

## 基于深度迁移学习的肠道息肉癌变风险评估<sup>①</sup>

程 珊<sup>②\*</sup> 叶 然<sup>③\*\*</sup> 曹 婧\* 姚佳锋\* 李 胜\* 何熊熊\*

(\* 浙江工业大学信息工程学院 杭州 310023)

(\*\* 丽水市人民医院 丽水 323400)

**摘 要** 针对肠道息肉分类中不同类型息肉难以区分的问题,提出了一种基于深度迁移学习的肠道息肉癌变风险评估方法。该方法首先在主干网络中引入通道注意力进行息肉特征提取;然后将任务网络注意力图分支输出的注意力图加权到主干网络输出的特征图上,获取更精确的病灶特征;最后将加权后的特征图输入到任务网络的感知分支,使用 2 个分支的损失函数以端到端的方式优化模型,以进行息肉癌变风险评估。本文方法对数据集进行了增强,并结合迁移学习进一步提升分类效果,最优模型的灵敏度、特异度、精准度、*F1* 分数及曲线下方的面积(AUC)分别为 99.38%、98.15%、98.17%、98.77% 和 0.9996。实验结果表明,本文方法能有效地对肠道息肉癌变风险进行评估,性能优于该领域流行的深度学习方法。

**关键词** 深度学习;迁移学习;通道注意力;注意力图;图像分类;肠道息肉

## 0 引言

结直肠癌是最常见和致命的癌症之一,根据《2020 年世界癌症报告》,结直肠癌已成为发病率第 3 的癌症<sup>[1]</sup>。肠道息肉是引发早期结直肠癌变的主要肠道疾病之一,且不同类型息肉的癌变风险不同。筛查肠道息肉和评估其癌变风险对于预防早期结直肠癌很重要,若筛查出高风险息肉应当尽快手术切除。然而做结肠镜检查时,医生凭借肉眼对息肉分类的准确性与医生的经验很相关,这种人工操作难度较大、费时耗力,且易造成漏诊和误诊<sup>[2]</sup>。因此,需要开发计算机辅助诊断技术来实现息肉癌变风险评估,以协助医生对癌变风险不同的息肉进行分类。

传统的息肉分类方法通常基于特征提取和分类器训练。Stehle 等人<sup>[3]</sup>利用腺瘤和增生息肉表面血管特征的区别来对息肉分类,通过相位对称和快速行进相结合对血管进行分割,以计算血管长度、血管

中心线处的平均灰度和平均血管周期。Tamaki 等人<sup>[4]</sup>采用高斯差分和网络采样等方案提取局部特征,使用表示局部特征的视觉单词包和支持向量机组成的识别系统对息肉进行分类。Mesejo 等人<sup>[5]</sup>利用二维颜色特征、二维纹理特征和三维形状特征对增生性病变、锯齿状腺瘤和腺瘤进行分类。传统方法基于手工特征工程,需要大量的工程技能,反复微调多个参数,且模型泛化能力差、适应性低<sup>[6]</sup>。

深度学习<sup>[7]</sup>在息肉分类任务中已具有出色的表现。Zhang 等人<sup>[8]</sup>开发了一种用于息肉检测和分类的自动系统,使用迁移学习,利用深度卷积神经网络从大型非医学数据集中学习到的特征参数来对息肉分类。Shaban 等人<sup>[9]</sup>提出了一种新的大肠组织学图像上下文感知学习框架。但这些深度学习方法或忽略了局部病变区域的细微特征,或依赖组织学数据对肠道息肉进行分类。因为医学图像采集困难,标记的数据集较小,息肉类内病灶形状差异大,

① 国家自然科学基金(61873239)和浙江省教育厅一般科研(Y202043133)资助项目。

② 女,1997 年生,硕士生;研究方向:智慧医疗;E-mail:2268980570@qq.com。

③ 通信作者,E-mail:849221400@qq.com。

(收稿日期:2021-09-14)

病变区域与正常组织之间高度相似,模型性能通常受到限制。针对上述问题,本文使用数据增强对原始数据集进行了扩展,基于深度学习,提出了引入通道注意力的主干网络用于息肉特征提取,使用注意力图聚焦局部病灶区域,并结合迁移学习来进一步提高息肉分类的效果,从而实现利用结肠内窥镜图像有效地对息肉癌变风险进行评估。

## 1 方法

### 1.1 数据增强

肠道息肉常见的有炎症性息肉、幼年性息肉、增生性息肉和腺瘤性息肉 4 种类型。各类型息肉的癌变风险不同,炎症性息肉、幼年性息肉和增生性息肉几乎不发生癌变,属于低风险息肉,腺瘤性息肉癌变风险较高。幼年性息肉可通过患者年龄来判别,因此本方法不对幼年性息肉进行计算机辅助分类。建立的结肠内窥镜息肉图像数据集包括炎症性息肉、增生性息肉和腺瘤性息肉 3 种类型的息肉图像,其中腺瘤性息肉分为管状腺瘤和绒毛状腺瘤。如图 1 所示。

