

DOI:10.16136/j.joel.2023.07.0346

基于 U-Net 的两阶段 SPECT 骨显像降噪方法研究

余泓^{1*}, 罗仁泽¹, 陈春梦², 郭亮³, 罗任权¹

(1. 西南石油大学 电气信息学院, 四川 成都 610500; 2. 宜宾市第二人民医院 核医学科, 四川 宜宾 644000; 3. 西南石油大学 地球科学与技术学院, 四川 成都 610500)

摘要: 进行单光子发射计算机断层成像(single-photon emission computed tomography, SPECT)骨显像检查时,为减少给病人带来的辐射伤害,医师常会减轻辐射剂量,导致骨显像信噪比、分辨率较低,严重影响诊断以及病灶自动检测效果。为提升骨显像质量,提出了一种基于 U-Net 的两阶段 SPECT 骨显像降噪方法。首先,设计了一种 U-Net 噪声估计网络来快速估计每张骨显像的噪声水平,为主干降噪网络提供噪声先验知识。其次,主干降噪网络同样以 U-Net 为基础框架,同时结合多尺度特征融合、通道-空间注意力机制结构来增强网络的噪声特征提取能力,预测出噪声图。最后,通过残差学习得到降噪骨显像。同时,为解决使用均方误差(mean square error, MSE)损失函数的重建图像过于平滑的问题,设计了一种复合损失函数,保留骨显像的原有细节信息。实验中,向训练集中的骨显像施加不同噪声水平进行数据扩充,并且采用迁移策略解决模型过拟合问题。结果表明,与目前主流算法相比,所提出的降噪方法能够有效降低骨显像噪声,并且保留病灶细节特征。此外,通过盲降噪能够改善原骨显像质量、提升病灶自动分割效果。

关键词: SPECT 骨显像; 深度学习; 噪声水平; 注意力机制; 残差学习

中图分类号: TP391 **文献标识码:** A **文章编号:** 1005-0086(2023)07-0771-14

Research on denoising method of two-stage SPECT bone imaging based on U-Net

YU Hong^{1*}, LUO Renze¹, CHEN Chunmeng², GUO Liang³, LUO Renquan¹

(1. School of Electrical Engineering and Information, Southwest Petroleum University, Chengdu, Sichun 610500, China; 2. Department of Nuclear Medicine, The No.2 People's Hospital of Yibin, Yibin, Sichun 644000, China; 3. College of Geoscience and Technology, Southwest Petroleum University, Chengdu, Sichun 610500, China)

Abstract: When performing single-photon emission computed tomography (SPECT) bone imaging examination, physicians often reduce radiation damage by reducing radiation dose, resulting in low signal-to-noise ratio and resolution of bone imaging, which seriously affects the diagnosis and automatic detection of lesions. In order to improve the quality of bone imaging, a two-stage SPECT bone imaging noise reduction method based on U-Net is proposed. Firstly, a U-Net noise estimation network is designed to quickly estimate the noise level of each bone image, providing noise prior knowledge for the backbone noise reduction network. Secondly, the backbone noise reduction network also uses U-Net as the basic framework, and combines multi-scale feature fusion and channel-spatial attention mechanism structure to enhance the noise feature extraction ability of the network and predict the noise map. Finally, denoised bone imaging is obtained through residual learning. At the same time, in order to solve the problem that the reconstructed image using the mean square error (MSE) loss function is too smooth, a composite loss function is designed to retain the original detailed information of bone imaging. In the experiments, differ-

* E-mail: 790622472@qq.com

收稿日期: 2022-05-12 修订日期: 2022-07-15

基金项目: 四川省科技计划项目(2019CXRC0027)资助项目

ent noise levels are applied to the bone images in the training set for data augmentation, and a transfer strategy is used to solve the problem of model overfitting. The results show that, compared with the current mainstream algorithms, the proposed noise reduction method can effectively reduce the noise of bone imaging and preserve the detailed features of the lesions. In addition, blind noise reduction can improve the imaging quality of the original bone imaging and improve the automatic segmentation effect of lesions.

Key words: SPECT bone imaging; deep learning; noise level; attention mechanism; residual learning

0 引言

在全球范围内,恶性骨转移肿瘤已经成为引起癌症患者死亡的重要原因。目前,医学影像诊断骨转移肿瘤的手段主要有电子计算机断层扫描(computed tomography, CT)^[1]、全身磁共振显像(magnetic resonance imaging, MRI)^[2]、正电子发射计算机断层显像(positron emission tomography, PET)和单光子发射计算机断层成像(single-photon emission computed tomography, SPECT)^[3]。其中,SPECT具有灵敏度高、成本低、成像视野广等优势,成为评价骨转移肿瘤不可或缺的检查手段。

SPECT骨显像技术通过获取患者体内的放射性追踪试剂的3D分布并投影至2D平面进行成像。为减轻病人受到的辐射伤害,放射性追踪试剂的用量必须尽可能低^[4]。此外,探测器上发射的光子数量有限,导致SPECT骨显像质量下降,容易引入噪声。为此,国内外学者对核医学影像的去噪算法展开了研究。

传统医学影像降噪算法主要分为迭代重建算法^[5]、投影域算法^[6]和图像后处理算法^[7]3类。ZHANG等^[8]在孤立噪声去除算法和噪声水平分割技术的基础上,提出了一种基于分割的自适应统计正弦平滑技术对低剂量CT图像进行去噪,并证明了该方法的有效性。但此类算法由于重复迭代导致运算量大,运算效率极低,难以进行实际应用。GUI等^[9]提出了一种模糊中值滤波算法,该算法能够有效抑制低剂量CT图像中的条纹伪影,提升CT图像质量。YU等^[10]结合了局部自适应和双边滤波算法,在不显著牺牲分辨率的同时保留了图像的边缘信息。这类算法便于系统集成且耗费时间较少,但需要大量CT投影域数据,另外在去噪过程中容易引入新的噪声。ZHANG等^[11]提出一种混合高斯马尔科夫随机场模型来重建低剂量CT图像,该方法通过调整模型来控制图像不同对比度区域的清晰度。尽管重建后的图像质量较高,但需要经过多次迭代,运算量较大。总体来说,传统医学影像去噪技术虽能在一定程度上降低图像噪声,但受限于设备算力和投

影数据等因素,并不能达到令人满意的效果。

近年来,随着深度学习的飞速发展,许多学者开始将其应用到医学影像去噪任务中。CHEN等^[12]将卷积神经网络以逐块的方式对低剂量CT影像进行训练,相比于传统算法,该方法在峰值信噪比(peak signal to noise ratio, PSNR)、均方根误差(root mean-square error, RMSE)、结构相似性(structural similarity, SSIM)指标上有了很大提升。KANG等^[13]利用深度卷积神经网络对低剂量CT图像进行小波变换来提取伪影的方向分量,以此抑制CT图像特有的噪声。RONNEBERGER^[14]提出了一种基于编码-解码结构的网络——U-Net,该网络通过下采样、上采样的方式获取图像的高层与低层的信息,采用跳跃连接结构实现特征的高效利用,有效减小了网络的运算复杂度。CHEN等^[15]提出将残差结构和U-Net相结合的卷积神经网络(RED-CNN)对CT图像进行降噪,实验结果表明,该方法在降低噪声与病变检测上效果良好。YANG等^[16]提出一种结合Wasserstein损失、预训练VGG模型和Charbonnier距离损失结构的生成对抗网络,实现对低剂量CT的降噪。KEISUKE^[17]等采用去噪卷积神经网络(DnCNN)对CT图像降噪,同时保持了图像清晰度,并通过裁剪CNN模型的方式防止了图像过度平滑。目前,基于深度学习的降噪算法主要关注的是MRI或CT影像,有关SPECT骨显像降噪方法的研究极少。

