

文章编号 1004-924X(2023)06-0936-14

基于ACGAN和迁移学习的骨显像分类方法

余泓^{1*}, 罗仁泽¹, 陈春梦², 唐祥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 计算机科学学院, 四川 成都 610500)

摘要: 由于骨显像存在样本数量有限、类别不平衡的问题, 导致骨显像分类存在较大困难。为提升骨显像的分类准确率, 本文提出了一种基于结合辅助分类器的生成对抗网络(ACGAN)数据生成和迁移学习的骨显像分类方法。首先, 为解决骨显像类别不平衡的问题, 设计了一种MU-ACGAN模型。该模型以U-Net为生成器框架, 同时结合密集残差连接和通道-空间注意力机制结构来提升骨显像细节特征生成, 判别器通过密集残差注意力卷积块提取骨显像特征进行判别; 然后, 结合传统数据增强方式进一步扩充数据量; 最后, 设计了一种多尺度卷积神经网络提取骨显像不同尺度的特征, 提升分类效果。在模型训练过程中, 采用两阶段迁移学习方式, 优化模型的初始化参数、解决过拟合的问题。实验结果表明, 本文提出方法分类准确率达到85.71%, 有效缓解了小样本骨显像数据集分类准确率不高的问题。

关键词: 骨显像; 结合辅助分类器的生成对抗网络(ACGAN); 迁移学习; 注意力机制; 数据增强

中图分类号: TP391 **文献标识码:** A **doi:** 10.37188/OPE.20233106.0936

Bone scintigraphic classification method based on ACGAN and transfer learning

YU Hong^{1*}, LUO Renze¹, CHEN Chunmeng², TANG Xiang³, LUO Renquan¹

(1. College of Electrical Engineering and Information, Southwest Petroleum University, Chengdu 610500, China;

2. Department of Nuclear Medicine, The No. 2 People's Hospital of Yibin, Yibin 644000, China;

3. College of Computer Science, Southwest Petroleum University, Chengdu 610500, China)

* Corresponding author, E-mail: 790622472@qq.com

Abstract: Owing to the limited availability of samples and unbalanced categories of bone images, it is difficult to classify these images. To improve the classification accuracy of bone images, this study developed a bone-image classification method based on auxiliary classifier generative adversarial network (ACGAN) data generation and transfer learning. First, an multi-attention U-Net-based ACGAN (MU-ACGAN) model was designed to address the imbalance of bone-image categories. The model uses U-Net as the generator framework and combines dense residual connection and channel-spatial attention mechanism to improve the generation of bone-image detail features. The discriminator extracts bone-image features by using a dense residual attention convolution block for discrimination. Next, the amount of data was further

收稿日期: 2022-07-19; 修订日期: 2022-08-22.

基金项目: 四川省科技计划资助项目(No. 2019CXRC0027)

expanded via combination with traditional data enhancement methods. Finally, a multi-scale convolutional neural network was designed to extract the features at different scales of bone imaging so as to improve the classification effect. In the model training process, a two-stage transfer learning method was adopted to optimize the initialization parameters of the model and address the problem of overfitting. Experimental results indicate that the classification accuracy of the proposed method reaches 85.71%, effectively alleviating the problem of low classification accuracy on small sample bone-image datasets.

Key words: bone scintigraphy; Auxiliary Classifier Generative Adversarial Networks(ACGAN); transfer learning; attention mechanism; data augmentation

1 引言

在全球范围内,骨转移肿瘤已经成为引起癌症患者死亡的重要原因。核医学中诊断骨转移肿瘤的手段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]和单光子发射计算机断层成像(Single-Photon Emission Computed Tomography, SPECT)^[3]。其中,SPECT骨显像具有成本低、灵敏度高、成像视野广等优势,成为判断骨转移肿瘤最为重要的检查手段。目前,医生主要依靠阅片经验和其他辅助信息(如文本信息)来判断骨显像病灶的类型,存在一定的主观性、重复性以及分析不足的问题。

为了能辅助医师进行快速、客观地诊断,近年来许多研究者开始将机器学习算法和数字图像处理技术应用至骨显像中来实现自动识别。May等人^[4]借助人工神经网络(Artificial Neural Network, ANN)对骨显像病灶进行分类,并应用至计算机辅助诊断系统(Computer Aided Diagnosis, CAD)。Aslantas等人^[5]提出一种CAD诊断系统——“CADBOSS”,该系统使用图像网格化方法提取病灶区域特征,然后借助ANN分类器对特征进行分类,判断是否存在骨转移。Mihaela等人^[6]结合边缘分割算法和lambda-schedule算法提取骨显像病灶的纹理及空间特征,并使用k-最近邻算法(k-Nearest Neighbor, KNN)和支持向量机(Support Vector Machine, SVM)对病灶类别进行判断。Elfarra等人^[7]运用平行六面体分类方法来生成放射性核素分布映射图,以此对骨显像进行分类。Wang等人^[8]使用SPSS和SPSS Modeler的统计软件包对骨显像进行数据分析和模型开发,比较了逻辑回归(Logistic Re-

gression, LR)、决策树(Decision Tree, DT)以及支持向量机(SVM)三种机器学习算法的分类效果。总之,传统数字图像处理技术虽能实现自动识别,但精度不高;传统机器学习算法分类准确率有所提升,但是计算效率、特异性较低。

近年来,深度学习在医学图像分析领域逐渐开始被研究与应用。其中,部分学者开始将卷积神经网络(Convolutional Neural Networks, CNN)^[9]应用至骨显像分类任务中。Zheng等^[10]通过旋转、镜像、平移的方式对小样本骨显像数据集进行数据增强,并对比了不同深度的VGG模型分类效果,准确率、召回率、F1得分均达到了90%以上。Nikolaos等人^[11]对CNN超参数进行微调,将模型性能调至最佳,然后采用VGG16; ResNet50; GoogleNet和MobileNet模型对前列腺癌患者的骨显像进行良、恶性判断。Charis等人^[12]为减轻网络计算量,以轻量级全卷积神经网络(LB-FCN)为基础框架,同时减少浮点运算来进一步减轻运算负担,相比于ResNet、VGG16、MobileNet,该框架在骨显像分类任务中取得了更出色的效果。Lin等人^[13]提出一种自定义CNN模型来识别骨显像种类(健康、骨转移或关节炎),并使用了传统数据增强和DCGAN进行数据增强,识别准确率达到70%以上。可以看出,目前的主要研究的针对的是单一疾病的分类方法。当对骨显像进行多分类(如健康、骨转移肿瘤、良性改变三类)时,存在准确率严重下降的问题。

由于SPECT骨显像数据稀缺、标注费时,难以建立大型的数据库。因此,自制的骨显像数据集存在类别不平衡、数据量有限的问题。数据增强^[14]和迁移学习^[15]两种技术能够极大程度缓解小样本数据集带来的影响。数据增强技术通过

图像几何变换、图像增强或生成对抗网络(Generative Adversarial Networks, GAN)^[16]等数据生成手段来扩充样本数量。迁移学习技术通过将其他样本中的信息特征迁移至目标数据集中,以此缓解小样本数据集中信息量有限的问题。