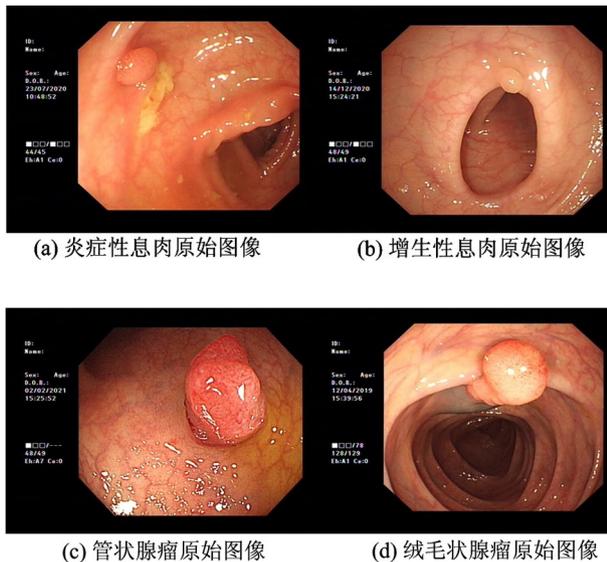


图 1 不同类型息肉原始图像

原始息肉图像周围的黑色区域含有实验并不需要的注释信息,使用 mask 技术盖住了这些无用的注释信息,再将图像大小裁剪为  $560 \times 480$  像素。由于原始数据集的样本数量较小,且各类别样本数量不

平衡,结肠内窥镜在拍摄肠道时角度随机,息肉图像有时因为曝光而过亮,有时因为光照不足过暗、对比度低,有时有泡泡、杂质等,因此本方法对原始息肉图像采用了亮度调整、对比度调整、随机翻转、随机旋转、加高斯噪声等一系列基本的数据增强操作,扩展了息肉样本数量,使类别的样本数量平衡,使训练的模型能够对质量较差的息肉图像进行分类。如图 2 所示,图 2(a)是遮盖无用信息并裁剪后的息肉图像,图 2(b)是亮度调整后的息肉图像,图 2(c)是对比度调整后的息肉图像,图 2(d)是随机翻转后的息肉图像,图 2(e)是随机旋转后的息肉图像,图 2(f)是加高斯噪声后的息肉图像。

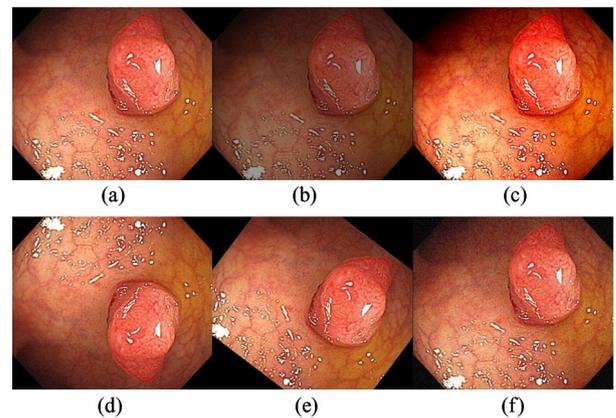


图 2 增强后的息肉图像

### 1.2 ResNet50 卷积神经网络

ResNet50<sup>[10]</sup>由很多个相似的残差块串联而成,残差块结构如图 3 所示,输入  $x$  经过残差路径的 3 个卷积层后输出  $F(x)$ ,再将  $F(x)$  与输入  $x$  相加,经过 ReLU 激活后得到整个残差块的输出。

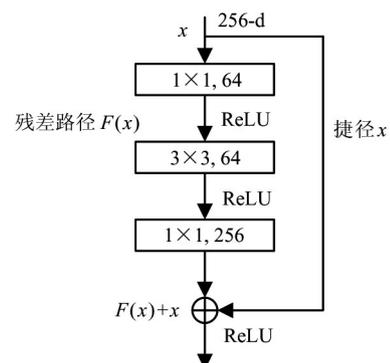


图 3 残差块结构

ResNet50 结构如图 4 所示,网络第 1 层是步长

为2、通道数为64的 $7 \times 7$ 卷积层,降采样后经过BN(batch normalization)层和ReLU激活层;第2层是步长为2的 $3 \times 3$ 最大池化层;接着就是分别具有3、4、6、3个残差块的卷积层 Conv2\_x、Conv3\_x、Conv4\_x、Conv5\_x,其中 Conv2\_x 的步长为1,其余卷积层的步长均为2;最后经过 $7 \times 7$ 平均池化层(average pooling, Avgpool)和全连接层(fully connected layers, FC)后输出类别。ResNet50在训练的

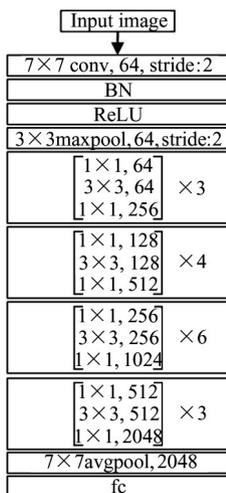


图4 ResNet50 结构

后向传播时更好地将深层梯度传到上一层,利于构建更深的网络。

### 1.3 基于 ResNet50 改进的网络结构

ResNet50在图像领域取得了显著成果,但是应对复杂的息肉图像进行分类时性能仍有提升空间。因此,针对不同类型息肉图像难以区分的特点,本方法提出了一种基于ResNet50改进的网络结构,用于对肠道息肉癌变风险的评估。

如图5所示,本方法改进的网络结构包括主干网络和任务网络2部分,主干网络基于ResNet50的前40层卷积层,任务网络包括2个并行执行任务的分支结构:注意力图分支和感知分支。本方法在整个网络的残差块结构中引入了通道注意力,同时将注意力图在特定通道方向上与主干网络输出的特征图进行加权,从而为网络引入空间注意力。本方法结合了通道注意力机制和空间注意力机制,能够更加准确地聚焦于息肉病灶区域。此外,生成的注意力图实现了模型对息肉区域聚焦作用的可视化,为网络做出的决策提供了直观的依据,同时能够在临床中对医生进行病灶位置提示。

