第 59 卷 第 6 期/2022 年 3 月/激光与光电子学进展 特邀研究论文

激光与光电子学进展 先进成像

# 基于BiT的早期胃癌内镜图像识别

李宏霄[1]，李姝[2]，石霞飞[1]，董晓曦[1]，金歌[2]，朱兰平[2]，李迎新[1]，阴慧娟[1*]

[1]中国医学科学院生物医学工程研究所，天津 300192；

[2]天津医科大学总医院消化内科，天津 300050

**摘要** 胃癌是我国主要的致死癌症之一，大部分患者发现时已处于进展期，如果能通过大规模筛查在早期发现胃癌，则可大大提高患者生存率。制约我国胃癌大规模筛查的障碍有二，其一是内镜的侵入性高，患者接受度低，其二是我国内镜医师相较于庞大的人口数量严重短缺。第一个障碍可通过胶囊内镜机器人得到缓解，第二个障碍则有望通过人工智能技术解决。将领域前沿的Big Transfer（BiT）技术迁移到一个小样本的早期胃癌内镜图像数据集上，构建了基于白光内镜图像的早期胃癌分类识别模型。在实现迁移过程中，对BiT的超参数调整规则进行了本地化适配，根据GPU内存的限制，选择了单批次数量；利用线性缩放规则，根据单批次数量动态调整了优化算法的初始学习率；固化小型目标数据集上的训练图像总量为256000张，在此基础上设置了迁移学习的其他超参数。对多个模型进行实验，所有模型的结构均为ResNet-v2，只是深度和宽度不同。最佳模型的深度为101，宽度为初始结构的3倍，在测试集上的准确率为97.14%，F1分值为94.77%，敏感度为90.67%，特异度为99.73%。此外，结果表明，单批次数量对模型训练效果的影响不具有显著的统计学差异。所提模型通过对BiT进行本地化适配，成功地在小规模内镜图像数据集上实现了较大模型的迁移，这将会促进大模型技术在内镜图像分析领域的应用，从而有助于早期胃癌大规模筛查的实现。

**关键词** 医用光学；早期胃癌；内镜图像；迁移学习；深度学习

**中图分类号** R573 **文献标志码** A **doi**：10.3788/LOP202259.0617028

# BiT-Based Early Gastric Cancer Classification Using Endoscopic Images

Li Hongxiao[1], Li Shu[2], Shi Xiafei[1], Dong Xiaoxi[1], Jin Ge[2], Zhu Lanping[2], Li Yingxin[1], Yin Huijuan[1*]

[1]*Institute of Biomedical Engineering, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin* 300192, *China*;

[2]*Department of Gastroenterology, General Hospital of Tianjin Medical University, Tianjin* 300050, *China*

**Abstract** Gastric cancer is one of the significant lethal cancers in China. Most patients are diagnosed at an advanced stage, and if gastric cancer can be detected at an early stage through large-scale screening, patient survival can be considerably improved. In China, there are two obstacles toward the large-scale screening of early gastric cancer. One is that endoscopy is overly invasive, resulting in low patient acceptance, and the other is that the number of endoscopists is too small compared with China's large population. A capsule endoscopic robot can alleviate the first obstacle, and the second obstacle is expected to be solved using artificial intelligence. We transferred the state-of-the-art Big Transfer (BiT) to a small dataset of early gastric cancer endoscopic images and

**收稿日期**：2021-11-15；**修回日期**：2021-12-17；**录用日期**：2022-01-07

**基金项目**：天津市科技计划(19YFZCSY00490)

**通信作者**：*yinzi490@163.com

built an early gastric cancer classification model based on white-light endoscopic images. We customized the BiT hyperparameter rules in transfer learning based on local situations. The batch size was selected according to the GPU memory limit, and based on the batch size, the linear scale rules were used to adjust the optimizer's initial learning rate dynamically. The total number of training images for the small dataset was set at 256000, on which other hyperparameters of the transfer learning were set. This study experimented with multiple models having the same structure of ResNet-v2 but different depths and widths. The best model has a depth of 101 and a width three times the original one. It has an accuracy of 97.14%, an F1 score of 94.77%, a sensitivity of 90.67%, and a specificity of 99.73% on the test set. Furthermore, the results show that the effect of batch size on the model training is statistically insignificant. This paper transferred a large model to a small dataset of endoscopic images with the BiT customization. This will promote the use of large-scale models in the field of endoscopic image analysis, which can help realize a large-scale screening of early gastric cancer.

**Key words** medical optics; early gastric cancer; endoscopic image; transfer learning; deep learning

## 1 引 言

胃癌是东亚地区的主要致死癌症之一[1]。根据全球癌症观察站(Global Cancer Observatory, gco.iarc.fr)2020年的数据,中国每10万人中平均胃癌患者人数为33.1人,平均胃癌死亡人数为25.8人,胃癌是中国排名第四的死亡相关癌症。胃癌的早诊早治可以有效提高预后生存率,从而降低癌症负担。消化道内镜检查是胃部疾病诊断的主要手段,但是由于该方法的复杂性和侵入性较高,在人群中的接受度较低,胶囊内镜技术具有无痛和非侵入性的特点,有望提高消化道肿瘤筛查的人群接受度。

