

基于深度学习的甲状腺癌病理图像分级方法

曹莉凌¹, 蒋珂宏¹, 曹守启¹, 蒋伏松²

1. 上海海洋大学工程学院, 上海 201306; 2. 上海交通大学医学院附属第六人民医院内分泌代谢科, 上海 200233

【摘要】针对日益增长的甲状腺癌早期诊断的需求, 基于深度学习方法, 在EfficientNet网络的基础上结合CA注意力机制, 进行甲状腺癌病理图像自动分级方法研究。实验结果显示, CA-EfficientNet网络模型的精确率达到96.6%, 证明了基于CA-EfficientNet网络的甲状腺癌病理图像自动分级算法的先进性, 基于该算法实现的自动辅助诊断系统具有实际应用性, 可有效降低病理医生工作负担, 并降低因疲劳等主观因素造成的人工诊断误诊率。

【关键词】深度学习; 甲状腺癌; 卷积神经网络; 全切片数字化图像; 图像分级

【中图分类号】R318; TP391

【文献标志码】A

【文章编号】1005-202X(2023)05-0580-09

Automatic grading of pathological images of thyroid cancer based on deep learning

CAO Liling¹, JIANG Kehong¹, CAO Shouqi¹, JIANG Fusong²

1. College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Department of Endocrinology and Metabolism, Shanghai Sixth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200233, China

Abstract: In response to the increasing demand for early diagnosis of thyroid cancer, a deep learning based method is proposed for the automatic grading of the pathological images of thyroid cancer through EfficientNet combined with CA-Net. The experimental results show that the accuracy of CA-EfficientNet model is up to 96.6%, which proves the algorithm superiority in the automatic grading of the pathological images of thyroid cancer. The automatic auxiliary diagnosis system implemented based on the proposed algorithm is applicable in practice for it can effectively reduce the workload of pathologists and reduce the rate of misdiagnosis caused by subjective factors such as fatigue.

Keywords: deep learning; thyroid cancer; convolutional neural network; whole slide image; image grading

前言

癌症是当今世界面临的主要公共卫生问题之一, 随着甲状腺细针穿刺精确诊断技术的普及, 近年来各地区甲状腺癌检出率急剧上升^[1-2]。由于癌症的隐蔽性, 大多患者一经确诊便是癌症晚期, 失去了最佳治疗时机, 因此早期正确诊断成为亟待解决的问题。甲状腺细针穿刺活检是一种传统的微创诊断技术, 可在手术前诊断出结节的病理性质, 为甲状腺疾

病的精准治疗提供依据^[3]。病理科医生借助显微镜对癌细胞进行观察仍是当前诊断甲状腺细针穿刺样本的主要方法。甲状腺细针穿刺活检是甲状腺肿瘤和结节的常规检查方法, 能为精准治疗提供依据, 能在不动手术的情况下有效地了解疾病的性质并指导治疗^[4-6]。穿刺前准备及操作流程^[7]: (1) 术前进行血常规检查, 保证凝血常规正常; (2) 家属陪伴签署知情同意书, 告知患者及家属穿刺的风险和注意事项; (3) 患者仰卧位, 垫靠枕头使颈部呈拉伸状; (4) 颈部常规消毒, 铺无菌洞布; (5) 使用经过无菌处理的超声探头对甲状腺扫描, 确认穿刺部位; (6) 局部消毒后进针, 超声引导下使用细针注射器, 超声下将针尖定位于病变部位, 对结节反复插入穿刺完成采样; (7) 取样本少量涂片并以95%酒精固定, 观察是否符合细胞学诊断要求(至少有6巢细胞团, 每巢细胞团至少有10个细胞), 合格后送病理科检查; (8) 穿刺结束, 患者在观察区留观30 min, 按压穿刺部位15 min

【收稿日期】2022-11-12

【基金项目】浦东新区科技发展基金(PKJ2019-Y03)

【作者简介】曹莉凌, 博士, 研究生导师, 研究方向: 智能检测与控制技术、无线通信安全技术, E-mail: llcao@shou.edu.cn

【通信作者】蒋伏松, 博士, 主任医师, 研究生导师, 研究方向: 糖尿病并发症的人工智能预测、图像识别技术在糖尿病足病和甲状腺眼病及甲状腺结节的应用, E-mail: hajfs@126.com

以上,并嘱咐术后注意事项。将患者手术后的组织送到病理科,病理科再次进行采样,为了保护组织免受细菌或细胞酶的影响,标本用固定剂处理。经过一系列步骤将样本制成载玻片,由病理医生通过显