本文充分考虑小样本骨显像数据集的特点,提出一种结合辅助分类器的生成对抗网络(Auxiliary Classifier Generative Adversarial Networks, ACGAN)^[17]和迁移学习的分类方法。首先,本文设计了一种 MU-ACGAN 对骨显像数据集进行过采样,解决骨显像不平衡的问题。其中, MU-ACGAN 的生成器以 U-Net 模型为框架,同时设计了密集残差连接和双输入的残差注意力门控制结构来提升骨显像的细节生成效果;判别器采用密集残差注意力卷积块进行特征提取。然后,同时结合传统数据增强方式扩充数据集量。最后,采用多尺度 CNN 模型进行分类识别。在模型训练过程中,采取两阶段迁移学习策略对模型进行训练,防止出现过拟合现象。在自建骨显像数据集上的实验结果表明,本文提出的方法能够显著提升分类效果,准确率达到了 85.71%。

2 相关原理

2.1 ACGAN 模型

GAN 是一种基于对抗学习的深度学习模型。GAN 通过向生成器输入随机噪声 z 来生成新数据 $G(z)$, 让判别器 D 判断 $G(z)$ 的真假, 并将结果重新反馈给生成器 G 与判别器 D 。在 G 与 D 相互对抗、博弈的过程中, 两个模型会不断优化自身参数, 引导 G 捕获真实样本 x 的分布规律, 生成能够以假乱真的新样本 $G(z)$, 最终达到平衡状态。整个过程可以表示为:

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}(x)}} [\log D(x)] + E_{z \sim P_{\text{data}(z)}} [\log (1 - D(G(z)))] \quad (1)$$

其中: z 表示输入噪声; x 表示真实样本; $G(z)$, $D(x)$ 分别表示生成器通过噪声 z 生成的样本和判别器判断其为真实样本的概率; $D(G(z))$ 表示判别器判别生成样本 $G(z)$ 为真实样本的概率。当 $D(G(z))$ 越接近 1 时, $V(D, G)$ 越小, 表明生成样本 $G(z)$ 越接近真实样本;

传统 GAN 模型只能生成某一类型的数据, 而 SPECT 骨显像分类问题包含健康、骨转移肿瘤、良性改变三种类别。判别器不仅需要判别输入骨显像的真实性, 还需要对其类别进行判断。因此, 本文采用 ACGAN 模型为基本框架, 其原理如图 1 所示。ACGAN 通过在生成器的输入端加入与样本相对应的类别信息来指导模型训练, 同时在判别器中新增了一个全连接层预测生成图像的类别。ACGAN 的目标函数如式(2)和式(3)所示:

$$L_S = E [\log P(S = \text{real} | X_{\text{real}})] + E [\log P(S = \text{fake} | X_{\text{fake}})] \quad (2)$$

$$L_C = E [\log P(C = c | X_{\text{real}})] + E [\log P(C = c | X_{\text{fake}})] \quad (3)$$

其中: L_S 表示判别真假的损失; L_C 表示判别类别的损失; X_{fake} 表示生成的骨显像; X_{real} 表示训练集中的真实骨显像。

$$L_G = L_S - L_C \quad (4)$$

$$L_D = L_C + L_S \quad (5)$$

在训练阶段, ACGAN 通过最大化 $L_S + L_C$ 来训练判别器 D 和最小化 $L_C - L_S$ 来训练生成器 G 。

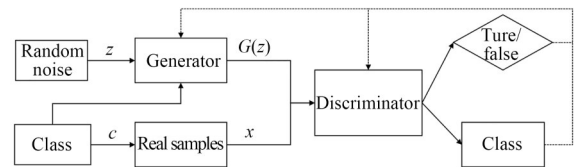


图 1 ACGAN 原理流程

Fig. 1 ACGAN principle flow

2.2 迁移学习

迁移学习是一种将源域的已有知识迁移至新目标域中的方法, 这种方法能够降低训练的成本、克服数据稀缺和训练耗时的问题。迁移学习可以定义为: 给定源域 D_s 、学习任务 T_s 以及相应目标域 D_t 和 T_t , 通过 D_s 和 T_s 已有的学习知识来提升 D_t 中的目标函数 $f(\bullet)$ 的效果。一种常用的迁移学习方法为参数迁移, 参数迁移分为两个阶段。在第一阶段中, 将模型放置源域进行预训练。在第二阶段中, 将从源域中学习到的模型参数迁移至新的数据集中, 通过小幅度调整让模型能够适用于新的数据集和任务。

3 本文方法

3.1 总体流程

本文提出方法的整体流程见图 2,主要包含 4 个部分:(1)GAN 过采样:设计了一种 MU-ACGAN 来生成高质量、不同类别的骨显像,均衡各类骨显像的比例。(2)传统数据增强:通过旋转、平移、镜像、加噪的方式来进一步扩充数据量。(3)多尺度 CNN 分类器:采用多尺度 CNN 模型提取骨显像不同尺度的特征,提升骨显像的分类效果。(4)迁移学习:使用两阶段迁移学习策略优化 MU-ACGAN 和多尺度 CNN 模型的训练过程。

3.2 改进 ACGAN 模型

ACGAN 模型虽然能够生成不同类别的骨显像,但其网络结构中只采用了单一的卷积与上采样方式,容易造成生成骨显像的细节特征粗糙、模糊。为此,本文设计了一种结合密集残差模块和残差注意力门控结构的 MU-ACGAN 模型。

3.2.1 MU-ACGAN

如图 3(a)所示为 MU-ACGAN 的生成器。为更好地重建出骨显像细节特征,生成器采用 4

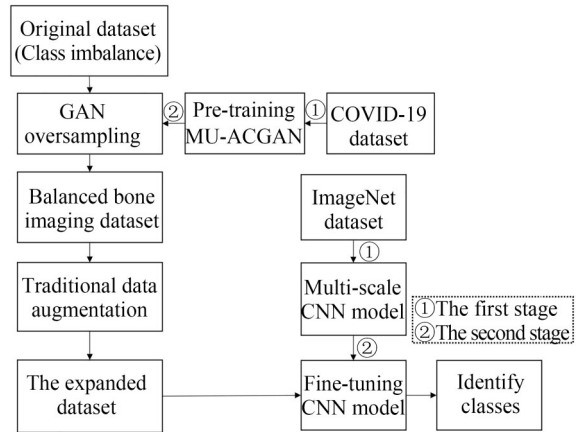


图 2 本文方法整体流程图

Fig. 2 Overall flow chart of the method in this paper

层 U-Net^[18]模型为基本框架,每层都采用密集残差注意力(Residual Dense Attention, RDA)卷积块进行特征提取。在输入端,首先将骨显像类别信息嵌入批次中,随后与噪声进行特征拼接,生成一个新的特征并重塑成 4 种不同尺度的特征,分别作为 U-Net 编码端每层的输入。在编码部分,使用 RDA 卷积块、最大池化层对特征图进行特征提取和降采样,逐步提取到图像的高层特征。同时,为了防止降采样时会丢失掉骨显像的