胶囊内镜是一种可吞咽的胶囊无线微型内窥镜,整个系统主要由微型摄像镜头、电池、无线传输系统和数据记录系统等组成。通常,胶囊内镜进入目标消化道腔后会以预设的频率拍摄腔内图像,并将图像数据实时传输至体外的数据记录系统,然后随着消化道肌肉群的运动自然排出体外。根据预设频率的差异,通常一个胶囊内镜全程可以拍摄数千至数万张不等的图像,由于图像数据量巨大,内镜医师的研判过程容易出现漏诊的情况。例如,最近一项在日本进行的随机临床实验中,临床医师人工检查胃癌的敏感度约为75%[2]。此外,我国内镜医师数量短缺也是早期胃癌筛查的主要障碍之一。

人工智能技术可以学习内镜医师的研判能力,帮助人类从事大量重复性的劳动,从而提高内镜检查的效率,有望早日实现大规模内镜筛查。人工智能技术目前正被广泛引入包括CT[3]、乳腺X射线[4]、MRI[5]等医学影像领域。在内镜图像分析领域,2017年Shichijo等[6]使用深度卷积神经网络GoogLeNet构建了两个基于内镜图像的幽门螺杆菌胃炎识别模型,模型一不考虑图像在胃内的位置,其准确率为83.1%,敏感度为81.9%,特异度为83.4%,模型二考虑了图像在胃内的位置,其准确率为87.7%,敏感度为88.9%,特异度为87.4%。2018年Hirasawa等[7]用深度卷积神经网络SSD对内镜图像进行识别,用于判断早期胃癌的存在情况,该方法的敏感度为92.2%,阳性预测率为30.6%。2019年他们又在内镜视频上进行了测试[8],获得了94.1%的检出率。2019年Wu等[9]利用VGG-16和ResNet-50两种深度卷积神经网络构建了早期胃癌的识别模型,该模型的准确率为92.5%,敏感度为94.0%,特异度为91.0%。2020年Zhang等[10]使用深度卷积神经网络DenseNet构建基于内镜图像的慢性萎缩性胃炎的识别模型,该模型的准确率为94.2%,敏感度为94.5%,特异度为94.0%。2021年Tang等[11]开发的基于ResNet-50的深度卷积神经网络模型可根据内镜图像诊断黏膜内胃癌,该模型的准确率为88.2%,敏感度为90.5%,特异度为85.3%。

但是,当前研究的局限性在于,一方面利用深度学习的内镜图像识别研究中未引入深度学习算法的最新前沿技术,分类识别的准确率受到算法性能的限制[12-13],人工智能的算法优势未被充分挖掘;另一方面,目前研究中表现较好的结果通常采用了数量巨大的图像数据库[14],对于小样本数据集来说,深度学习模型的训练难度较大,严重限制相关研究的发展。为了配合本课题组研发的光动力诊疗一体胶囊内镜的工作需求,借助领域前沿的Big Transfer(BiT)技术[15]开发了基于迁移学习策略的早期胃癌内镜图像分类识别模型。该模型充分利用了BiT骨干网络的强大性能和BiT超参数调节规

则的便捷特性，成功将大模型迁移到小规模的内镜图像数据集上，在有限的数据集上快速构建了早期胃癌的分类识别模型。该模型可以从内镜图像中快速准确地定位可疑图像，以便内镜医师进一步研判，为人工智能辅助的早期胃癌内镜图像的快速分类识别提供了一种可能的方案。同时，该成果将会促进大模型技术在内镜图像分析领域的应用，从而有助于早期胃癌大规模筛查的实现。

## 2 材料和方法

### 2.1 数据获取

采用最为常见和易获得的普通白光内镜图像作为输入数据。所述实验经过天津医科大学总医院伦理审查委员会批准（批件号为 IRB2021-YX-007-01）。从天津医科大学总医院回顾性地收集了 374 位早期胃癌患者的白光内镜图像数据，此外收集了作为对照组负样本的 446 位慢性胃炎患者的白光内镜图像数据。经过图像质量控制筛选，去掉内容模糊不清的病例或图像，最终保留来自 284 位早期胃癌患者的 1003 张内镜图像纳入数据集，赋予每张图像癌症标签，保留来自 404 位慢性胃炎患者的 2498 张内镜图像纳入数据集，赋予每张图像非癌症标签。

### 2.2 预处理

对全部 3501 张图像随机分层分割为训练集（约 70%，包含 1748 张非癌标签、702 张胃癌标签，共计 2450 张图像）、验证集（约 15%，包含 376 张非癌标签、151 张胃癌标签，共计 527 张图像）和测试集（约 15%，包含 374 张非癌标签、150 张胃癌标签，共计 524 张图像）。训练集用于模型参数更新，验证集用于微调时选取超参数和最优模型，测试集用于最后的模型验证。由于原始的图像为矩形，且尺寸不同，为了统一图像尺寸以便后续的模型训练，对每张图像都以图像中心为基准点切割为正方形，正方形的长度是原图像的短边长度，然后再将图像缩放为 384×384 像素大小。为了增加图像数据的多样性，基于内镜图像采集过程中可能包含的图像变换类型，使用 imgaug 库[16]对训练集的图像进行数据增强。增强方法包括：左右翻转；上下翻转；90°、180°或者 270°旋转，以模拟内镜成像时的角度旋转；在 0.8～1.5 之间随机缩放每个像素的强度值，以模拟图像亮度的变化；采用高斯模糊模拟图像采集中的电子和光场噪声，其中的 sigma 设为[0.0, 5.0]之间的随机值。在实际运算时会从这五类增强方法中随机选择 0～2 种应用于图像。之后，为了提高模型训练的稳定性，图像的每个像素的强度值都除以 255，归一化到[0,1]的区间范围。