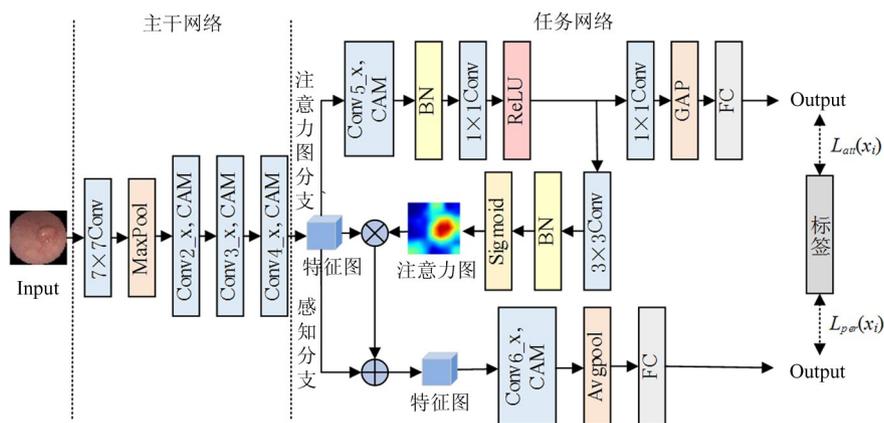


图5 整体模型网络结构

主干网络的输入为 $224 \times 224$ 像素的结肠内窥镜息肉图像,其相当于一个特征提取器,用于息肉图像的细微特征提取,以克服不同类型息肉的类内病灶形状、大小差异大的问题,同时输出特征图。主干网络的结构基于改进的ResNet50前40层卷积层,首先是 $7 \times 7$ 卷积层、BN层、ReLU激活层和最大池化层MaxPool,接下来就是引入通道注意力模块

(channel attention module, CAM)的残差块卷积层 Conv2\_x、Conv3\_x、Conv4\_x,它们分别具有3、4、6个引入CAM的残差块。通过在这些残差块中引入CAM,使主干网络能够自动学习特征图中各通道的权重,关注重要的特征,抑制不相关的特征,以便主干网络更有效地提取特征。

任务网络的注意力图分支接收来自主干网络输

出的特征图,并通过引入基于响应的视觉解释模型<sup>[11]</sup>来生成注意力图。注意力图分支首先将特征图输入到 Conv5\_x 以提取更深层的特征,Conv5\_x 具有 3 个引入 CAM 的残差块,经过 BN 层、1×1 卷积层后进行 ReLU 激活,ReLU 激活后分为 2 个小分支:一个小分支经过 3×3 卷积层、BN 层后通过 Sigmoid 激活输出注意力图——可视化图像中具有高度响应的区域,从而聚焦于病灶;另一个小分支经过 1×1 卷积层、14×14 全局平均池化层(global average pooling, GAP)和全连接层(FC)后,对息肉癌变风险高低的类别进行判断。本方法利用注意力机制的思想,将注意力图分支生成的注意力图在特定通道上与主干网络输出的特征图进行点积,再与特征图相加得到新特征图,即将注意力图分支已学习到的空间方向上像素的权值加到主干网络输出的特征图上,相当于给网络引入了空间注意力。

感知分支类似于一个分类器,将特征图通过具有 3 个引入 CAM 的残差块卷积层 Conv6\_x、7×7 平均池化层(Avgpool)、FC 后,对息肉癌变风险高低的类别进行判断。使用注意力图分支和感知分支 2 个分支损失函数的加权和作为总训练损失函数,来对整个模型进行端到端的优化训练,克服不同类型息肉相似性大的问题,以提高模型对肠道息肉癌变风险评估的性能。

#### 1.4 通道注意力残差模块

近年来,注意力机制已被广泛应用于计算机视觉应用的深度学习中,其核心思想是模拟人类视觉的感知机制,聚焦特征图中最重要的局部,从而消除计算机视觉应用中的冗余部分<sup>[12]</sup>。因此,为聚焦结肠内窥镜息肉图像中的息肉部分,本方法提出了一种有效的通道注意力机制,它使网络自动确定通道方向特征的重要性,过滤性地选择特征,使网络能够更有效地提取特征,且能以较低的计算量成本提高网络的分类性能,其结构如图 6 所示。

在图 6 中,将图像输入到残差块,经过 1×1 卷积、3×3 卷积、1×1 卷积后输出特征图  $F_i$ ,其中  $C$  表示特征图的通道数目, $H$  和  $W$  分别代表特征图的高和宽。将  $F_i$  作为输入,经过 CAM 后输出  $C \times 1 \times 1$  的通道注意力因子(channel attention factors, CAF)。