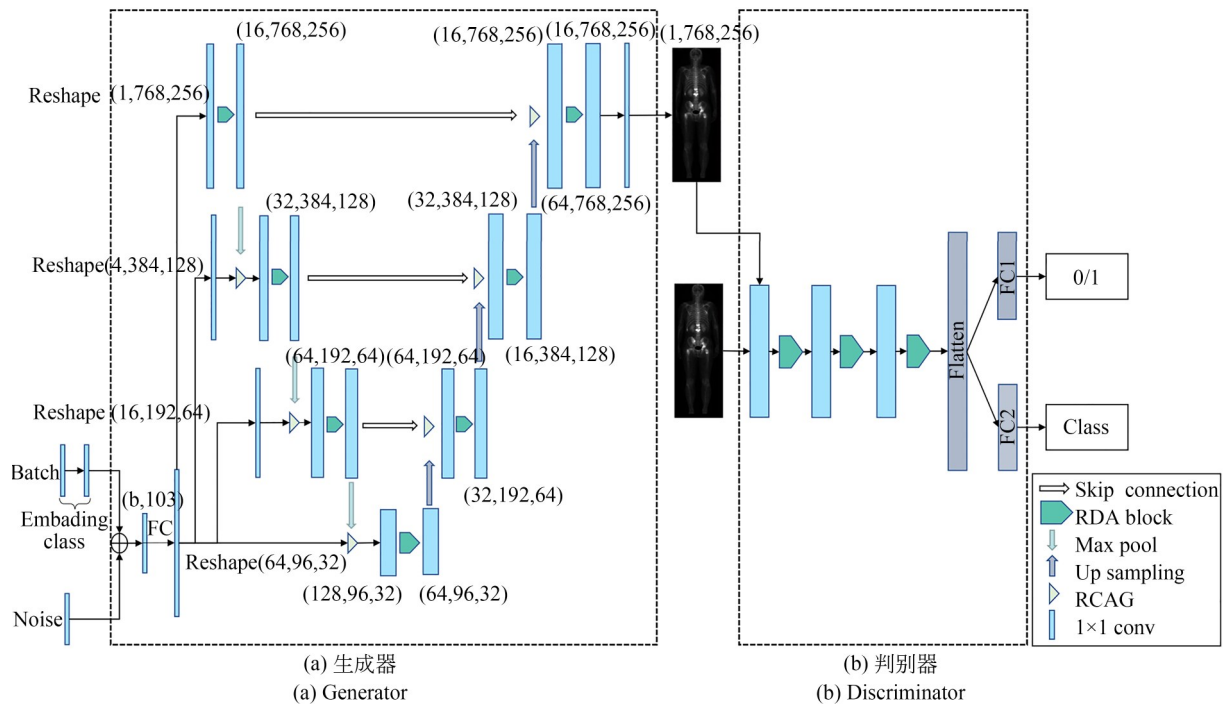


图 3 改进网络模型

Fig. 3 Improved network models

细节信息,将下采样后的特征和输入端重塑后的特征一起送入残差注意力门控结构(Residual Coordinate attention Gate, RCAG)。在解码部分,使用RDA卷积块和线性插值(避免棋盘效应)的方式对特征图进行特征提取和上采样,逐层将特征图恢复至原图大小。此外,在跳跃连接处采用RCAG注意力门控结构来指导高层全局特征与低层细节信息进行高效融合。最后,使用一个 1×1 卷积输出重建后的骨显像。

判别器模型如图3(b)所示。首先,向模型输入生成骨显像或真实骨显像。然后,通过3个RDA卷积块来充分提取骨显像的特征信息。接着,将提取到的特征矩阵进行展平,通过全连接层1(Fully Connected Layer1, FC1)输出其真假信息、全连接层2(FC2)输出骨显像的生成/真实骨显像的类别信息。

3.2.2 双输入残差注意力门控结构(RCAG)

协调注意力机制(Coordinate Attention, CA)^[19]模块能够学习模型“感兴趣”的特征。与其他注意力结构不同的是,该模块通过降维的方式将特征图中的详细位置信息嵌入通道注意力中,在学习特征图空间位置信息与通道关系的同时,不会造成额外的运算负担,见图4。

在CA模块的基础上,本文设计了一种双输入残差注意力门控结构对骨显像的高层全局特征与低层细节信息进行有效融合,见图5。首先,将高层特征 x_1 和低层特征 x_2 进行拼接得到一个

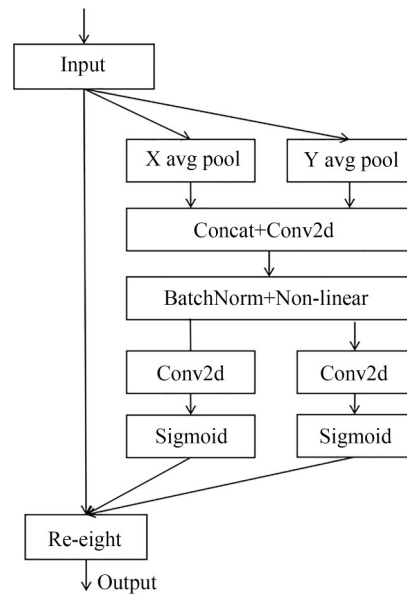


图4 CA模块

Fig. 4 CA modules

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,将得到的权重矩阵后乘以 x_c 得到更新后的特征,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、批量归一化(Batch Normalization, BN)和PReLU激活函数输出模块。具体原理如式(6):

$$z_c = \text{concatenate}(x_1 + x_2), \quad (6)$$

其中: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;concatenate表示进行特征拼接; z_c 表示拼接后的特征。

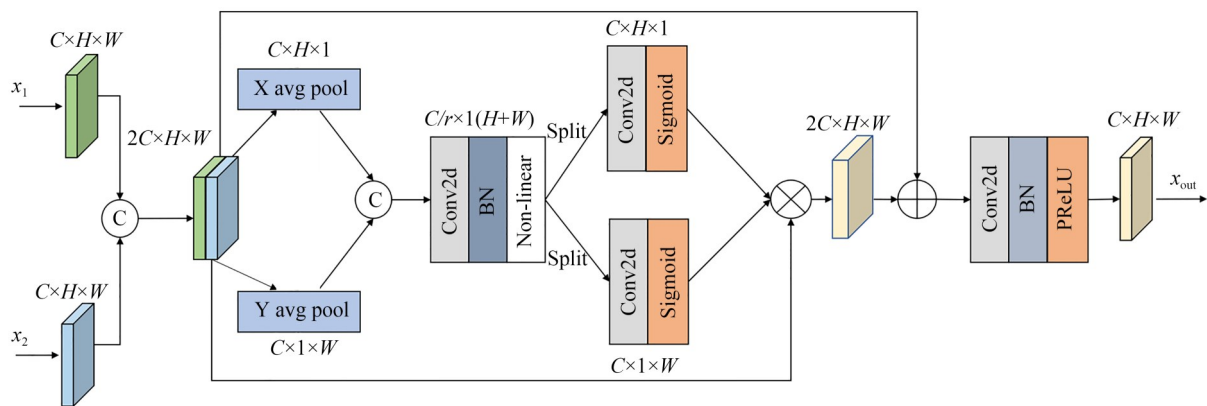


图5 残差门控注意力结构

Fig. 5 Residual gated attention architecture

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq w} x_c(h, i), \quad (7)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(j, w), \quad (8)$$

其中: $z_c^h(h)$ 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。

$$f = \delta(F1([z_c^h, z_c^w])), \quad (9)$$

其中: $F1([z_c^h, z_c^w])$ 表示将 $z_c^h(h)$ 和 $z_c^w(w)$ 进行特征拼接,并进行 1×1 卷积运算; δ 表示非线性变换,即BN和ReLU激活函数的组合。

$$g_c^h = \sigma(F_h(f^h)), \quad (10)$$

$$g_c^w = \sigma(F_w(f^w)). \quad (11)$$

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 映射(Sigmoid激活函数)。

$$x'_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (12)$$

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。

$$x_{out} = \delta(Conv_{1 \times 1}(x_c + x'_c)), \quad (13)$$

其中: $x_c + x'_c$ 表示将 x_c 和更新后的 x'_c 进行特征

融合; $Conv_{1 \times 1}$ 表示 1×1 卷积运算; x_{out} 表示整个模块的输出特征。

3.2.3 密集残差注意力模块(RDA)

