

DOI:10.16136/j.joel.2022.10.0060

基于改进 U-Net 的 SPECT 骨显像病灶分割研究

余泓^{1*}, 罗仁泽¹, 陈春梦², 罗任权¹, 李华督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)

摘要: 在核医学中, 单光子发射计算机断层(single-photon emission computed tomography, SPECT)骨显像是辅助医师诊断癌症的重要手段。针对骨显像图像信噪比低、边界模糊、病灶小难以提取和人工勾画病灶耗时等问题, 提出一种基于改进 U-Net 网络的骨显像病灶自动分割算法。该算法在 U-Net 的原卷积块基础上, 采用了多尺度密集连接(multi-scale dense connection, MDC)的方式来提高对小病灶特征的提取能力, 同时解决了网络加深后出现的梯度消失问题。其次, 为提取病灶的细节特征, 在密集连接和跳跃连接处引入了注意力机制结构。最后, 针对使用小样本数据集, 模型难以收敛的问题, 采用迁移学习的方法, 优化了模型的初始参数, 提升模型的泛化能力和分割效率。此外, 为了降低计算量、进一步提高分割效果, 对数据集进行了裁剪和去噪。同时, 将处理后的图像采用旋转、镜像等方法进行了数据扩充。实验结果表明, 改进的 U-Net 的识别精确率(*precision*)、平均交并比(mean intersection-over-union, *mIoU*)分别能达到 0.7352、0.4673, 效果优于目前主流的分割算法, 具有一定实际应用价值。

关键词: SPECT 骨显像; 多尺度密集连接(MDC); 图像处理; 注意力机制; 迁移学习

中图分类号: TP391 **文献标识码:** A **文章编号:** 1005-0086(2022)10-1110-11

SPECT bone imaging lesion segmentation based on improved U-Net

YU Hong^{1*}, LUO Renze¹, CHEN Chunmeng², LUO Renquan¹, LI Huadu¹

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

Abstract: In nuclear medicine, single-photon emission computed tomography (SPECT) bone imaging is an important means to assist physicians in diagnosing diseases. Aiming at the problems of low signal-to-noise ratio, blurred boundaries, small lesions, and time-consuming manual lesion delineation in bone imaging images, an automatic segmentation algorithm for bone imaging lesions based on improved U-Net network was proposed. Based on the original convolution block of U-Net, the algorithm adopts a multi-scale dense connection (MDC) method to improve the extraction ability of small lesion features, and at the same time solves the problem of gradient disappearance after the network is deepened. Second, to extract detailed features of lesions, an attention mechanism structure is introduced at dense and skip connections. Finally, in view of the problem that the model is difficult to converge when using a small sample dataset, the transfer learning method is used to optimize the initial parameters of the model and improve the generalization ability and segmentation efficiency of the model. In addition, in order to reduce the amount of computation and further improve the segmentation effect, the dataset is cropped and denoised. At the same time, the processed images are augmented by rotation, mirroring and other methods. The experimental results show that the improved U-Net's recognition *precision* and mean intersection-over-union ratio (*mIoU*) can reach 0.7352 and 0.4673, respectively, which are better than the current mainstream segmentation algorithms, and have certain practical application value.

Key words: single-photon emission computed tomography (SPECT) bone imaging; multi-scale dense connections (MDC); image preprocessing; attention mechanism; transfer learning

* E-mail: 790622472@qq.com

收稿日期: 2022-01-25 修订日期: 2022-03-10

基金项目: 四川省科技计划项目(2019CXRC0027)资助项目

1 引言

随着医疗技术的发展,核医学影像识别技术成为了疾病诊断的重要手段。目前,核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)、全身磁共振显像(magnetic resonance imaging, MRI)^[1]和单光子发射计算机断层成像(single-photon emission computed tomography, SPECT)。其中, SPECT 全身骨显像具有灵敏性高、实用性强、成本低、成像视野大的特点,能够在早期显示出恶性肿瘤病灶,成为评价骨转移肿瘤最为重要的检查手段。

医生依靠丰富的阅片经验,可以通过 SPECT 全身骨显像和其他辅助信息来判断影像中的骨转移肿瘤病灶,但同时也存在着主观性强、重复性低以及分析不够等问题。为了能够解决这些问题,近年来许多研究者开展了图像处理技术^[2]和机器学习等^[3]方法的研究。HUANG 等^[4]提出了一种基于模糊集直方图阈值法和解剖学知识的图像分割方法,能够自动对原始图像数据进行分析 and 量化,定位出各个骨骼区域的关键点,并且识别病灶可能会出现的位置。SAJN^[5]提出了一种具有鲁棒性的、基于医学知识的全身骨显像分割算法,并且能够支持标准图像处理算法,可以利用模式分类算法对分割出的骨骼区域进行参数化,然后采用机器学习算法对病变区域进行分类,能够检测,自动对病变区域进行评估。GAO 等^[6]将 256×1024 的骨显像裁剪成 256×256 的图像后,然后采用 K-means 聚类法、区域生长法、和 C-V 模型算法对 SPECT 骨显像图像中的病灶进行微调参数分割,实验结果表明这几种算法对于 SPECT 骨显像图像病灶分割是可行的。ASLANTAS 等^[7]提出了一种 CAD 诊断系统——“CADBOSS”,该系统能够利用人工神经网络(artificial neural network, ANN)对 SPECT 骨显像图像进行分类,判断是否存在骨转移。基于数字图像处理的方法识别精度不高,传统机器学习算法虽然识别精度得到了提高,但计算效率和特异性较低。

深度学习作为机器学习的分支广泛应用于语音识别、计算机视觉、医疗等领域。近几年,相继出现了 Alexnet^[8]、VGG(visual geometry group)^[9]、残差网络(ResNet)^[10]、Googlenet^[11]等高效的卷积神经网络(convolutional neural networks, CNN)框架,并且在一些全身骨显像图像分类研究中也得到了较好的识别效果。NIKOLAOS 等^[12]通过对 CNN 超参数进行选择 and 微调,将全身骨显像图像分为恶性和健康两类,分类精确度和灵敏度分别

达到了 97.38%、95.80%。CHARIS 等^[13]使用 778 张前列腺癌男性的骨显像作为数据集,通过浮点运算和改进的 CNN 减轻运算量,使用改进的 CNN 与 VGG16、ResNet、Inception 等结构进行对比,证明算法在识别骨转移方面均优于当时的主流算法。ZHENG 等^[14]通过对数据进行旋转、平移、镜像等处理来扩充数据,并对比了不同深度的 VGG 网络框架,判断良性与恶性病例的准确率、召回率(Recall)、F1 得分均达到了 99%。使用 CNN 对 SPECT 骨显像分类的准确率、特异性、Recall 均高于传统算法。目前,基于深度学习的方法主要关注的是骨显像图像分类研究,有关骨显像病灶分割方法的研究极少。