### 2.3 BiT 技术

BiT 是谷歌大脑研究团队提出的一种具有深度挖掘迁移学习策略[17-18]潜力的模型构建技术。它将迁移学习策略所具有的小样本需求和简化调参能力提升到了当前技术的新水平，并且 BiT 只需要一次预训练，再配合 BiT 的超参数调整策略即可实现快速迁移，大大降低了迁移成本。

BiT 技术所传递的核心思想是：用更大的模型和更大的数据集训练得到性能更优的预训练模型，这些预训练模型可以便捷地迁移至其他只具有小样本数据集的下游任务中，从而快速得到表现出色的下游任务模型。BiT 采用 5 种骨干网络结构：r50×1、r101×1、r50×3、r101×3、r152×4，这些网络结构全部采用了 ResNet-v2 架构[19]，只是网络的深度（层数）和宽度（每层的节点/单元数量）不同。网络结构中字母 r 代表 ResNet-v2 架构，字母 r 后的数字代表网络的深度，乘号后的数字代表网络的宽度。宽度为 1 代表了 ResNet-v2 架构的原始宽度，宽度为 3 则表示将原来 ResNet-v2 架构中每层的节点/单元数增加到 3 倍。

利用 BiT，分别在三个不同大小的数据集上比较上述 5 种模型结构的性能，这三个数据集分别是 ILSVRC-2012[20]、ImageNet-21k[21]和 JFT-300M[22]。三者包含的图像数量分别是 1281167、14197122 和约 300000000；这三个数据集的数量级依次递增，代表了从小到大的数据量变化。BiT 测试结果的总体结论是，在 JFT-300M 上预训练得到的 r152×4 模型在大多数下游任务中都具有更为出色的表现，即大模型和大数据的结合具有更出色的表现。BiT 的作者已经将预训练的模型公开发布在 Tensorflow Hub 上[23]，据此可以利用迁移学习快速构建一个专用于内镜图像的早期胃癌分类识别模型，并可以期待得到较好的表现。

### 2.4 模型选择

由于 JFT-300M 是专有数据集，因此 BiT 的作者并没有公开在 JFT-300M 上预训练的模型，他们只发布了在 ILSVRC-2012 和 ImageNet-21k 两个公开数据集上预训练的模型。由于 ImageNet-21k 的图像数量比 ILSVRC-2012 多出一个数量级，因此在

ImageNet-21k上预训练的模型会比ILSVRC-2012上预训练的模型学习到更多的“基础知识”，在迁移到下游任务后更容易获得出色的表现，因此，本文采用在ImageNet-21k上预训练的模型作为骨干网络。BiT的另一个结论表明，越大的模型需要越大的预训练数据集，以完全发挥出大模型的性能，较大的模型在较小的数据集可能无法完全发挥性能。此外，在迁移到下游任务时，不同结构模型的表现也可能会出现波动，因此无法简单地根据模型大小从中等数据集ImageNet-21k上预训练的5个模型中选取最好的一个，本课题组的方案就是通过实验来确定最佳的模型。模型的结构是“骨干网络＋分类器”，其中骨干网络是BiT在ImageNet-21k上预训练的5个模型，分类器是一层全连接层网络。

### 2.5 模型训练

BiT迁移学习的超参数调整内容主要包括：数据训练量、优化器学习率、分辨率调整和MixUp[24]数据增强。BiT迁移学习的超参数调整规则（以下简称为“规则”）因数据训练量的不同而有所差异，规则中将迁移学习的目标数据集按照带标签图像的数量$N$分为三个等级：$N<20000$为小数据集，$20000<N<500000$为中数据集，$N>500000$为大数据集。本文的3501张图像属于小数据集。根据BiT的实践经验，在实验中采用了256000作为小数据集上模型训练遍历图像的总数量$N_{\text{total}}$，单次训练遍历的图像数量Batchsize则选取GPU内存允许的上限，使用的GPU为英伟达Titan Xp，包含11 GB显卡内存。5种骨干网络结构选取的Batchsize如表1所示。网络模型参数更新一次称为一个训练步骤，则总的训练步骤数$N_{\text{step}}=N_{\text{total}}/N_{\text{Batchsize}}$，整个训练集的2450张图像被完全遍历一次称为一个训练纪元，则总的纪元数量$N_{\text{Epoch}}=N_{\text{total}}/2450$，每个训练纪元包含的训练步骤数为$a=N_{\text{step}}/N_{\text{Epoch}}$。

优化器采用随机梯度下降算法，初始学习率在BiT经验的基础上依据Batchsize进行线性缩放[25]，BiT迁移学习的初始学习率设为0.03，由于采用的Batchsize小于512，所以等比例缩小初始学习率，将其设为$0.03\times(N_{\text{Batchsize}}/512)$，其中512是BiT在迁移学习中的单批次图像数量。在此基础上采用学习率衰减技术，分别在总迭代步数的30%、60%和90%位置将学习率衰减到之前一步的1/10。由于本文数据集中的图像尺寸接近400×400像素，所以不需要在迁移学习时额外放大分辨率。BiT原文建议对中、大数据集采用MixUP数据增强技术，鉴于本文数据集的图像数量较小，因此没有采用MixUp对数据进行增强。