为了提升卷积的利用率,本文设计了一种结合CA模块的注意力密集残差卷积块代替原模型中的 3×3 卷积,见图6。整个模块由 3×3 卷积、 1×1 卷积、PReLU激活函数、批量归一化以及CA模块构成。首先,该模块通过3个 3×3 卷积提取统一尺度下的不同的骨显像特征,每次卷积后都运用了BN批量归一化和PReLU激活函数来增强模型的非线性表达能力。其次,模块通过密集连接方式来汇聚统一尺度下的不同特征信息,提升卷积层的利用效率、并解决网络因过深而产生的梯度消失问题。同时,在每次特征汇聚后都采用了一个 1×1 卷积来压缩通道数,减轻网络运算量。第 i 层卷积的输入如式(14)所示:

$$x_i = H_i([x_0, x_1, \dots, x_{i-1}]), \quad (14)$$

式中: x_i 表示第 i 层的输入, $[x_0, x_1, \dots, x_{i-1}]$ 表示第 i 层之前的所有特征图拼接后的特征。 H_i 表示非线性映射,即BN和PReLU激活函数运算的组合。

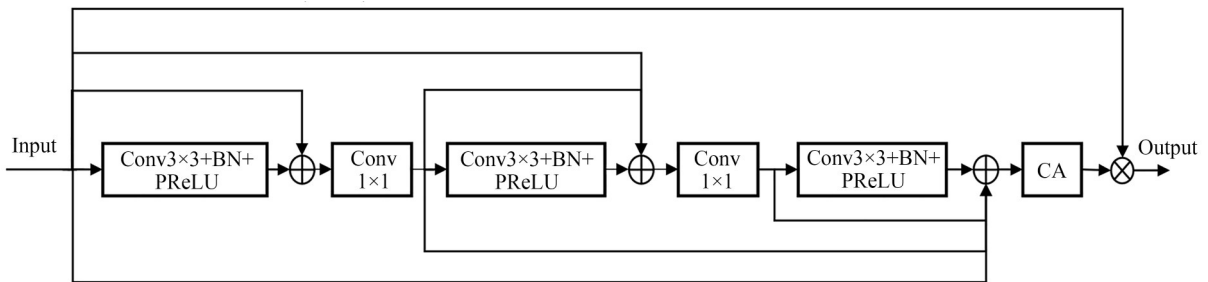


图6 密集残差注意力模块

Fig. 6 Dense residual attention module

为有效利用密集连接结构所提取到的特征,在第3次特征拼接后,使用了一个CA模块来抑制提取到的无用特征,让模型重点关注细节特征的生成。最后,采用残差结构将模块提取特征与输入特征进行融合。

3.3 多尺度CNN分类模型

为了准确地对骨显像进行分类,本文设计了一个多尺度CNN分类模型,见图7。模型采用在ImageNet上已经预训练好的DenseNet121,

VGG16, ResNext50 3个模型同时提取骨显像的特征。其中,DenseNet121模型采用密集连接方式,解决了梯度消失的问题,随着网络层数加深,能够提取到骨显像的更深层的特征。VGG16网络层数较浅,可以提取到骨显像的浅层特征。ResNeXt50采用了多尺度和残差的连接方式,能够提取不同尺度的骨显像特征。将3个模型(去除全连接层)提取到的特征通过全局平均池化降采样至相同大小后,通过特征拼接并展平输入全

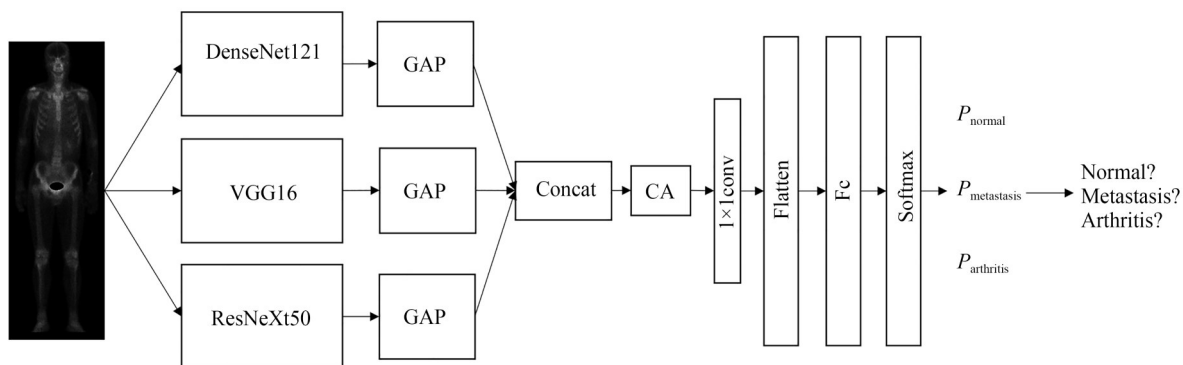


图7 CNN分类模型

Fig. 7 CNN classification model

连接层。最后,经过 softmax 层输出每类概率值,预测骨显像类别。

3.4 迁移学习

随着网络的层数加深、结构更为复杂时,模型的参数量也随之增加。由于骨显像的数据量较少,直接使用 MU-ACGAN 和多尺度 CNN 模型训练容易出现过拟合现象。为此,本文采用迁移学习方法现在大型数据集上进行预训练,将训练好的模型再迁移至骨显像数据集进行分类训练,以此来解决骨显像数据量较少的问题。在第一阶段迁移中,分别采用 COVID19 数据集和 ImageNet 数据集对 MU-ACGAN、多尺度 CNN 进行预训练。其中,COVID19 数据集中有 20 000 张肺部图像,包含新型冠状病毒肺炎、肺混浊、正常、病毒性肺炎四类。在第二阶段迁移时,将训练好的 MU-ACGAN 模型在迁移至骨显像数据集进行微调,生成新的骨显像样本。然后将 MU-ACGAN 生成的骨显像和传统数据增强生成的骨显像一起送入已经预训练好的多尺度 CNN 模型,微调模型参数,最后预测骨显像类别。

4 预处理及模型训练

4.1 数据集介绍

数据集一共包含 178 例 SPECT 骨显像,收集于四川省宜宾市第二人民医院核医学科。每例骨显像通过 DICOM 文件进行储存。为将其转化为能送入模型的图像,首先通过归一化的方式转化为灰度图,每例骨显像包含正面、背部两张图片。然后,通过自动裁剪方式得到头部、至脚

部区域。最后,筛选除去受污染严重的图片,一共得到 346 张骨显像作为本次实验的数据样本。骨显像数据集的训练、测试的数量及比例如表 1 所示。

表 1 骨显像训练、测试的数量与比例

Tab. 1 Number and ratio of training and testing of bone imaging

Index	Health	Metastases	Benign changes
	Train/Test	Train/Test	Train/Test
Quantity	130/32	101/25	43/15
Total	162	126	58
Ratio	46.82%	36.42%	16.76%

为防止过拟合、增加数据量,本文采用旋转、平移、镜像、泊松加噪的传统数据增强对训练集骨显像进行了扩充。其中,患者拍片姿势可能不标准(如斜躺、未处于检查设备中央)会带来一定差异,故采用旋转、平移方式进行扩充;每例骨显像包含正面、背部两张图片,两张具有一定联系,故采用镜像方式进行扩充;进行检查时医师常通过减少管电流、电压来减少核辐射,会为骨显像带来泊松噪声,故采用泊松加噪方式进行扩充,消除噪声干扰。扩充后的骨显像数据量见表 2。