在骨显像的相关研究中,研究者通常采用高斯滤波、中值滤波或阈值去噪等传统算法进行降噪。该类算法虽在一定程度上有利于后续自动分类、分割任务,但会破坏骨显像病灶、骨骼边缘细节特征。因此,为保留原骨显像的细节特征,同时防止新的噪声引入,本文提出了一种基于U-Net的两阶段SPECT骨显像降噪方法。首先,采用一个3层的U-Net模型对骨显像的噪声水平进行快速估计。其次,主干降噪网络以U-Net模型为基本框架,同时结合通道-空间注意力机制、多尺度特征融合、密集残差连接等方式提升模型提取骨显像噪声特征的能力。最后,通过残差学习得到降噪骨显像。同时,设计了一种复合损失函

数来保留骨显像的细节特征。实验中,首先向每张骨显像加入不同水平的噪声,扩充数据量。接着,为解决小样本数据集训练易出现过拟合的现象,采用迁移学习策略对主干降噪网络进行预训练,优化模型初始化参数。实验结果表明,与目前主流的降噪算法相比,本文提出的模型能对不同噪声水平的骨显像进行降噪,且均取得了较好的降噪效果。此外,通过盲降噪能在一定程度改善原图质量,新的骨显像在自动分割实验中能够取得更精确的分割效果。

1 方法原理

1.1 问题描述及算法流程

1.1.1 问题描述

在进行 SPECT 骨显像检查时,常通过降低设备的管电流或扫描时间来降低核辐射强度,减轻给病人来带的辐射伤害。但这导致探测器上的光子数量减少,在骨显像中引入近似泊松分布的噪声^[18]。泊松分布如式(1)所示:

$$n = P(x_i | z_i) = \frac{z_i^{x_i}}{x_i!} e^{-z_i}, z_i = 0, 1, 2, \dots, \quad (1)$$

式中, z_i 代表像素点 i 在泊松噪声环境下的大小, x_i

为原始无噪声污染的骨显像像素点 i , n 表示骨显像中的噪声。

将含噪骨显像、原骨显像分别作为模型的输入与标签,通过深度学习模型学习两者之间的映射关系,如式(2)所示。但这种训练方式效率不高,导致模型去噪效果欠佳,而残差学习可以解决这一问题。残差学习通过训练含噪图像与噪声之间的映射关系来降低模型训练难度,提升去噪效果,原理如式(3)所示:

$$y = x + n, \quad (2)$$

$$n = y - x, \quad (3)$$

式中, $y \in r^{w \times h}$ 为包含泊松噪声、大小为 $w \times h$ 的骨显像, $x \in r^{w \times h}$ 表示相对应的干净无噪声的骨显像。

1.1.2 整体算法流程

骨显像降噪的整体算法流程如图 1 所示。首先,将骨显像加入不同水平的泊松噪声并送入噪声估计网络,评估出每张图像的噪声水平。然后,将噪声水平图和含噪骨显像一起输入主干降噪网络,输出噪声水平图。最后,将输出的噪声水平图进行残差运算得到降噪骨显像。同时,噪声估计网络、主干降噪网络分别通过交叉熵损失函数与复合损失函数优化模型的降噪效果。

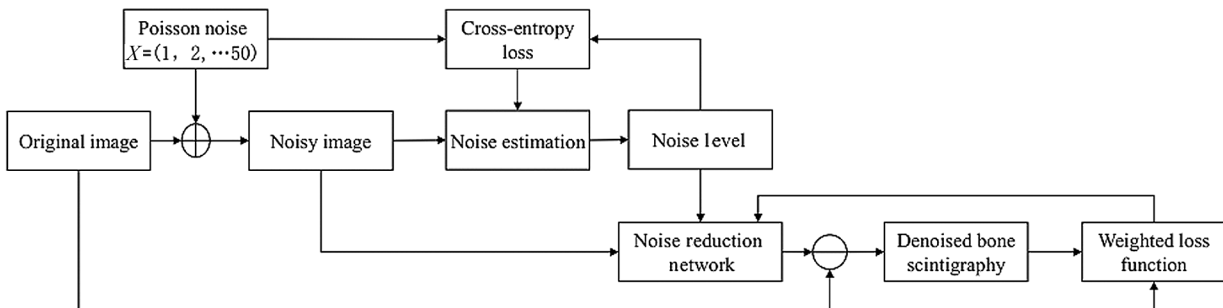


图 1 骨显像降噪算法流程图

Fig. 1 Flow chart of noise reduction algorithm for bone imaging

1.2 降噪网络模型

为抑制骨显像中的噪声,本文设计了一种基于 U-Net 的两阶段 SPECT 骨显像降噪模型 EMANet,如图 2 所示。该模型将降噪任务分为两个阶段:噪声水平估计与噪声抑制。

在第一阶段中,首先向每张骨显像加入不同水平的泊松噪声。然后,设计了一个 3 层 U-Net 模型对骨显像的噪声水平进行快速评估。

在第二阶段中,将第一阶段输出的噪声水平作为先验噪声,与含噪骨显像拼接后一起送入主干降噪网络。主干降噪网络以 U-Net 模型为基本框架,

输出骨显像中的噪声分布图。最后,通过残差学习得到降噪骨显像。

噪声估计网络和主干降噪网络均以 U-Net 为基本框架,并设计了密集残差模块(RDC block)、多尺度空间卷积模块(MD block)、残差协调注意力门控结构(RCAG)、协调输出模块(CF block)4 个模块来提升网络的噪声提取能力,在降低骨显像噪声的同时,也保留了骨显像的细节信息、避免了新的噪声引入。

1.2.1 噪声估计网络

噪声估计网络采用了一个 3 层的 U-Net 模型。

在编码部分,输入大小为 $256 \times 256 \times 3$ 的骨显像,通过卷积块、最大池化层提取含噪骨显像高层特征。在解码部分,利用上采样将骨显像恢复至原图大小,

结合跳跃连接结构融合图像低层与高层的噪声特征。最后,经过全局平均池化和全连接层输出噪声水平的种类概率,如图 2(a)所示。

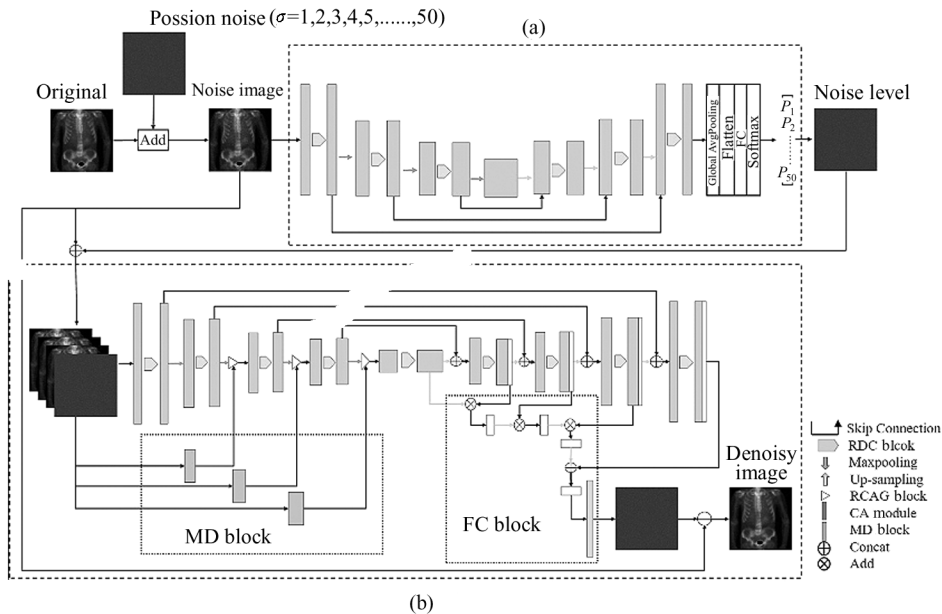


图 2 降噪模型整体框架:(a) 噪声估计网络;(b) 主干降噪网络

Fig. 2 Overall framework of noise reduction model: (a) Noise estimation network; (b) Backbone noise reduction network