微镜观察病理样本并得出病理诊断报告。要实现计算机自动分析,需要将病理样本进行数字化扫描并分级。如图1所示为数字化病理学图像生成步骤。

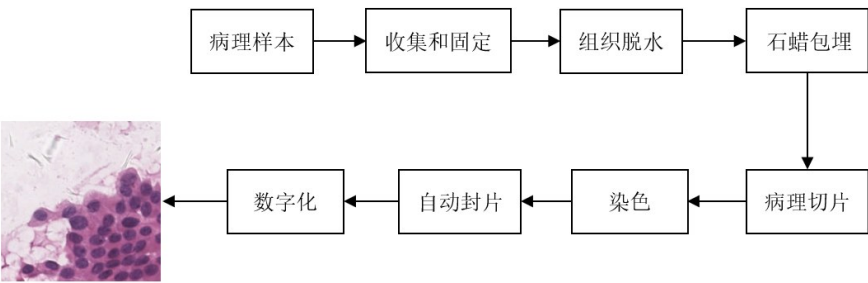


图1 数字化病理学图像
Figure 1 Digitalizing pathological images

数字化病理学^[8-11]使用全切片数字扫描技术对整个组织病理学或细胞病理学玻片进行数字化处理得到高分辨率图像,即全切片数字化图像(Whole Slide Image, WSI)。最初用于研究目的的数字化病理学已经成为一种成像和处理高倍率病理切片图像的方法,既降低了病理学医师日常繁琐的分析工作,提高诊断能力,减少病理医师因技术差异而引起的误诊,节省诊断时间,还能实现远程病理会诊和信息共享,更快地获取院外专家的意见,实现数字化病理切片的管理^[12]。但是,病例的增长速度远远超过该领域专家的培养速度,将有限且有重要价值的人力资源用于重复病理图像的识别和诊断是一个巨大的浪费。近年来,随着人工智能的进步以及用于初级诊断的病理切片的日益数字化,数字化病理学开始逐步满足医生对病理图像更准确检测、分类和预测的需求^[13-15]。

1 实验方法

1.1 基于病理诊断标准的特征提取

深度学习网络通过学习病理图像的细胞特征来判断甲状腺结节的良恶性级别。Bethesda 报告系统^[16-17]将甲状腺癌病理诊断标准分为6个级别,由上海市第六人民医院临港院区的病理科医生对甲状腺癌病理图像进行专业诊断分级。甲状腺癌病理图像和深度学习特征图如表1所示。

病理医生通过甲状腺 Bethesda 报告系统对病理样本进行分级,但是计算机并不能直接通过这个分级标准提取病理图像的细胞特征,只有将图像的抽象特征转换为计算机能识别的亮度、边缘、纹理和色彩等自然特征,因此需要将 Bethesda 系统分级标准简化为计算机能学习的特征分级标准,如表2所示。许

多信息并不方便用数值量化,因此通过图示的方式展现,部分描述信息如图2所示。

1.2 改进 EfficientNet-B0 网络

2019年谷歌提出了 EfficientNet 模型的基本网络架构^[18],如图3所示。该架构通过一个复合系数自适应统一缩放网络的宽度、深度和分辨率来优化神经网络结构,减少训练参数,具有更高的准确性和效率,同时降低了计算复杂度。

EfficientNet-B0 网络的 SE-Net^[19-22]对通道进行权重分配来重新衡量每个通道的重要性而忽略了位置信息,但是甲状腺癌细胞病理图像对细胞位置信息同样关注,细胞的分布会直接影响病理医生对病理图像的诊断结果。针对 SE-Net 网络的这一问题,本研究引入 Coordinate Attention^[23],即 CA 注意力机制,它将通道注意力分解拆分为两个空间方向,分别捕获通道依赖关系和保留精确的位置信息。使用迁移学习策略,在原模型网络基础上,替换原来的 SE-Net 注意力机制,这个模型称为改进的 EfficientNet-B0,即 CA-EfficientNet 模型。

通道注意力机制通过全局池化生成二维特征映射再转换为输出特征向量,与通道注意力不同,CA 注意力是在两个空间方向上对特征进行聚类的一维特征编码过程,这允许在两个空间方向上分别学习通道依赖关系和保留精确位置信息。然后,将得到的特征映射编码为含有方位和位置敏感的特征映射对,将其互补地应用于输入特征映射以增强关注对象的表示。CA 注意力机制结构如图4所示。其中, r 是用来控制模块大小的缩减率。CA 注意力机制对位置信息和通道关系进行编码包括坐标信息嵌入和坐标注意力生成。

表 1 甲状腺癌病理图像及特征图
Table 1 Pathological images and characteristic graphs of thyroid cancer

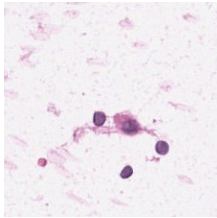
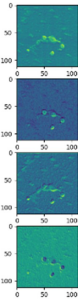
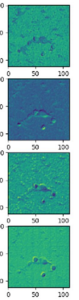
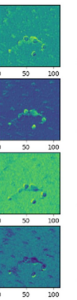
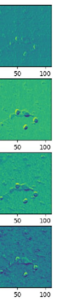
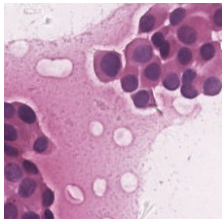
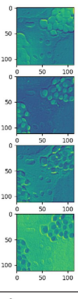
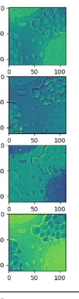
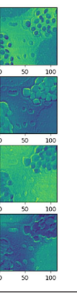
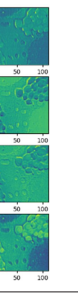
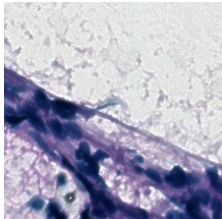
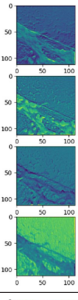
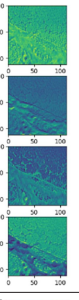
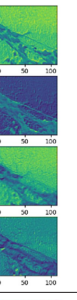
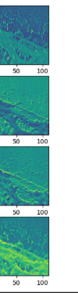
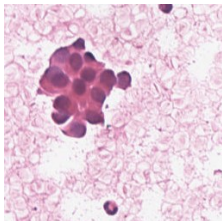
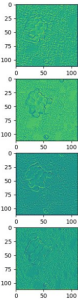
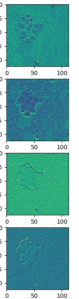
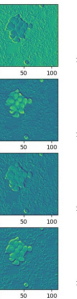
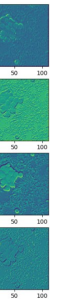
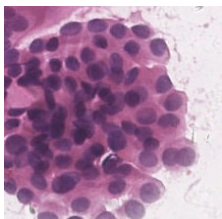
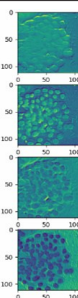
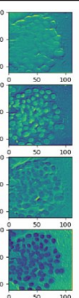
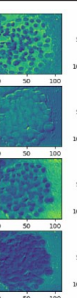
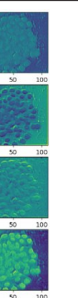
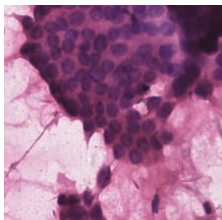
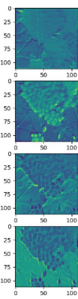
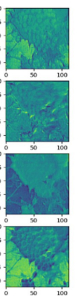
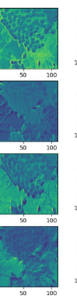
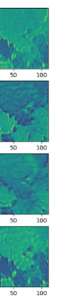
癌症级别	甲状腺癌病理图像	神经网络特征图			
I					
II					
III					
IV					
V					
VI					