在模型训练时采用两种参数调整策略：局部微调和全局微调。局部微调只更新最后一层分类器网络的参数，骨干网络的参数处于锁定状态，在训练时不会改变；全局微调则同时更新整个网络中所有的参数，骨干网络和分类器网络的参数同步调整。局部微调时，骨干网络参数被锁定，单纯作为一个特征提取模块被使用，没有根据具体任务进行适配；全局微调时，在模型训练过程中会根据具体任务的数据对骨干网络的参数进行微调，使模型更专注于特定任务。因此，全局微调训练出的模型通常会比局部微调的模型在特定任务上有更好的表现。

表1 五种BiT骨干网络模型的单批次数量和参数数量

Table 1 Batchsizes and parameter amounts of the five BiT backbone networks

| Parameter | 50×1 | 101×1 | 50×3 | 101×3 | 152×4 |
|---|---|---|---|---|---|
| Batchsize | 32 | 16 | 8 | 4 | 4* |
| Number of parameters | 23504450 | 42496578 | 211186370 | 381802178 | 928356610 |

*Because the 152×4 backbone network has too much parameters that its requirement of memory exceeds the upper limit of our GPU during the global fine-tuning. This value is the batchsize used during the local fine-tuning where the backbone's parameters are untrainable.

### 2.6 评价指标

主要使用4种评价指标，分别是准确率、F1分值、敏感度和特异度。它们的计算公式分别为

$$A=\frac{N_{\text{TP}}+N_{\text{TN}}}{N_{\text{TP}}+N_{\text{TN}}+N_{\text{FP}}+N_{\text{FN}}},\tag{1}$$

$$F_1=\frac{2N_{\text{TP}}}{2N_{\text{TP}}+N_{\text{FP}}+N_{\text{FN}}},\tag{2}$$

$$S_{\text{sensitivity}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FN}}},\tag{3}$$

$$S_{\text{specificity}}=\frac{N_{\text{TN}}}{N_{\text{TN}}+N_{\text{FP}}},\tag{4}$$

式中：$N_{\text{TP}}$、$N_{\text{TN}}$、$N_{\text{FP}}$、$N_{\text{FN}}$分别表示真阳性（true positive）样本数、真阴性（true negative）样本数、假阳性（false positive）样本数、假阴性（false negative）样

本数。设置胃癌标签为阳性，非癌标签为阴性。

## 3　结果和讨论

### 3.1　全局微调和局部微调的对比

对以5种骨干网络为基础的模型分别进行了全局微调和局部微调的训练实验，其中，由于152×4骨干网络的参数太多，在全局微调时需要的内存空间超过了GPU设备的上限，因此无法对以152×4骨干网络为基础的模型进行全局微调训练。图1展示了5种网络结构的局部微调模型（方形较小的点）和4种网络结构的全局微调模型（圆形较大的点）在测试集上的四种评价指标的对比结果，在多数情况下具有相同网络结构的全局微调模型比局部微调模型具有更出色的评价指标。例外的两种情况分别是：101×1结构的局部微调模型在敏感度方面超过了其全局微调模型；50×3结构的局部微调模型和全局微调模型在特异度方面的表现不分伯仲。图2(a)展示了5种网络结构的局部微调模型在测试集上的受试者工作特征（ROC）曲线及其曲线下面积（AUC），图2(b)展示了4种网络结构的全局微调模型在测试集上的ROC曲线及其AUC。对比图2(a)和图2(b)发现，全局微调模型优于局部微调模型。图3是多个不同网络结构的全局微调模型和局部微调模型在测试集上预测结果的混淆矩阵，展示了每个模型在测试集预测结果中的真阳性、假阳性、假阴性和真阴性样本的数量占比。由这些值可计算得到四种评价指标，如表2所示，可得图3和表2的结论与图1的结论相同。综上所述，图1～3和表2的结果表明，在多数情况下，全局微调的模型优于局部微调的模型。