为提升噪声估计网络的噪声学习能力,设计了一种 RDC block 代替原 U-Net 卷积块对骨显像噪声特征进行提取。整个 RDC block 由 1×1 、 3×3 卷积和协调注意力机制 (coordinate attention, CA)^[19] 构成,采用密集残差连接方式提升卷积层的利用效率,如图 3(b)所示。首先,该结构通过 3×3 卷积提取骨显像中的噪声特征。其次,通过密集残差连接方式和 1×1 卷积汇聚统一尺度下的不同特征信息,同时起到降低网络计算量的作用。最后,采用一个 CA 注意力模块,帮助模型重点关注感兴趣的特征,抑制无用信息。其中,CA 模块是一种能够同时关注特征图通道与空间关系的注意力结构。其通过降维的方式,将空间特征图拆分为两个不同方向的分量,分别学习通道之间的关系。然后,将特征图的空间位置信息嵌入通道注意力中,以此来挖掘通道信息与空间信息之间的联系。此外,由于采用了降维的方式,该模块不会耗费过多计算量,有助于提升模型的训练效率,具体结构见图 3(a)。

1.2.2 主干降噪网络

主干降噪网络同样采用 U-Net 模型,通过 RDC block、下采样、上采样提取骨显像的噪声特征。此外,为加强模型的特征提取能力,设计了 MD block、RCAG、CF block 3 个模块来提取不同尺度的噪声特

征并抑制冗余信息。首先,在每层的输入端设计了 MD block 提取不同尺度的输入特征。然后,将 MD block 输出的特征与池化后的特征一起输入 RCAG,抑制无用的信息。解码端,由于在上采样的过程中,存在细节信息丢失的情况,对此,本文设计了 CF block 来融合每个解码层输出的特征,避免细节信息的丢失。下面具体介绍 MD block、RCAG、CF block 3 个模块。

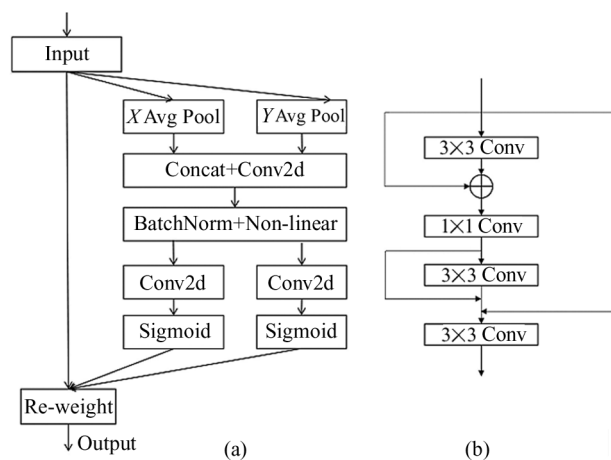


图 3 (a) CA 模块;(b) RDC block

Fig. 3 (a) CA module; (b) RDC block

MD block 通过空洞卷积来提升卷积核的感受野,但常规的空洞卷积容易引入棋盘伪影噪声。因此,本文采用多尺度空洞卷积的方式进行提取,具体结构见图 4。文献[20]中验证当采用多尺度空洞卷积时,扩张率组合的公约数越大,缺失越明显,分支扩张率为全奇数或奇偶混合时,棋盘伪影效应能够降低。因此,将 3 个 MD block 中的多尺度空洞卷积扩张率分别设为 5、7、9、4、5、6 与 2、3、5。然后,提取到的不同尺度的特征后进行聚合,输入 CA 模块抑制冗余信息,并使用 1×1 卷积压缩参数、降低模型计算量。最后,经过批量归一化和激活函数后,采用全局平均池化实现下采样并输出特征图。

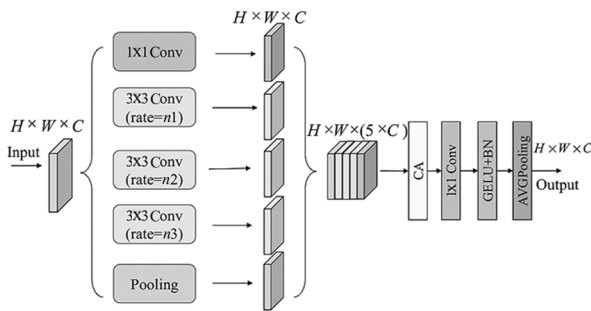


图 4 多尺度空洞卷积模块

Fig. 4 Multiscale dilated block (MD block)

为将 MD block 提取到的多尺度特征与 U-Net 池化后的特征充分融合,在 CA 模块的基础上设计了 RCAG 结构,如图 5 所示。该结构能够同时关注两个输入特征的空间与通道的重要程度。首先,将经过最大池化后的特征 x_1 和 MD 输出特征 x_2 进行拼接得到一个新的特征 x_c 。然后,通过 CA 模块学习新特征图的空间、通道之间的联系,将得到的权重乘以 x_c 得到更新后的特征 x'_c ,并与 x_c 进行残差融合,保留输入特征的部分信息。最后,经过 1×1 卷

积、BN层和 BeLU 激活函数输出模块。具体原理如下:

$$z_c = concatenate(x_1 + x_2), \quad (4)$$

式中,设两个输入特征的大小均为 $C \times H \times W$, x_1, x_2 表示输入特征, z_c 表示经过特征拼接后的特征。

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i), \quad (5)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w), \quad (6)$$

式中, $z_c^h(h)$ 表示在通道 c 中,对高度 h 的第 w 行特征信息求和并取平均,式(6)同理。

$$f = \sigma(F1([z_c^h, z_c^w])), \quad (7)$$

式中, $F1([z_c^h, z_c^w])$ 表示将上一步得到的 $z_c^h(h)$ 和 $z_c^w(w)$ 进行拼接,再通过 1×1 卷积得到大小为 $C/r \times 1 \times (W + H)$ 的特征图, σ 表示进行批量归一化和 BeLU 非线性变换。

$$g_c^h = \sigma(F_h(f^h)), \quad (8)$$

$$g_c^w = \sigma(F_w(f^w)). \quad (9)$$

经过 split 操作后,得到 f^h, f^w , F_h, F_w 表示再次经过 1×1 卷积,将特征恢复至拼接之前的大小,然后进行 σ 非线性变换。

$$x'_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (10)$$

然后,将 g_c^h, g_c^w 作为注意力权重,通过 $x_c(i, j) \times g_c^h(i) \times g_c^w(j)$ 的方式对输入的特征进行权重更新; $x'_c(i, j)$ 表示经过权重更新后的特征。

$$x_{out} = \sigma(Conv_{1 \times 1}(x_c + x'_c)), \quad (11)$$

式中, $Conv_{1 \times 1}(x_c + x'_c)$ 表示将 x_c 和更新后的 x'_c 进行特征融合,并使用 1×1 卷积将通道数压缩至与输入特征 x_1, x_2 相同, x_{out} 表示整个模块的输出特征。

本文设计的 CF block 通过特征融合和 CA 模块相结合的方式,将解码部分每层的输出噪声特征进行融合,避免在上采样的过程中丢失骨显像细节信

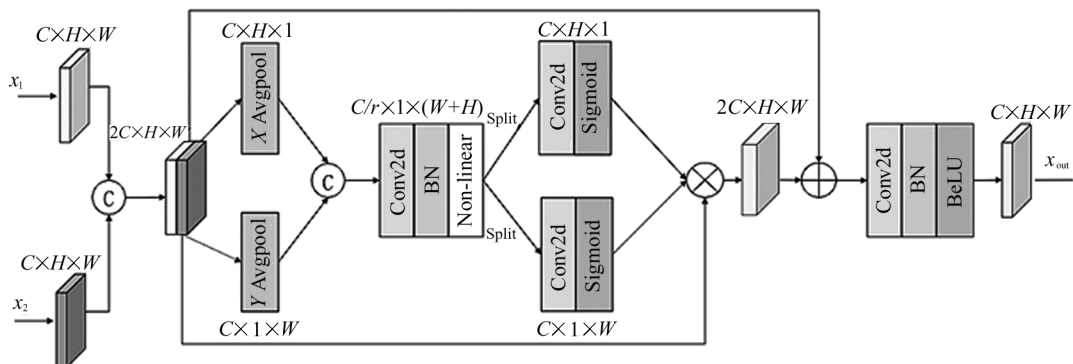


图 5 残差协调注意力门控结构(RCAG)

Fig. 5 Residual coordinated attention gating structure (RCAG)