表 2 扩充后骨显像数据量

Tab. 2 Amount of bone imaging data after the expansion

Class	Health	Metastases	Benign changes
Quantity	1 950	1 515	645

4.2 评价指标及参数设置

本文采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、特异性(Specificity)、F1分数(F1 Score)作为分类模型的评价指标,见式(15)~式(19)。其中,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。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (15)$$

$$Precision = \frac{TP}{TP + FP}, \quad (16)$$

$$Recall = \frac{TP}{TP + FN}, \quad (17)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (18)$$

$$F-1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (19)$$

本次实验中,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^[20]优化器。在MU-ACGAN生成实验中,学习率设为0.0002, batch size为2,迭代次数为500。在多尺度CNN分类实验中batch size设为4,迭代次数为500,损失函数采用交叉熵损失。

5 实验结果及与分析

5.1 MU-ACGAN生成效果

5.1.1 迁移策略效果

为验证迁移学习对模型训练的作用,本文将直接训练与采用COVID19数据集进行预训练后的MU-ACGAN生成效果进行了对比,模型的训练损失、准确率变化曲线如图8所示。可以看出,在直接训练模型的情况下,生成器损失值波动较大,难以得到稳定;判别器损失值在前400轮下降缓慢,在第400轮后逐渐稳定;判别器准确率上升同样缓慢,在接近第500个epoch时才逐渐稳定在80%左右。采用迁移学习策略后,生成器损失在第150轮后基本稳定;判别器损失在接近200轮时便得到收敛,速度提升了近50%,并且收敛后的损失值更小;判别器训练准确率在第150轮时便达到了90%以上,速度提升了近70%,准确率提升了近10%。由此可知,迁移学习对于加速MU-ACGAN模型损失收敛、提高训练准确率具有明显的促进作用。模型生成骨显像的对比效果如图9。未使用迁移策略时,骨显像的细节纹理特征生成效果较差,在头部、肋骨及肩胛骨处伴随有伪影生成;采用迁移学习策略后,骨显像细节特征的生成更加清晰、更接近真实图像,并且伪影消失。

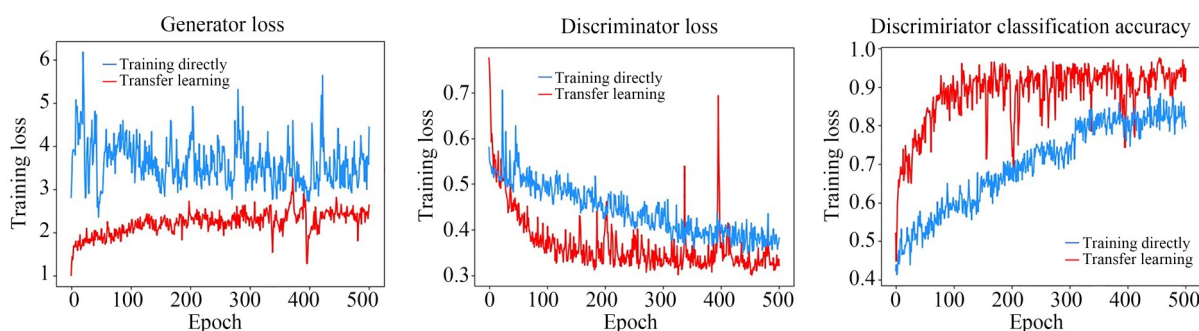


图8 MU-ACGAN训练过程对比

Fig. 8 Comparison of MU-ACGAN training process

5.1.2 主客观对比分析

为展示MU-ACGAN模型的生成效果,选取了第1轮、100轮、200轮、500轮4个轮次的骨显像生成图进行对比分析。由图10可见:模型的输

入为不同的随机噪声;在迭代次数为1次时,生成的图像中还残留着部分COVID19数据集中的CT肺部特征;在迭代次数为100次时,模型能生成骨显像的基本轮廓和病灶特征,但细节特征刻

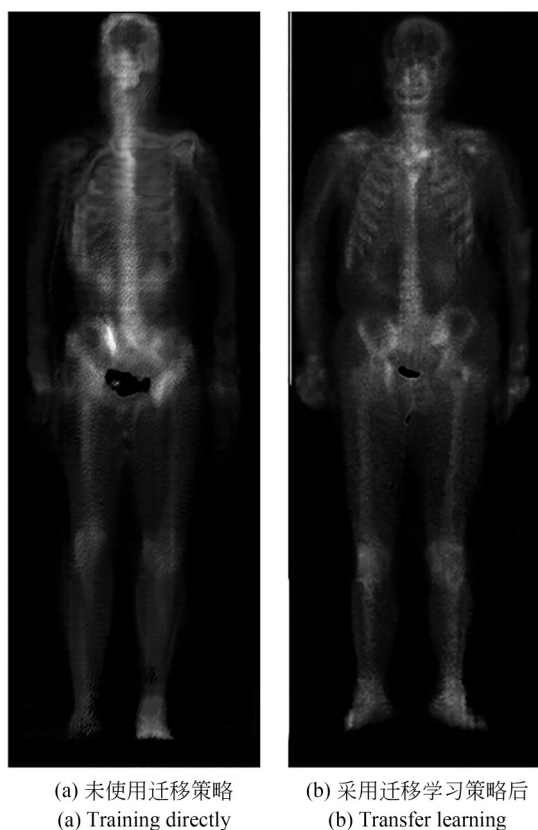


图 9 迁移学习效果对比

Fig. 9 Comparison of transfer learning

画非常模糊,出现了伪影现象;在迭代次数为 200 时,生成骨显像的伪影和模糊区域面积减少;在迭代次数为 500 时,生成骨显像的伪影、模糊区域

已基本消失,并且肩胛骨、肋骨、髌骨等关节细节特征的清晰度较好。

为了进一步展示 MU-ACGAN 的生成效果,将 MU-ACGAN 和 ACGAN 的生成骨显像与真实骨显像进行了对比。如图 11 所示,从上往下依次为 ACGAN, MU-ACGAN, 真实骨显像,从左至右为正常、骨转移、良性改变骨显像,每组包含正面与背部两张图片。可以看出,ACGAN 由于采用单一的上采样和卷积,虽然能够生成大致骨显像的轮廓,但是在肋骨、肩胛骨以及病灶区域等细节特征生成上并不理想。而骨显像的类别判定往往需要根据病灶的区域大小和病灶分布的对称性等信息来进行判定。因此,ACGAN 生成的骨显像可能为后续的分类任务带来一定影响。本文设计的 MU-ACGAN 通过特征融合的方式结合了网络提取到的高层全局信息与低层细节特征,并且采用注意力门控结构来指导二者进行高效融合,抑制无用信息,注重细节特征的提取。相比于 ACGAN, MU-ACGAN 的生成骨显像在肋骨、髌骨、肩胛骨、头骨以及病灶区域的生成有了极大提升,图像质量比较接近真实骨显像。此外, MU-ACGAN 更加关注胸腔特征(如肋骨、肩胛骨、肾脏和病灶)的生成,而真实骨显像的病灶主要集中在胸腔中,病理特征与真实骨显像更为相近。