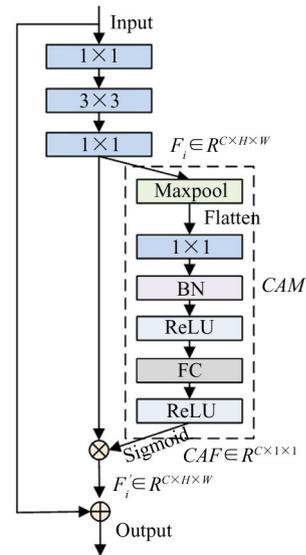


图 6 具有通道注意力的残差模块

在 CAM 中,输入的特征图首先通过全局 MaxPool 进行下采样,以保留最重要的信息,并降低计算量。接着将下采样后的特征图展平为一维向量,再经过一层 1×1 卷积,以从向量中提取特征并调整其维数。然后使用 BN 层和 ReLU 激活层对提取的特征进行处理,以提高模型训练的收敛速度和非线性表示能力。最后,经过一层全连接层 FC、ReLU 激活和 Sigmoid 激活函数输出  $C \times 1 \times 1$  的 CAF,CAF 的值范围在 0~1 之间,其大小反映了通道的重要性。经过 CAM 后,将 CAF 与特征图  $F_i$  逐元素相乘进行融合,输出融合后的特征图  $F_i'$ ,其过程表示为式(1)和式(2),最后将  $F_i'$  与残差块的输入相加后作为输出,再进入到下一个具有 CAM 的残差块。

$$CAF = CAM(F_i) \quad (1)$$

$$F_i' = CAF \otimes F_i \quad (2)$$

#### 1.5 迁移学习

本方法基于监督学习对结肠内窥镜息肉图像进行分类。监督学习需要大量标注过的数据,而原始结肠内窥镜息肉图像的获取代价大、采集困难,所以标记的息肉图像数据集通常较小。虽然卷积神经网络具有很强的表征能力,但由于数据集规模的受限,模型的性能依然会受到限制。迁移学习可以用来解决这个问题,迁移学习将神经网络在大规模源数据集上已训练好的模型(预训练模型)参数复制到目标网络中,来帮助新的模型训练<sup>[13-14]</sup>,其过程示意

图如图 7 所示。

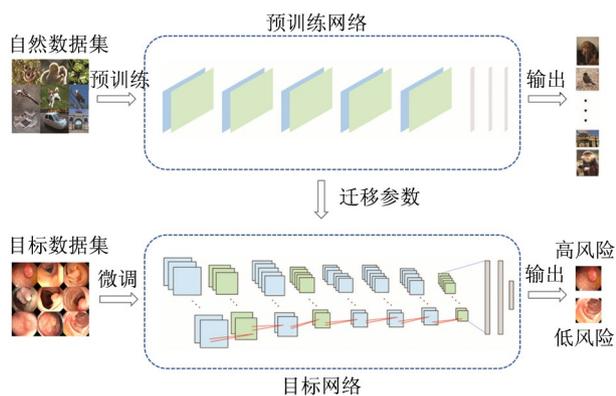


图 7 迁移学习

在图 7 中,预训练网络是已经在 ImageNet 自然数据集上训练好的 ResNet50 分类网络模型,其在 ImageNet 数据集上学到的一些空间特征可以有效地帮助学习息肉图像中与自然图像共有的特征。本研究迁移了 ResNet50 预训练模型的参数,去初始化与本研究方法网络共有层的权重,从而加快并优化了网络的学习收敛效率,得到了性能更优的肠道息肉癌变风险评估模型。

### 1.6 损失函数

本文方法同时使用任务网络的注意力图分支和感知分支 2 个分支的损失函数加权和以端到端的方式对模型进行优化训练。每个分支的训练损失由结合标签平滑<sup>[15]</sup>的交叉熵损失来计算,交叉熵损失函数  $L$  表达为式(3),其中  $N$  表示总样本数量,  $y_i$  表示样本  $i$  的标签,  $p_i$  表示样本  $i$  预测为正类的概率。分类问题中,常认为训练数据中标签向量的目标类别概率应为 1,非目标类别概率应为 0。传统 one-hot 编码的标签向量  $y_i$  为式(4)。标签平滑是一种正则化方法,它结合均匀分布,用新的标签向量  $\hat{y}_i$  来代替  $y_i$ ,  $\hat{y}_i$  表示为式(5),其中  $\varepsilon$  是一个很小常量,本方法经过实验设置  $\varepsilon = 0.05$ ,  $K$  是类别数量,本文  $K = 2$ 。标签平滑后的分布相当于给真实分布加入噪声,避免模型对正确标签过于自信,降低对正负样本的预测差别,从而可以显著地提升神经网络的学习速度和泛化能力,避免过拟合。

$$L = \frac{1}{N} \sum_i - [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (3)$$

$$y_i = \begin{cases} 1 & i = target \\ 0 & i \neq target \end{cases} \quad (4)$$

$$\hat{y}_i = \begin{cases} 1 - \varepsilon & i = target \\ \varepsilon / (K - 1) & i \neq target \end{cases} \quad (5)$$

总训练损失函数  $L(x_i)$  是注意力图分支和感知分支 2 个分支损失函数的加权和,其公式表示为式(6)。在反向传播过程中,通过传递注意力图分支和感知分支的梯度来优化整个模型。用  $L_{att}(x_i)$  表示输入样本为  $x_i$  时的注意力图分支损失,  $L_{per}(x_i)$  表示输入样本为  $x_i$  时的感知分支损失,经过实验,设置  $\alpha$  为 0.1,  $\beta$  为 0.9 时模型最优。

$$L(x_i) = \alpha L_{att}(x_i) + \beta L_{per}(x_i) \quad (6)$$

## 2 实验

### 2.1 实验整体流程

实验的整体流程如图 8 所示,首先建立包括训练集、验证集和测试集的数据集,并通过数据增强对数据集进行扩展,搭建基于深度迁移学习的两分支网络框架,在训练集上学习权重建立模型,通过验证集来调整模型参数,在测试集上测试已经训练好的模型的性能。