息,具体结构见图 2(b)。该结构通过线性插值和 1×1 卷积的方式进行上采样,将解码端每层输出的特征大小恢复至与下一层的输出相同。然后,通过特征融合的方式来结合不同尺度的特征。此外,每次融合之后都采用了一个 CA 模块来重点关注对模型降噪有益的特征信息。经过层层融合后,最终得到预测噪声图像。

1.2.3 损失函数

使用均方误差(mean square error, MSE)作为损失函数,能够让降噪后的图像获得更优秀的评价指标,见式(12)。但逐像素计算损失的方式也让图像趋于平滑^[21]。

$$L_{mse} = \frac{1}{\omega \times h} \sum_{i=1}^N \|y_i - x_i\|_2^2, \quad (12)$$

式中, ω 、 h 表示宽和高, y_i 、 x_i 分别代表降噪图与原图。

感知损失(perceptual loss)能够重建出更多的图像细节特征。其通过从已经预训练好的 CNN 模型里面提取不同层次的特征来计算损失值,原理见图 6。本文采用预训练好的 VGG-16 模型来计算感知损失大小,见式(13):

$$L_{prec} = \frac{1}{\omega \times h} \sum_{i=1}^N \|\varphi(y_i) - \varphi(x_i)\|_2^2. \quad (13)$$

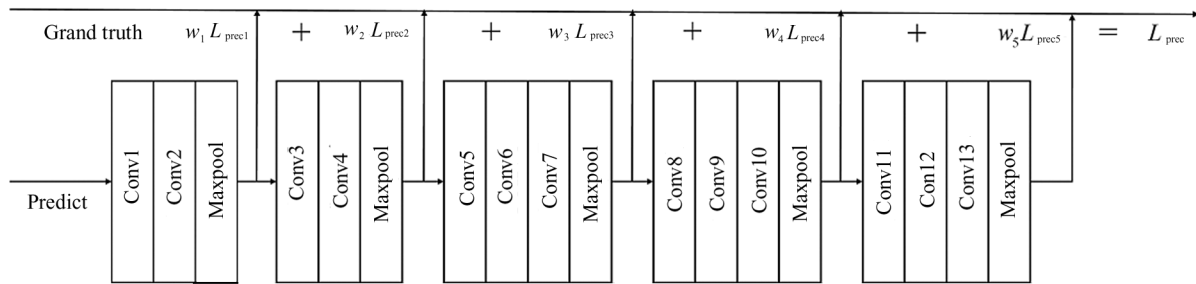


图 6 感知损失原理

Fig. 6 Principle of perceived loss

为充分发挥 MSE 损失和感知损失各自的优势,在取得优秀的评价指标同时,也能保持骨显像的病灶细节特征,设计了一种结合 MSE 损失和感知损失的复合损失函数。由于需要保留骨显像的病灶细节特征(即低层特征),选择 VGG-16 前两层的输出特征来计算感知损失 L_H :

$$L_H = (1 - \lambda)L_{mse} + \lambda L_{prec}, \quad (14)$$

式中, λ 为常数,用于平衡两项损失函数, L_{mse} 、 L_{prec} 分别表示 MSE 损失和感知损失, $\varphi(\cdot)$ 表示 VGG-16 模型提取到的特征。

2 数据处理和模型训练

2.1 数据集及数据预处理

数据集一共包含 158 例 SPECT 骨显像,每例骨显像通过 DICOM 文件进行储存,无法直接送入深度学习模型。为将其转化成能够输入模型的图像,首先通过归一化的方式转化为灰度图,每例包含人体正面、背部两张图片。其次,为提高骨显像信噪比,采用自动裁剪的方式得到肩部至盆骨区域。最后,筛选除去受污染严重的图像,一共得到 310 张骨显像作为实验数据样本。实验中,选取 270 张进行训

练,40 张进行测试。在训练模型时,向每张骨显像添加 1—50(间隔为 1,为整数)的泊松噪声,测试集中骨显像加入 5—40(间隔为 5,为整数)的泊松噪声,骨显像数据量变化如表 1 所示。

表 1 数据增强前后的数据量

Tab. 1 The amount of data before and after data enhancement

Data types	Training set	Testing set	Total
Before expansion	270	40	310
After expansion	13 500	320	13 820

进行增强后,在原数据集中挑选了 2 张骨显像进行相似度对比,其中一张噪声较低,另一张噪声较高,均调至亮度、对比度相近,具体见图 7。然后,将不同噪声水平的低噪骨显像与高噪声骨显像进行比较,其矩形框区域(即噪声部分)的 SSIM 变化见表 2。可见在噪声水平 0—20 范围内,随着噪声加入,噪声 SSIM 有所提升;当噪声水平大于 20 时,SSIM 开始下降。

由于骨显像初始数据样本较少,为防止模型过拟合,采用了迁移学习^[22]策略对主干降噪网络进行

预训练。此外,为进一步得到更优的模型初始化参数,预训练数据集采用医学数据集——MAYO数据集^[23](CT影像)。选取其中1000对图像进行训练,

每对图像为全剂量CT(干净图像)和低剂量CT(含噪图像)。预训练结束后,再使用骨显像数据集进行

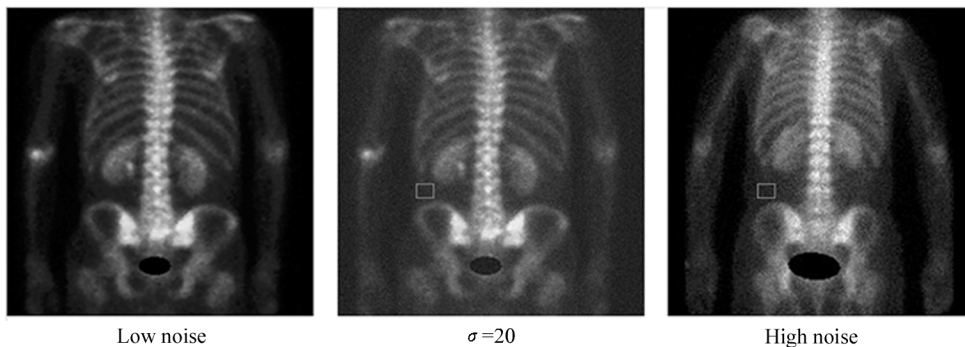


图7 不同噪声水平骨显像

Fig. 7 Bone imaging with different noise levels

训练,微调EMANet模型参数。

表2 结构相似性变化

Tab. 2 Structural similarity change

Noise level	0	5	10	15	20	30	40
SSIM	0.382	0.406	0.440	0.461	0.465	0.380	0.311

2.2 评价指标

实验选用PSNR、SSIM、RMSE 3种客观指标对降噪骨显像质量进行评估。其中,PSNR能反映出模型的降噪效果,当PSNR越高时,模型降噪效果越好。SSIM代表图片降噪后的失真程度,范围在[0,1]之间,其值越接近1,表示失真程度越小。

RMSE =

$$\left[\frac{1}{WH} \sum_{w=1}^M \sum_{h=1}^N (X(i,j) - Y(i,j))^2 \right]^{\frac{1}{2}}, \quad (15)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) =$$

$$20 \cdot \log_{10} \left(\frac{MAX_I}{RMSE} \right), \quad (16)$$

$$SSIM = \frac{(2\mu_x\mu_y)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (17)$$

式中, $W \times H$ 表示骨显像大小, $X(i,j)$ 、 $Y(i,j)$ 表示降噪前后的骨显像在 (i,j) 处的像素值; MAX_I 表示骨显像中的最大像素值, σ_x^2 、 σ_y^2 表示骨显像降噪前后的灰度方差, σ_{xy} 表示协方差; μ_x 、 μ_y 表示像素点均值, c_1 、 c_2 为常数。

2.3 模型性能影响因素分析

本次骨显像降噪实验中,使用的硬件配置为:AMD Ryzen5 3600 CPU;12 GB Nvidia GeForce

RTX 3060 显卡;软件环境配置:Python3.6.12、pytorch1.8.0、CUDA11.1、cuDNN8.1.0;使用RAdam优化器,学习率设为0.001,batch size为8,迭代次数为200。