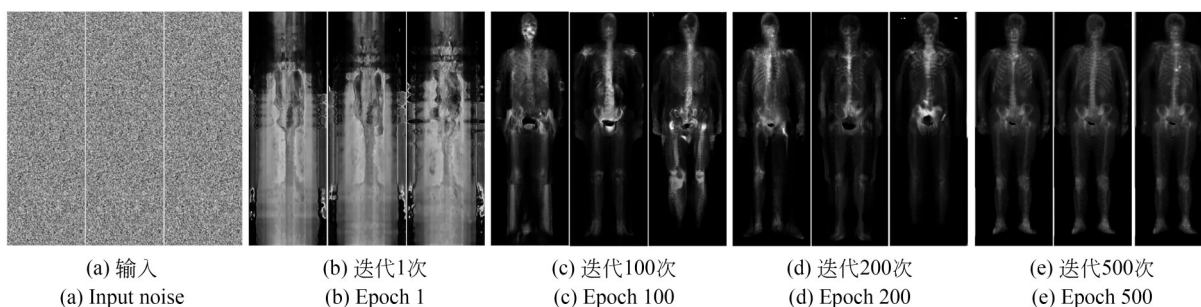


图 10 MU-ACGAN 生成的骨显像

Fig. 10 Bone images generated by MU-ACGAN

为客观评价生成骨显像的质量,实验中使用结构相似度 (Structural Similarity, SSIM) 指数评估 ACGAN 和 MU-ACGAN 生成效果。SSIM 表现了两幅图在亮度、对比度、结构三方面的相似程度,其数值越高(0~1)表明两幅图越相似。首

先,将原图与生成图像缩放至相同尺寸,再将训练集中所有骨显像的对比度、亮度调至相近。然后,用每张生成骨显像与训练集中的骨显像依次计算 SSIM 值,在求和并取平均值后,结果见表 3,表中对比了 ACGAN, MU-ACGAN 不同类别

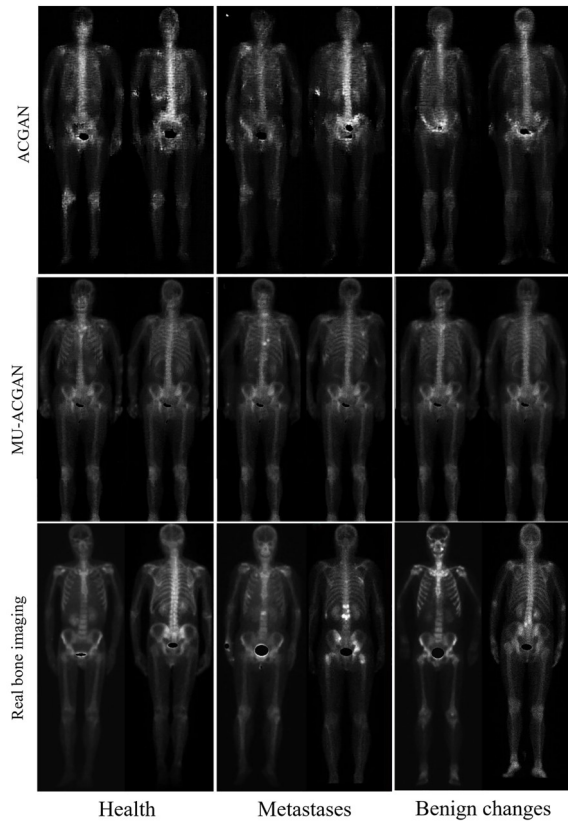


图 11 生成骨显像和真实骨显像

Fig. 11 Generative bone imaging and real bone imaging

生成图的 SSIM 值。可以看出,ACGAN 和 MU-ACGAN 生成骨显像 SSIM 均值在 0.45~0.55 之间。为了生成更丰富的骨显像数据,骨显像的部分结构以及细节纹理特征会发生部分改变。因

此,生成图像与训练集中图像的 SSIM 值即不能太高也不能过低。此外,相比于 ACGAN 模型, MU-ACGAN 生成的健康、骨转移与良性变化骨显像的 SSIM 分别提升了 0.080 5 (17.6%), 0.041 2 (8.1%), 0.041 7 (8.2%), 与训练集中的骨显像结构更为相似。

表 3 骨显像质量评价

Tab. 3 Quality assessment of bone imaging

Image type	Health	Metastases	Benign changes
ACGAN	0.458 5	0.516 0	0.506 8
MU-ACGAN	0.539 0	0.557 2	0.548 5

此外,从 ACGAN 和 MU-ACGAN 生成的各类骨显像中随机选取了 1 张骨显像与测试集骨显像进行相似度对比,其变化曲线如图 12 所示。可以看出, MU-ACGAN 的图片质量优于 ACGAN。首先, ACGAN 和 MU-ACGAN 的 SSIM 变化趋势基本相同,而 MU-ACGAN 的 SSIM 数值略高,与原骨显像的结构相似度更近。其次,测试集真实骨显像中的恶性骨转移、良性变化的 SSIM 变化趋势与 ACGAN, MU-ACGAN 基本吻合,并且与 MU-ACGAN 的变化更加接近。

总体来说, MU-ACGAN 生成的骨显像更接近真实骨显像,能够为训练集的数据增强提供高质量的生成样本。

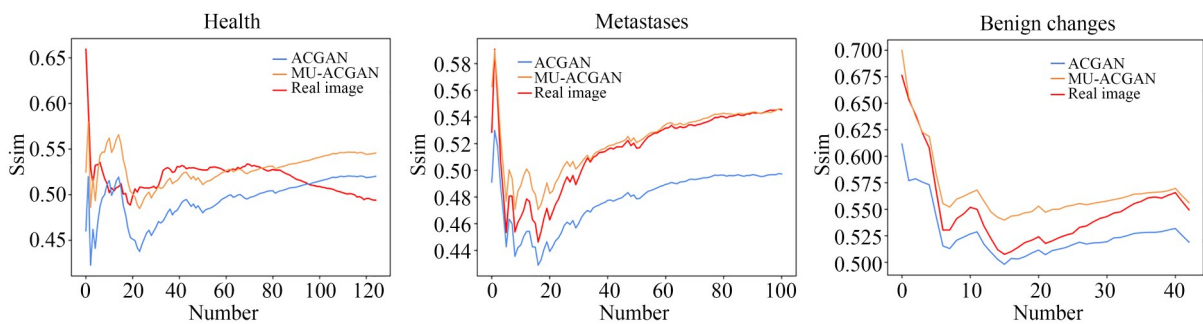


图 12 相似度对比

Fig. 12 Similarity comparison

5.1.3 数据增强效果对比

为了探究 MU-ACGAN 数据增强对骨显像分类结果的影响,本文设置了 4 个对照组。其中:数据集 1 为原始数据集;数据集 2 为 ACGAN 均

衡后的数据集;数据集 3 为 MU-ACGAN 均衡后的数据集。此外,加入传统数据增强方式:数据集 4 为传统数据增强方式扩充数据集;数据集 5 为 MU-ACGAN+传统扩充数据集。所有数据

集均采用 ResNet34 模型进行分类。

表 4 为对比了 ResNet34 在 5 种数据集上的分类表现。可以看出:数据集 1 的分类精度仅达到 0.701 3;数据集 2 和 3 采用 ACGAN 生成新骨显像均衡类别后,分类精度有所提升。并且在扩充数据量、分类模型相同的情况下,数据集 3 的各

项指标提升均略高于数据集 2;数据集 4 通过几何变换的方式扩充了数据量,消除了拍摄过程中带来的影响,分类准确率提升了 0.07。数据集 5 将 MU-ACGAN 和传统数据增强方式进行结合,准确率、召回率、特异性分别提升了 0.103 9, 0.063, 0.111 2。