表2 分级标准
Table 2 Classification criteria

癌症级别	病理医生分级诊断标准	计算机特征分级标准
I	细胞数量过少,无法诊断	细胞核数量稀少;细胞核间距较大
II	细胞之间平铺,分布比较均匀,几乎没有重叠;细胞核显著增大,核浆比(细胞核的距离比细胞核边缘到细胞膜的距离)显著增大;细胞比较细腻	细胞核增大;细胞核数量适量;细胞核间距适中;核浆比增大
III	细胞形状模糊不清,但是细胞显著增大且有明显的细胞重叠现象	细胞核增大;细胞核数量适量;细胞核堆叠导致边界模糊
IV	苏木素伊红染色嗜酸性,遇酸性环境显粉色,碱性环境显紫色;癌细胞显酸性,细胞的胞浆呈粉红色	细胞核增大;细胞核数量适量;细胞核间距偏小;呈粉色环境
V	细胞显著增大并重叠明显,能观察到部分细胞出现核沟,有少量的细胞呈毛玻璃状	细胞核增大;细胞核数量很多;细胞核间距偏小甚至重叠;部分细胞核呈明暗交替的马赛克形状
VI	细胞显著增大并重叠明显,能观察到部分细胞明显核沟,细胞形态明显不规则,有大量的细胞呈毛玻璃状	细胞核增大;细胞核数量很多;细胞核间距偏小甚至重叠;大部分细胞核呈明暗交替的马赛克形状