近年来, U-Net^[15]在医学影像分割任务中已经取得了较好的效果,其结构借鉴了 FCN(full convolution network)^[16]模型。但是由于设备的限制,拍摄的 SPECT 全身骨显像图像存在信噪比低、边界模糊,造成了不同患者之间的病灶差异较大。此外,良性病灶(关节炎、创伤、感染等)也会影响恶性肿瘤的识别,因此在进行图像分割时难度较大。此外,病人的病灶可能较多,人工勾画需要耗费大量时间成本。针对这些问题,提出了一种基于改进 U-Net 网络的骨显像病灶自动分割方法。首先,采用多尺度密集连接(multi-scale dense connection, MDC)卷积块替代 U-Net 中原有卷积块。其次,在密集连接和 U-Net 的跳跃连接处引入了注意力机制结构来关注病灶的细节特征提取。实验中,首先将 DICOM 文件使用归一化方法转化为 1024×256 pixels 大小的 RGB 图像。接着对骨显像图像进行裁剪,提取出肩部到盆骨区域,再使用图像去噪和数据增强方法对图像进行预处理,将处理后的图像输入改进的 U-Net 模型,最后使用迁移学习来初始化模型参数,缓解小样本数据集带来的影响。实验结果表明,改进 U-Net 在分割病灶时,平均交并比(mean intersection-over-union, mIoU)、精确率(precision)分别达到了 0.7352、0.4673,相比于 U-Net 原模型以及其他语义分割模型的效果有显著提升。

2 网络结构

2.1 基于注意力机制的门控结构(CAG)

注意力机制(coordinate attention, CA)能够学习“感兴趣”的特征^[17],该模块是一种将特征图中的详细位置信息引入通道注意力的结构。相比于其他轻量级的注意力模块,CA 不仅能够学习通道之间的关系,还能捕获特征图的空间位置信息与通道联系,并且通过降维的方式,大量减少了计算

量。具体原理如下:

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

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(w, j), \quad (2)$$

式中, $z_c^h(h)$ 表示在通道 c 中, 对高度 h 的第 i 行信息求和并取平均, 式(2)同理。

$$f = \sigma(F1([z_c^h, z_c^w])), \quad (3)$$

式中, $F1([z_c^h, z_c^w])$ 表示对上一步得到的两个特征进行拼接后, 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非线性变换。

$$g_c^h = \sigma(F_h(f^h)), \quad (4)$$

$$g_c^w = \sigma(F_w(f^w)). \quad (5)$$

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。

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (6)$$

最后, 为了将得到的空间特征信息作为注意力权重, 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, $y_c(i, j)$ 表整个 CA 模块的输出, 具体结构见图 1。

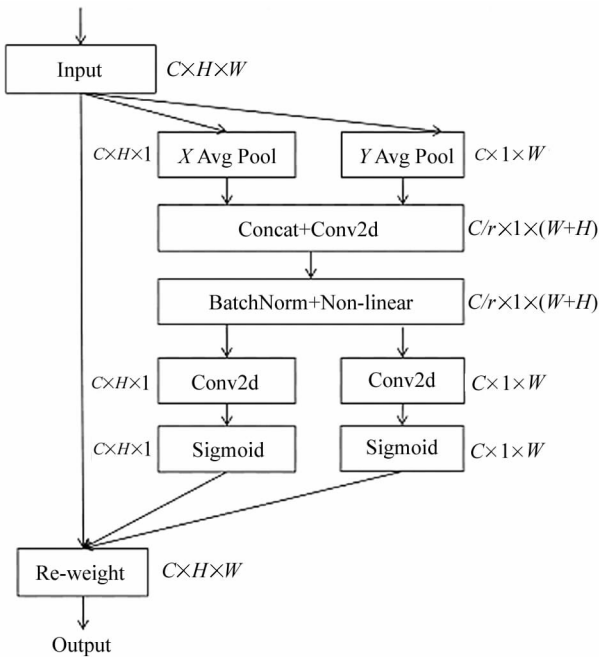


图 1 CA 模块

Fig. 1 CA module

传统 U-Net 模型中, 为了提取信息, 使用了跳跃连接的方式来结合编码部分和解码部分的信息, 但这种方式不能充分利用上下文信息, 由于骨显像中病灶属于小目标, 其纹理信息薄弱、边缘模糊, 分割

较为困难。为此, 在改进 U-Net 的跳跃连接处加入了结合 CA 的注意力门控结构 (CAG), 该结构能够同时关注两个输入特征图的通道和空间的重要程度。首先, 将两个相同大小的特征图 x_i (上采样后) 和 g_i (跳跃连接) 输入 CAG 并相加, 然后经过 CA 模块结构, 学习不同通道之间的特征信息, 并将空间特征信息嵌入其中, 将得到的特征权重大小别分乘以 g_i, x_i , 最后将权重更新后的两个输入特征图进行拼接并输出, 具体见图 2。

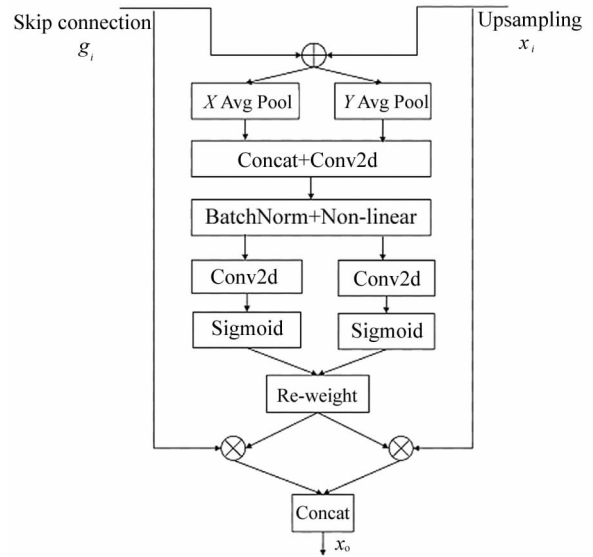


图 2 注意力门控模块 (CAG)

Fig. 2 Attention gating module

2.2 多尺度密集连接 (MDC Block)

在医学影像的语义分割任务中, 使用不同卷积核大小、不同扩张率的方法并行运行能够加强模型对小病灶特征的提取能力, 因此采用 Inception Res-net (IR)^[18] 结构替代原 U-Net 中的 3×3 卷积, 以提取不同尺度的特征信息, 其结构见图 3。