表 4 不同数据增强方式的分类效果

Tab. 4 Classification effects of different data augmentation methods

Number	Accuracy	Precision	Recall	Specificity	F1-score
1	0.701 3	0.695 3	0.665 7	0.680 1	0.841 6
2	0.727 3	0.730 1	0.687 8	0.708 3	0.845 2
3	0.740 3	0.737 5	0.675 2	0.705 0	0.850 8
4	0.779 2	0.803 2	0.692 8	0.743 9	0.868 8
5	0.805 2	0.866 2	0.728 4	0.791 3	0.882 5

以上数据表明,采用 GAN 和传统几何变换进行数据增强的方法均能提升模型分类效果。其中, MU-ACGAN 生成的骨显像质量高于 ACGAN, 能够更好地提升 ResNet34 的分类精度。当同时采用 MU-ACGAN 和传统几何变化进行增强时,模型分类效果最佳。

5.2 分类效果对比

5.2.1 迁移策略效果

本节实验使用未进行预训练的 ResNet34 模型分别对未扩充、扩充后的骨显像数据集进行了测试。从表 5 可以看出在直接训练 ResNet34 模型的情况下,模型没有在大型数据集上学习到丰富的样本特征,加上数据量有限,骨显像分类准确率和召回率分类下降了 0.207 8, 0.200 8。在使用数据集 5 进行扩充的情况下,准确率和召回率也下降了 0.064 9, 0.062 3。

5.2.2 多尺度 CNN 消融实验

为验证本文设计的多尺度 CNN 模型对骨显像分类结果的作用,设置了一组消融实验。表 6

表 5 迁移学习策略效果对比

Tab. 5 Comparison of transfer learning strategies

Type	Accuracy	Recall
Transfer learning	0.805 2	0.728 4
Dataset5	0.662 3	0.589 9
Training directly	0.597 4	0.527 6

对比了多尺度 CNN 中的单个分类模型在骨显像数据集(数据集 4)中的识别效果,黑色加粗部分表示最好的评价。可见,多尺度 CNN 模型在 Accuracy, Recall, F1-score 3 个指标上取得了最好的评价。相比于单独的 VGG16, DenseNet121, ResNeXt50 模型, accuracy 分别提升了 0.065, 0.026, 0.065。

表 6 多尺度 CNN 模型消融实验评估

Tab. 6 Experimental evaluation of multi-scale CNN model ablation

Models	Accuracy	Recall	F1-score
VGG16	0.766 2	0.666 3	0.743 6
DenseNet121	0.805 2	0.755 0	0.767 7
ResNext50	0.766 2	0.696 8	0.736 7
Multi-scale CNN	0.831 2	0.763 9	0.817 8

为了更好地理解上述 4 种模型的识别过程,采用 Grad-CAM(Gradient-weighted Class Activation Mapping)技术来生成测试集中骨显像的可视化热力图。红色表示模型对此区域的关注度越高(即权重越高),黄色次之,蓝色表示模型对该区域的特征不感兴趣。骨显像的病灶区域通常集中在肩胛骨、胸骨、肋骨、盆骨等位置,因此模型应更关注这部分区域。从图 13 可以看出,所有模型对背景的关注程度并不高,对病灶区域的

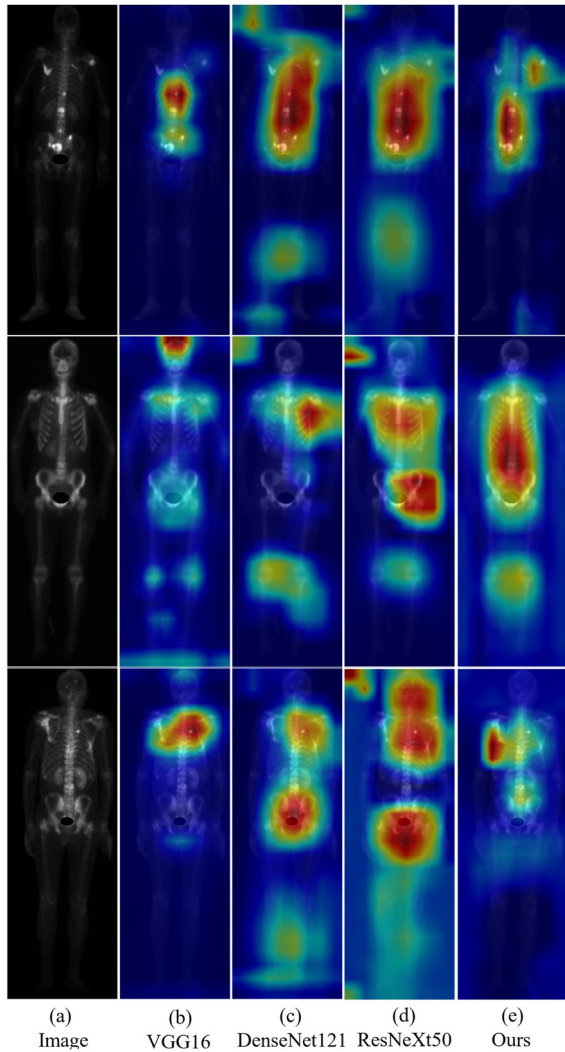


图 13 骨显像识别热力图

Fig. 13 Bone scintigraphy to identify heatmaps

关注程度为:多尺度 CNN \approx ResNeXt50>DenseNet121>VGG16。其中,ResNeXt50 和 DenseNet121 的红色区域主要集中在盆骨和胸腔中,有少许集中在膝盖(可能会出现关节炎)部分。VGG16 红色、黄色区域较少,红色集中在头部、胸腔,黄色出现在各个关节处,蓝色区域相较于其他 3 种模型更多,容易忽略对病灶区域的关注(彩图见期刊电子版)。多尺度 CNN 将 VGG16, DenseNet121, ResNext50 关注的区域进行了结合,通过 CA 注意力模块帮助模型关注病灶区域,热力图中的红色、黄色区域基本覆盖了各个病灶的位置,更有利于骨显像分类。

5. 2. 3 分类模型比较

为了进一步验证多尺度 CNN 模型的有效性,将其与近年来流行的图像分类网络 ResNet34, MobileNetV3^[21], InceptionV3^[22] 以及 EfficientNet^[23] 进行了对比。每种网络均在 ImageNet 上进行了预训练,使用 5. 1. 3 节中的数据集 4 进行模型参数微调,结果见表 7。

从表 7 可以看出,多尺度 CNN 的分类性能最优,相较于其他 4 种网络,分类准确率提高了 0. 026 (3. 23%) 以上。主要原因是多尺度 CNN 通过并行的 DenseNet121, VGG16, ResNext50 能够提取到骨显像深层、浅层、不同尺度的特征,提升了骨显像特征的利用率。

表 7 不同 CNN 模型的实验结果对比

Tab. 7 Comparison of the experimental results of different CNN models

Models	Accuracy	Precision	Recall	Specificity	F1-score
ResNet34	0. 779 2	0. 803 2	0. 692 8	0. 868 8	0. 743 9
MobileNetV3	0. 779 2	0. 785 7	0. 745 9	0. 875 6	0. 765 2
InceptionV3	0. 792 2	0. 847 7	0. 750 6	0. 874 2	0. 796 2
EfficientNet	0. 805 2	0. 868 4	0. 741 8	0. 880 6	0. 800 1
Ours	0. 831 2	0. 879 9	0. 763 9	0. 899 2	0. 817 8

