

文章编号: 1671-7449(2025)01-0072-09

基于CAP-Net的多粒度乳腺癌病理图像识别模型

张丹蕾, 白艳萍*, 程蓉, 续婷

(中北大学 数学学院, 山西 太原 030051)

摘要: 在医学图像识别领域, 图像的特征提取与图像的放大倍数有着紧密的联系, 因此, 多数乳腺癌图像识别模型都会在不同放大倍数下进行实验。但在实际应用中希望能够综合不同倍数的图像信息来全面评估疾病特征, 提升患者治疗效果。针对上述问题以及医学图像中肿瘤分类的挑战, 聚焦于关注肿瘤类别而不依赖于特定放大倍数, 提出了基于卷积神经网络(Convolutional Neural Networks, CNN)和上下文感知注意力池化(Context-aware Attentional Pooling, CAP)的分类模型。首先通过CNN提取图像的卷积特征, 然后结合CAP模块综合考虑4种级别的特征上下文信息(包括像素级、小区域、大区域和图片级)进行分类。使用DenseNet121、MobileNetV2和Xception 3种CNN网络结合CAP在BreCaHis数据集上进行实验, 将同一类别4种不同放大倍数的数据合并起来, 对8类乳腺癌病理图像进行识别。该模型的准确率达到96.87%, 验证了其在医学图像分类中的有效性。

关键词: 上下文感知注意力池化; 乳腺癌病理图像; 图像识别; 卷积神经网络; 多粒度图像识别

中图分类号: TP391 **文献标识码:** A **doi:** 10.62756/csjs.1671-7449.2025011

引用格式: 张丹蕾, 白艳萍, 程蓉, 等. 基于CAP-Net的多粒度乳腺癌病理图像识别模型[J]. 测试技术学报, 2025, 39(1): 72-80.

ZHANG Danlei, BAI Yanping, CHENG Rong, et al. Multi-granularity breast cancer pathological image recognition model based on CAP-Net[J]. Journal of Test and Measurement Technology, 2025, 39(1): 72-80.

Multi-Granularity Breast Cancer Pathological Image Recognition Model Based on CAP-Net

ZHANG Danlei, BAI Yanping*, CHENG Rong, XU Ting

(School of Mathematics, North University of China, Taiyuan 030051, China)

Abstract: In the field of medical image recognition, the feature extraction of images is closely related to the magnification of the image, so most models of breast cancer image recognition will perform experiments at different magnifications. However, in practical applications, it is hoped that different magnifications of image information can be comprehensively utilized to comprehensively evaluate disease features and improve patient treatment effectiveness. In response to the above issues and the challenges of tumor classification in medical images, a classification model based on convolutional neural networks (CNN) and context-aware attentional pooling (CAP) is proposed, focusing on tumor categories without relying on specific magnifications. Firstly, the convolutional features of the image are extracted through CNN, and

收稿日期: 2024-05-09

基金项目: 国家自然科学基金资助项目(61774137); 山西省基础研究计划资助项目(202103021224195, 202103021224212, 202103021223189, 20210302123019)

作者简介: 张丹蕾(1999—), 女, 硕士生, 主要从事图像处理与图像分类研究。E-mail: 1625748445@qq.com。

* 通信作者: 白艳萍(1962—), 女, 教授, 博士, 主要从事现代优化理论与方法、神经网络算法及应用研究。E-mail: baiyp@nuc.edu.cn。

then the four levels of feature context information (including pixel-level, small-region, large-region and image-level) are comprehensively considered by combining them with the CAP module for classification. Using DenseNet121, MobileNetV2 and Xception three CNN networks combined with CAP, experiments were carried out the on BreakHis dataset. Four data of the same category with different magnifications were combined to identify eight types of breast cancer images. The accuracy of the model reached 96.87%, verifying its effectiveness in medical image classification.

Key words: context-aware attentional pooling; breast cancer pathological images; image recognition; convolutional neural network; multi-granularity image recognition

0 引言

据世界卫生组织最新数据,乳腺癌已取代肺癌成为全球第一大肿瘤。2023年4月14日,中国疾控中心周报发布的我国女性乳腺癌数据分析报告指出,到2030年,全国将有超40万例新诊断病例和超10万例死亡^[1]。庞大的数据说明乳腺癌的早期发现和诊断尤为重要,高效快速的组织病理学诊断方法已成为必需。然而,目前用于指导治疗决策的组织病理学、肿瘤大小和分级等技术并不完善。因此,迫切需要更快、更好的能够识别出肿瘤类型的模型,以此为病理学家提供客观科学的诊断依据。

随着人工智能的蓬勃发展,深度学习在各种场景中的应用大大增加,生物学也逐渐成为使用深度学习取得显著进步的热门领域。Spanhol等^[2]在提出BreakHis数据集的论文中使用Alexnet对4种放大倍数的图像分别进行了良恶性二分类,得出的精度范围为80%~85%,这为评估基于卷积神经网络(Convolutional Neural Networks, CNN)的乳腺癌图像分类算法提供了基准。深度残差网络(Residual Neural Network, ResNet^[3])的诞生使得卷积网络成为乳腺癌图像分类的主力军。Koné等^[4]使用Resnet50模型对乳腺癌图像进行分类,四分类精度达到81%;Brancati等^[5]将3种ResNet组合在一起,取得了较好的分类效果;Rakhlin等^[6]将ResNet50、vgg16和Inception_v3 3种模型结合,四分类任务的准确率达到87.2%,二分类任务的准确率为93.8%;Campilho等^[7]融合了Resnet与Inception-v2模型,取得四分类任务的准确率为76%。