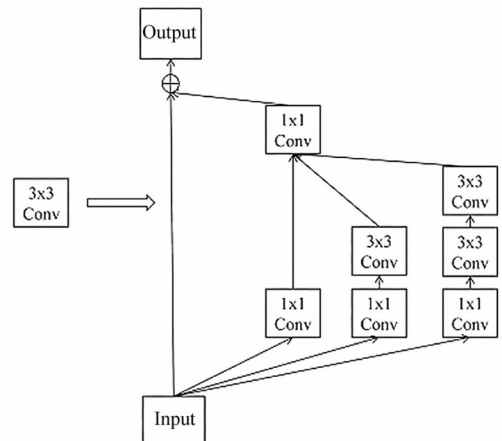


图 3 IR 模块

Fig. 3 IR module

IR 模块结构中包含 3 个分支,3 个分支分别使用 1×1 、 3×3 、 5×5 的卷积运算(由 2 个 3×3 卷积代替,减轻计算量)来提取不同尺度的病灶特征。每个卷积后都使用了 ReLU 激活函数和批量归一化来防止过拟合,然后将 3 个分支卷积后的结果进行特征拼接,并使用 1×1 卷积来压缩通道数,减少网络运算量。最后利用残差结构将经过 1×1 卷积的输入的特征与拼接压缩后的特征进行融合来恢复原有的部分特征,模块中从输入到输出的计算流程如下:

$$x_1 = \text{Conv}_{1 \times 1}(x_{\text{in}}), \quad (7)$$

$$x_2 = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(x_{\text{in}})), \quad (8)$$

$$x_3 = \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(x_{\text{in}}))), \quad (9)$$

$$x_c = \text{concat}(x_1 + x_2 + x_3), \quad (10)$$

$$x_{\text{out}} = x_{\text{in}} + x_c, \quad (11)$$

式中, $\text{Conv}_{1 \times 1}$ 、 $\text{Conv}_{3 \times 3}$ 分别代表 1×1 卷积和 3×3 卷积, x_{in} 为 IR 模块的输入, x_1 、 x_2 、 x_3 分别为 3 个通道的输出特征图, x_{out} 为 IR 模块的输出。

为了增加病灶特征的利用率,模型采用了密集连接的方式,见图 4。这种结构能够将本层的信息特征与之前的卷积层所提取的信息通过特征拼接的方式进行结合,加强特征提取的能力。并且每次卷积后都使用了批量归一化层和 ReLU 激活函数,增加模型的运行效率和非线性表达。在重复利用特征信息的同时,该连接的方式也带来了非常大的运算量。因此,在每次 IR 模块后都使用了一个 1×1 的卷积来压缩通道数,减少网络运算量。第 i 层卷积的输入如下式所示:

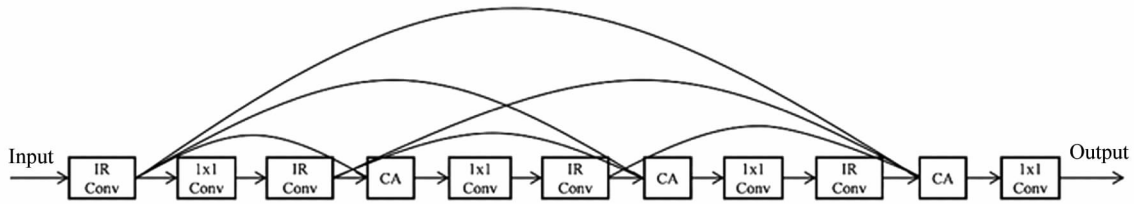


图 4 多尺度密集连接结构(MDC)

Fig. 4 Multi-scale dense connection structure

$$x_i = H_i([x_0, x_1, \dots, x_{i-1}]), \quad (12)$$

式中, x_i 代表第 i 层的输入, $i \in [1, 4]$, $[x_0, x_1, \dots, x_{i-1}]$ 代表第 i 层之前的所有特征图拼接后的特征。 H_i 代表非线性映射,即批量归一化和 ReLU 激活函数运算的组合。

另外,为了更加有效地利用密集连接结构,在每次特征拼接后都使用了一次 CA 注意力模块,来筛选更有利于提取病灶的特征。

2.3 改进 U-Net 网络结构

如图 5 所示为改进后的 U-Net 模型,每个 MDC 卷积块中使用了 4 个 IR 卷积、4 个 1×1 卷积以及 3 个 CA 模块。使用多尺度卷积的结构代替原有的 3×3 卷积,其中 Inception 结构能够提取出不同尺度的病灶特征,残差结构又能保持部分输入的特征信息。在每个卷积之间使用了密集连接的方式来加强对病灶特征的提取,并用 1×1 卷积来压缩通道数,减轻运算量,并且在每次密集连接时使用一次 CA 模块来让模型重点学习病灶特征。

在编码部分,使用 MDC 卷积块和最大池化层来对特征图进行特征提取和降采样,得到图像的高层特征。在解码部分,使用 MDC 卷积块和转置卷积对特征图进行特征提取和上采样,将特征图恢复至输入大小,并在跳跃连接处引入 CAG 来提取病灶的细节信息。为了更好地利用局部上下文信息,在输出之前,将解码部分每层的输出通过转置卷积恢复为原图尺度大小,进行多尺度特征融合后经过一个 CA 模块来关注有利于病灶分割的特征。最后,使用 1×1 卷积输出预测像素类别,分割出病灶区域。

由于数据样本较少,改进 U-Net 模型泛化性能较差。因此,在模型训练时采用了迁移学习^[19]的方法来学习网络。迁移学习可以将已有模型应用到新的或者有关联的领域中,极大缓解有深度学习中因样本不足而引起的过拟合问题。目前,迁移学习主要有模型迁移、参数迁移、特征迁移、样本迁移 4 种方法。为了赋予模型较好的初始化参数,本次实验采用参数迁移来实现迁移学习。