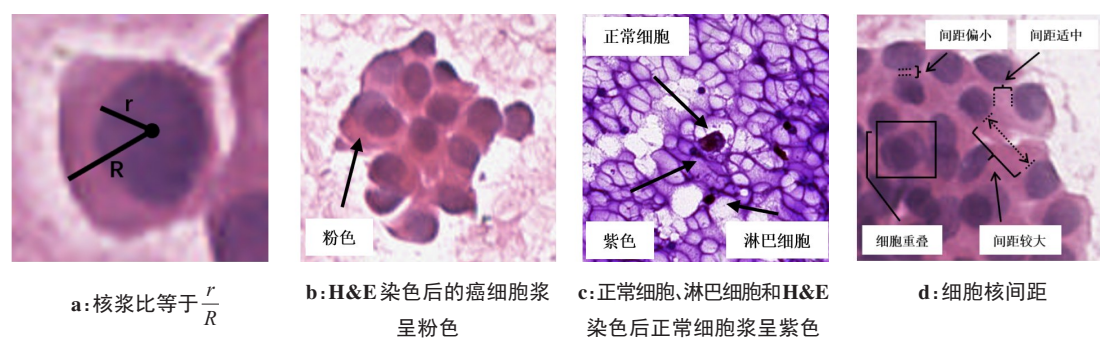


图2 部分描述信息
Figure 2 Some of the description information

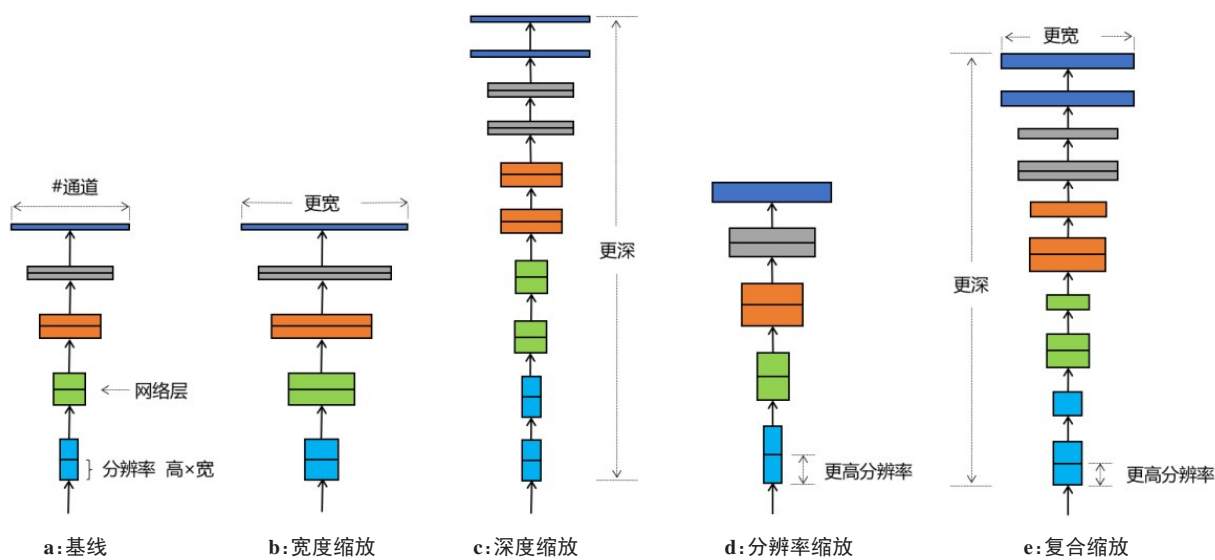


图3 模型缩放
Figure 3 Model scaling

(1)坐标信息嵌入:空间信息一般采用全局平均池化的方法对信道进行编码,但是会将编码信息压缩到一个信道描述子中,位置信息难以保存。为了使注意力模块能够以正确的位置信息捕捉远程空间

的相互作用,将全局平均池化分解为一对一维的特征性编码操作,使用的公式如下:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i,j)$$

(1)

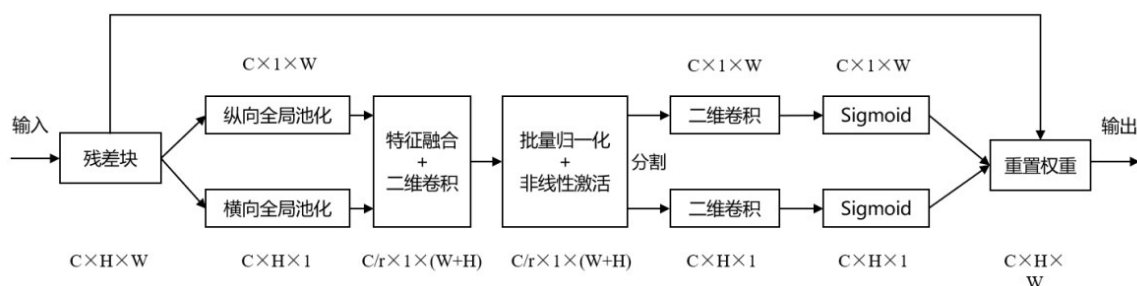


图4 CA注意力机制结构

Figure 4 Structure of CA mechanism

参数 W 是图像的宽度, H 是高度, C 是通道的个数

首先使用 $(H, 1)$ 或 $(1, W)$ 大小的池化核从水平和垂直坐标编码所有信道, 因此高度为 h 的第 c 通道的输出可以表示为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (2)$$

同样, 宽度为 w 的第 c 通道的输出可以写成:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (3)$$

不同于 SE 模块只产生单一的特征向量, 上述两次变换所得特征分别在两个空间方向上聚合得到, 返回一对方向感知特征映射。两种变换允许注意力模块在两个不同方向分别捕捉沿空间通道关系和精确位置信息, 使网络更精确地定位数据特征。

(2) 坐标注意力生成: 为了更好地利用上面产生的一对具有全局感受野和精确位置信息的特征映射, 提出了坐标注意力生成, 要求如下: 在轻量型网络中新的转换应该简单易用; 必须要充分利用得到的位置信息, 使得细胞位置能被精确捕获; 最后还要能够有效地捕获通道间的关系。将上面两个信息嵌入变换进行拼接操作, 然后通过变换函数 F_1 对其进行 1×1 卷积操作:

$$f = \delta(F_1([z^h, z^w])) \quad (4)$$

式中, $[z^h, z^w]$ 为将坐标和通道进行拼接操作, δ 为 sigmoid 激活函数, f 为水平和垂直空间信息编码的中间特征映射。然后沿着空间方向将 f 分解为 $f^h \in R^{C/r \times H}$ 和 $f^w \in R^{C/r \times W}$ 。利用变换 F_h 和 F_w 分别将 f^h 和 f^w 变换为张量, 得到:

$$g^h = \sigma(F_h(f^h)) \quad (5)$$

$$g^w = \sigma(F_w(f^w)) \quad (6)$$

通常采用适当的缩减比 r 减少通道数来降低模型复杂度和计算开销, 然后将输出 g^h 和 g^w 作为坐标和通道的注意力权重, 坐标注意力模块的输出可以用公式表达为:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

1.3 实验指标

本研究所有实验都使用 Win10 操作系统、Geforce RTX 1650 显卡、CUDA11.0、Anaconda3 的科学计算环境和 Pycharm 的 IDE 开发环境。神经网络是在 Pytorch 框架上进行搭建, 对甲状腺癌病理图像进行训练验证和测试。

深度学习模型在不同任务中的表现需要用定量指标进行实验才能进行横向比较。本研究通过精确率、召回率、特异性、F1-score、混淆矩阵、ROC 曲线和 Kappa 系数进行评估和检测性能^[24-25]。

(1) 混淆矩阵: 混淆矩阵用于观察模型在各类别上精确率和召回率等方面的表现。绘制深度学习模型的混淆矩阵来计算各个类别的特异性、召回率、精确率以及 F1 值, 直观地展现模型的可靠性。以二元分类为例的混淆矩阵结构分为真正类 (TP)、真反类 (TN)、假正类 (FP) 和假反类 (FN)。

准确率 (Accuracy): 准确率是预测正确的图片张数占总测试图片的张数, 适用于二分类和多分类任务, 其公式如下:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