将CNN应用于医学图像分类的确加快了临床诊断速度,但是在对乳腺癌图像进行分类时存在这样一个事实:病理学家在诊断时,需要先通

过低倍镜确定疑似区域,再用高倍镜找到病灶位置并判断肿瘤的良恶性以及等级,因此,需要对不同放大倍数的数字病理图像进行分类。而不同放大倍数的图像具有不同的空间尺度,这将导致网络难以捕捉到所有尺度下的重要特征。较大的尺度可能会丢失较小尺度的细节,反之亦然。如果只训练特定尺度的图像,那么模型可能无法很好地泛化到其他尺度。此外,不同尺度的图像可能需要不同的特征提取器,原始的神经网络可能无法有效地适应这种多尺度的需求,导致在某些尺度上失去关键信息。

注意力机制的出现一定程度上缓解了上述问题。在传统的卷积神经网络中加入注意力机制,通过动态调整不同位置的权重,以此来突出有效信息,使模型更加关注图像中的关键区域,从而提高分类的准确性^[8]。但在使用注意力机制时也会存在不足之处,如注意力机制可能过于集中于图像的某些区域,导致其他区域的信息被忽略。这种情况下,模型可能会错过一些关键特征,特别是在不同放大倍数下图像中可能包含的关键信息。另外,在训练注意力机制时,可能需要更大规模的数据集来确保模型能够准确地学习到图像中不同区域的权重。对于医学图像领域的的数据,公开且大量的数据集获取是较为困难的。

在实际应用中,无论放大倍数为多少,病理学家都需要精准地定位病灶位置,所以,构建一个与放大倍数无关的乳腺癌图像识别模型具有实际意义。如何能使模型感兴趣的区域不受放大倍数的影响,即便是在不同尺度的图像下,模型的表现力仍然能够达到一致是非常重要的。为了使计算机能够真正地“读懂”一幅图,我们需要以更细致的方式去描述。考虑到图像生成过程是从像素到物体再到场景,因此不仅要调节图片中某一部分及其空间排列,还要用多个局部描述来定义它们的外观以及它们在区分细微变化中的重要

性。这些局部描述应该是丰富的,相互补充的,以提供对该图像的完整描述。那么模型就可以避免因过度关注某一区域而丢失关键信息的可能,这将有利于寻找到病灶的位置。

总结上述内容,采用多粒度的乳腺癌病理图像识别模型能够更准确地定位和识别肿瘤,为乳腺癌的分类提供更高效的手段,特别是在早期诊断阶段。本文提出的方法基于BreakHis数据集,与传统方法不同的是不限于特定放大倍数,而是在基础CNN网络提取图像特征的基础上,引入了上下文注意力池化进行八分类任务。实验结果表明,相对于以往的分类方法,本文所提方法在分类准确率上有显著提升,更符合临床实际需求,有望为乳腺癌的早期诊断提供有力支持。

1 乳腺癌图像分类网络

本文提出基于卷积神经网络和上下文感知注意力池化(Context-aware Attentional Pooling, CAP^[9])的乳腺癌图像分类网络。图片数据输入网络模型后,经过主干CNN网络提取特征并输出,再将卷积网络输出的特征作为输入,在CAP模块中学习调整特征中不同区域的重要性,从而得出局部区域的丰富外观特征及其空间特征,最后进行准确的分类。主要包括3个部分:基础CNN网络 $F_b(,; \theta_b)$ 、CAP模块 $F_c(,; \theta_c)$ 和分类模块 $F_d(,; \theta_d)$ 。CAP-Net模型总体结构如图1所示。

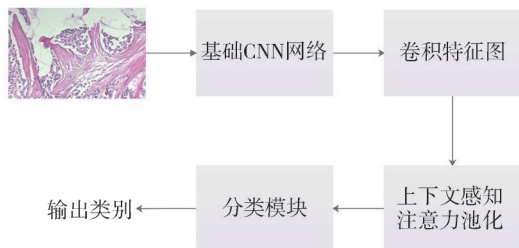


图1 CAP-Net模型总体结构图

Fig. 1 Overall structure diagram of CAP-Net model

1.1 基础卷积神经网络

本文分别使用DenseNet121^[10]、MobileNetV2^[11]、Xception^[12]3种卷积神经网络作为基础CNN网络模块。

DenseNet提出了一个密集连接机制,将所有的层互相连接,具体来说就是每个层都要接受该层前面所有层作为其额外的输入。在传统的 l 层卷积神经网络中,共有 l 个连接,每个连接只作用

于当前网络层和其后的网络层。而在DenseNet中,对于每一层网络层,其之前所有网络层的特征图都会作为此层的输入,而其自身的特征图会作为其后每一层网络层的输入,实现特征重用,提升效率。DenseNet-121是DenseNet模型的一个变种,其网络深度为121层。DenseNet-121的网络结构共包含4个稠密块(Dense Block),各个稠密块之间通过过渡块(Transition Layer)连接。稠密块示意图如图2所示,其中 $[X_0, \dots, X_{l-1}]$ 表示第 $0, \dots, l-1$ 层特征图集合, H_l 可视作BN、Relu、卷积等函数的累加。