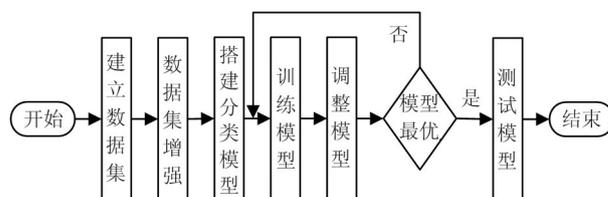


图 8 实验流程

### 2.2 数据集

#### 2.2.1 癌变风险数据集

从丽水市人民医院消化内科 405 个病例收集了 405 张原始结肠内窥镜图像,所有受试者均签署了知情同意书,所有图像均由经验丰富的医生进行了手动标注,并按病理活检结果进行了分类。由于炎症性息肉和增生性息肉几乎不发生癌变,腺瘤性息肉有一定的癌变风险,所以特别需要对腺瘤性息肉进行判别。因此,将管状腺瘤和绒毛状腺瘤划分为高风险,炎症性息肉和增生性息肉划分为低风险。这样原始数据集有 135 张高风险和 270 张低风险息

肉图像,将原始数据集按照 3:1:1 的比例随机划分为训练集、验证集和测试集,即原始数据集的训练集包含 81 张高风险和 162 张低风险息肉图像,验证集和测试集均包含 27 张高风险和 54 张低风险息肉图像,这 3 个子数据集互相独立不重复。利用数据增强分别对原始数据集的 3 个子数据集进行了扩展。由于类别不平衡,原始高风险数据集扩展了 12 倍,原始低风险数据集扩展了 6 倍,数据集信息见表 1。

### 2.2.2 Kvasir 公开数据集

为进一步验证本方法所提出网络的泛化能力和稳定性,本文从最大的公开胃肠图像数据集——Kvasir 公开数据集中选取了一些正常图像和息肉图像作为分类数据集,以进行对比实验。该数据集包括 1000 张正常图像和 1000 张息肉图像,也按照 3:1:1 的比例随机地划分为训练集、验证集和测试集,数据集信息见表 1。

表 1 数据集的统计信息

数据集	类别	训练集	验证集	测试集
癌变风险数据集	高风险	972	324	324
	低风险	972	324	324
Kvasir 公开数据集	正常	600	200	200
	息肉	600	200	200

### 2.3 实验细节

本文实验基于 1.6.0 + cu101 版 PyTorch 深度学习框架和 Python 3.7 编程语言。硬件配置为 Intel(R) Xeon(R) Silver 4210R CPU@2.40 GHz 2.39 GHz(2 处理器),64 GB 内存,3 块 NVIDIA GeForce RTX 2080 SUPER GPU 并行计算,提高训练速度。本实验训练模型时批尺寸(batch size)设为 32,使用 Adam 优化器来优化模型参数,其参数为  $lr = 0.0001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $eps = 1e-8$ ,并使用 StepLR 机制来调整学习率,实验对模型进行了 100 迭代(epochs)训练。训练前,将所有读入的图像的像素尺寸均调整为  $224 \times 224$  像素后再输入到网络。

为了评估网络的分类性能,本研究使用常应用在医学图像二分类算法的评价指标——灵敏度(sensitivity)、特异度(specificity)、精准度(precision)以及 F1 分数来定量评估模型的性能。这 4 个评价

指标的计算方式表示如下。

$$sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$specificity = \frac{TN}{TN + FP} \times 100\% \quad (8)$$

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (9)$$

$$F1-score = \frac{2 \cdot precision \cdot sensitivity}{precision + sensitivity} \quad (10)$$

其中,真阳性(true positive, TP)表示被模型预测为阳性的阳性样本,真阴性(true negative, TN)表示被模型预测为阴性的阴性样本,假阳性(false positive, FP)表示被模型预测为阳性的阴性样本,假阴性(false negative, FN)表示被模型预测为阴性的阳性样本。此外,本方法使用了受试者工作特征(receiver operating characteristic, ROC)曲线下方的面积(area under the ROC curve, AUC)来进一步评估癌变风险评估模型的准确性。

### 2.4 实验结果与分析

#### 2.4.1 样本扩展前后对比

训练卷积神经网络时,训练集的样本数量越多,所得模型测试的性能会更好,本方法采用了一系列基本的数据增强技术对数据集进行了扩展,并对数据增强前后进行了对比实验分析,结果见表 2。实验结果表明,经过数据增强后,模型的灵敏度、精准度和 F1 分数得到了有效的提升,因此在后续不同模型性能的对比实验中,本文使用扩展后的癌变风险数据集。

表 2 在癌变风险数据集上的对比结果

模型	数据	灵敏度	特异度	精准度	F1 分数
本文	原始数据集	92.59%	98.15%	96.15%	94.34%
方法	扩展数据集	99.38%	98.15%	98.17%	98.77%

#### 2.4.2 在癌变风险数据集上的消融对比

表 3 展示了本方法所提出的基于 ResNet50 改进的两分支网络 T-Res50、引入通道注意力 CAM、使用标签平滑 Label-smooth、结合迁移学习后模型的消融实验性能比较。从实验结果可以发现,在引入 CAM 后模型的灵敏度、特异度、精准度和 F1 分数分别提升了 1.86%、1.85%、1.91% 和 1.88%;引入

