

网络出版时间: 2023-01-20 16:07:21 网络出版地址: <https://kns.cnki.net/kcms/detail//34.1065.R.20230119.1742.020.html>

基于卷积神经网络的白光内镜下早期胃癌检测

晋晶^{1,2}, 张倩倩^{1,2}, Bill Dong³, 马涛³, 王曦^{1,2}, 梅雪灿^{1,2}, 宋绍方⁴, 彭杰⁴, 吴艾久⁴, 董兰芳³, 孔德润^{1,2}

摘要 目的 开发一种基于区域卷积神经网络(Mask R-CNN)的内镜下自动检测早期胃癌(EGC)系统。方法 首先从安徽医科大学第一附属医院获得3 579张和892张EGC白光图像(WLI),分别进行训练和测试。随后前瞻获取10个WLI实时视频以测试动态性能。另外再随机选取400张WLI图像,用于Mask R-CNN系统和内镜医师对照。诊断能力以准确率、敏感度、特异度、阳性预测值(PPV)和阴性预测值(NPV)进行评估。结果 在WLI图像诊断中,Mask R-CNN系统准确率、敏感度和特异度分别为90.25%、91.06%和89.01%,与病理诊断差异无统计学意义。在WLI视频中,诊断EGC的准确率为90.27%,实时测速可达35帧/s。在对照实验中,Mask R-CNN系统的敏感度明显高于高年资组医师(93.00% vs 80.20%, $\chi^2 = 7.059$, $P < 0.001$),特异度高于低年资组医师(82.67% vs 71.87%, $\chi^2 = 9.955$, $P < 0.001$),总体准确率高于中年资组医师(85.25% vs 78.00%, $\chi^2 = 7.009$, $P < 0.001$)。结论 在WLI下,Mask R-CNN系统检测EGC的性能良好,在实际临床应用中有较大潜力。

关键词 人工智能; 区域卷积神经网络; 白光内镜; 早期胃癌
中图分类号 R 573.9

2022-12-22 接收

基金项目: 安徽省重点研究与开发计划项目(编号: 2022e07020048)

作者单位: ¹ 安徽医科大学第一附属医院消化内科, 合肥 230022

² 安徽省消化病重点实验室, 合肥 230022

³ 中国科学技术大学计算机科学与技术学院, 合肥 230027

⁴ 合肥中纳医学仪器有限公司, 合肥 230088

作者简介: 晋晶, 女, 硕士研究生;

孔德润, 男, 教授, 主任医师, 博士生导师, 责任作者, E-mail: kongderun@ahmu.edu.cn

文献标志码 A 文章编号 1000-1492(2023)02-0285-07
doi: 10.19405/j.cnki.issn1000-1492.2023.02.020

胃癌是世界上最常见的恶性肿瘤之一^[1], 早期胃癌(early gastric cancer, EGC)的5年生存率可以达到95%, 而进展期胃癌(advanced gastric cancer, AGC)的5年生存率只有30%^[2]。因此, 胃癌的早期发现、诊断及早期根除是降低胃癌死亡率的重要措施。人工智能(artificial intelligence, AI)基于机器的计算和学习能力, 能高效地解决问题, 在医学图像识别领域得到广泛应用和发展。在内镜领域, 既往文献^[3]报道了AI在识别食管癌等方面的应用。目前内镜图像检测多采用基于深度学习的目标检测算法, 有训练样本量大、训练时间长、计算体量大、表达精度低等局限性^[4]。该研究开发了一种基于掩膜区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)的AI系统作为内镜下辅助诊断EGC的新技术, 并测试其性能, 与内镜医师对照, 评估该系统的临床适用性。

1 材料与方法

1.1 仪器和软件 CV-290型内镜主机、GIF Q260J型胃镜、GIF H260Z型胃镜(Olympus Corporation), LabelMe软件(版本1.1, IBM Corporation)、SPSS(版本26.0, IBM Corporation)、Pair软件(版本2.0, 深圳度影医疗科技有限公司)。