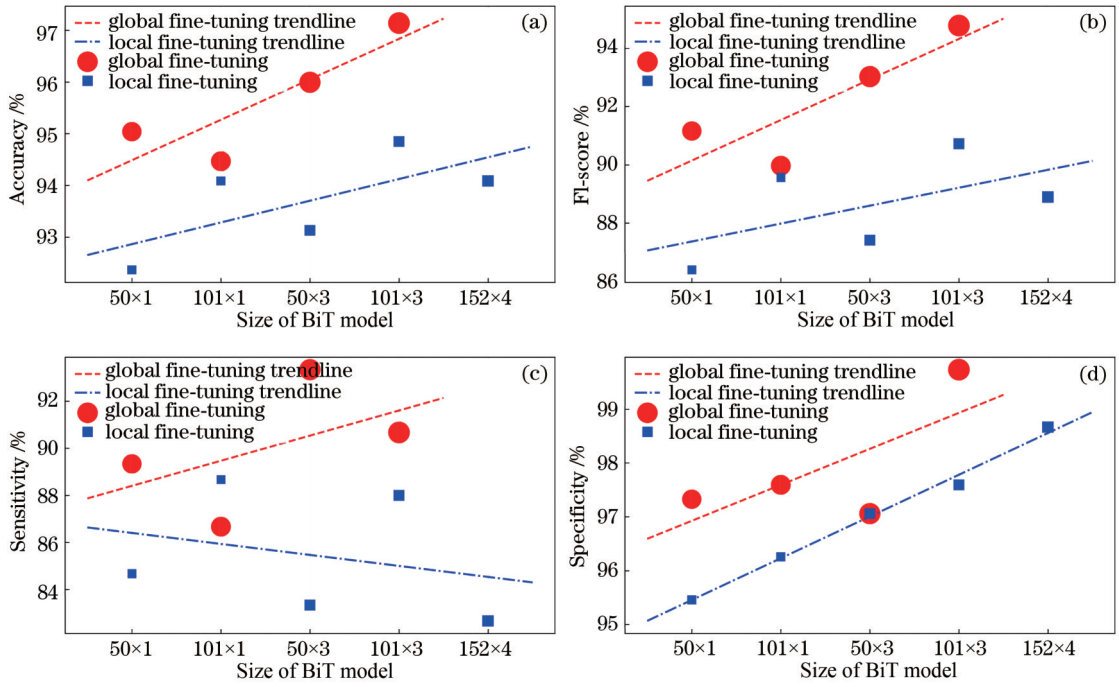


图1 不同骨干网络为基础的各模型的评价指标对比，散点的尺寸代表每个模型可训练参数的数量。(a)准确率；(b) F1分值；(c)敏感度；(d)特异度

Fig. 1 Comparison of four metrics among the models with different backbone networks, the spot size represents the number of trainable parameters of each model. (a) Accuracy; (b) F1-score; (c) sensitivity; (d) specificity

### 3.2　模型尺寸比较

从图1的结果中观察到：四种网络结构的全局微调模型在四种评价指标方面均呈现出相似的趋势，即随着网络结构的增大，模型的表现在变好；五种网络结构的局部微调模型也基本呈现出类似的趋势，只有一个例外，那就是在敏感度方面，观察到随着网络结构的增大，模型的表现在变差。这个现象或许意味着所使用的阳性胃癌图像与预训练模型使用的ImageNet-21k图像存在一定程度上的特征分布差异，理由是：同样的网络结构在进行全局微调时表现出了模型性能随着网络增大而提升的现象，而局部微调时则表现出模型性能随着网络增大而下降的现象。通常认为随着网络的增大，模型对数据集特征的学习程度会加深，或者也可以理解

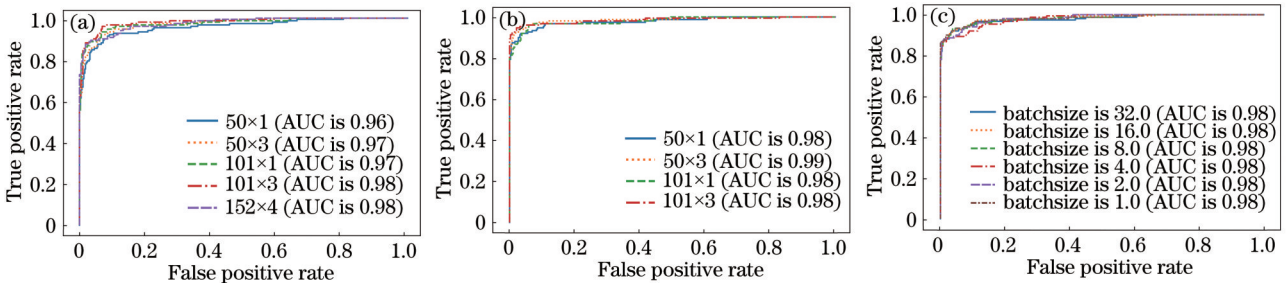


图 2　各种模型的 ROC 曲线及其 AUC。(a) 5 种使用不同骨干网络(参数不可训练)的模型的 ROC 曲线;(b) 4 种使用不同骨干网络(参数可训练)的模型的 ROC 曲线;(c) 6 个使用了 50×1 骨干网络(参数可训练)的模型在不同单批次数量设定下的 ROC 曲线

Fig. 2　ROC curves and AUC of different models. (a) ROC curves of five models with different untrainable backbone networks; (b) ROC curves of four models with different trainable backbone networks; (c) ROC curves of six models with a trainable 50×1 backbone network under different batchsizes

表 2　不同骨干网络为基础的各模型在测试集上的评价指标

Table 2　Testing metrics of the models with different backbone networks on test set

| Size of backbone network | Trainable state | Accuracy /% | Sensitivity /% | Specificity /% | F1-score /% |
|---|---|---|---|---|---|
| 50×1 | TRUE | 95.04 | 89.33 | 97.33 | 91.16 |
| 50×1 | FALSE | 92.37 | 84.67 | 95.45 | 86.39 |
| 101×1 | TRUE | 94.47 | 86.67 | 97.59 | 89.97 |
| 101×1 | FALSE | 94.08 | 88.67 | 96.26 | 89.56 |
| 50×3 | TRUE | 95.99 | **93.33** | 97.06 | 93.02 |
| 50×3 | FALSE | 93.13 | 83.33 | 97.06 | 87.41 |
| 101×3 | TRUE | **97.14** | 90.67 | **99.73** | **94.77** |
| 101×3 | FALSE | 94.85 | 88.00 | 97.59 | 90.72 |
| 152×4 | FALSE | 94.08 | 82.67 | 98.66 | 88.89 |