精确率 (Precision): 预测正确图片数相对于真正类预测总数的比例, 其公式如下:

$$P = \frac{TP}{TP + FP} \quad (9)$$

召回率 (Recall): 又称查全率, 确定了正类被预测为正类图片数占所有标注图片的个数, 其计算公式为:

$$R = \frac{TP}{TP + FN} \quad (10)$$

特异度 (Specificity): 负类样本被分类器判断为负类的比例, 其公式如下:

$$S = \frac{TN}{TN + FP} \quad (11)$$

F1-score: 该分类指标主要是针对召回率和精确率之间不一致性, 通过两者的加权调和平均得到。

F1-score 取值范围为[0,1],其中,1表示网络性能最好,0代表网络性能最差。F1-score 公式如下:

$$F1 = \frac{2TP}{2TP + FP + FN}$$

(12)

(2)ROC:ROC 曲线是反映敏感度和特异度的综合指标,曲线越接近左上角表示分类器假阳率低且真阳率高,意味着模型性能越好,公式如下:

$$TPR = \frac{TP}{TP + FN}$$

(13)

$$FPR = \frac{FP}{TP + FN}$$

(14)

(3)Kappa 系数:Kappa 系数是评价分类器一致性的方法之一,用于测定多分类中模型准确率的方法。一般值范围为[0, 1],ROC 曲线不会出现下凸曲线,系数值越大,模型得到的分类精度越高。其公式如下:

$$k = \frac{P_o - P_e}{1 - P_e}$$

(15)

其中, P_o 为总体分类精度。不妨设样本总数为 n ,真实样本分别为 a_1, a_2, \dots, a_i ,预测出来的分别为 b_1, b_2, \dots, b_i ,则 P_e 用公式表示为:

$$P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_i \times b_i}{n \times n}$$

(16)

Kappa 系数按照数值分为 5 个等级:0~0.2 之间为极低;0.21~0.40 为一般;0.41~0.60 为中等;0.61~0.80 为高度;0.81~1.00 为几乎完全一致。

2 实验结果与分析

2.1 参数优化

超参数选择的重点是尽快学习数据的结构,同时保证模型不会对训练数据集过拟合和欠拟合。优化参数一般是指优化批样本数量、学习率、不同优化器的参数等,这些参数能极大地帮助神经网络收敛,提升模型效果。本研究由于数据集的数量限制,实验仅分出部分甲状腺病理图像作为辅助诊断系统测试。其余数据集以 7:2:1 随机划分作为神经网络的训练集、验证集和测试集,其中训练集会进行数据增广扩充。本节将对学习率、卷积核、批样本数量和训练集/验证集等参数进行优化实验。

(1)学习率优化:首先对不同初始学习率 lr 和学习衰减率 lrf 进行比较。由于分级数量过多,因此只计算所有类别的平均值,对比如表 3 所示。从表 3 可以看出第 5 组的精确率、召回率、特异性和 F1 数值明显高于其他 8 组的数据。因此设置初始学习率 lr 为 0.01 和学习衰减率 lrf 为 0.01 能得到效果更好的网络模型,也有研究指出初始学习率设置为 0.01 效果相对更好^[26]。

表3 不同学习率lr和学习衰减率lrf比较
Table 3 Comparison among different learning rates lr and learning decay rates lrf

学习率和 学习衰减率	损失值	精确率	召回率	特异性	F1 得分
lr=0.001;lrf=0.1	0.256	0.930	0.927	0.985	0.930
lr=0.001;lrf=0.01	0.273	0.929	0.928	0.986	0.928
lr=0.001;lrf=0.001	0.263	0.921	0.920	0.984	0.922
lr=0.01;lrf=0.1	0.079	0.950	0.940	0.981	0.945
lr=0.01;lrf=0.01	0.081	0.952	0.951	0.990	0.950
lr=0.01;lrf=0.001	0.087	0.950	0.949	0.989	0.948
lr=0.1;lrf=0.1	0.543	0.813	0.790	0.957	0.768
lr=0.1;lrf=0.01	0.410	0.903	0.899	0.979	0.898
lr=0.1;lrf=0.001	0.382	0.905	0.901	0.980	0.898

(2)批样本数量优化:批样本数量也会影响模型的优化程度和速度。由于本研究所用电脑性能的限制,batch size 取值不能大于 32。batch size 对网络的影响如表 4 所示。从表 4 可以看到,在迭代次数固定时,更大 batch size 的损失值确是最低的,但是精确率达不到最优解。当 batch size=16 时,损失值较小且其他指标数值都在最高点,因此 batch size=16 时,模型的性能最出色。

表4 不同批样本数量的比较
Table 4 Comparison among different batch sizes

批样本数量	损失值	精确率	召回率	特异性	F1 得分
4	0.340	0.902	0.899	0.980	0.898
8	0.247	0.921	0.920	0.984	0.918
16	0.132	0.962	0.961	0.992	0.960
32	0.102	0.944	0.944	0.989	0.942

(3)其他优化:有些参数在不建议训练模型的时候进行修改,例如 Adam 优化器中的 β_1 设置为 0.9, β_2 设置为 0.999。除了上述参数外,还有一些常规参数需要研究者优化调节,比如说卷积核大小、训练集/验证集比率。①卷积核大小:将大小为3×3、5×5 和 7×7 的卷积核进行横向比较,对比结果如表 5 所示。在损失值相差不大的情况的,首选精确率高的 7×7 卷积核。②迁移学习:迁移学习对大多数网络都有显著的增益效果,提高网络的学习效率。表 6 为迁移学习对深度学习网络的影响。本实验使用在 ImageNet 上训练出的预训练权重导入卷积神经网络中,以减少训练上万张甲状腺癌病理数据集的时间。两个实验