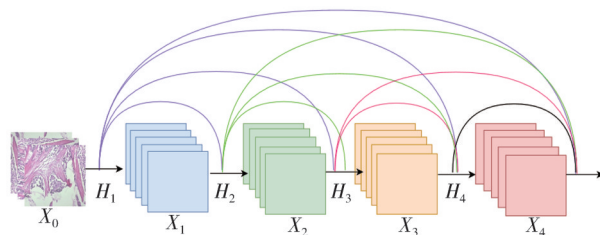


图2 稠密块示意图

Fig. 2 Schematic diagram of dense blocks

MobileNetV2是谷歌推出的一款轻量级深度神经网络模型,主要用于移动端和嵌入式设备上。MobileNetV2在保持高性能的同时,相比于传统的深度神经网络模型具有更小的参数量和更低的计算复杂度,因此更加适合在资源有限的设备上运行。

MobileNetV2主要通过两个方面来提升性能和降低计算复杂度:倒残差结构(Inverted Residuals)和线性瓶颈层(Linear Bottlenecks)。在传统的ResNet残差结构中,先使用 1×1 卷积进行降维,再使用 3×3 卷积,最后通过 1×1 卷积实现升维,即两头大中间小。而在MobileNetV2中,将降维和升维的顺序进行了调换,并且将 3×3 卷积换为 3×3 的逐通道卷积(Depthwise Convolution)。逐通道卷积的原理如图3所示,一个卷积核负责其中一个通道,即一个通道只会被一个卷积核卷积。这种倒残差结构可以使网络中间层的通道数更多,从而提升特征提取能力。同时,由于倒残差结构在输入和输出之间增加了恒等映射,使得网络更容易训练。另外,最后一层的卷积层采用线性激活函数,而不是ReLU激活函数。倒残差结构示意图如图4所示,由于倒残差结构是两头小中间大的结构,因此输出的是一个低维的特征信息。使用线性激活函数可以避免特征信息的损失。

Xception是对Inception v3网络的改进,主要是将原来Inception v3中的卷积操作作用深度可分离

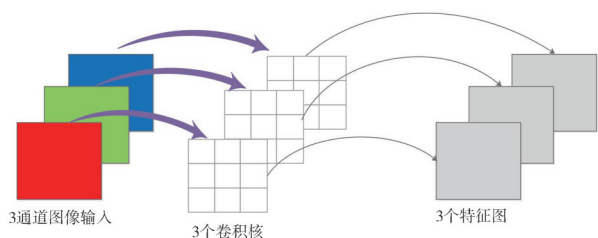


图 3 逐通道卷积(以三通道图像为例)

Fig. 3 Depthwise convolution (using a three channel image as an example)

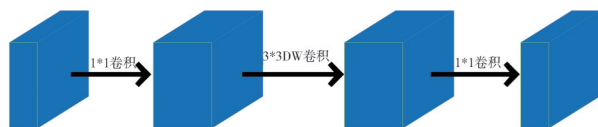


图 4 倒残差结构

Fig. 4 Inverted residuals

卷积(Depthwise Separable Convolution)替换。深度可分离卷积主要分为两部分:逐通道卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)。它通过将多个 1×1 的卷积层堆叠在 3×3 的卷积层之后,实现了更深的网络结构,同时保

持了较小的模型大小和计算复杂度。这种结构可以有效地提高模型的特征提取能力,从而提高模型的性能。

1.2 上下文感知注意力池化

上下文感知注意力池化(CAP)通过聚焦于一个给定的区域,并以其他区域和该区域为条件,通过双线性池化,有效地捕捉了亚像素梯度在每个区域的细微变化,并学习关注信息积分区域及其在区分不同子类别中的重要性。此外,CAP模块提出了一种新的特征编码方式,对网络的空间排列和外观特征进行编码,通过考虑积分区域的信息量与其空间结构之间的内在一致性来捕获积分区域之间的语义相关性。最后还引入了一个可学习的池化算法,自动选择循环网络的隐藏状态,学习并编码整体区域潜在的空间排布,并据此来推断子类别。