为“过拟合”的程度会加深,当将这种“过拟合”的预训练模型应用在与训练图像不同的图像数据集时,就容易出现性能下降的情况,而且,模型的“过拟合”程度越深,这种性能下降就越显著。而全局微调可以调整网络整体的参数使其适配新的数据集,从而将“过拟合”的状态从预训练数据集转变到新的数据集上,以此获得对特定任务的优秀解决方案。此外,图 1 和表 2 的结果表明,101×3 骨干网络为基础的模型在准确率(97.14%)、F1 分值(94.77%)、特异度(99.73%)三个指标上都具有最佳的表现,只是在敏感度方面,101×3 模型比 50×3 模型差一些(90.67% vs 93.33%),因此最佳模型是以 101×3 骨干网络为基础的模型。

### 3.3　单批次数量比较

为了考察单批次数量对模型的训练效果是否有影响,对 50×1 骨干网络模型进行了多种不同单批次数量的全局微调训练实验,选用的单批次数量分别为 32,16,8,4,2、1。图 2(c)展示了这 6 个模型在测试集上的 ROC 曲线和 AUC,观察到这 6 个模型的 AUC 没有差别。表 3 展示了这 6 个模型在测试集上的 4 种评价指标的数值,观察到除了单批次数量为 4 的评价指标外,其他模型的各项评价指标基本都持平。表 4 展示了这 6 个模型的 4 种评价指标与单批次数量之间的 Pearson 相关系数分值和 p 值,观察到各项指标与单批次数量之间的相关系数很小,同时 p 值远大于 0.05。因此,图 2(c)、表 3 和表 4 的结果表

表 3　6 个使用了 50×1 骨干网络(参数可训练)的模型在不同单批次数量设定下的四种评价指标

Table 3　Four metrics of the six models with a trainable 50×1 backbone network under different batchsizes

| Batchsize | Accuracy /% | Sensitivity /% | Specificity /% | F1-score /% |
|---|---|---|---|---|
| 32 | 95.04 | 89.33 | 97.33 | 91.16 |
| 16 | 95.23 | 89.33 | 97.59 | 91.47 |
| 8 | 95.23 | 90.00 | 97.33 | 91.53 |
| 4 | 93.89 | 89.33 | 95.72 | 89.33 |
| 2 | 95.23 | 88.67 | 97.86 | 91.41 |
| 1 | 95.42 | 90.00 | 97.59 | 91.84 |

表 4　表 3 中的单批次数量与各项评价指标之间的 Pearson 相关系数及其 p 值

Table 4　The Pearson correlation scores and their p-values between the four metrics and the batchsize in Table 3

| Parameter | Accuracy | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|
| Correlation score | 0.0835 | −0.0784 | 0.1056 | 0.0753 |
| p-value | 0.8750 | 0.8826 | 0.8423 | 0.8873 |

明，单批次数量与以 50×1 为骨干网络的模型的评价指标之间不存在显著的统计学相关性。注意到有关单批次数量对模型训练效果的影响的研究指出[26-27]：当优化算法的学习率不变时，较小的单批次数量容易得到更好的模型训练效果；但是采用线性缩放规则，根据单批次数量动态调整优化算法的学习率时则可以在不损失训练效果的前提下增大单批次数量[25]，从而加快训练速度，该策略正是本文在迁移学习中所采用的策略之一，这在一定程度上解释了有关单批次数量对训练效果无显著影响的结果。此外，所采用的 5 种不同的骨干网络具有相同的 ResNet-v2 架构，只是模型深度和宽度不同，这意味着模型在优化训练时梯度传播的路径是相似的。综上所述，推测单批次数量对 5 种不同骨干网络为基础的各个模型的训练效果的影响可以忽略。

### 3.4　识别结果

测试集由 374 张带有非癌标签的阴性样本和 150 张带有胃癌标签的阳性样本图像组成，所有模型在训练阶段不以任何形式接触测试集。按照模型尺寸和使用的单批次数量的区别，总共训练了 14 个模型，这些模型在测试集上的混淆矩阵如图 3 所示。所有模型的假阳性样本数在 10 到 26 之间，假阳性样本的并集包含 43 个样本，假阳性样本交集包含 4 个样本；所有模型的假阴性样本数在 1 到 17 之间，假阴性样本的并集共包含 45 个样本，假阴