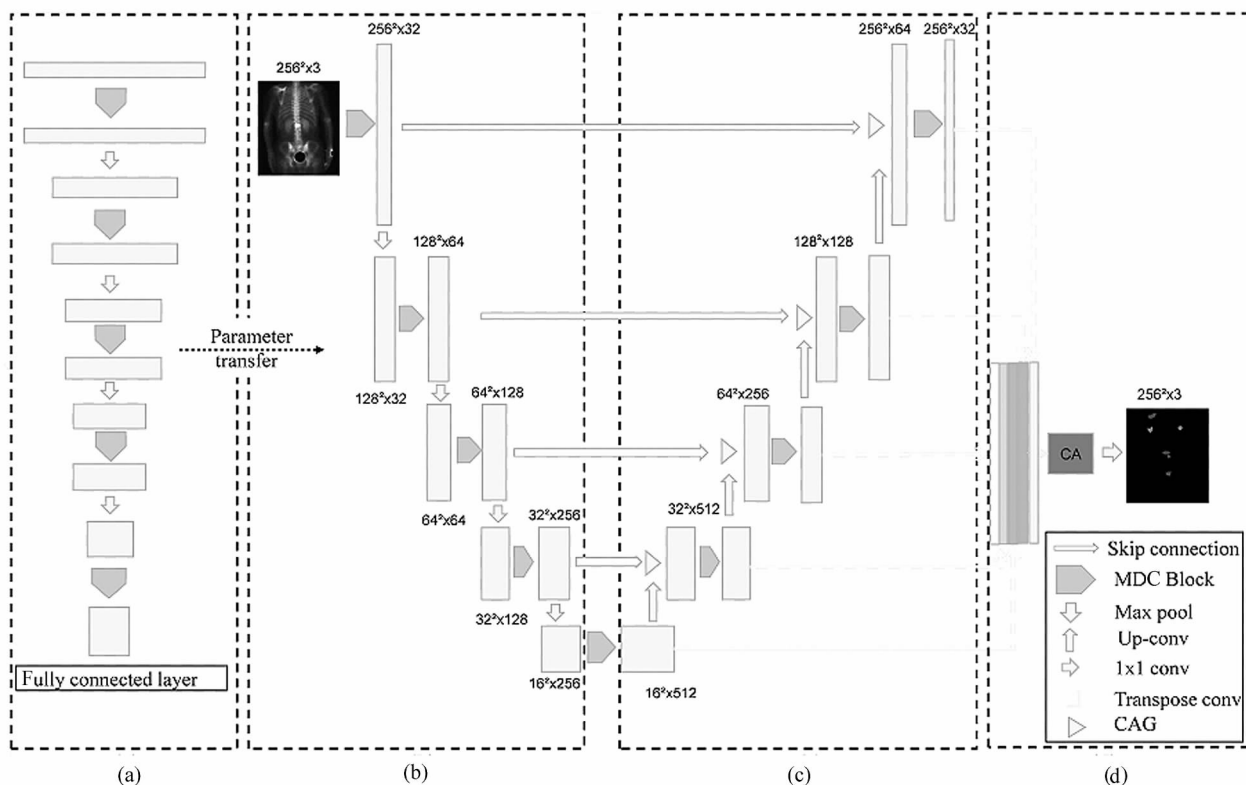


图5 改进 U-Net 结构: (a) 预训练模块; (b) 编码模块; (c) 解码模块; (d) 特征融合输出模块

Fig. 5 Improve U-Net network structure: (a) Pre-training module;

(b) Encoding module; (c) Decoding module; (d) Feature fusion output module

3 数据处理和模型训练

3.1 骨显像数据预处理流程及评价指标

SPECT 骨显像的原始数据是以 DICOM 格式进行存储的,无法直接输入到模型进行训练,需要转化

为 RGB 格式的图像。同时,将原图和标签文件进行裁剪来增加病灶像素所占比例。接着使用亮度对比度调节、去噪的方法增强图像的信噪比。最后,通过旋转、镜像、平移、缩放等图像增强的方式来扩充数据集,具体预处理流程见图 6。

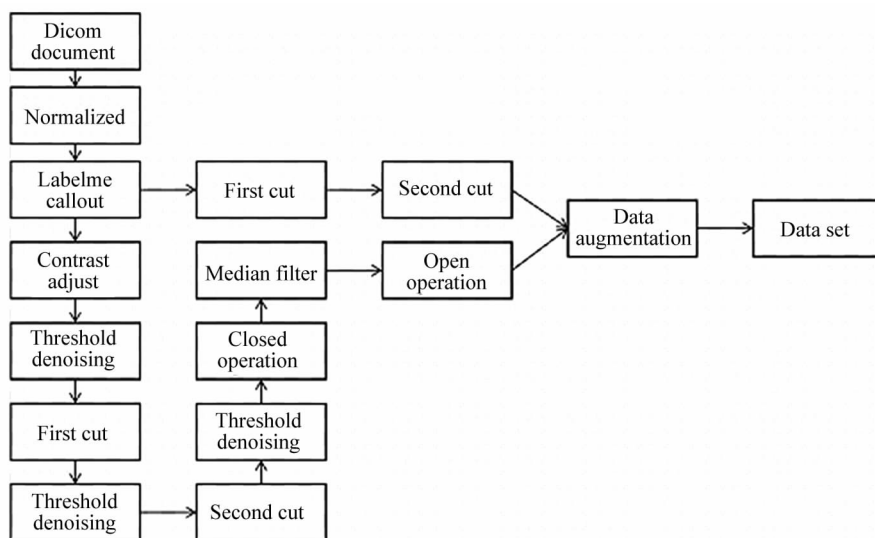


图6 骨显像图像预处理流程

Fig. 6 Preprocessing process of bone scan image

1) 图像格式转化。获取的 SPECT 全身骨显像数据是将大小为 1024×256 的元素矩阵储存起来的 DICOM 文件。DICOM 文件是医学图像和相关信息的国际标准,每个元素代表辐射剂量。其中,病灶区域的辐射剂量相比于周围区域更高,呈现出亮点特征。为了转换成能够送入 U-Net 的 RGB 图像,先通过归一化的方式将原文件中的元素剂量归一化为像素值在 $[0, 255]$ 之间的 RGB 图像,再使用图像标注软件 Labelme 对图像中的病灶进行人工标注。最后,将生成的 json 文件通过 python 代码转化为掩膜图像。

2) 图像裁剪、去噪。SPECT 骨显像中,人体病灶部位存在像素占整幅图比例较小、形状各异、信噪比低的问题。为解决以上问题,数据预处理中先通过手动阈值去噪,去除人体周围含有的少量噪声。接着分别从图像上部和下部进行扫描,去除像素值全为 0 的行数,即除头顶上部和脚下部的区域。然后根据全身骨显像图像中左右两侧到颈部和肩部的距离有突变的特点,定位出肩部和盆骨的高度,利用得到的高度值对原图和掩膜图像进行裁剪,得到胸腔和盆骨区域。经过裁剪后,病灶像素点所占图像比例得到了提升。最后,将裁剪得到的图像通过阈值法、中值滤波、以及开、闭运算 4 种滤波方法对裁剪后的图像进行去噪,具体见图 7。

