(编辑:陈丽霞)