1.2 患者选择 选取2017年10月—2022年3月在安徽医科大学第一附属医院接受放大内镜和窄带

Results The average DSC value between the deep learning prediction plan and the manual plan was greater than 0.92 in isodose surface, and the average Hausdorff distance HD_{95} of the isodose surface was 0.58-0.62 cm; The V_{20} , V_{30} , D_{mean} of total lung were slightly lower than those of manual plan ($P < 0.05$) for the prediction model, meanwhile, the D_2 , D_{50} , D_{mean} , HI of the target area and V_{30} of total lungs were better than those of manual plan ($P < 0.05$) for Auto-Plan; Three-dimensional dose distribution of the three groups and the corresponding DVH curve showed that the three-dimensional dose distribution of the three groups had a little differences, and the DVH curves of the target area and organs at risk had a good agreement. **Conclusion** Auto-Plan can realize the design of automated plan for postoperative esophageal cancer, while the deep learning prediction model can realize the accurate prediction of the 3D dose distribution.

Key words esophageal cancer; automated plan; deep learning; dosimetry; intensity modulate radiotherapy

成像(narrow band imaging , NBI) 检查 ,并对病灶获取活检的患者 ,以及接受内镜下黏膜下剥离术为初步治疗的患者。两名具有 10 年以上经验的内镜医师 ,使用 NBI 回顾性诊断胃黏膜病变 ,并根据血管和表面分类诊断 EGC^[5]。所有活检或切除标本均由病理科医师据维也纳胃肠道上皮肿瘤分类^[6]对其进行组织学评估 ,其中 4 型(高级别黏膜肿瘤)和 5 型(浸润性肿瘤)中未侵及肌层的病变被诊断为 EGC ,而 1 型(肿瘤阴性)、2 型(肿瘤不确定)和 3 型病变(低级肿瘤)病变被诊断为非癌性病变。当内镜诊断与病理结果不一致时 ,以病理结果为金标准。回顾性资料来源于既往临床诊疗 ,患者知情同意得到豁免(安徽医科大学第一附属医院 ,伦理编号: PJ2021-08-12)。前瞻性实验的所有患者均签署了知情同意书。

1.3 图像筛选 既往研究^[7]显示 ,由于炎症细胞浸润 ,EGC 与胃炎的鉴别较为困难 ,往往导致 EGC 的漏诊。本研究针对 EGC 与胃炎、胃糜烂等胃部疾病的鉴别开发了 Mask R-CNN 系统。

1.3.1 图片纳入标准 ① 非癌性图像只纳入胃炎、胃糜烂白光图像(white light images , WLIs); ② 癌性图像包括 EGC 和 AGC WLI 图像 ,癌性病灶需占整个图像的 30% 以上 ,并且可被内镜医师判断识别。③ 根据维也纳病理诊断标准 ,明确诊断为 EGC 的图片。

1.3.2 图片排除标准 ① 含有息肉、溃疡或溃疡瘢痕、胃癌残留和其他病变的图像; ② 有黏液、血液或其他异物附着的图像; ③ 由于光环、气泡、未对焦

或过度模糊而难以评估的图像; ④ 靛胭脂等染料染色图像。

1.4 数据准备 获取纳入患者的所有 WLI 图像和视频 ,另外纳入部分正常黏膜、胃炎、糜烂和 AGC 的图片 ,作为对照训练。所有图像都已匿名化。

由两名接受过培训的医师共同标注 ,一名医师先用 LabelMe 或 Pair 软件对实际病变区域进行勾边标注 ,另一名医师再检查修正 ,两名医师轮流进行标注工作。只有在两名医师达成共识后 ,才能确认图像分类和标注。随后 ,由一位安徽医科大学第一附属医院的专家(具有至少 10 年内镜检查工作经验)确认这些图像是否符合纳入标准和排除标准(图 1)。

1.5 算法训练及模型识别流程

1.5.1 训练数据 回顾性获得来自安徽医科大学第一附属医院 3 579 张 WLI 图像 ,其中包括 1 182 张炎症、糜烂图像、2 184 张 EGC 图像及 213 张 AGC 图像。