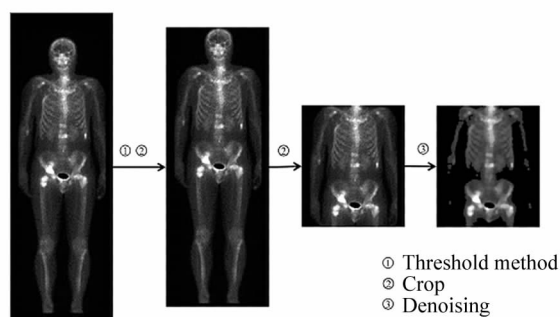


图 7 裁剪、去噪过程图

Fig. 7 Diagram of cropping and denoising process

3) 图像数据增强。模型进行训练,通过不断更新模型参数来使输出的预测图像与真实的掩膜标签相对应。如果训练图像越多,模型的泛化能力也就越好,分割出的病灶也就更准确。针对减少小样本训练带来的影响,在数据预处理时,对裁剪、去噪后的 RGB 图像采用了旋转、平移、镜像、缩放 4 种图像增强的方法,将训练集图像数量进行扩充。

本次实验采用 $precision$ 、 $mIoU$ 、 $Recall$ 作为分

割模型的评价指标,见式(13)—(15):

$$precision = \frac{TP}{TP + FP}, \quad (13)$$

$$mIoU = \frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP}, \quad (14)$$

$$Recall = \frac{TP}{TP + FN}, \quad (15)$$

式中, TP 为真阳性,表示预测为正样本的正样本; FP 为假阳性,表示预测为负样本的正样本; FN 为真阴性,表示预测为负样本的正样本; TN 为假阴性,表示预测为负样本的负样本。

3.2 数据准备及迭代设置

首先将改进 U-Net 的编码部分提取出来,在最后一层卷积后加入全连接层进行图像分类预训练。预训练数据集使用的 COVID-19 Radiography 数据集,此数据集中包含 4 类(冠状病毒病、可疑肺炎、正常、病毒性肺炎)共 21 165 张 X 光胸片,尺寸大小均为 299×299 。预训练结束后,保存训练好的模型参数并用这部分参数替换掉编码部分的初始化参数,然后冻结(即学习率为 0)。最后,通过已经获取的编码部分参数去训练解码部分参数。

通过迁移学习完成模型编码、解码部分的训练后,将 124 例 SPECT 骨显像病例(每例病例中包含人体正面、背部两张图像),共 242 张图像作为原始数据样本。其中,含有骨转移病灶的图像有 66 张,没有骨转移病灶的图像有 176 张。测试集、训练集样本数分别为 217 和 25 张。在预处理流程中,使用了阈值去噪、闭运算、中值滤波、开运算方法。在把图像送入模型之前,将每张图像统一为 256×256 的尺寸大小。训练集图像样本通过旋转、平移、镜像、缩放的方式扩充至 2 170 张,进行数据增强前后的数据量变化如表 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 type	Training set	Test set	Total
Before enhancement	217	25	242
After enhancement	2 170	25	2 195

本次实验的硬件配置为: AMD Ryzen5 3600 CPU; 12 GB Nvidia GeForce 3060 显卡; 软件环境为 Python3.6、pytorch1.8.0、CUDA11.1、cuDNN8.1.0; 使用 RAdam^[10] 优化器,学习率设为 0.001, batch

size 为 4。本实验采用的损失函数为交叉熵函数,公式如下:

$$H(p, q) = - \sum_{i=1}^C P(x_i) \log(q(x_i)), \quad (16)$$

标签分为背景、良性病灶(退行性改变、关节炎等)、恶性病灶(骨转移肿瘤)3类,因此本实验采用多分类交叉损失函数。其中, p 代表真实标签, q 代表预测概率, $H(p, q)$ 代表 p, q 之间的差异。

在训练时,将 5 种模型进行了对比。具体为 FCN, U-Net, Deeplab v3+^[20], 只加入密集连接结构的 Dense_U-Net^[21] 以及改进的 U-Net。在迁移学习中,将迭代次数设为 100, 获取编码部分的参数,再进行解码部分的训练。解码部分训练的 Loss 随迭代次数的变化见图 8。改进 U-Net 大约迭代 200 次后达到收敛,收敛速度略高于 U-Net 和 Dense_U-Net,但慢于 FCN 和 Deeplab v3+。总体上,所有模型在 500 次迭代后基本达到收敛状态。

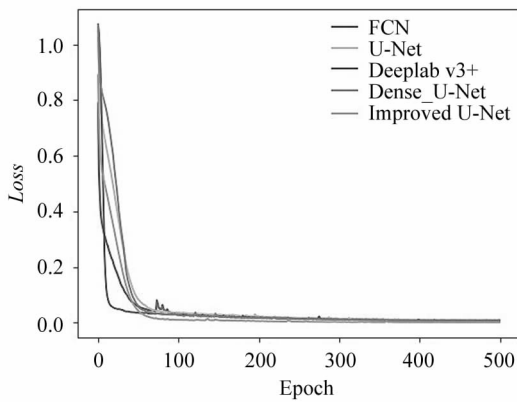


图 8 5 种模型的 Loss 曲线
Fig. 8 Loss curve of five models

3.3 模型参数、结构性能训练

3.3.1 训练样本预处理对模型的影响

由于 SPECT 图像信噪比较低,且病灶属于小目标,数据量较少,影响分割的效果。因此,实验采用裁剪、去噪和数据增强的方法对图像进行预处理。为验证裁剪、去噪和数据增强的方式是否对最终的分割效果有提升,使用 U-Net 模型分别对裁剪、去噪和数据增强图像进行训练,然后用测试集的图像测试分割效果,病灶分割后的 $precision$ 和 $mIoU$ 值见表 2。

由表 2 可以看出,原图的 $precision$ 和 $mIoU$ 较低。使用裁剪后的 SPECT 图像进行训练, $precision$ 和 $mIoU$ 分别提升了 0.102、0.122 4,主要是因为裁剪只保留了肩部到盆骨部位的区域,病灶大多集中

在此部分。在去除头部、腿部区域后,病灶占整幅图的像素比例得到了提高,使 U-Net 能够更加容易地进行特征提取。去噪后,分割结果的 $precision$ 和 $mIoU$ 相比于去噪前分别提升了 0.105 3、0.098 1,图 9 为提升亮度前后的效果对比,可以看出,人体内部和外部都存在部分噪声,通过阈值去噪能够去掉这部分噪声,闭运算可以填充骨架轮廓内缺少的像素。同时,使用中值滤波和开运算能够去除阈值去噪时引入的少量噪声,经过这些处理后使得病灶特征更加明显。最后通过数据增强的方式扩充数据量后, $precision$ 和 $mIoU$ 相比于数据增强前提升了 0.09、0.299,能够一定程度上减少小样本训练带来的影响,优化了模型的性能。

表 2 3 个阶段图像的分割效果对比
Tab. 2 Comparison of Image Segmentation effects in three stages