的各类指标数值接近,但是加入了迁移学习的训练迭代次数是没有加入迁移学习训练迭代次数的四分之一。

表5 不同卷积核比较
Table 5 Comparison among different convolutional kernels

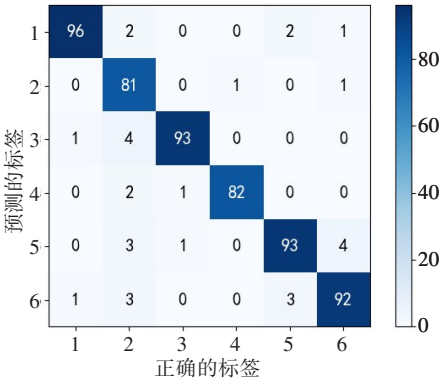
卷积核尺寸	损失值	精确率	召回率	特异性	F1得分
3×3	0.130	0.953	0.952	0.991	0.952
5×5	0.136	0.958	0.958	0.992	0.957
7×7	0.132	0.962	0.961	0.992	0.960

表6 迁移学习对比
Table 6 Effect of transfer learning on the network

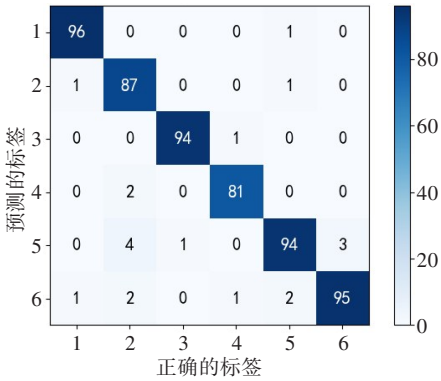
是否使用 迁移学习	迭代次数	精确率	召回率	特异性	F1得分
是	50	0.962	0.961	0.992	0.960
否	200	0.966	0.964	0.993	0.965

表7 各项数据比较
Table 7 Comparison of diagnostic efficiency

癌症级别	EfficientNet-B0				CA-EfficientNet			
	精确率	召回率	特异性	F1得分	精确率	召回率	特异性	F1得分
I	0.960	0.969	0.991	0.960	0.990	0.980	0.998	0.980
II	0.936	0.926	0.987	0.930	0.978	0.916	0.996	0.950
III	0.979	0.989	0.996	0.980	0.989	0.989	0.998	0.990
IV	0.988	0.964	0.998	0.980	0.976	0.976	0.996	0.980
V	0.940	0.959	0.987	0.950	0.922	0.959	0.983	0.940
VI	0.969	0.959	0.994	0.960	0.941	0.969	0.987	0.950
平均	0.962	0.961	0.992	0.960	0.966	0.968	0.993	0.965



a: EfficientNet-B0 的混淆矩阵



b: CA-EfficientNet 的混淆矩阵

图5 混淆矩阵

Figure 5 Confusion matrix

2.2 实验结果

基于2.1节进行参数优化后,针对EfficientNet-B0模型和改进后的CA-EfficientNet模型对甲状腺癌病理图像数据集进行训练测试,其训练过程中精确率、召回率、特异性和F1数据变化如表7所示,CA-EfficientNet模型在单个参数有的不如EfficientNet-B0模型,但是大部分数据的数值更大,且平均数都超过了EfficientNet-B0模型。从上述对比可以看出,CA-EfficientNet模型在甲状腺癌病理图像数据集上的表现更好。

混淆矩阵用于检验分类器分类的准确程度。本实验采用通过混淆矩阵计算得出网络各类评价指标对模型性能进行评估,根据EfficientNet-B0和CA-EfficientNet模型所生成的混淆矩阵。混淆矩阵表格的正对角上是模型正确识别的数量,正对角上的数据占比越大模型性能越好。从图5对比可以看出,CA-EfficientNet以微弱的数量优势超过EfficientNet-B0,模型具有更好的性能。

ROC 曲线也是衡量模型的一种重要方式。图 6 是 EfficientNet-B0 和改进后的 CA-EfficientNet 基于本研究数据集进行训练过程中的 ROC 曲线对比。ROC 曲线越接近左上角表示分类器假阳率低且真阳率高,换句话说就是模型将甲状腺癌病理数据集中第 i 类中,其他类

别识别为第 i 类的概率低;在第 i 类中,识别为第 i 类的概率高。从图 6 可以看出,这两个模型性能都比较好,但改进后的 CA-EfficientNet 在总的 ROC 曲线下面积比 EfficientNet-B0 总的 ROC 曲线下面积更大,因此 CA-EfficientNet 表现更优秀。