1.5.2 模型建立方法 针对 EGC 识别任务 ,训练了 Mask R-CNN 目标检测神经网络模型。输入多批次的训练图像 ,使用反向传播算法 ,迭代更新网络参数 ,直到损失函数收敛。

Mask R-CNN 基于目标检测框架 Faster R-CNN 的改进实现 ,其架构主要包括三个部分。第一部分是主干网络 ,用于从输入图像中提取特征图。此模型采用的骨干结构是 ResNet-50 ,具有 50 层神经元。该结构先经过卷积层、批标准化层、ReLU 激活函数和最大池化层的计算 ,然后接入了 4 个具有瓶颈层

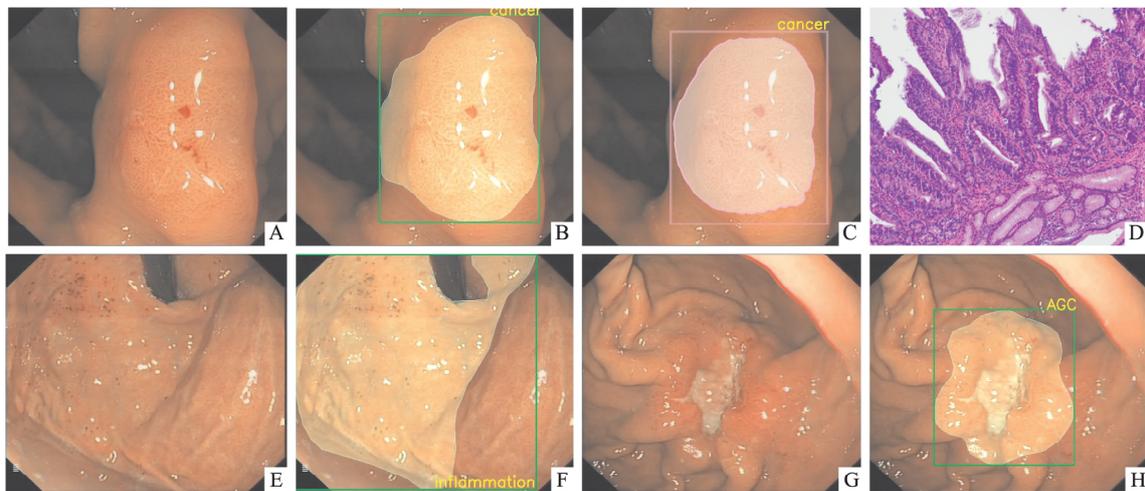


图 1 内镜医师及 Mask R-CNN 系统标注病灶

A: 含有 EGC 病灶的白光图片; B: 内镜医师使用 LabelMe 软件勾画 EGC 病灶; C: Mask R-CNN 系统勾画 EGC 病灶; D: EGC 病理图片 HE × 100; E: 含有胃炎病灶的白光图片; F: 内镜医师使用 LabelMe 软件勾画胃炎病灶; G: 含有 AGC 病灶的白光图片; H: 内镜医师使用 LabelMe 软件勾画 AGC 病灶