CAP模块综合考虑像素级、小区域、大区域以及图片级这4种级别的特征上下文信息进行分类。图5为CAP模块的结构图。

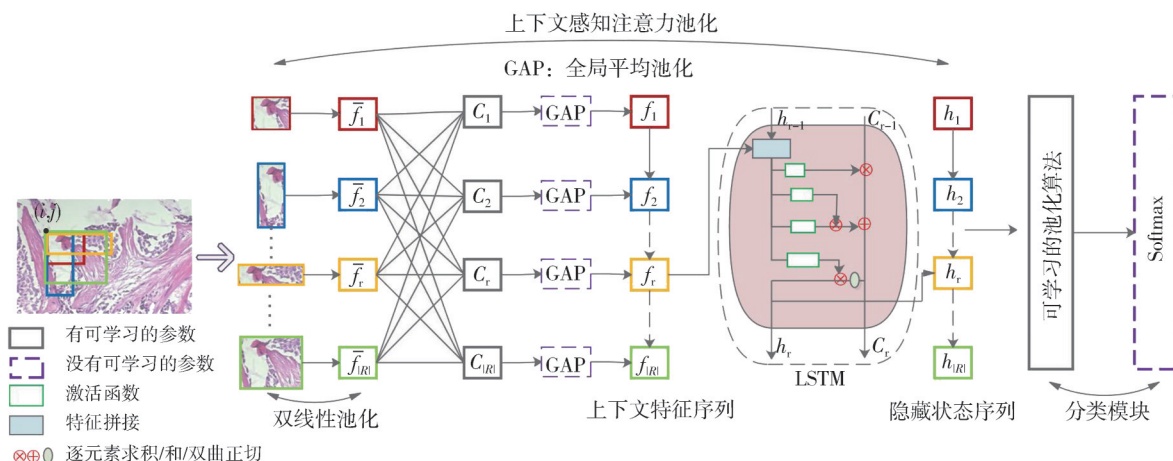


图 5 CAP 的详细架构

Fig. 5 Detailed architecture of CAP

已知输入 N 个图像 $I = \{I_n | n = 1, \dots, N\}$ 和它们各自的类别标签。该模块的目的是找到一个映射函数 F ,预测标签为 $\hat{y}_n = F(I_n)$,对应真实的标签为 y_n 。最终的目标是最小化损失函数 $L(y_n, \hat{y}_n)$,定义卷积网络输出的特征为 $x = F_b(I_n; \theta_b)$ 。

1.2.1 像素级关系

像素级特征的上下文信息主要是学习到像素间的关联度 $p(x_i | x_j; \theta_p)$ 。即使像素点相距很远(即 $\forall i \neq j, 1 \leq i, j \leq W \times H$),都要计算像素之间的关

联度。像素的远近关系如图6所示。在计算 j 位置的输出时,根据关联度来综合所有其他像素特征,特征转化时使用 1×1 卷积。这一步直接对主干网络输出的特征进行操作,但没在整体流程图中体现。

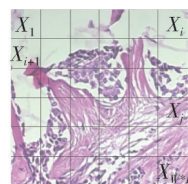


图 6 像素级关系

Fig. 6 Pixel-level relationships

1.2.2 提出积分区域

为了更高效地学习上下文信息,需要在不同特征图上定义不同粒度级别的区域,粒度级别由区域的大小决定。假设 (i,j) 位置上的最小区域为 $r(i,j,\Delta_x,\Delta_y)$,可通过放大它的宽和高衍生出一系列区域 $R=\{r(i,j,m\Delta_x,n\Delta_y)\}$,其中 $i<i+m\Delta_x\leq W,j<j+n\Delta_y\leq H$ 。在不同的位置产生相似的区域合集 R ,得到最终的区域合集 $\mathfrak{R}=\{R\}$ 。 \mathfrak{R} 可以提供全面的上下文信息和不同层级图片的细微特征。

1.2.3 双线性池化

在特征图上得到 $|\mathfrak{R}|$ 个区域后,其大小为从最小的 $\Delta_x\times\Delta_y\times C$ 到最大的 $W\times H\times C$,我们的目标是将不同大小的区域表示为固定大小的特征,这里主要采用了双线性插值法。定义 $T_\psi(Y)$ 为坐标转换函数, $Y=(i,j)\in\mathbb{R}^2$ 为区域坐标,对应的特征值为 $R(Y)\in\mathbb{R}^C$,则转换后的图片 \tilde{R} 的 \tilde{Y} 坐标上的值为

$$\tilde{R}(\tilde{Y})=\sum_y R(T_\psi(Y))K(\tilde{Y},T_\psi(Y)), \quad (1)$$

式中: $R(T_\psi(Y))$ 为采样函数; $K(\cdot)$ 为核函数。这里采用将目标坐标映射回原图,取最近的4个点,按距离进行输出的方法,最终得到池化后的固定特征 $\bar{f}(w\times h\times C)$ 。

1.2.4 上下文注意力机制

根据 \bar{f}_r 与其他特征 $\bar{f}_{r'}$ 的相似性进行加权输出 $(r,r'\in\mathfrak{R})$,使得模型能够选择性地关注更相关的区域,从而产生更全面的上下文信息。通过查询项 $q(\bar{f}_r)$ 和一组关键词项 $k(\bar{f}_{r'})$,输出上下文向量 C_r 。

$$C_r=\sum_{r'=1}^{|\mathfrak{R}|}\alpha_{r,r'}\bar{f}_{r'}, \quad (2)$$

$$\alpha_{r,r'}=\text{softmax}(W_\alpha\beta_{r,r'}+b_\alpha), \quad (3)$$

$$\beta_{r,r'}=\tanh(q(\bar{f}_r)+k(\bar{f}_{r'})+b_\beta), \quad (4)$$

$$q(\bar{f}_r)=W_\beta\bar{f}_r \text{ and } k(\bar{f}_{r'})=W_{\beta'}(\bar{f}_{r'}). \quad (5)$$

参数矩阵 W_β 和 $W_{\beta'}$ 用来将输入特征转换为查询项和关键词项, W_α 为非线性组合, b_α 和 b_β 为偏置项,整体的可学习参数为 $\{W_\beta,W_{\beta'},W_\alpha,b_\alpha,b_\beta\}\in\theta_c$,而注意力项 $\alpha_{r,r'}$ 则代表两个特征之间的相似性。这样,上下文向量 C_r 能够代表区域 \bar{f}_r 蕴含的上下文信息,这些信息是根据该区域与其他区域的相关程度获得的,

整体的计算思想与自注意力机制基本相似。图7及图8中用不同形状来代表不同大小的区域。

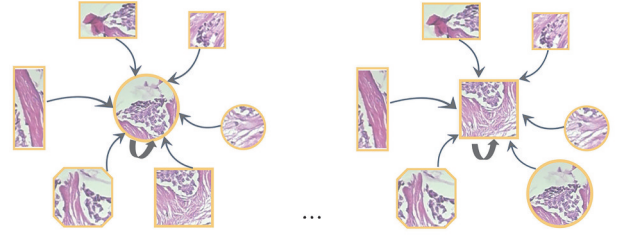


图7 来自完整区域的关注上下文信息(周围上下文)