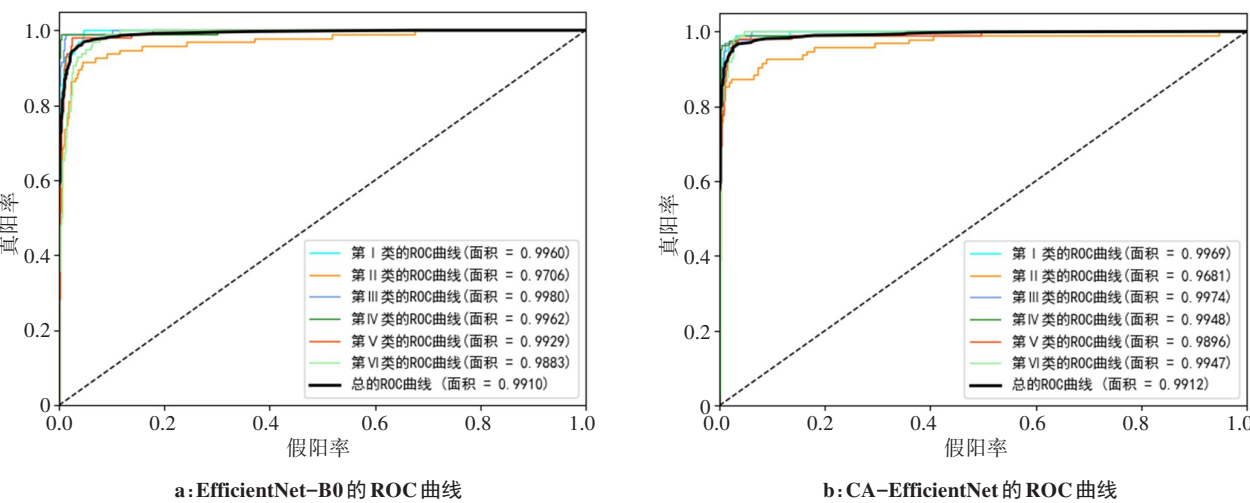


图 6 ROC 曲线
Figure 6 ROC curves

EfficientNet-B0 的 Kappa 系数为 0.950, CA-EfficientNet 的 Kappa 系数为 0.958。两个模型的分数非常高且接近,两个模型的预测结果和实际分级结果一致性极高,但 CA-EfficientNet 的 Kappa 得分比 EfficientNet-B0 高了 0.008。

2.3 辅助诊断系统测试

人工智能融入甲状腺癌病理学医生繁琐的分析工作,不仅能减轻病理医生的工作量,还能提高甲状腺癌诊断的准确性和客观性。本节将以甲状腺细针穿刺活检病理的癌细胞分级算法为核心,参考相关量表,测试深度学习算法模型的准确率,并根据实验结果,确定甲状腺结节临床辅助诊断系统的可靠性。flask 框架是一种小巧便捷的 Web 框架,不仅可基于 Python 语言快速编写搭建小型网页,也可作为 APP 的 API,拓展性极高且第三方库的选择面广,开发时可轻松结合最流行最强大的第三方库实现开发需求。使用 Pycharm 编译器,基于 Python 语言,得到 flask 可视化界面。受限于甲状腺癌病理图像数量的限制,能用于辅助系统测试的补丁级图像数量每一级不超过 20 张,总共 112 张。在 flask 可视化界面中(图 7),随机挑选一张照片进行预测,图中预测结果和病理医生预测结果一致。分别对剩下的甲状腺病理图像进行测试,从表 8 可以看出,对比病理医生的诊断结果,辅助诊断系统仅将一张 IV

级图像误诊断为 III 级,误诊率为 0.9%,但单次检测时间可以忽略不计,而病理医生 1~2 s 才能判断一张补丁级病理图像的分级结果,证明了基于深度学习的甲状腺癌病理图像分级辅助诊断系统的可靠性。



图 7 flask 可视化界面
Figure 7 flask visualization interface

表 8 医生和辅助诊断系统对比
Table 8 Comparison of diagnostic results between physician and auxiliary diagnosis system

诊断方法	甲状腺细针穿刺涂片 Bethesda 分类					
	I	II	III	IV	V	VI
医生诊断结果	20	18	18	16	20	20
辅助诊断系统结果	20	18	19	15	20	20

3 结 论

本研究主要研究利用甲状腺癌病理细胞的WSI,将诊断细胞学的Bethesda分级报告系统转换为计算机能识别的自然特征分级标准,证明深度学习网络能学习到该细胞特征,再建立以EfficientNet网络为核心的甲状腺癌病理图像自动分级辅助诊断系统。对EfficientNet-B0网络改进,加入CA注意力机制和ELU激活函数,ELU激活函数计算相对简单,对负输入有很好的鲁棒性;CA注意力机制利用对通道关系和精确位置信息两个方向的特征聚合,增强神经网络对细胞位置的关注,使得模型在训练过程中学习到更多的细胞特征,比EfficientNet-B0网络提升了0.4%的精确率。

综上所述,CA-EfficientNet网络模型的精确率达到96.6%,基于该算法的辅助诊断系统和经验丰富的病理医生的诊断结果不相上下,证明了基于CA-EfficientNet网络的甲状腺癌病理图像自动分级算法的先进性,辅助诊断系统具有实际应用性,在减轻病理医生工作负担的同时降低人工诊断因疲劳和主观因素的误诊率。本研究方法对甲状腺癌的病理图像分级精度很好,但仍有很多不足的地方:首先,本研究简化了甲状腺癌病理Bethesda分级报告系统的内容,去掉了一些不常见的细胞;其次,采集数据集分为3个步骤:①甲状腺细针穿刺获取样本;②一系列操作将病理样本制成病理玻片;③用专业的扫描仪将病理样本扫描为高分辨率图像。但这显然是极其繁琐且耗费时间和人力的,因此下一步研究希望能在第一步穿刺获取样本之后就能用辅助诊断系统诊断出可靠的结果,实现当场采样当场出结果。