的残差模块(图2),每个模块同样具有卷积层和ReLU 激活函数。使用残差结构可以增加训练网络的深度,学习更加复杂的图像特征,同时消除可能降低学习效果的因素。

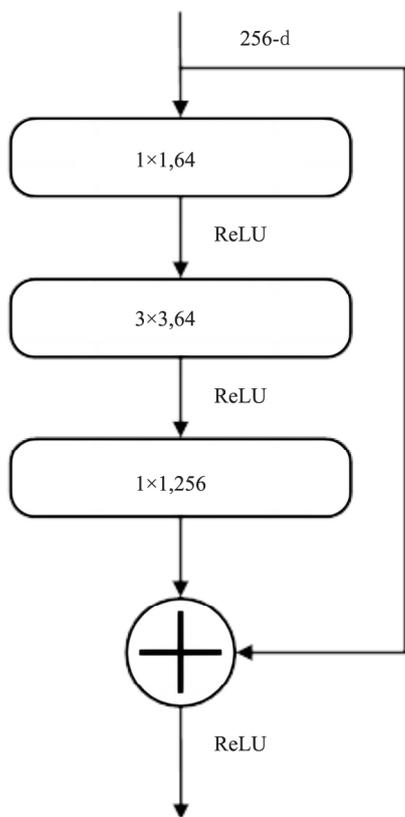


图2 ResNet 残差模块

使用 ImageNet 预训练的权重作为骨干网络所有神经元的训练初始值,因为从结果来看,这样做实验表现更好。骨干网络会提取图像特征为一系列向量给整个架构的第二部分即 ROI Align 层。其根据提取图像特征,并得到大致的候选框。第三部分是 ROI head,作为决策层,该阶段除了对这些候选框的内容进行种类和包围框位置的精准预测以外,还加入全卷积网络的分支,对进行图像二值掩膜,得到像素级的图像分割结果。在计算 Mask R-CNN 模型输出的结果时包含对分类的置信度即分类概率和包围框的交并比(IoU)。IoU 的定义如下:

$$IoU = \frac{area_{pr} \cap area_{gt}}{area_{pr} \cup area_{gt}}$$

其中 $area_{pr}$ 为神经网络预测得到的包围框, $area_{gt}$ 为真实标签的包围框。Mask R-CNN 整体框架如图3所示。

1.6 测试实验

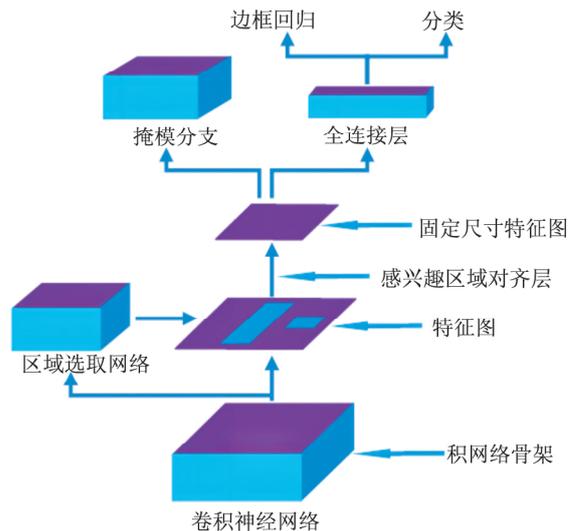


图3 Mask R-CNN 架构示意图

1.6.1 在白光静态图片中测试 Mask R-CNN 系统
参考测试集分类及数据比例,同时按照训练集:测试集 = 8 : 2,共准备 892 张 WLI 图像作为静态图片测试集,其中胃炎、糜烂:EGC:AGC = 324 : 534 : 34。当 Mask R-CNN 系统识别到 EGC 时则判读正确,识别到胃炎、糜烂或 AGC 病灶时则判读错误。

1.6.2 在白光实时视频中测试 Mask R-CNN 系统
前瞻性获得 10 例 EGC 患者的 10 个内镜视频,对获取的 EGC 视频进行剪辑拼接。实时动态检测 10 个视频的平均时长为 270.9 s,最短的为 105 s,最长的为 422 s(图4),视频检测速度达 35 帧/s。按照相应帧数将视频转化为图像后,由内镜医师筛选出包含 EGC 病灶的图片并分类,再使用 Mask R-CNN 系统对全部图片进行识别。Mask R-CNN 系统加入了投票筛选模块,即对一定时间段内视频内容的预测处理结果进行投票,规定只要在半秒内有超过 1/5 的视频帧检测到 EGC 目标,即可认为在这段时间内正确识别出目标。当一张图片上同时有多种病