Fig.7 Attention-focused contextual information from integral regions (surrounding context)

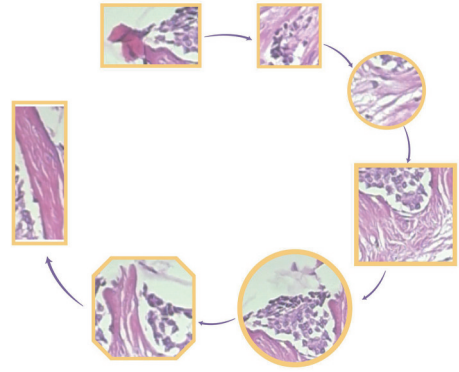


图8 捕捉空间排布

Fig.8 Capturing spatial arrangements

1.2.5 空间结构编码

上下文向量 $C=\{C_r|r=1,\dots,|\mathfrak{R}|\}$ 描述了区域的关键程度和特点,为了加入与空间排列相关的结构信息,首先将区域的上下文向量 C 转为区域序列(按从上到下、从左到右的顺序),然后输入到循环神经网络中,使用循环神经网络的隐藏单元 $h_r\in\mathbb{R}^n$ 来表达结构特征。区域 r 的中间特征可表示为 $h_r=F_h(h_{r-1},f_r;\theta_h)$, F_h 采用长短期记忆网络(LSTM)^[13], $\theta_h\in\theta_c$ 包含了LSTM的相关参数。为了增加泛化能力和减少计算量,上下文特征 $f_r\in\mathbb{R}^{1\times C}$ 由 $C_r\in\mathbb{R}^{w\times h\times C}$ 进行全局平均池化得到,最终输出上下文特征序列 $f=(f_1,f_2,\dots,f_r,\dots,f_{|\mathfrak{R}|})$ 所对应的隐藏状态序列 $h=(h_1,h_2,\dots,h_r,\dots,h_{|\mathfrak{R}|})$,后续用于分类模块中。

1.2.6 分类模块

在此模块中使用可学习的池化操作,通过组合响应相似的隐藏层 h_r 来整合特征信息,以此引导模型分辨细微的变化。用可导的聚类方法对隐藏层的响应值进行转换,首先计算隐藏层响应对类簇 k 的相关性,再加权到类簇 k 的VLAD encoding中。

$$\gamma_k(\mathbf{h}_r) = \frac{e^{\mathbf{W}_i^T \mathbf{h}_r + b_i}}{\sum_{i=1}^K e^{\mathbf{W}_i^T \mathbf{h}_r + b_i}}, \quad (6)$$

$$N_v(o, k) = \sum_{r=1}^{|\mathcal{R}|} \gamma_k(\mathbf{h}_r) \mathbf{h}_r(o), \quad (7)$$

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}_N N_v). \quad (8)$$

每个类簇都有可学习的参数 \mathbf{W}_i 和 b_i , 整体思想基于 Softmax。将隐藏层的响应值按 Softmax 的权重分配到不同的类簇中。在得到所有类簇的编码向量后, 使用可学习的权值 \mathbf{W}_N 和 Softmax 进行归一化, 图 9 为隐藏层的可学习聚合分类过程。因此, 分类模块 F_d 的可学习参数为 $\theta_d = \{\mathbf{W}_i, b_i, \mathbf{W}_N\}$ 。

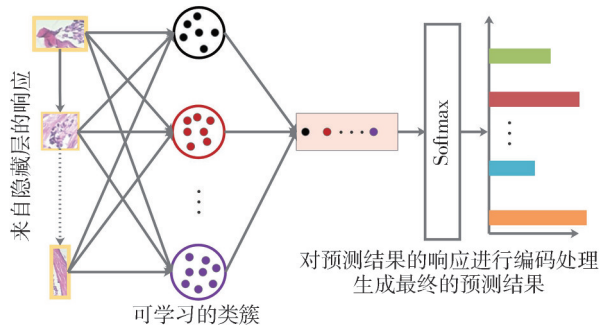


图 9 对类预测的聚集隐藏状态响应

Fig. 9 Aggregated hidden state responses to class prediction

2 实验

2.1 实验环境及评价指标

本文所有实验都在一台 Intel Core i7-5820K CPU、GPU 处理器为 NVIDIA GeForce GTX 1080 Ti、运行内存为 16 G 的计算机上执行。编程语言使用 Python, 网络模型的实现基于 TensorFlow 深度学习框架。训练设置批大小为 6, 初始学习率为 0.000 3, 使用随机梯度下降优化器 (SGD)。在实验中使用预训练网络进行特征提取, 以提高分类性能。

评价指标为图像级准确率。 N_I 为分类正确的图片数量, N 为测试集样本数量, 则识别准确率为

$$acc = \frac{N_I}{N}. \quad (9)$$

2.2 数据集

BreaKHis 数据集由 8 种不同类别的乳腺肿瘤图像组成, 这些图像来自 2014 年巴西 P&D 实验室进行的临床研究。该研究获得了机构审查委员

会的批准, 并给予所有参与患者书面知情同意。该数据集一共包括 7 909 张乳腺癌病理检测图像, 其中包括良性肿瘤 2 480 例, 恶性肿瘤 5 429 例。良性子类包括腺病、纤维腺瘤、叶状肿瘤和管状腺瘤, 恶性子类包括导管癌、小叶癌、粘液癌和乳头状癌。实验对该数据集中的 8 类乳腺癌肿瘤图像进行识别, 训练集和测试集的比例为 7:3。表 1 为该数据集的具体分布。图 10 给出了该数据集中部分代表性的示例。