Sample processing	$precision$	$mIoU$
Original map	0.354 7	0.143 5
After cropping	0.456 7	0.265 9
After denoising	0.562 0	0.364 0
After data augmentation	0.652 0	0.393 9

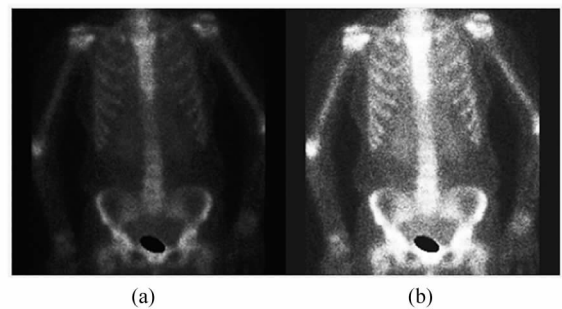


图 9 亮度提升前后:(a) 原图;(b) 亮度提升后
Fig. 9 Before and after the brightness increase:
(a) Original picture; (b) After brightness increase

3.3.2 MDC 结构对比

为验证多尺度卷积对病灶特征的提取效果,实验分别设置了 4 组不同尺度的卷积进行对比。单通道采用单个 3×3 卷积,双通道在单通道的基础上多增加了一个 1×1 卷积分支,并且将提取的特征图与 3×3 卷积的提取结果进行拼接,三通道为 IR 卷积模块,四通道使用了 4 个卷积大小的卷积进行特征提取,分别为 1×1 卷积、 3×3 卷积、 5×5 卷积、 7×7 卷积,为了降低运算量,使用两个 3×3 卷积代替 5×5 卷积,3 个 3×3 卷积代替 7×7 卷积,图 10 为各个通

道的提取结构,通过 *precision*、*mIoU* 和完成一个 epoch 耗时间 t 对各个模块进行比较,见表 3。

由表 3 可以看出,随着通道和卷积数目增多,模型能够提取到更多的特征信息,最后识别效果越来越好。三通道的运行时间虽略高于单通道和双通道,但 *precision* 和 *mIoU* 分别 0.6537、0.4264,均高于单通道和双通道。随着通道数的增加,运算量也随之增大,运行时间也逐渐增加,四通道的 *precision*、*mIoU* 略微高于三通道,但耗费的时间是三通道

的 2 倍多。综合考虑运行时间和识别效果,采用三通道的多尺度卷积是对病灶特征提取最为合理。

使用多尺度卷积代替卷积可以提升病灶识别效果,但同时会带来更多的参数。在小样本数据集中,多尺度卷积很容易造成梯度消失的情况。在卷积块中加入残差结构可以有效解决这一问题,而密集连接相比于残差结构对特征信息的利用率更高。表 4 对比了是否采用密集连接和残差结构对病灶的分割效果。

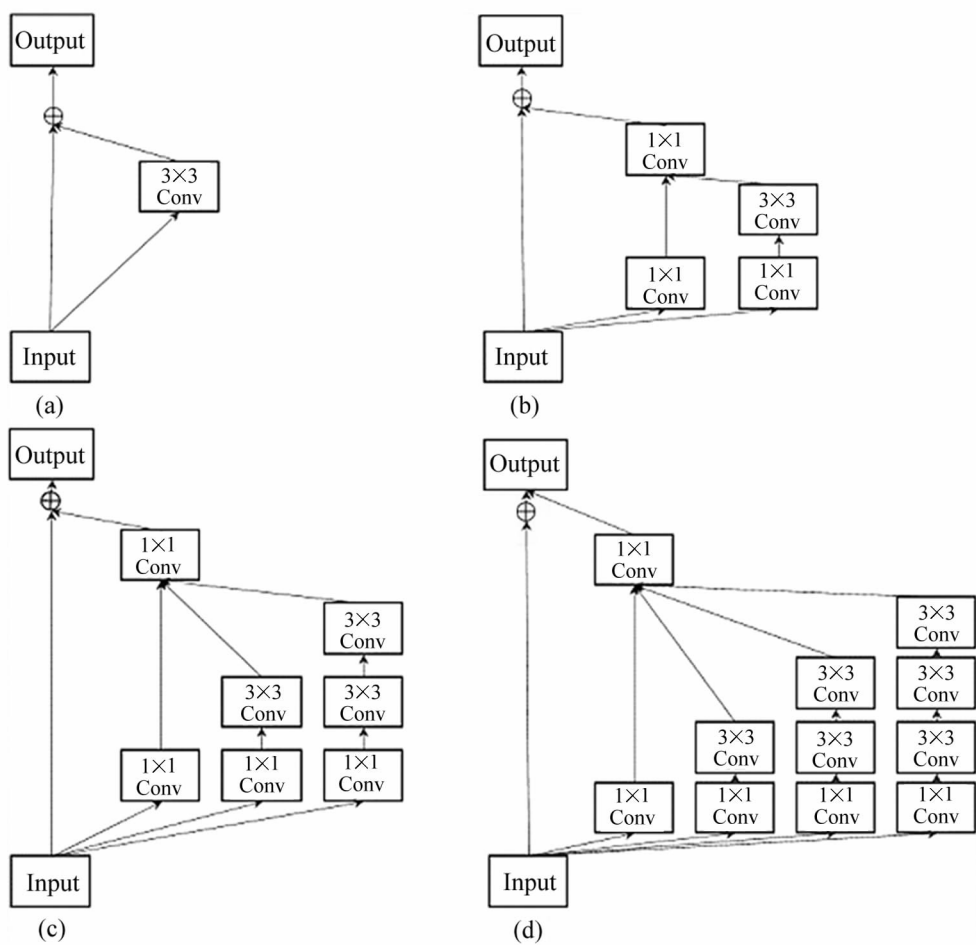


图 10 不同尺度卷积提取:(a) 单通道 1×1 卷积模块;(b) 双通道 3×3 卷积模块;
(c) 三通道 3×3 模块;(d) 四通道 3×3 卷积模块

Fig. 10 Different scale convolution extraction: (a) Single channel 1×1 convolution module;
(b) Dual channel 3×3 convolution module; (c) Three channel 3×3 module; (d) Four channel 3×3 convolution module

表 3 不同多尺度卷积效果对比

Tab. 3 Comparison of the effects of different multi-scale convolution			
Channel number	<i>precision</i>	<i>mIoU</i>	t/min
1	0.6597	0.4164	2.151
2	0.5854	0.4153	4.620
3	0.6537	0.4264	7.629
4	0.6234	0.4292	14.986

表 4 不同连接效果对比

Tab. 4 Comparison of different connection effects		
Structure	<i>precision</i>	<i>mIoU</i>
Residual connection	0.6091	0.4057
Dense connection	0.6661	0.4199