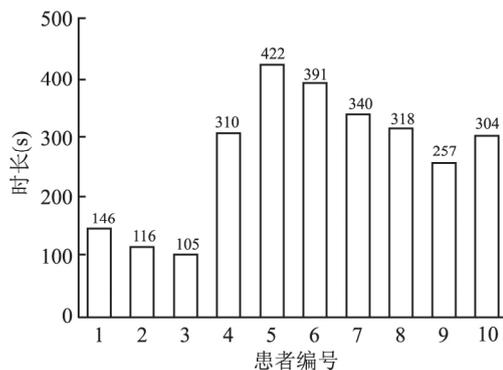


图4 白光视频测试集数据情况

变时,最优先识别 EGC,且仅以识别出 EGC 为阳性结果。

1.6.3 Mask R-CNN 系统与内镜医师对照 准备 743 张正常黏膜和 858 张包含 EGC 或胃炎、糜烂的 WLI 图像,从中按照正常:胃炎、糜烂:EGC=2:1:1 的比例随机抽取共 400 张图像,即非癌性图像:癌性图像=3:1。邀请 15 名内镜医师,分为高年资组(内镜经验>10 年)、中年资组(内镜经验>5 年)和低年资组(内镜经验>1 年),每位医师独立完成读图。同时用 Mask R-CNN 系统识别图像,对照 Mask R-CNN 系统和内镜医师的表现。

1.6.4 主要评价指标 包括准确率、敏感度、特异度、阳性预测值(positive predictive value, PPV) 和阴性预测值(negative predictive value, NPV)。准确率是 Mask R-CNN 正确识别数占所有图片的比例。敏感度指 Mask R-CNN 识别为 EGC 占实际 EGC 的比例,敏感度越高,则漏诊越少。特异度指 Mask R-CNN 识别出的非 EGC 占真实非 EGC 的比例,特异度越高,则误判越少。PPV 指真实 EGC 占 Mask R-CNN 识别为 EGC 的比例,PPV 越高,Mask R-CNN 识别为 EGC 的真实确诊率越高。NPV 指真实非 EGC 占识别为非 EGC 的比例,NPV 越高,Mask R-CNN 识别为非 EGC 的真实非患病率越高。

1.7 统计学处理 采用受试者工作特征(receiver operating characteristic, ROC) 曲线分析,计算 ROC 曲线下的区域面积(area under curve, AUC),以得到最佳阈值,评估 Mask R-CNN 系统在图片测试集和视频测试集中的诊断能力。当 Mask R-CNN 的置信值输出大于该阈值时,Mask R-CNN 系统识别 EGC 并绘制病灶边界。

Mask R-CNN、病理学诊断和医师之间的对照采用 χ^2 检验 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 WLI 图片测试 WLI 图片测试包含胃炎或胃糜烂、EGC 和 AGC 三组病变。以 10% 为区间,Mask R-CNN 系统设置了从 10% 到 90% 的 9 个不同执行阈值。Mask R-CNN 系统检测到病灶后会生成包围框,并评估该病灶是 EGC 的可能性(即置信度)。只有当包围框的置信度超过执行阈值时,才会判定该病灶为 EGC。使用 Mask R-CNN 系统对 WLI 图片进行测试,得到各阈值下 Mask R-CNN 系统的表现,根据 ROC 曲线(图 5)得出在 80% 的阈值下,Mask R-CNN 系统表现最佳,识别 EGC 的准确率为

90.25%, 敏感度为 91.06%, 特异度为 89.01%, PPV 为 92.61%, NPV 为 86.81%, AUC 为 0.94。识别胃炎的准确率、敏感度和特异度分别为 83.86%、88.50% 和 81.01%, 识别 AGC 的准确率、敏感度和特异度分别为 98.32%、82.35% 和 98.95%。勾勒边界的精度评估使用检测边界和真实标注边界的交并比(intersection over union, IoU) 计算,基于当 IoU > 0.5 时算作判断正确的基准,勾勒边界的综合准确率为 61.04%。