将本文设计的分类模型应用至数据集 5 中,分类准确率、召回率、特异性如表 8 所示。可以看出,经过 MU-ACGAN 过采样后,模型的 Accuracy, Recall, Specificity 提升了 0. 025 9, 0. 040 1, 0. 012 8。图 14 为模型的混淆矩阵,颜色越深表

示概率越高。总体来说,健康骨显像的分类效果最好,分类准确率达到 90% 以上;其次为骨转移骨显像,错误的病例均划分为了健康图片,原因是部分骨转移病例病灶个数较少、面积较小,容易导致分类错误;良性改变容易误分为健康图片

或骨转移肿瘤,结合骨显像特点来看是因为部分图片其与健康图片差异不大,从而造成误分。部分良性改变图片包含了感染、关节炎、骨折等类型,可能出现高亮区域,与骨转移病灶特征相近。

表 8 不同 CNN 分类模型的评价指标

Tab. 8 Evaluation metrics of different CNN classification models

Dataset	Accuracy	Recall	Specificity
Dataset4	0.831 2	0.763 9	0.899 2
Dataset5	0.857 1	0.804 0	0.912 0

6 结 论

本文针对骨显像数据集存在的分类精度不高的问题,提出了一种基于 ACGAN 和迁移学习的骨显像分类方法。首先,设计了一种 MU-ACGAN 模型对骨显像进行过采样,解决类别不均衡的问题。然后,通过旋转、平移、镜像和加噪的

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: 4153211.
- [2] 张永平, 兰朋训, 周兆霞. MRI 和骨显像诊断骨转移瘤的临床价值分析 [J]. *医学影像学杂志*, 2017, 27(10): 2040-2042.
ZHANG Y P, LAN P X, ZHOU ZH X. Clinical value of MRI and bone scintigraphy in the diagnosis of bone metastases [J]. *Journal of Medical Imaging*, 2017, 27(10): 2040-2042. (in Chinese)
- [3] MADSEN M T. Recent advances in SPECT imaging [J]. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, 2007, 48(4): 661-673.
- [4] SADIK M, HAMADEH I, NORDBLOM P, et al. Computer-assisted interpretation of planar whole-body bone scans [J]. *J Nucl Med*, 2008, 49(12): 1958-1965.
- [5] ASLANTAS A, DANDIL E, SAĞLAM S, et al.

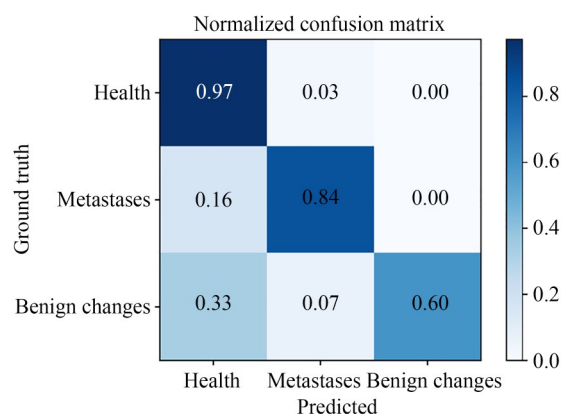


图 14 混淆矩阵

Fig. 14 Confusion matrix

传统数据增强方式进一步扩充骨显像数据量。最后,采用并行卷积神经网络提取骨显像不同尺度的特征,提升分类效果。实验中,采用迁移学习策略优化模型训练,解决小样本训练信息量有限的问题。结果表明,本文提出的方法分类准确率能够达到 85.71%,缓解了小样本骨显像分类精度不高的问题,能有效辅助医生进行诊断、提高诊断效率。

- CADBOSS: a computer-aided diagnosis system for whole-body bone scintigraphy scans [J]. *Journal of Cancer Research and Therapeutics*, 2016, 12(2): 787-792.
- [6] ANTONINA M, CALIN . Object-oriented classification approach for bone metastasis mapping from whole-body bone scintigraphy [J]. *Physica Medica*, 2021, 84: 141-148.
- [7] ELFARRA F G, CALIN M A, PARASCA S V. Computer-aided detection of bone metastasis in bone scintigraphy images using parallelepiped classification method [J]. *Annals of Nuclear Medicine*, 2019, 33(11): 866-874.
- [8] WANG Z Y, WEN X T, LU Y H, et al. Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases [J]. *Oncotarget*, 2016, 7(11): 12612-12622.
- [9] LU L M, ZHANG C L, CAO K, et al. A multi-channel CNN-GRU model for human activity recognition [J]. *IEEE Access*, 2022, 10: 66797-66810.
- [10] ZHENG Z W, LIU L X, CHEN X Y, et al. Construction of Bisection Model of SPECT Bone Scan

- Image Based on VGGNet[C]. 2021 *IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)*. May 28-30, 2021, Guangzhou, China. IEEE, 2021: 150-154.
- [11] PAPANDRANOS N, PAPAGEORGIOU E, ANAGNOSTIS 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) : e0237213.
- [12] NTAKOLIA C, DIAMANTIS D E, PAPANDRANOS N, *et al.* A lightweight 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.
- [13] LIN Q, CAO C G, LI T T, *et al.* dSPIC: a deep SPECT image classification network for automated multi-disease, multi-lesion diagnosis [J]. *BMC Medical Imaging*, 2021, 21(1): 122.
- [14] ABDOLLAHI B, TOMITA N, HASSANPOUR S. Data augmentation in training deep learning models for medical image analysis[J]. *Intelligent Systems Reference Library*, 2020: 167-180.
- [15] PADMAVATHI KORA. Transfer learning techniques for medical image analysis: a review [J]. *Biocybernetics and Biomedical Engineering*, 2022, 42(1): 79-107.
- [16] TEHSEEN ZIA. VANT-GAN: adversarial learning for discrepancy-based visual attribution in medical imaging [J]. *Pattern Recognition Letters*, 2022, 156: 112-118.
- [17] WU C M, ZENG Z. A fault diagnosis method based on Auxiliary Classifier Generative Adversarial Network for rolling bearing [J]. *PLoS One*, 2021, 16(3): e0246905.
- [18] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [J]. *Lecture Notes in Computer Science*, 2015: 234-241.
- [19] HOU Q B, ZHOU D Q, FENG J S. 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. IEEE, 2021: 13708-13717.
- [20] XU J, WANG J, XU X, *et al.* Image recognition for different developmental stages of rice by RAdam deep convolutional neural networks [J]. *Non-gye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, 2021, 37 (8): 143-150.
- [21] HOWARD A, SANDLER M, CHEN B, *et al.* Searching for MobileNetV3 [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019, Seoul, Korea (South). IEEE, 2020: 1314-1324.
- [22] LIU K, YU S T, LIU S D. An improved InceptionV3 network for obscured ship classification in remote sensing images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 4738-4747.
- [23] TAN M, LE Q V. EfficientNet: Rethinking Model Scaling for Convolution Neural Networks [C]. 36th *International Conference on Machine Learning (ICML 2019)*, Los Angeles, USA: *Proceedings of Machine Learning Research*, 2019: 6105-6114.

作者简介:



余泓(1997—),男,四川自贡人,硕士研究生,2020年于西南石油大学获得学士学位,主要从事医学影像识别、计算机视觉、深度学习等方面的研究。
E-mail:790622472@qq.com

导师简介:



罗仁泽(1973—),男,四川内江人,博士,四川省二级教授,1992年于西南石油大学获得学士学位,1999年于西南石油大学获得硕士学位,2005年于电子科技大学获得博士学位,主要从事信号处理与检测、地震资料处理、人工智能等方面的研究。E-mail:lrzsm1th@126.com