表 1 BreaKHis 数据集图像分布

Tab. 1 BreaKHis dataset image distribution

类别	子类	放大倍数				合计
		40	100	200	400	
良性	腺病	114	113	111	106	444
	纤维腺瘤	253	260	264	237	1014
	叶状肿瘤	109	121	108	115	453
	管状腺瘤	149	150	140	130	569
恶性	导管癌	864	903	896	788	3451
	小叶癌	156	170	163	137	626
	粘液癌	205	222	196	169	792
	乳头状癌	145	142	135	138	560

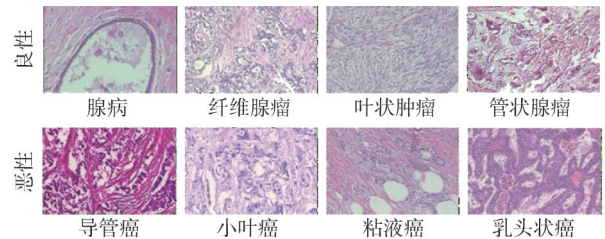


图 10 BreaKHis 数据集代表性示例

Fig. 10 Representative example of BreaKHis dataset

2.3 主干网络调参

本文选取 DenseNet121、MobileNetV2、Xception 3 种网络作为基础卷积神经网络。为了提高模型的训练效果和预测精度, 需要对预训练网络调整一些参数, 其中包括调整学习率、激活函数、优化器等参数。

本文中采取动态学习率 (Dynamic Learning Rate), 初始学习率为 0.000 3, 每 50 个 epoch 学习率进行一次衰减, 为初始学习率的 0.1, 一共进行 200 次 epoch 训练。与固定学习率 (Fixed Learning Rate) 相比, 准确率有所提升。图 11 通过描绘 Xception 的准确率曲线和损失曲线, 验证了动态学习率分类效果更好。图中 Xception-Flr 和 Xception-Dlr 分别代表 Xception 使用固定学习率和动态学习率。

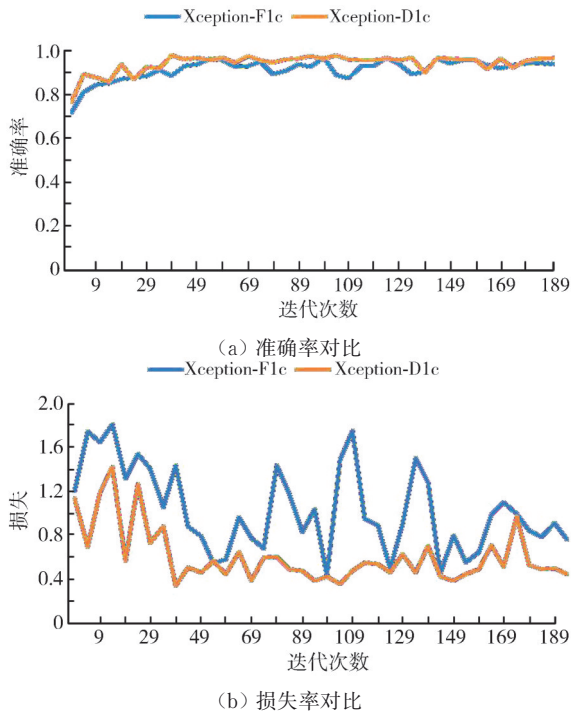


图 11 准确率与损失率对比

Fig. 11 Comparison of accuracy and loss rate

2.4 实验结果与分析

将 CAP 模块分别应用于 DenseNet121、MobileNetV2、Xception，通过训练集和验证集的准确率和损失率来评价各个模型的性能，该三类模型训练集的准确率曲线和损失曲线见图 12。

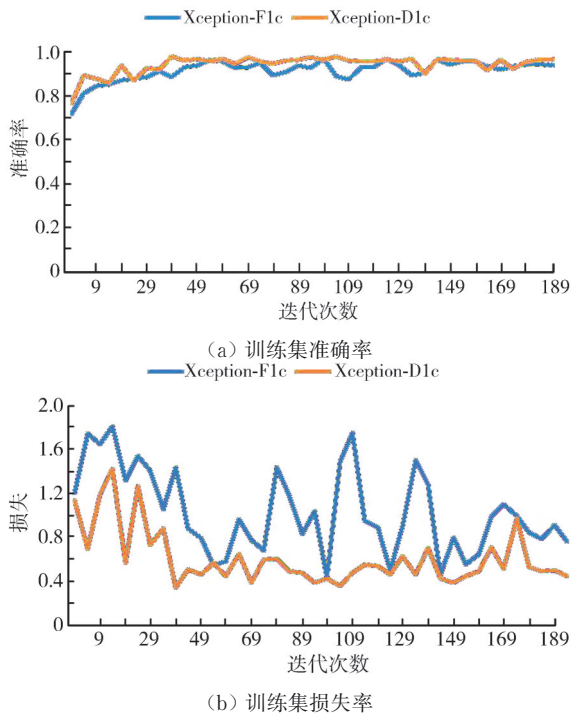


图 12 训练集的准确率和损失率

Fig. 12 Training set accuracy and loss rate

从图 12 中可以看出，CAP-Xception 网络模型的准确率最高，损失率最小。此外，经过实验发现，在训练过程的前 50 次迭代中损失率下降比较快，而在 100 次迭代后趋于缓慢。