2.3.1 迁移策略效果

为验证迁移学习对模型训练的提升作用,本文将直接训练与采用迁移学习策略的效果进行了对比,结果见表3。由表3可以看出,经过迁移策略预训练后的模型得到的PSNR、SSIM分别提升了7.7、0.17。此外,对训练集中的噪声图像进行了可视化,如图8所示。可以看出,直接训练模型的方式造成模型输出的噪声图杂乱无章,而经过预训练的模型能够学习到规律的噪声分布。

表3 采用迁移策略效果对比(噪声水平 $\sigma=20$)

Tab. 3 Comparison of the effect of adopting migration strategy

Pre-training dataset	With transfer learning	PSNR	SSIM
/	×	22.1	0.66
Mini Imagenet	✓	29.8	0.83
Mayo	✓	30.5	0.85

出现上述现象的原因是骨显像的初始数据量较少、模型结构复杂、参数量较多。虽然复杂的模型能够得到更好的非线性表达能力、更强的特征拟合能力,但由于初始骨显像数据量较少,尽管进行了数据扩充,骨显像本身包含的特征信息量并未增加。训练时,模型在优化迭代的过程中,容易陷入局部最优值,出现过拟合现象。当模型采用迁移学习策略后,能够将其他数据集学习到的先验知识迁移到骨显像数据集中,得到更优的初始化参数,模型泛化能力增

强,降噪性能也随之提升。为进一步分析不同源域数据集对模型性能的影响,将 Mayo 数据集与 Mini Imagenet 数据集进行了对比。经过两个数据集预训

练后的模型均能得到收敛,但 Mayo 同样处理的是医学降噪任务,能为骨显像降噪提供更为准确的先验知识,其 PSNR 和 SSIM 分别提升了 0.7、0.02。

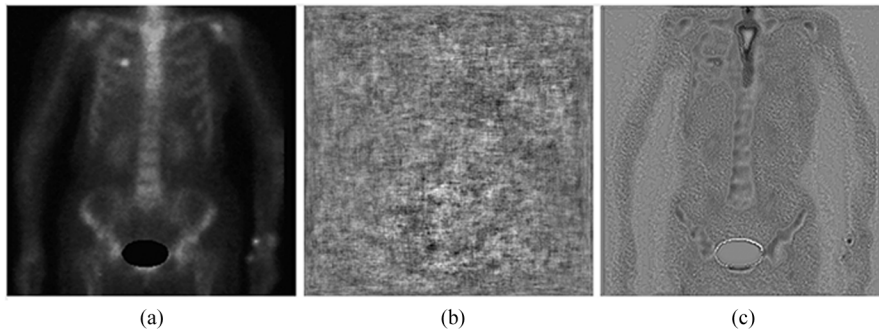


图 8 噪声图对比

Fig. 8 Noise picture comparison

2.3.2 模块消融实验

为分析本文所设计的 RDC block、RCAG、MD、CF block 4 种模块对主干降噪网络的作用,设置了一组消融实验。使用测试集中的含噪骨显像进行测试,降噪后骨显像的平均 PSNR 和 SSIM 如表 4 所示。可以看出,4 个模块对主干降噪网络性能均有提升作用。其中,RDC block 通过密集残差连接方式重复利用骨显像特征信息,避免了卷积神经网络随网络层数增加而导致的梯度消失问题,CA 模块进一步提高了 RDC block 特征利用效率。MD 模块采用多尺度空洞卷积,增加了感受野范围,让模型学习到了更丰富的骨显像噪声特征。RCAG 作为一种双输入的通道-空间注意力结构,能够对两个输入特征的空间、通道信息进行权重调整,抑制无用信息,提高模型计算效率。CF block 将多尺度结构与 CA 模块进行结合,高效地融合了解码层信息,避免骨显像丢失细节信息,从而提高了降噪效果。

表 4 改进模块效果对比

Tab. 4 Comparison of the effects of improved modules

Modules	PSNR	SSIM
/	26.95	0.77
RDC block	27.55	0.80
RDC block + MD	28.77	0.82
RDC block + MD + RCAG	29.80	0.85
RDC block + MD + RCAG + CF block	30.72	0.86

2.3.3 损失函数性能对比

为验证复合损失函数的有效性,本文对比了不同损失函数对模型性能的影响,结果如表 5 和图 9 所

示。首先,从表 5 中的平均 PSNR 和 SSIM 来看, MSE 损失函数取得了最高评价,复合损失函数评分略低于 MSE 损失,而感知损失函数评分远低于 MSE 损失和复合损失函数。其次,由图 9 可见, MSE 损失重建的骨显像能基本除去骨显像噪声,但病灶边缘细节特征模糊。感知损失的重建图像能够保留骨显像的病灶边缘细节特征,而高层特征却并不明显。本文设计的复合损失函数将二者各自的优、劣势进行互补。在小幅度牺牲 PSNR 和 SSIM 指标的情况下,既对骨显像进行了降噪,又保留了病灶边缘和肋骨处的细节特征。

表 5 损失函数性能对比

Tab. 5 Performance comparison of loss function

Loss function	PSNR	SSIM
L_{mse}	31.334 4	0.845 6
L_{prec}	19.184 4	0.501 3
L_H	30.190 2	0.836 0

为进一步提升模型性能,针对复合损失函数中的权重 λ 进行调优,测试结果的平均 PSNR 和 SSIM 值随 λ 的变化情况如图 10 所示。可以直观看出, PSNR 与 SSIM 呈现先减小再增大、最后减小的趋势,当 λ 取值 0.2 时,效果最好。

总体来说,迁移学习策略优化了模型初始化参数,防止了过拟合现象的出现。各个改进模块采用的密集残差连接、多尺度特征融合方式和注意力机制结构,对模型的降噪性能有一定提升。复合损失函数能够保留骨显像的细节特征,并且不同损失函数应当设置合适的权重比,进一步提升降噪效果。

综上所述,本次主干降噪模型设置迭代次数为 200。经过预处理后,主干降噪网络采用 RDC 卷积模块、RCAG 门控注意力结构、MAD、CF-Block 结构对骨

显像噪声进行特征提取。训练时,通过迁移学习策略预训练EMANet模型,利用复合损失函数优化降噪效果。

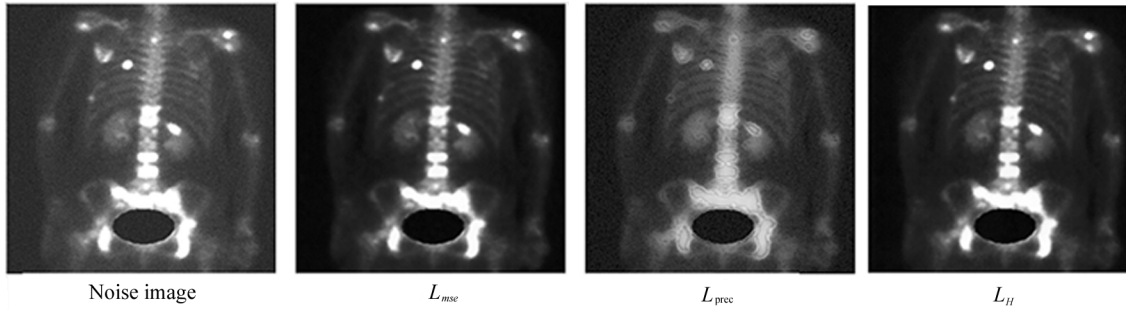


图 9 降噪效果对比

Fig. 9 Comparison of noise reduction effect

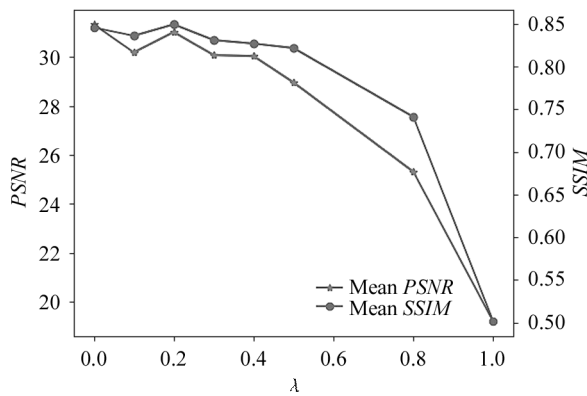


图 10 平均 PSNR 和 SSIM 变化曲线

Fig. 10 Mean PSNR and SSIM change curves

3 实验及结果分析

3.1 噪声估计网络评估

为证明本文设计的噪声估计网络的有效性,选取基于直方图的噪声估计算法 HBNE、离散小波变换噪声估计算法 DWTNE 与本文设计的噪声估计模型进行了比较。3 种算法的噪声估计值和 RMSE 的平均值如表 6 所示。