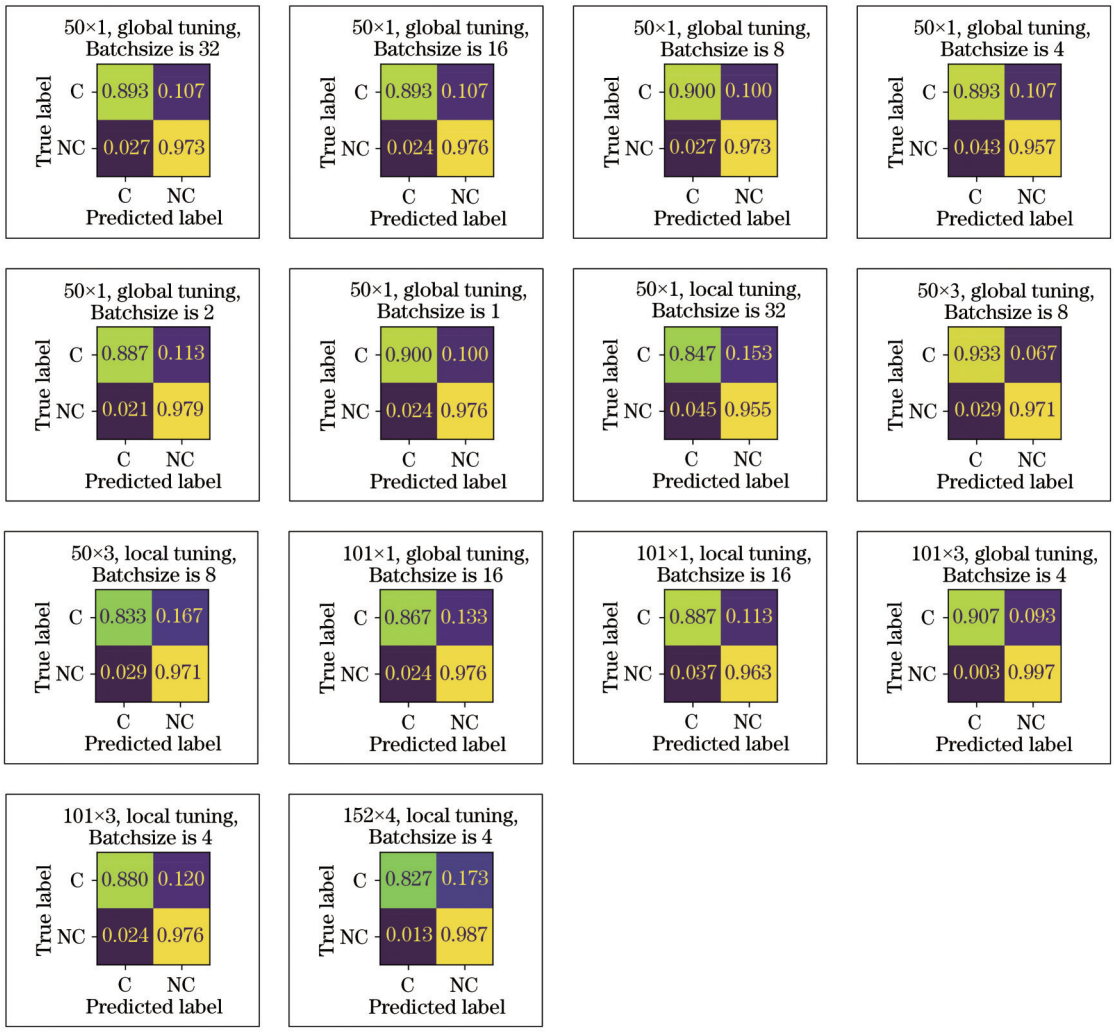


图 3　所有模型在测试集上的混淆矩阵，C 代表胃癌标签，NC 代表非癌标签

Fig. 3　Confusion matrices of all the models applied on the test set, C represents the cancer label, NC represents the non-cancer label

性样本的交集是空集。这些结果表明：总体而言，这些模型辨识出阴性样本的能力高于辨识出阳性样本的能力；各个模型之间的假阳性样本重合度高于假阴性样本的重合度。除去43个假阳性样本和45个假阴性样本，称剩余的374－45＋150－43＝436个样本为简单样本，它们从未被任何一个模型识别错误；同时，将4个反复出现在每个模型预测结果中的假阳性样本称为最具挑战性的困难样本。图4展示了4个困难阳性样本和随机抽取出的简单阳性和阴性样本各4个。