CAP-Xception 网络模型混淆矩阵见图 13。每一个小方块中上面的数字代表图片数量，下面的数字代表图片数量所占总数的比例。从混淆矩阵中可以看出导管癌类别的召回率最高，并且其他类别的图片被错误预测为导管癌的图片数量较多。原因可以归为 BreakHis 数据集中各类别图像数量不平衡，在该数据集中导管癌这一类别的图片数量最多。另外，对每类图片的精确率、召回率以及 F1 分数进行了统计，如表 2 所示。从表 2 数据可以看出，导管癌这一类别的精确率、召回率以及 F1 分数均为最高，分别达到了 97.34%、98.75% 和 0.9804。

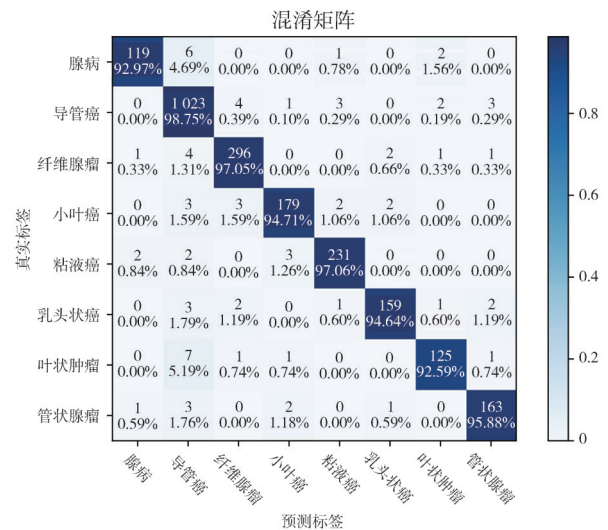


图 13 CAP-Xception 模型的混淆矩阵

Fig. 13 Confusion matrix of CAP-Xception model

表 2 不同类别的精确率、召回率、F1 分数对比

Tab. 2 Comparison of precision, recall, and F1 Scores for different categories

类别	精确率/%	召回率/%	F1 分数
腺病	96.75	92.97	0.9482
导管癌	97.34	98.75	0.9804
纤维腺瘤	96.73	97.05	0.9689
小叶癌	96.24	94.71	0.9547
粘液癌	97.06	97.06	0.9706
乳头状癌	96.95	94.64	0.9578
叶状肿瘤	95.42	92.59	0.9398
管状腺瘤	95.88	95.88	0.9588

为了验证 CAP 模块的有效性，将本文提出的 CAP-Xception 模型与不加入 CAP 模块的 Xception 模型同时使用梯度加权类激活映射 (Gradient-weighted

Class Activation Mapping, Grad-CAM)^[14]技术进行特征可视化。设定热力图的红色为高表达区域,一定范围内红色区域占比越大,说明该模型提取到的有效信息越多。对生成的热力图提取红色通道并将

其转化为灰度图像,设置阈值为0.7,统计每幅热力图中大于该阈值的像素点,得出提取特征面积占总图像面积的比例,从而对模型在特征提取方面进行定量分析,具体结果如图 14 所示。

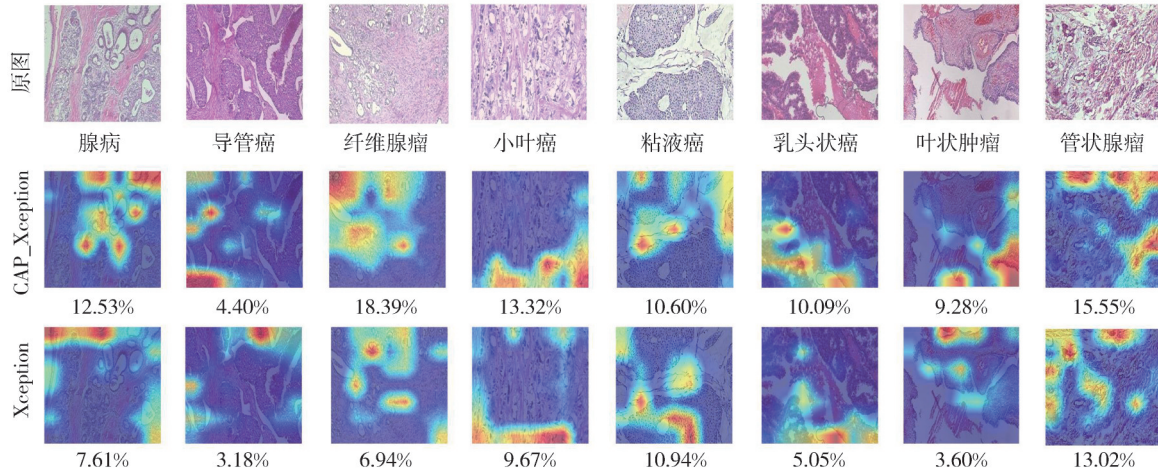


图 14 GradCam 可视化

Fig. 14 GradCam visualization

图 14 分别展示了 CAP-Xception 模型与 Xception 模型对不同类别的乳腺癌病理检测图像生成的热力图,图片下方的数字表示热力图中红色区域面积所占的比例。从图中可以看出,CAP-Xception 模型与 Xception 模型在提取特征时关注的图像特征范围以及所占比例均存在一定差异。导管癌和粘液癌这两类图像中的红色区域占比接近,差异较小;而对于腺病、纤维腺瘤以及乳头状癌等其他类图像,占比差异较大,说明加入 CAP 模块后模型的性能有了显著提高。总体来说,本文所提模型比原始模型能关注到的有用信息更多,因此本文所提模型是有效的。