从表 6 可以看出,本文设计的噪声估计模型效果最好。首先,传统算法 HBNE、DWTNE 对不同噪声水平的骨显像有不同的效果。当噪声水平在 5—25 范围内,HBNE、DWTNE 处于合理的范围内。当噪声水平大于 25 时,HBNE 和 DWTNE 的估计值与

表 6 不同算法噪声评估效果

Tab. 6 Noise evaluation effect of different algorithms

Noise level	HBNE		DWTNE		Ours	
	Estimation	RMSE	Estimation	RMSE	Estimation	RMSE
5	4.78	0.32	4.98	0.18	5.05	0.22
10	9.52	0.74	10.21	0.38	9.90	0.32
15	16.21	0.52	16.21	0.62	15.15	0.39
20	22.04	0.91	21.17	0.59	20.20	0.78
25	23.82	1.08	23.97	0.87	25.05	0.81
30	28.29	1.96	29.01	1.30	30.30	0.95
35	38.81	2.07	33.21	1.79	35.40	0.77
40	42.83	2.19	42.09	2.01	40.10	1.10

实际值差距突然增大,模型性能迅速下降。其次,随着骨显像噪声水平增加,HBNE、DWTNE 的估计效果降低, RMSE 逐渐增大。本文设计的模型在所有噪声水平下均能给出较为准确的估计值,并且 RMSE

稳定在 1.2 以内。造成上述现象的原因是本文设计的模型借助了卷积神经网络强大的特征提取能力,并且结合了密集残差连接和注意力机制结构,进一步提升了模型性能,具有较好的鲁棒性。

3.2 模型降噪效果对比

为了更好地评价 EMANet 整体的降噪效果,将其与 BM3D、DnCNN^[17]、RED-CNN^[15]、WGAN 4 种经典医学降噪模型进行了对比, PSNR、SSIM 的平均值见表 7,黑色加粗部分表示最好的评价。可以看出,随着噪声强度提升,除 EMANet 外的降噪模型的 PSNR、SSIM 均呈现下降趋势。其中, BM3D 受噪声的影响较大。4 种深度学习模型推理出的降噪骨显像在 PSNR、SSIM 指标均高于 BM3D。DnCNN 利用 CNN 和残差学习的方式,获得了比 BM3D 更好的降噪效果。RED-CNN 采用编码—解码结构和跳

跃连接方式加强了上下文信息的利用,学习到了更深层骨显像噪声信息,取得的 PSNR、SSIM 值略高于 DnCNN。而 WGAN 的指标数值略低于 DnCNN。本文设计的 EMANet 模型受噪声水平大小的影响较小,除 10、15、20 3 种噪声水平外,均取得了最好的评价。原因是 EMANet 采用了残差学习和编码—解码的模型框架,同时结合了噪声估计模块、通道空间—注意力结构和多尺度特征融合方式,模型降噪性能得到了提升,能够更好地学习含噪骨显像与原骨显像之间的映射关系,鲁棒性更强。

图 11 为骨显像加入 25 噪声水平时,各模型的降

表 7 模型降噪效果对比

Tab. 7 Comparison of noise reduction effect of model

Model	Noise-level (σ)							
	5	10	15	20	25	30	35	40
Noisy picture	29.36/0.85	27.72/0.78	24.34/0.71	21.92/0.65	20.02/0.60	18.46/0.56	17.15/0.52	16.01/0.49
BM3D	29.93/0.78	28.68/0.70	25.90/0.65	21.87/0.60	20.26/0.56	18.95/0.53	17.84/0.51	17.37/0.56
DnCNN	31.55/0.87	31.03/0.86	30.15/0.83	29.66/0.82	28.37/0.79	27.02/0.77	25.86/0.74	24.62/0.71
RED-CNN	32.11/ 0.90	33.96/0.92	33.01/0.90	31.60/0.87	29.65/0.83	28.21/0.80	25.99/0.77	24.49/0.73
WGAN	30.23/0.87	31.43/0.85	29.75/0.82	28.30/0.79	26.99/0.76	25.88/0.69	24.74/0.69	23.05/0.66
EMANet(Ours)	32.91/0.88	30.95/0.84	29.16/0.83	29.43/0.83	31.33/0.85	31.92/0.85	32.44/0.86	31.51/0.85

噪效果。整体来看,各模型均能一定程度上对含噪骨显像进行降噪。BM3D 去除了部分噪声,但是降噪后的骨显像较为模糊,严重丢失骨病灶边缘细节信息。DnCNN、RED-CNN 借助残差学习和编码—解码结构的优势,去除了大部分噪声,但骨骼周围仍有部分噪声残留。WGAN 采用生成对抗思想和

Wasserstein 损失指导骨显像重建,虽一定程度降低了骨显像中的噪声水平,但噪声残留量多于另外 4 种模型,效果并不明显。而本文设计的 EMANet 能够基本去除骨显像周围的噪声。

细节上,分别从肩胛骨、胸骨、肋骨 3 处的降噪效果进行分析。在肩胛骨处(图(a)),加噪骨显像周

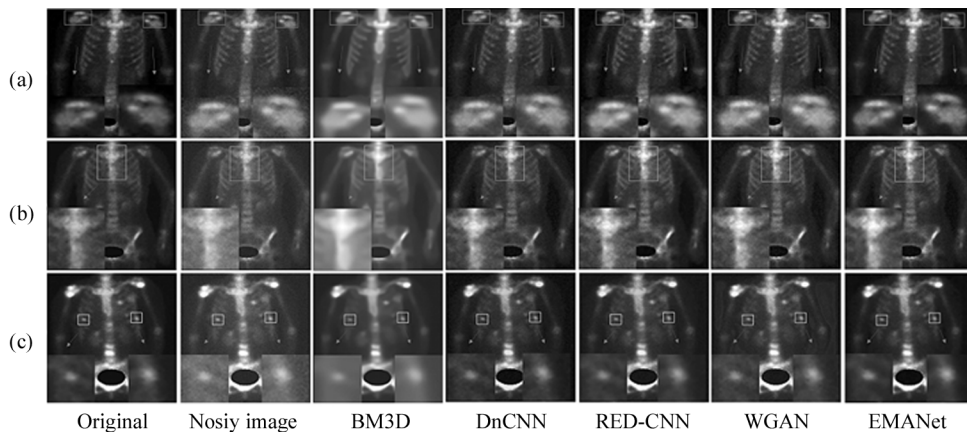


图 11 降噪效果对比

Fig. 11 Comparison of noise reduction effect

围含有大量泊松噪声。BM3D 的降噪骨显像病灶特征遭到模糊化,边缘纹理特征丢失严重,效果较差。DnCNN 和 WGAN 的降噪骨显像周围仍残留部分噪

声,原因是这两种模型提取到的噪声特征信息不够充分。RED-CNN 采用编码—解码框架和跳跃连接结构,提高了信息的利用效率,病灶周围残留的噪声

更少,但在上、下采样的过程中丢失了部分信息,并且 MSE 损失函数使得降噪骨显像丢失了细节纹理特征。在胸骨处(图(b)),经过深度学习模型重建后的病灶特征比原图更明显。EMANet 相比于前几种模型,病灶周围残留的噪声更少,并且 EMANet 利用多尺度融合、空间-通道注意力结构以及复合损失函数让模型学习到了更丰富的噪声特征,弥补了上采样和 MSE 损失函数带来的问题,保留了骨显像的细节纹理特征。在肋骨病灶处(图(c)),噪声残留量为 $BM3D > DnCNN > WGAN > RED-CNN > EMANet$ 。由此可以看出, BM3D 噪声抑制效果依然不佳。DnCNN、WGAN 保留了原有的边缘纹理和细节特征,但只抑制了病灶周围的部分噪声。RED-CNN 噪声抑制能力优于 DnCNN 与 WGAN,但病灶的边缘纹理特征与细节特征比较模糊。EMANet 噪声抑制能力最佳,同时也保留了细节纹理特征。