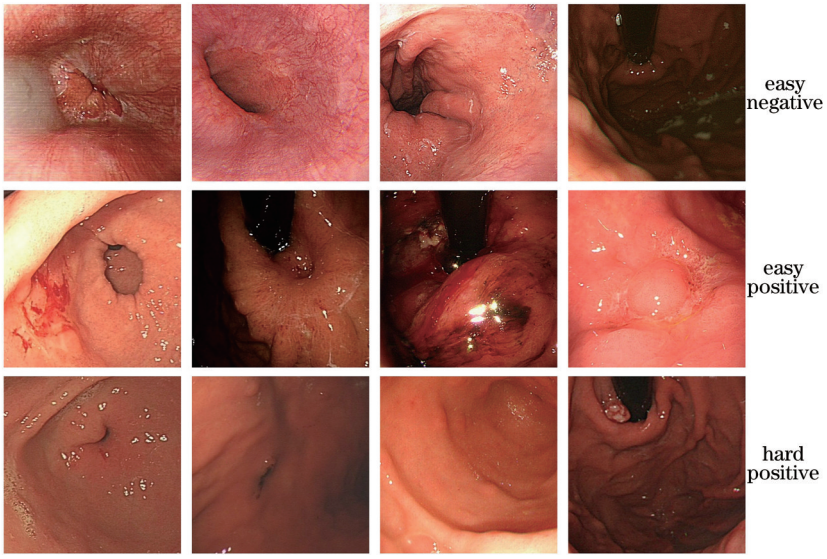


图4　测试图像示例

Fig. 4　Examples of testing images

## 4　结　　论

在小样本早期胃癌的内镜图像数据集上快速构建并测试了基于BiT骨干网络的分类识别模型，最佳的模型是以101×3 ResNet-v2骨干网络为基础的模型，在全局微调训练后，其准确率达到97.14%，F1分值为94.77%，敏感度为90.67%，特异度为99.73%。在医学领域，获得有标签的数据是一项成本较高的工作，因此，像BiT这种能够在小样本集有高性能表现且可以快速部署的建模技术很适于医学研究的应用场合。此外，通过实验对比发现，更大的骨干网络模型往往会有更高的识别性能，但需要在迁移到下游任务时采用全局微调的训练策略，即在下游任务的数据集上微调骨干网络的全部参数，通过将大模型适配到下游任务数据空间发挥其最大效能。此外，虽然实验结果表明单批次数量对模型性能的影响不具有显著的统计学差异，但是，对于具有更大内存的GPU，可以采用更高的单批次数量参数，从而加快模型的训练收敛速度。实验证明BiT骨干网络这样的大模型对内镜图像分析处理的效果很出色，因此，下一步的计划是将BiT骨干网络应用于语义分割模型中，用于癌变部位的精准分割。

## 参　考　文　献

[1] Sung H, Ferlay J, Siegel R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. A Cancer Journal for Clinicians, 2021, 71(3): 209-249.

[2] Yoshida N, Doyama H, Yano T, et al. Early gastric cancer detection in high-risk patients: a multicentre randomised controlled trial on the effect of second-generation narrow band imaging[J]. Gut, 2021, 70(1): 67-75.

[3] Liu Y M, Xiao Z Y. Automatic segmentation algorithm of liver tumor based on feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1417001.
刘一鸣, 肖志勇. 基于特征融合的肝脏肿瘤自动分割方法[J]. 激光与光电子学进展, 2021, 58(14): 1417001.

[4] Sun Y J, Qu Z Y, Li Y H. Study on target detection of breast tumor based on improved mask R-CNN[J]. Acta Optica Sinica, 2021, 41(2): 0212004.
孙跃军, 屈赵燕, 李毅红. 基于改进的Mask R-CNN的乳腺肿瘤目标检测研究[J]. 光学学报, 2021, 41(2): 0212004.

[5] Zhao X, Wang X, Wang H K. End-to-end

segmentation of brain white matter hyperintensities combining attention and Inception modules[J]. Acta Optica Sinica, 2021, 41(9): 0910002.
赵欣, 王欣, 王洪凯. 融合注意力和Inception模块的脑白质病变端到端分割[J]. 光学学报, 2021, 41(9): 0910002.

[6] Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of helicobacter pylori infection based on endoscopic images[J]. EBioMedicine, 2017, 25: 106-111.

[7] Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images[J]. Gastric Cancer, 2018, 21(4): 653-660.

[8] Ishioka M, Hirasawa T, Tada T. Detecting gastric cancer from video images using convolutional neural networks[J]. Digestive Endoscopy: Official Journal of the Japan Gastroenterological Endoscopy Society, 2019, 31(2): e34-e35.

[9] Wu L L, Zhou W, Wan X Y, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots[J]. Endoscopy, 2019, 51(6): 522-531.

[10] Zhang Y Q, Li F X, Yuan F Q, et al. Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence[J]. Digestive and Liver Disease, 2020, 52(5): 566-572.

[11] Tang D H, Zhou J, Wang L, et al. A novel model based on deep convolutional neural network improves diagnostic accuracy of intramucosal gastric cancer (with video)[J]. Frontiers in Oncology, 2021, 11: 622827.

[12] Cho B J, Bang C S, Park S W, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network[J]. Endoscopy, 2019, 51(12): 1121-1129.

[13] Li L, Chen Y S, Shen Z, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging[J]. Gastric Cancer, 2020, 23(1): 126-132.

[14] Luo H Y, Xu G L, Li C F, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study[J]. The Lancet Oncology, 2019, 20(12): 1645-1654.

[15] Kolesnikov A, Beyer L, Zhai X H, et al. Big Transfer (BiT): general visual representation learning [M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12350: 491-507.

[16] Jung A B, Wada K, Crall J, et al. Imgaug[EB/OL]. [2021-11-11]. https://github.com/aleju/imgaug#citation.

[17] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[18] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3: 9.

[19] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in networks and systems. Cham: Springer, 2016, 9908: 630-645.

[20] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[21] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE Press, 2009: 248-255.

[22] Sun C, Shrivastava A, Singh S, et al. Revisiting unreasonable effectiveness of data in deep learning era [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 843-852.

[23] Google. TensorFlow Hub of BiT[EB/OL]. (2021-11-11)[2021-11-11]. https://tfhub.dev/google/collections/bit/1.

[24] Zhang H Y, Cissé M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization[C]//6th International Conference on Learning Representations, ICLR 2018, April 30-May 3, 2018, Vancouver, BC, Canada. La Jolla: ICLR, 2018.

[25] Goyal P, Dollár P, Girshick R, et al. Accurate, large minibatch SGD: training ImageNet in 1 hour [EB/OL]. (2017-06-08) [2021-05-09]. https://arxiv.org/abs/1706.02677.

[26] Keskar N S, Mudigere D, Nocedal J, et al. On large-batch training for deep learning: generalization gap and sharp minima[C]//5th International Conference on Learning Representations, ICLR 2017, April 24-26, 2017, Toulon, France. La Jolla: ICLR, 2017.

[27] Kandel I, Castelli M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset[J]. ICT Express, 2020, 6(4): 312-315.