CAM 并使用标签平滑后,模型的灵敏度、精准度、F1 分数分别继续提升了 2.47%、0.31% 和 1.41%;经过迁移学习后,各项指标比随机初始化的模型都提升了 10% 以上。可见,利用大规模自然图像预训练的网络参数初始化网络,可以进一步提升模型性能,有效地改善深度学习模型因为数据不足而导致性能低的问题。

表 3 在癌变风险数据集上的消融对比

模型	灵敏度	特异度	精准度	F1 分数
T-Res50	83.02%	86.11%	85.67%	84.33%
CAM + T-Res50	84.88%	87.96%	87.58%	86.21%
CAM + label-smooth + T-Res50	87.35%	87.96%	87.89%	87.62%
本文方法	99.38%	98.15%	98.17%	98.77%

表 4 展示了本文方法在癌变风险数据集的测试集上生成的混淆矩阵,在 324 张高风险息肉样本中,有 322 张得到正确的分类;在 324 张低风险息肉样本中,有 318 张得到正确的分类。

表 4 在癌变风险数据集上分类结果的混淆矩阵

混淆矩阵	预测标签		
	高风险	低风险	
真实标签	高风险	322	2
	低风险	6	318

图 9 展示了本文方法任务网络的注意力图分支在癌变风险数据集的测试集上生成的注意力图,可视化了注意力机制的效果。图 9(a) 是高风险息肉原图,图 9(b) 是高风险息肉的注意力图,图 9(c) 是低风险息肉原图,图 9(d) 是低风险息肉的注意力图。所有息肉原图的像素为  $560 \times 480$ ,生成的注意力图像素为  $224 \times 224$ 。从图 9 中的注意力图可以看出,本方法的注意力图分支可以较准确地聚焦于图像中的息肉病灶区域,高响应区域的边界也比较清晰,将这些注意力加权到主干网络输出的特征图上再输入到感知分支,可以进一步地提高网络的分类效果。从图 9(d) 中可以看出,对于体积较小的低风险癌变息肉图像,本网络也能较好地聚焦其病灶区域。

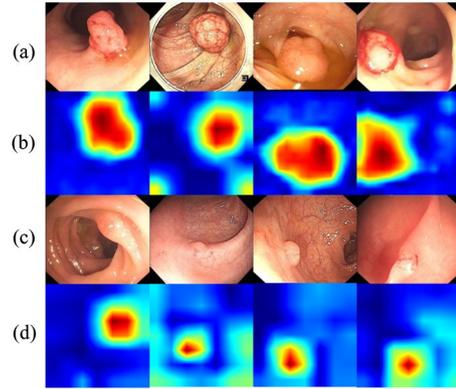


图 9 注意力图

图 10 展示了本文方法训练的准确率和损失曲线。从图 10 中可以看出,模型的准确率曲线很快达到饱和,且饱和时的数值接近 1,损失曲线震荡较小,损失收敛数值接近 0。

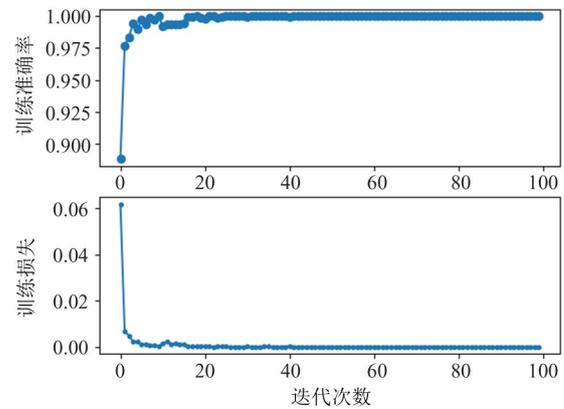


图 10 准确率和损失曲线

### 2.4.3 不同方法在癌变风险数据集上的对比

为了验证本方法对肠道息肉癌变风险评估的有效性,在癌变风险数据集上,本文将所提出的基于 ResNet50 改进的方法与本领域国内外流行的几个深度学习方法: ResNet50<sup>[10]</sup>、ResNext50<sup>[16]</sup>、DenseNet121<sup>[17]</sup>、SENet50<sup>[18]</sup>、EfficientNet-b0<sup>[19]</sup>、Triple-Attention<sup>[20]</sup> 等进行了息肉分类对比实验。经过与几个深度学习方法对比实验,本方法客观地展现出其优越性。表 5 展示了不同方法在癌变风险数据集上的性能对比。从表中可以看出,其他几个深度学习方法的 3 个评价指标均不超过 90%,而基于 ResNet50 改进的卷积神经网络结合迁移学习以后以 99.38% 的灵敏度、98.15% 的特异度、98.17% 的精准度、98.77% 的 F1 分数实现了对高风险息肉的识别,能有效地对肠

道息肉癌变风险进行评估。

表 5 在癌变风险数据集上的对比结果

模型	灵敏度	特异度	精准度	F1 分数
ResNet50 <sup>[10]</sup>	82.10%	84.57%	84.18%	83.13%
ResNext50 <sup>[16]</sup>	80.86%	87.35%	86.47%	83.57%
DenseNet121 <sup>[17]</sup>	81.48%	84.57%	84.08%	82.76%
SENet50 <sup>[18]</sup>	83.64%	81.48%	81.87%	82.75%
EfficientNet-b0 <sup>[19]</sup>	75.31%	79.32%	78.46%	76.85%
Triple-Attention <sup>[20]</sup>	75.93%	82.10%	80.92%	78.34%
本文方法	99.38%	98.15%	98.17%	98.77%