3.3 盲降噪效果分析

由于无法获得全剂量(即完全干净)的骨显像,因此,本文从主观和客观两个角度对盲降噪后的骨显像进行了深入分析。主观上,通过盲降噪的骨显像和模型输出的噪声图的视觉评价来分析降噪效果。客观上,借助骨显像病灶自动分割实验进行客观评估。

3.3.1 视觉效果

各个模型的盲降噪骨显像如图 12 所示。从图(b)来看,4 种深度学习模型的降噪图像与原图差异不大,但却增强了病灶特征(如矩形框所示)。由图(a)和图(c)可见, BM3D 的降噪图仍比较模糊,病灶细节损失严重。相比原图, RED-CNN 去除了少量噪声,但在图(a)的胸骨剑突处和图(c)右侧肋骨处引入

了新的伪影噪声。WGAN 也只去除了少许噪声,图(b)右侧手臂上也引入了少量伪影,降噪效果不明显。DnCNN 能够去除骨骼周围的大部分噪声,并且将图(a)中因肾脏残留注射试剂而引入的噪声也进行了去除,但同时手臂骨骼特征也遭到了破坏。与其他几种模型相比,本文设计的 EMANet 模型降噪整体效果最好,将病灶和骨骼周围的噪声进行了部分去除,并且保留了骨显像高频细节特征、防止了新的噪声引入。

细节上,选取图(c)中的胸腔(矩形框区域)进行观察,如图 13 所示。可以直观看出,5 种算法的降噪效果为: $DnCNN > EMANet > RED-CNN > WGAN > BM3D$ 。BM3D 和 WGAN 算法的重建图都丢失了部分高频细节特征。DnCNN 去除了肋骨间的部分噪声,效果较好。RED-CNN 的重建图像与原图相比,无明显变化。EMANet 也能去除肋骨间的少量噪声,效果不如 DnCNN,但优于其余 3 种模型。

为进一步验证 EMANet 的降噪性能,本文通过原骨显像与各模型的降噪骨显像差值图(即噪声图)像来进行分析。如图 14 所示,不同降噪算法的噪声学习能力为: $EMANet > WGAN > DnCNN > RED-CNN$ 。从图(a)和图(c)可以看出, RED-CNN、WGAN 算法的噪声图中包含较多病灶的边缘特征,造成了降噪骨显像边缘比较模糊、平滑。从图(b)可以发现, DnCNN 的噪声图提取到的病灶边缘特征较少,但将部分病灶信息当作了噪声,造成降噪骨显像病灶的对比度略微下降。相比上述算法,本文设计的 EMANet 预测的噪声主要分布在骨骼以及病灶周围,虽含有部分病灶边缘特征,但相对较少,噪声分布更为合理。

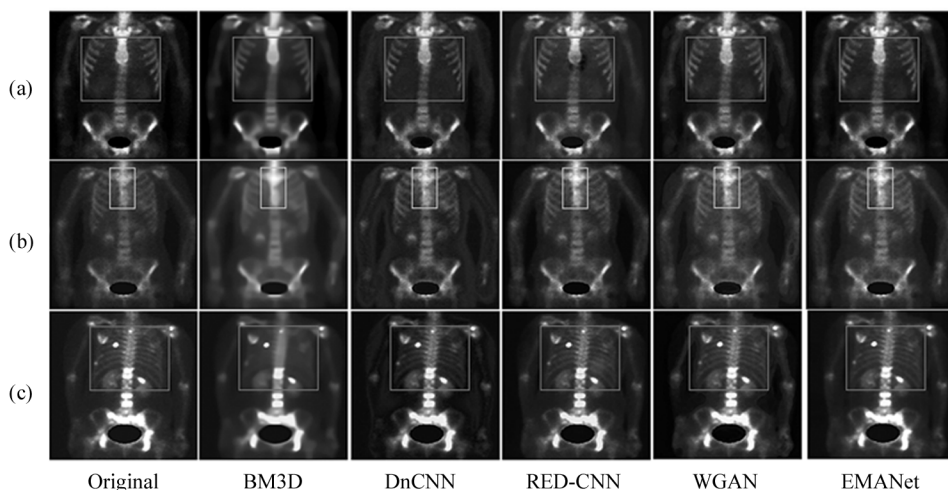


图 12 盲降噪效果对比

Fig. 12 Comparison of blind noise reduction effect

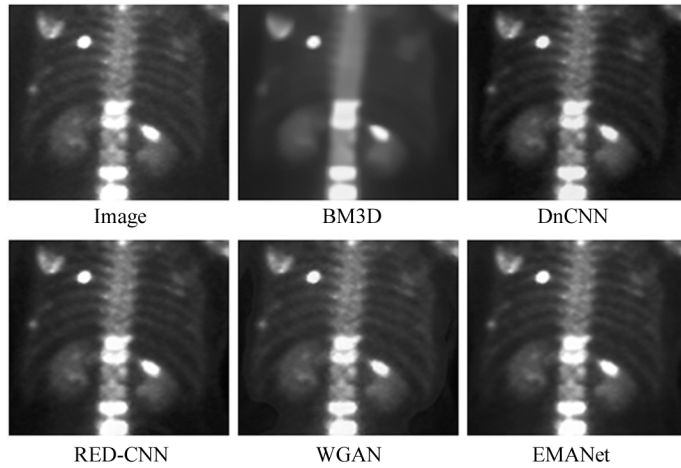


图 13 放大图 12(c) 中矩形框区域

Fig. 13 Enlarging the rectangle area in Fig. 12(c)

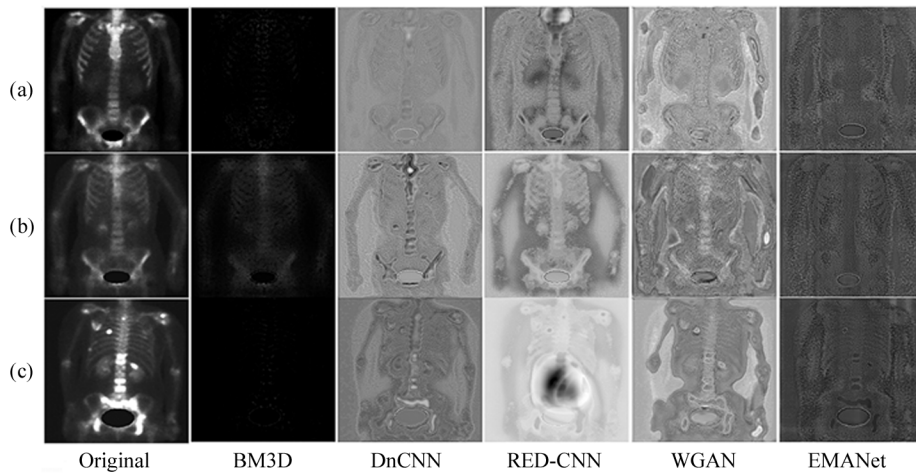


图 14 噪声图对比

Fig. 14 Noise picture comparison

3.3.2 病灶自动分割实验

为了测试本文降噪方法的骨显像病灶分割效果,选用了带残差结构和 SE 模块的 U-Net 模型对上述生成结果进行了病灶自动分割实验(二值分割)。采用精确率 (*precision*)、交并比 (*IoU*)、召回率 (*recall*) 3 种指标来评估分割效果,对比结果见表 8 和图 15。可以看出,4 种深度学习算法的降噪图分割效果均有提升。其中,EMANet 效果最好,相比于原图, *presion*、*IoU*、*recall* 3 项指标分别提升了 0.023、0.013、0.023。由图 15 可见,U-Net 模型基本能分割出 5 种算法降噪图的大多数病灶区域,并且病灶形态无太大差异。从矩形框中区域来看,EMANet 降噪图在肋骨小病灶和胸骨病灶区域有着更加精细的分割,更加接近专家手动标定的标签图。这表明本文设计的 EMANet 模型生成的降噪骨显像能