最后为了检验该方法的可行性和准确性,将数据集以相同的比例随机划分为训练集和测试集。训练设置批大小为 64,初始学习率为 0.0003,使用随机梯度下降优化器(SGD),最大迭代次数为 50。在对 8 个基础卷积神经网络进行训练后,将测试集送入这 8 个模型中进行乳腺癌病理图像八分类实验,并与本文所提出的增加 CAP 模块后的模型实验结果进行对比。对比结果见表 3。

表 3 的消融实验说明了增加 CAP 模块后的模型相较于原始的网络模型,分类准确率大幅提升,因此该方法有效可行的。

同时,选用近年来针对不区分放大倍数的乳腺癌图像多分类相关文献中的结果进行对比,如表 4 所示。

表 3 基础网络模型与 CAP-Net 模型的多分类识别结果

Tab. 3 Subclass recognition results of the basic network model and the CAP-Net model

模型	图像级准确率/%
AlexNet ^[15]	88.75
VGG16 ^[16]	87.62
ResNet50 ^[17]	91.87
GoogLeNet ^[18]	88.83
InceptionV3 ^[19]	90.10
DenseNet121 ^[10]	83.75
MobileNetV2 ^[11]	90.90
Xception ^[12]	85.99
CAP-DenseNet121	94.62
CAP-MobileNetV2	92.25
CAP-Xception	96.87

表 4 其他文献的多分类识别结果

Tab. 4 Subclass recognition results of other literatures

文献	训练集和测试集比例	图像级准确率/%
文献[20]	7:3	63.6(恶性子类)
文献[21]	5折交叉验证	95.95
文献[22]	8:2	92.03
文献[23]	7:3	94.78
文献[24]	8:2	95.50
本文模型	7:3	96.87

从表 4 中可以看出,本文所提方法相较于以往研究,除去 Benhammou 等^[15]关于恶性子类的研究以外,准确率分别提升了 0.92、4.84、2.09、1.37 百分点,说明了该方法的有效性。此外,近年来关于不针对特定倍数的乳腺癌图像多分类研究较少,本文所提方法可对乳腺癌肿瘤分类的临床应用做出一定贡献。

3 结 论

对乳腺癌图像进行多分类识别研究是为了在给模型输入一张未标注的图像时,它能够准确地判断出肿瘤类别及其等级。因此,提出基于卷积神经网络和上下文感知注意力池化的乳腺癌图像分类方法。通过在BrecaKHis数据集上的实验可以表明,该方法比起以往的研究方法既高效又保证了较好的分类效果,距离实际临床应用又近了一步。但是,由于可用在多分类研究的乳腺癌公开数据集非常少,未来可在丰富医学图像数据、提高标记数据可用性等方面开展研究。

参考文献:

- [1] SUN K, LEI L, ZHENG R, et al. Trends in incidence rates, mortality rates, and age-period-cohort effects of female breast cancer-China, 2003-2017[J]. *China CDC Weekly*, 2023, 5(15): 340-346.
- [2] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. A dataset for breast cancer histopathological image classification[J]. *IEEE Transactions on Biomedical Engineering*, 2016, 63(7): 1455-1462.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [4] KONÉ I, BOULMANE L. Hierarchical ResNeXt models for breast cancer histology image classification[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 796-803.
- [5] BRANCATI N, FRUCCI M, RICCIO D. Multi-classification of breast cancer histology images by using a fine-tuning strategy[C]//International Conference Image Analysis and Recognition. Cham: Springer, 2018: 771-778.
- [6] RAKHLIN A, SHVETS A, IGLOVIKOV V, et al. Deep convolutional neural networks for breast cancer histology image analysis[C]//International Conference Image Analysis and Recognition. Cham: Springer, 2018: 737-744.
- [7] FERREIRA C A, MELO T, SOUSA P, et al. Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet V2[M]//Lecture Notes in Computer Science. Cham: Springer, 2018: 763-770.
- [8] 梁彥彦, 张倩楠, 任竹云, 等. 基于注意力机制的结肠癌病理学图像识别研究[J]. *测试技术学报*, 2022, 36(2): 93-100.
- LIANG Meiyan, ZHANG Qiannan, REN Zhuyun, et al. Research on identification of colon pathology image based on attention mechanism[J]. *Journal of Test and Measurement Technology*, 2022, 36(2): 93-100. (in Chinese)
- [9] BEHERA A, WHARTON Z, HEWAGE P R P G, et al. Context-aware attentional pooling (CAP) for fine-grained visual classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(2): 929-937.
- [10] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261-2269.
- [11] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 4510-4520.
- [12] CHOLLET F. Xception: deep learning with depth-wise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1800-1807.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9: 1735-1780.
- [14] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision (ICCV)*, 2020, 128(2): 336-359.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. *Computer Science*, 2014. DOI: 10.48550/arXiv.1409.1556.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.

(下转第95页)