通过对比结果可以看出,采用密集连接方式时,模型的 *precision* 和 *mIoU* 相比残差连接更高,*preci-*

sion 和 $mIoU$ 分别提升了 0.0570、0.0142,更有利于病灶的分割,因此在卷积块中采用密集连接的方式对病灶特征进行提取。

3.3.3 注意力机制结构对比

在结合 MDC 的基础上,为验证使用的注意力机制结构对病灶特征提取的提升效果,设置了 4 组对比。第 1 组不加入注意力模块,第 2 组为在密集连接处加入 CA 模块,第 3 组为在跳跃连接处采用 CAG 结构,第四组将 CA 模块和 CAG 结构一起加入。

由表 5 可以看出,第 4 组实验分割的 $precision$ 和 $mIoU$ 均高于其他 3 组。由此可看出,CA 模块结合 MDC 结构能够更加高效地利用提取的特征,CAG 有助于 U-Net 充分利用上下文信息,提升病灶的分割效果。

表 5 注意力机制结构对比

Tab. 5 Structure comparison of attention mechanism

Structure	$precision$	$mIoU$
—	0.6537	0.4264
CA	0.6791	0.4296
CAG	0.6810	0.4321
CAG+CA	0.7148	0.4416

总体来说,采用样本数据处理以及密集连接能有效地提高模型的性能。对于不同尺度卷积通道数的选择应结合时间成本和运行设备的因素,适当地选取通道数。CA 模块和 CAG 能够进一步提升模型性能。综上所述,本次骨显像分割改进的 U-Net 模型训练的迭代次数设置为 500 次。同时样本经过预处理后,模型采用 MDC 的方式结合 CA 模块和 CAG 对病灶特征进行提取。

4 实验与应用

为了更好地评价改进 U-Net 的分割效果,选取了 10 张没有经过专家标定的病灶图用训练完成的模型进行了应用,并与经典的语义分割算法进行了效果对比。病灶分割结果的平均值见表 6。从各个指标数据来看,改进 U-Net 模型的 $precision$ 、 $mIoU$ 和 $Recall$ 要高于其他模型,原因是多尺度特征结构能够提取到不同尺度的病灶特征。同时,CA 模块、CAG 以及解码部分输出后的多尺度特征融合模块对网络的性能有一定提升。这些结构使模型在训练时能够得到图像不同尺度的感受野以及学习到不同层次的语义特征,进一步加强对病灶特征的提取。

表 6 模型分割效果对比

Tab. 6 Comparison of segmentation effects of each model

Model	$precision$	$mIoU$	$Recall$
FCN8s ^[16]	0.6102	0.3867	0.5667
U-Net ^[15]	0.6323	0.4001	0.5517
Deeplab v3+ ^[20]	0.5930	0.4119	0.5736
Dense_U-Net ^[21]	0.6531	0.4160	0.5462
Improved U-Net	0.7352	0.4673	0.5744

此外,将 10 张原始病灶图送给专家进行标定。其中 3 张专家和人工智能模型标定的结果见图 11。图中,由上到下依次为:(a)为重度病灶图,(b)为中度病灶图,(c)为轻度病灶图。其中,重度病灶图中全部为恶性肿瘤,中度病灶图中包含恶性和良性两种肿瘤,轻度病灶图中只有良性肿瘤。从整体来看,轻度病灶图容易分割,其原因在于轻度病灶图中病灶数量较少,容易识别。

对于重、中度病灶图,各个模型容易分割出错的地方主要集中在肩部、胸骨、肋骨较暗区域和髂骨翼,见图 11 红色矩形框。当肩部的病灶成镜像对称时,重、中度病灶图的各个模型的分割结果显示:FCN、U-Net、Deeplab v3+三个模型容易将肩部上的恶性病灶分割成良性病灶,原因是训练集中肩部的恶性病灶数较少,良性病灶数多且病灶面积小,造成了这 3 种模型对恶性病灶特征的分割效果不是很理想。而 Dense_U-Net 和改进 U-Net 中卷积之间使用了密集连接方式,能够重复利用卷积提取到的恶性病灶特征,较好地分割出了肩部上的恶性病灶,并且改进 U-Net 模型中使用了 CA 结构使分割结果比 Dense_U-Net 更加精细。

颈部和胸骨两处部位多数情况下的变异为良性变异,尤其是胸骨处,良性变异的情况较多,呈现出亮点。U-Net 和 Deeplab v3+模型将这部分分割成了良性病灶,FCN、Dense_U-Net 能够部分将恶性病灶分割出来,而改进 U-Net 分割出病灶区域的面积更大,原因是采用了多尺度卷积提取到了不同尺度的病灶特征,但将胸骨柄中的病灶也分割成了恶性病灶。在第 1 张原图的肋骨处存在几个亮度较暗和面积较小的恶性病灶,前 4 种模型没能将这部分病灶分割出来,而改进 U-Net 模型也是因为有多尺度卷积提取的优势,成功将这部分病灶分割出来。在髂骨翼处,由于左右病灶不对称,改进 U-Net 能够将这部分区域特征提取到。但分类出现错误,而前 4 种模型能够成功分割出这部分区域。