够提升病灶的自动分割效果。

表 8 病灶自动分割效果

Tab. 8 Effect of automatic segmentation of lesions

Method	<i>Precision</i>	<i>IoU</i>	<i>Recall</i>
Original	0.725 5	0.498 2	0.599 4
BM3D	0.670 1	0.459 3	0.593 3
DnCNN	0.734 2	0.505 7	0.624 0
RED-CNN	0.738 8	0.497 9	0.608 6
WGAN	0.736 3	0.501 7	0.605 3
EMANet	0.748 4	0.511 2	0.622 9

4 结 论

针对 SPECT 骨显像质量不高、易受噪声干扰的问题,本文提出了一种基于 U-Net 的两阶段 SPECT 骨显像降噪模型 EMANet。首先,设计了一种 U-

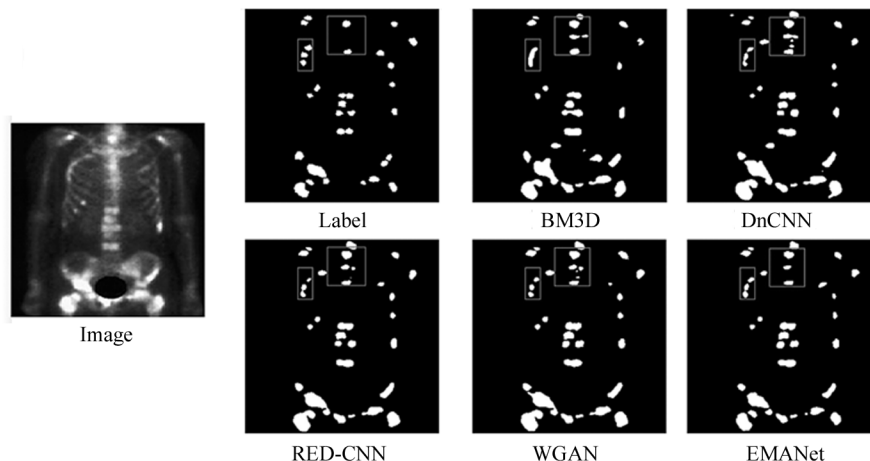


图 15 病灶分割效果图

Fig. 15 Lesion segmentation renderings

Net 噪声模型自适应估计出每张骨显像的噪声水平, 并作为主干网络的先验噪声。其次, 主干降噪网络以 U-Net 为基本框架, 采用多尺度特征融合、通道-空间注意力机制以及密集残差连接方式的方法提升模型噪声提取能力, 输出更精确的噪声图像。同时, 结合残差学习、复合损失函数优化训练效果、保留病灶细节特征, 并采用迁移学习策略提升模型泛化能力。实验结果表明, 本文设计的降噪模型能够有效降低骨显像中的噪声, 并保留原图中的病灶细节特征, 优于目前的主流算法。此外, 通过盲降噪能够有效改善原骨显像质量, 对病灶自动分割效果有一定提升。

参考文献:

- [1] FENG J X, JIANG J. Deep learning-based chest CT image features in diagnosis of lung cancer[J]. *Computational and Mathematical Methods in Medicine*, 2022, 2022: 1-7.
- [2] RONG Y, REN H, DING X J. MRI and bone scintigraphy for breast cancer bone metastase: A meta-analysis[J]. *Open Medicine*, 2019, 14(1): 317-323.
- [3] MADSEN M T. Recent advances in SPECT imaging[J]. *Journal of Nuclear Medicine*, 2007, 48(4): 661-673.
- [4] CHO S G, KIM J, SONG H C. Radiation safety in nuclear medicine procedures[J]. *Nuclear Medicine and Molecular Imaging*, 2017, 51(1): 11-16.
- [5] COORAY S, TAKEUCHI T T, AKAHDRI T, et al. An iterative reconstruction algorithm for Faraday tomography[J]. *Monthly Notices of the Royal Astronomical Society*, 2020, 500(4): 5129-5141.
- [6] HE X Y, ZHANG A Q, ZHAO S M. Quantum locality preserving projection algorithm[J]. *Quantum Information Processing*, 2022, 21(3): 1-22.
- [7] ZHENG W F, YANG B, XIAO Y, et al. Low-dose CT image post-processing based on learn-type sparse transform[J]. *Sensors*, 2022, 22(8): 1-17.
- [8] ZHANG Y K, ZHANG J Y, LU H B. Statistical sinogram smoothing for low-dose CT with segmentation-based adaptive filtering[J]. *IEEE Transactions on Nuclear Science*, 2010, 57(5): 2587-2598.
- [9] GUI Z G, LIU Y. Noise reduction for low-dose X-ray computed tomography with fuzzy filter[J]. *Optik*, 2012, 123(13): 1207-1211.
- [10] YU L F, MANDUCA A, TRZASKO J D, et al. Sinogram smoothing with bilateral filtering for low-dose CT[C]// *Proceedings of SPIE. Medical Imaging 2008: Physics of Medical Imaging*, February 16-21, 2008, Bellingham, Washington, USA: SPIE, 2008, 6913: 1-8.
- [11] ZHANG R Q, DONG H Y, PAL D, et al. A Gaussian mixture MRF for model-based iterative reconstruction with applications to low-dose X-ray CT[J]. *IEEE Transactions on Computational Imaging*, 2016, 2(3): 359-374.
- [12] CHEN H, ZHANG Y, ZHANG W H, et al. Low-dose CT via convolutional neural network[J]. *Biomedical Optics Express*, 2017, 8(2): 679-694.
- [13] KANG E, MIN J, YE J C. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction[J]. *Medical Physics*, 2017, 44(10): 360-375.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[M]// *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2015: 234-241.

- [15] CHEN H, ZHANG Y, KALRA M K, et al. Low-dose CT with a residual encoder-decoder convolutional neural network [J]. *IEEE Transactions on Medical Imaging*, 2017, 36(12): 2524-2535.
- [16] YANG Q S, YAN P K, ZHANG Y B, et al. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss [J]. *IEEE Transactions on Medical Imaging*, 2018, 37(6): 1348-1357.
- [17] KEISUKE U, KOICHI O, MASAMI G, et al. Quantitative evaluation of deep convolutional neural network-based image denoising for low-dose computed tomography [J]. *Visual Computing for Industry, Biomedicine, and Art*, 2021, 4(1): 21-26.
- [18] WEN J H, LI Y C, WANG W. Wavelet-based denoising and its impact on analytical SPECT reconstruction with nonuniform attenuation compensation [J]. *International Journal of Imaging Systems and Technology*, 2013, 23(1): 36-43.
- [19] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C] // 2021 IEEE CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 13708-13717.
- [20] LIU Q, HE Z F, ZHANG Y H. Semantic segmentation of mechanical workshop scenes with branch-atrous convolutional neural networks [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2021, 33(1): 126-141.
- 刘强, 何自芬, 张印辉. 分支空洞卷积神经网络的机加工车间场景语义分割 [J]. *计算机辅助设计与图形学学报*, 2021, 33(1): 126-141.
- [21] HANG Z, GALLO O, FROSIO I, et al. Loss functions for image restoration with neural networks [J]. *IEEE Transactions on Computational Imaging*, 2017, 3(1): 47-57.
- [22] HUAN E Y, WEN G H. Transfer learning with deep convolutional neural network for constitution classification with face image [J]. *Multimedia Tools and Applications*, 2020, 79(17-18): 11905-11919.
- [23] MCCOLLOUGH C H, BARTLEY A C, CARTER R E, et al. Low-dose CT for the detection and classification of metastatic liver lesions [J]. *Medical Physics*, 2017, 44(10): 339-352.

作者简介:

罗仁泽 (1973—), 男, 博士, 教授, 硕士、博士生导师, 主要从事信号处理与人工智能等方向的研究。