在二分类算法的评价中,AUC 是常用的评价指标,被定义为 ROC 曲线下的面积,AUC 值较高的算法通常被认为更好,图 11 展示了不同方法模型在癌变风险数据集上测试的 ROC 曲线和 AUC。

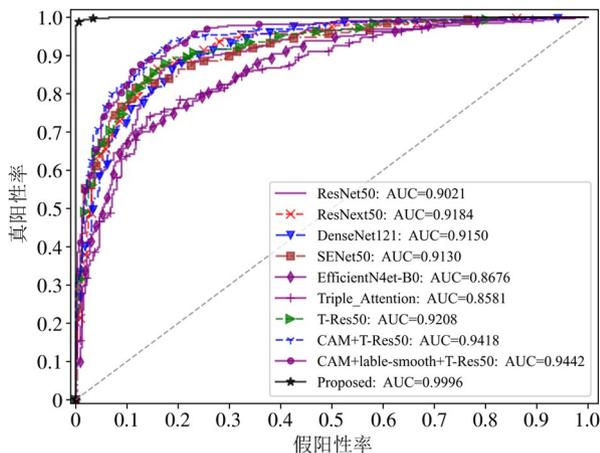


图 11 不同方法识别高风险息肉的 ROC 曲线

从图 11 中可以看出,本文提出的两分支网络 T-Res50 的 AUC 为 0.9208,高于其他几个深度学习方法,引入 CAM、使用标签平滑并结合迁移学习后,其 AUC 达到 0.9996,进一步证明该模型对高风险息肉分类的准确性,对肠道息肉癌变风险评估的有效性。

#### 2.4.4 不同方法在癌变风险数据集上的测试对比分析

在癌变风险数据集的测试集上,本文分别对 10 种不同的模型进行了参数量和 GPU 测试时间的对比分析,实验结果如表 6 所示,所有模型均采用 Batch-size 为 32 的批量来测试。

表 6 不同模型的参数量和对比

模型	参数量/M	GPU 时间/s	灵敏度
ResNet50 <sup>[10]</sup>	23.51	13	87.00%
ResNext50 <sup>[16]</sup>	22.98	17	87.50%
DenseNet121 <sup>[17]</sup>	6.96	14	89.50%
SENet50 <sup>[18]</sup>	26.03	18	89.00%
EfficientNet-b0 <sup>[19]</sup>	4.01	10	87.00%
Triple-Attention <sup>[20]</sup>	23.51	12	86.50%
T-Res50	38.49	17	91.00%
CAM + T-Res50	42.60	18	91.50%
CAM + lable-smooth + T-Res50	42.60	18	92.00%
本文方法	47.74	18	98.50%

从表 6 中的实验结果中可以发现,本文方法的参数量为 47.74M, GPU 测试时间为 18 s,相比于其他方法模型,虽然本文提出方法的模型参数量和 GPU 测试时间有所增长,但是在时间损耗可以考虑的情况下,本文方法的灵敏度具有更加明显的优势。

#### 2.4.5 不同方法在 Kvasir 公开数据集上的对比

为了证明本文方法模型的泛化能力和稳定性,本研究在 Kvasir 公开数据集上将本文方法与其他深度学习方法进行了对比。表 7 展示了不同方法在 Kvasir 公开数据集上的性能对比的实验结果。从表 7 可以看出,本文方法相较于其他对比的深度学习方法也具有更好的效果,其灵敏度、特异度、精准度、F1 分数分别为 98.50%、99.00%、98.99% 和 98.75%。

表 7 在 Kvasir 公开数据集上的对比结果

模型	灵敏度	特异度	精准度	F1 分数
ResNet50 <sup>[10]</sup>	87.00%	90.50%	90.16%	88.55%
ResNext50 <sup>[16]</sup>	87.50%	93.00%	92.59%	89.97%
DenseNet121 <sup>[17]</sup>	89.50%	91.00%	90.86%	90.18%
SENet50 <sup>[18]</sup>	89.00%	91.50%	91.28%	90.13%
EfficientNet-b0 <sup>[19]</sup>	87.00%	91.50%	91.10%	89.00%
Triple-Attention <sup>[20]</sup>	86.50%	90.00%	89.64%	88.04%
T-Res50	91.00%	90.50%	90.55%	90.77%
CAM + T-Res50	91.50%	93.50%	93.37%	92.42%
CAM + lable-smooth + T-Res50	92.00%	93.50%	93.40%	92.70%
本文方法	98.50%	99.00%	98.99%	98.75%

说明本方法所提出的网络具有很好的泛化能力和稳定性。

### 3 结论

针对结肠内窥镜息肉图像数据集的样本数量有限,息肉类内图像病灶大小形状差异大,病变区域与正常组织之间高度相似等问题,本文首先通过数据增强对结肠内窥镜息肉图像数据集进行了扩展,并基于深度学习和迁移学习,提出了一种新的肠道息肉癌变风险评估的方法。本文方法对 ResNet50 进行了改进,建立了一个包括主干网络和任务网络 2 部分的两分支网络;在整个网络的残差块中引入通道注意力,自动学习通道的权重,增强网络的特征表达能力;通过将注意力图分支生成的注意力图与主干网络输出的特征图加权,为网络引入空间注意力;通过生成的注意力图增加了深度学习方法的解释性;使用 2 个标签平滑的交叉熵损失的加权和以端到端地方式对模型进行训练,提高了模型的泛化能力和稳定性。另外,为证明其泛化能力和稳定性,本文在公开数据集上进行了对比实验。实验结果表明,本方法的性能优于一些流行的深度学习方法,能有效地对结肠内窥镜息肉图像进行分类,为肠道息肉癌变风险评估提供了一种有效的新途径,具有一定的临床实用价值。