总体来说,各个模型都能分割出大多数的病灶

区域,分割出病灶分布的总体形态没有太大的差别。但改进 U-Net 分割结果图更加精细,对分割出的病

灶区域的种类分类更加准确,更接近专家标记的结果。

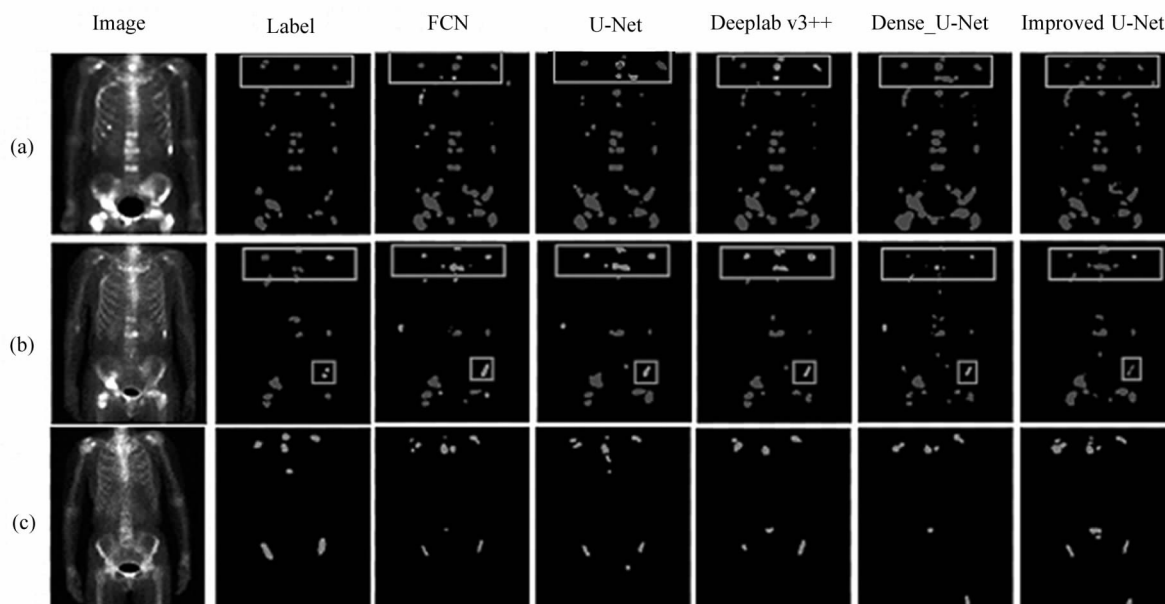


图 11 分割效果对比:(a) 重度病灶图;(b) 中度病灶图;(c) 轻度病灶图

Fig. 11 Segmentation effect comparison:(a) Severe lesions picture;(b) Moderate lesions picture;(c) Mild lesions picture

5 结 论

本文提出了一种基于改进 U-Net 的骨显像病灶自动分割方法。该模型采用 MDC 卷积块提取不同尺度的病灶特征,提高模型提取小病灶特征能力。同时,在密集连接和跳跃连接处引入了注意力机制结构,有助于模型充分结合上下文信息来重点关注病灶特征,增强模型的病灶分割效果。在训练时,针对小样本训练容易过拟合的问题,使用迁移学习初始化网络参数,加快了模型的收敛速度。在数据集的处理过程中,使用了裁剪、去噪等预处理方法来让病灶特征更加容易被改进 U-Net 提取。此外,采用数据增强的方式对数据集进行了扩充,缓解小样本对模型性能的影响。实验结果表明,经过数据集处理和模型改进后,分割出的病灶更加精准,模型分割的 *precision*、*mIoU* 分别达到了 0.735 2、0.467 3,相比于其他语义分割算法的分割效果,有一定提升。该模型更有利于辅助医生发现病灶区域,提高诊断效率。

参考文献:

[1] WANG Q. Image segmentation and registration for bone scintigraphy images[D]. Shanghai: Shanghai Jiao Tong

University, 2012:15-25.

王强. 骨扫描图像分割及配准算法研究[D]. 上海: 上海交通大学, 2012:15-25.

[2] MADSEN M T. Recent advances in SPECT imaging[J]. Journal of Nuclear Medicine, 2007, 48(4): 661-673.

[3] ZHANG X H, SHI Q L, WANG B, et al. Review of machine learning algorithms in traditional Chinese medicine[J]. Computer Science, 2018, 45(11A): 32-36.

张晓航, 石清磊, 王斌, 等. 机器学习算法在中医诊疗中的研究综述[J]. 计算机科学, 2018, 45(11A): 32-36.

[4] HUANG J Y, KAO P F, CHEN Y S. A set of image processing algorithms for computer-aided diagnosis in nuclear medicine whole body bone scan images[J]. IEEE Transactions on Nuclear Science, 2007, 54(3): 514-522.

[5] SAJN L, KONONENKO I, MILCINSKI M. Computerized segmentation and diagnostics of whole-body bone scintigrams[J]. Computerized Medical Imaging and Graphics, 2007, 31(7): 531-541.

[6] GAO R T, LIN Q, LUO M Y, et al. Automatic segmentation of metastasis lesions in SPECT bone scan images[C]// 2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS), December 11-13, 2020, Chongqing, China. New York: IEEE, 2020, 214-219.

- [7] ASLANTA A, DANDIL E, SAGLAM, et al. CADBOSS: a computer-aided diagnosis system for whole-body bone scintigraphy scans[J]. *Journal of Cancer Research & Therapeutics*, 2016, 12(2): 787-792.
- [8] ERGUN H, SERT M. Feature-level fusion of convolutional neural networks for visual object classification[C]//24th IEEE Signal Processing and Communication Application Conference (SIU), May 16-19, 2016, Zonguldak, Turkey. New York: IEEE, 2016: 2173-2176.
- [9] WANG K, CHEN Z X, WU Q M J, et al. Face recognition using AMVP and WSRC under variable illumination and pose[J]. *Neural Computing & Applications*, 2019, 31(8): 3805-3818.
- [10] VAKERIANI D, POLI R. Cyborg groups enhance face recognition in crowded environments[J]. *Plos One*, 2019, 14(3): 1-17.
- [11] LEE Y, NAM S. Performance comparisons of AlexNet and GoogLeNet in cell growth inhibition IC50 prediction[J]. *International Journal of Molecular Sciences*, 2021, 22(14): 1-15.
- [12] NIKOLAOS P, ELPINIKI P, ATHANASIOS A, et al. Bone metastasis classification using whole body images from prostate cancer patients based on convolutional neural networks application[J]. *Plos One*, 2020, 15(8): 1-29.
- [13] CHARIS N, DIMITRIOS E, NIKOLAOS P, et al. A light-weight convolutional neural network architecture applied for bone metastasis classification in nuclear medicine: A case study on prostate cancer patients[J]. *Healthcare*, 2020, 8(4): 493.
- [14] ZHENG Z, LIU L X, CHEN X Y, et al. Construction of bi-section model of SPECT bone scan image based on VG-GNet[C]//IEEE International Conference on Artificial Intelligence and Industrial Design (AIID), May 28-30, 2021, Guangzhou, China. New York: IEEE, 2021: 150-154.
- [15] 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.
- [16] REN X H, AHMAD S, ZHANG L C, et al. Task decomposition and synchronization for semantic biomedical image segmentation[J]. *IEEE Transactions on Image Processing*, 2020, 29: 7497-7510.
- [17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 13708-13717.
- [18] SZEGEDY C, IOFFE S, VANHOUCKE V. Inception-v4, Inception-ResNet and the impact of residual connections on learning[C]//Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. Menlo Park: AAAI, 2017, 12: 4278-4284.
- [19] 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.
- [20] FU H, FU B H, SHI P L, et al. An improved segmentation method for automatic mapping of cone karst from remote sensing data based on DeepLab V3+ Model[J]. *Remote Sensing*, 2021, 13(3): 1-16.
- [21] CAO Y, LIU S G, PENG Y L, et al. DenseUNet: densely connected UNet for electron microscopy image segmentation[J]. *IET Image Processing*, 2019, 14(12): 2682-2689.

作者简介:

余泓 (1997—), 男, 硕士, 从事医学影像识别、深度学习方向的研究。