【参考文献】

- [1] Kitahara CM, Sosa JA. The changing incidence of thyroid cancer[J]. Nat Rev Endocrinol, 2016, 11(12): 646-653.
- [2] Fitzmaurice C, Allen C, Dicker DJ, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease Study [J]. JAMA Oncol, 2017, 3(4): 524.
- [3] 陈健, 江亮. 精确病理诊断在甲状腺癌精准医疗中的意义[J]. 中国肿瘤临床, 2017, 44(4): 181-185.
Chen J, Jiang L. Accurate pathological diagnosis of thyroid cancer in the era of precision medicine [J]. Chinese Journal of Clinical Oncology, 2017, 44(4): 181-185.
- [4] 蔡晓频, 徐建红, 靳霞, 等. 超声引导下甲状腺结节粗针穿刺活检临床应用价值[J]. 中华医学超声杂志(电子版), 2012, 9(7): 641-644.
Cai XP, Xu JH, Jin X, et al. A clinical study of ultrasonography-guided core-needle biopsy of thyroid nodules [J]. Chinese Journal of Medical Ultrasound (Electronic Edition), 2012, 9(7): 641-644.
- [5] 王文涵, 詹维伟, 周伟, 等. 超声引导下细针抽吸活检对甲状腺小结节的诊断价值[J]. 中华医学超声杂志(电子版), 2014, 11(8): 12-15.
Wang WH, Zhan WW, Zhou W, et al. The diagnostic value of ultrasound-guided fine needle aspiration biopsy in small thyroid nodules [J]. Chinese Journal of Medical Ultrasound (Electronic Edition), 2014, 11(8): 12-15.
- [6] 田文, 姚京. 重视甲状腺结节规范化诊治[J]. 中国实用外科杂志, 2015, 35(6): 579-583.
Tian W, Yao J. Paying attention to the standardization of diagnosis and treatment of thyroid nodules [J]. Chinese Journal of Practical Surgery, 2015, 35(6): 579-583.
- [7] 田文, 孙辉, 贺青卿. 超声引导下甲状腺结节细针穿刺活检专家共识及操作指南(2018版)[J]. 中国实用外科杂志, 2018, 38(3): 241-244.
Tian W, Sun H, He QQ. Ultrasound-guided expert consensus and operating guidelines for fine needle needle biopsy of thyroid nodules (2018 edition) [J]. Chinese Journal of Practical Surgery, 2018, 38(3): 241-244.
- [8] Jara-Lazaro AR, Thamboo TP, Teh M, et al. Digital pathology: exploring its applications in diagnostic surgical pathology practice [J]. Pathology, 2010, 42(6): 512-518.
- [9] Bera K, Schalper KA, Rimm DL, et al. Artificial intelligence in digital pathology-new tools for diagnosis and precision oncology [J]. Nat Rev Clin Oncol, 2019, 16(11): 703-715.
- [10] Balazs Acs M, David L, Rimm MP. Not just digital pathology, intelligent digital pathology [J]. JAMA Oncol, 2018, 4(3): 403-404.
- [11] Barisoni L, Lafata KJ, Hewitt SM, et al. Digital pathology and computational image analysis in nephropathology [J]. Nat Rev Nephrol, 2020, 16(11): 669-685.
- [12] Williams BJ, Bottoms D, Treanor D. Future-proofing pathology: the case for clinical adoption of digital pathology [J]. J Clin Pathol, 2017, 70(12): 1010-1018.
- [13] Wang L, Ding L, Liu Z, et al. Automated identification of malignancy in whole-slide pathological images: identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning [J]. Br J Ophthalmol, 2020, 104(3): 318-323.
- [14] Ibrahim A, Gamble P, Jaroensri R, et al. Artificial intelligence in digital breast pathology: techniques and applications [J]. Breast, 2020, 49: 267-273.
- [15] Iizuka O, Kanavati F, Kato K, et al. Deep learning models for histopathological classification of gastric and colonic epithelial tumours [J]. Sci Rep, 2020, 10(1): 1504.
- [16] Syed A, Edmunds C. 甲状腺细胞病理学Bethesda报告系统: 定义、标准和注释 [M]. 北京: 北京科学技术出版社, 2010.
Syed A, Edmunds C. The Bethesda system for reporting thyroid cytopathology: definitions, criteria and explanatory notes [M]. Beijing: Beijing Science and Technology Press, 2010.
- [17] 曲澜涛. 超声引导下粗针穿刺活检鉴别诊断甲状腺结节良恶性的分析 [J]. 影像研究与医学应用, 2020, 4(5): 115-116.
Qu LT. Analysis of differential diagnosis of benign and malignant thyroid nodules by ultrasound-guided crude needle needle biopsy [J]. Journal of Imaging Research and Medical Applications, 2020, 4(5): 115-116.
- [18] Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks: in proceedings of the international conference on machine learning [C]. 2019.
- [19] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(8): 2011-2023.
- [20] Zhao C, Shuai R, Ma L, et al. Improving cervical cancer classification with imbalanced datasets combining taming transformers with T2T-ViT [J]. Multimed Tools Appl, 2022: 1-36.
- [21] Ge R, Shen T, Zhou Y, et al. Convolutional squeeze-and-excitation network for ECG arrhythmia detection [J]. Artif Intell Med, 2021, 121: 102181.
- [22] Roy AG, Navab N, Wachinger C. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks [J]. IEEE Trans Med Imaging, 2019, 38(2): 540-549.
- [23] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [24] Downey TJ, Meyer DJ, Price RK, et al. Using the receiver operating characteristic to assess the performance of neural classifiers [C]// International Joint Conference on Neural Networks. 1999: 3642-3646.
- [25] Tang W, Hu J, Zhang H, et al. Kappa coefficient: a popular measure of rater agreement [J]. Shanghai Arch Psychiatry, 2015, 27(1): 62-67.
- [26] Smith LN. Cyclical learning rates for training neural networks [C]. 17th IEEE Winter Conference on Applications of Computer Vision (WACV). Santa Rosa: IEEE 2017.

(编辑:陈丽霞)