本文方法还存在不足,目前的结肠内窥镜图像数据集有限、数据量较少,且距离实际应用于临床还远远不够,在后续的工作中,不仅要构建更完善的数据集,还要进一步提升模型的准确度。

#### 参考文献

- [ 1 ] WILD C P, WEIDERPASS E, STEWART B W. World cancer report: cancer research for cancer prevention [M]. Lyon: International Agency for Research on Cancer, 2020:1-613.
- [ 2 ] LEONARDO Z, BROCK C, BURT A D, et al. Computer-aided diagnosis for characterising colorectal lesions: interim results of a newly developed software[J]. Gastrointestinal Endoscopy, 2018,87(6):AB245.
- [ 3 ] STEHLE T, AUER R, GROSS S, et al. Classification of colon polyps in NBI endoscopy using vascularization features[C]//Medical Imaging 2009: Computer-Aided Diagnosis. Lake Buena Vista:SPIE, 2009:72602S.
- [ 4 ] TAMAKI T, OSHIMUTA J Y, KAWAKAMI M, et al. Computer-aided colorectal tumor classification in NBI endoscopy using local features[J]. Medical Image Analysis, 2013,17(1):78-100.
- [ 5 ] MESEJO P, PIZARRO D, ABERGEL A, et al. Computer-aided classification of gastrointestinal lesions in regular colonoscopy[J]. IEEE Transactions on Medical Imaging, 2016,35(9):2051-2063.
- [ 6 ] 范姗姗,刘士臣,曹鸷,等. 无线胶囊内窥镜图像小肠息肉的自动识别[J]. 中国生物医学工程学报,2019,38(5):522-532.
- [ 7 ] 赵新秋,贺海龙,杨冬冬,等. 基于改进的卷积神经网络在图片分类中的应用[J]. 高技术通讯,2018,28(Z2):50-56.
- [ 8 ] ZHANG R, ZHENG Y, MAK T, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain [J]. IEEE Journal of Biomedical and Health Informatics, 2017,21(1):41-47.
- [ 9 ] SHABAN M, AWAN R, FRAZM M, et al. Context-aware convolutional neural network for grading of colorectal cancer histology images [J]. IEEE Transactions on Medical Imaging, 2020,39(7):2395-2405.
- [ 10 ] HE K, ZHANG X, RENS, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:770-778.
- [ 11 ] OQUAB M, BOTTOU L, LAPTEV I, et al. Is object localization for free? -Weakly-supervised learning with convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015:685-694.
- [ 12 ] HAFIZ A M, PARAH S A, BHAT R. Attention mechanisms and deep learning for machine vision: a survey of the state of the art[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Online:IEEE, 2021:1-24.
- [ 13 ] SOEKHOE D, PUTTEN P, PLAAT A. On the impact of data set size in transfer learning using deep neural networks[C]//International Symposium on Intelligent Data

- Analysis. Stockholm: Springer, 2016:50-60.
- [14] ZHAO W. Research on the deep learning of the small sample data based on transfer learning[C]//Proceedings of American Institute of Physics Conference. Chongqing: AIP, 2017: 020018.
- [15] MÜLLER R, KORNBLITH S, HINTON G E. When does label smoothing help? [J]. Advances in Neural Information Processing Systems, 2019,32:1-10.
- [16] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:1492-1500.
- [17] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:4700-4708.
- [18] HU J, SHEN L, SUNG. Squeeze-and-excitation networks [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018:7132-7141.
- [19] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks [ C ] // International Conference on Machine Learning. Long Beach: ICML, 2019: 6105-6114.
- [20] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: convolutional triplet attention module [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Waikoloa: IEEE, 2021: 3139-3148.

## Risk assessment of intestinal polyp canceration based on deep transfer learning

CHENG Shan<sup>\*</sup>, YE Ran<sup>\*\*</sup>, CAO Jing<sup>\*</sup>, YAO Jiafeng<sup>\*</sup>, LI Sheng<sup>\*</sup>, HE Xiongxiang<sup>\*</sup>  
(<sup>\*</sup> College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)  
(<sup>\*\*</sup> Lishui People's Hospital, Lishui 323400)

### Abstract

To accurately distinguish different types of polyps, a novel risk assessment method based on deep transfer learning is proposed. Firstly, a channel attention is introduced into the backbone network to extract the polyp features. Then, the attention map from the attention map branch of task network is weighted to the feature map from the backbone to obtain more accurate lesion features. Finally, the weighted feature map is fed into the perception branch of the task network, and loss functions of the two branch is used to optimize the model in an end-to-end manner to determine the canceration risk of polyps. The method enhances the established dataset, and employ transfer learning to further improve the classification results. The sensitivity, specificity, precision,  $F1$ -Score and area under the ROC curve (AUC) of the optimal model are 99.38%, 98.15%, 98.17%, 98.77%, 0.9996, respectively. The experimental results show that this method can effectively evaluate the canceration risk of intestinal polyps with better performance compared with the popular deep learning methods.

**Key words:** deep learning, transfer learning, channel attention, attention map, image classification, intestinal polyps