经与病理结果比较,Mask R-CNN 系统与病理学金标准诊断结果差异无统计意义,见表 1。

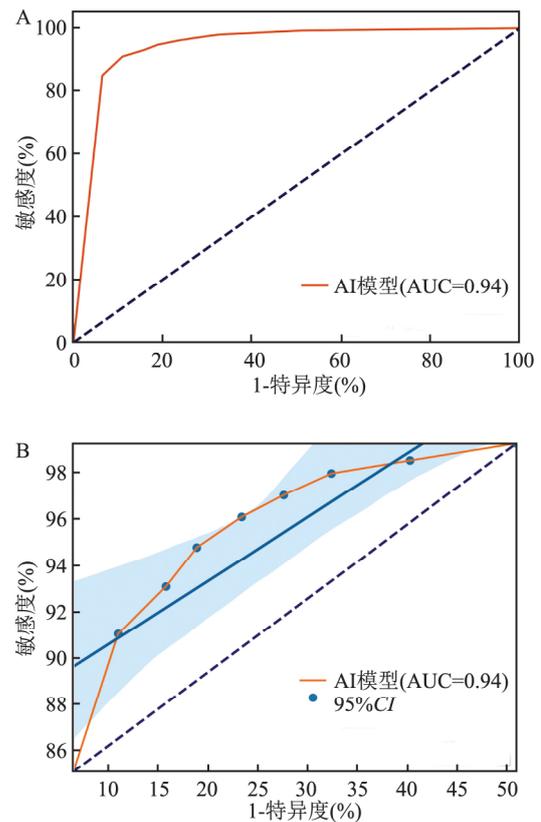


图 5 白光图片测试 ROC 曲线

A: Mask R-CNN 在白光图片测试的整体 ROC 曲线; B: Mask R-CNN 在白光图片测试的局部 ROC 曲线

表 1 Mask R-CNN 系统在图片测试集中的表现

Mask R-CNN	病理		
	EGC(个)	非 EGC(个)	合计
EGC(个)	489	39	528
非 EGC(个)	48	316	364
合计	537	355	892

2.2 WLI 视频测试 Mask R-CNN 系统可实时采集最大为 35 帧/秒的视频,并输出结果,最高延迟 80 ms,能够满足大多数临床环境需求。

视频实验的结果及绘制的 ROC 曲线如图 6 所示,当阈值为 90% 时,Mask R-CNN 系统在视频流上取得较优结果,诊断 EGC 的准确率、敏感度和特异度为 90.27%、84.86% 和 91.87%,PPV 和 NPV 为 75.47% 和 95.37%,AUC 为 0.93。

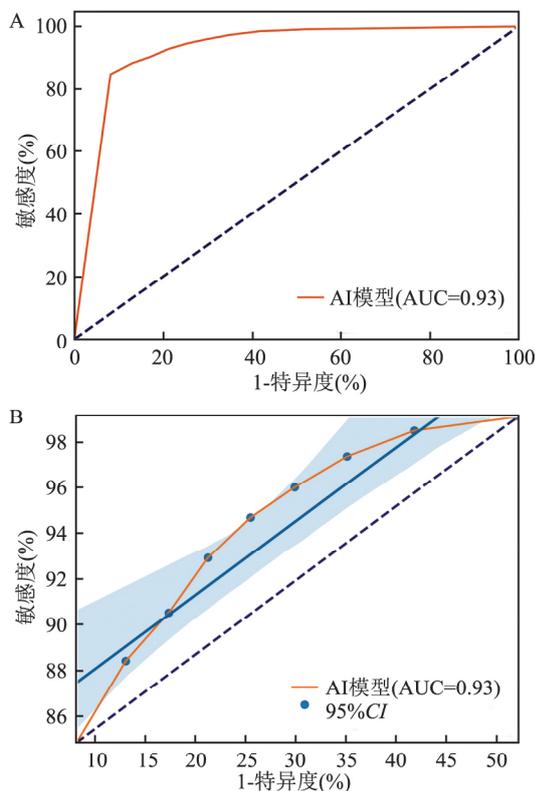


图 6 白光视频测试 ROC 曲线

A: Mask R-CNN 在白光视频测试的整体 ROC 曲线; B: Mask R-CNN 在白光视频测试的局部 ROC 曲线

特异度高于低年资医师组(82.67% vs 71.87%, $\chi^2 = 9.955$, $P < 0.001$),总体准确率高于中年资医师组(85.25% vs 78.00%, $\chi^2 = 7.009$, $P < 0.001$)。见表 2。

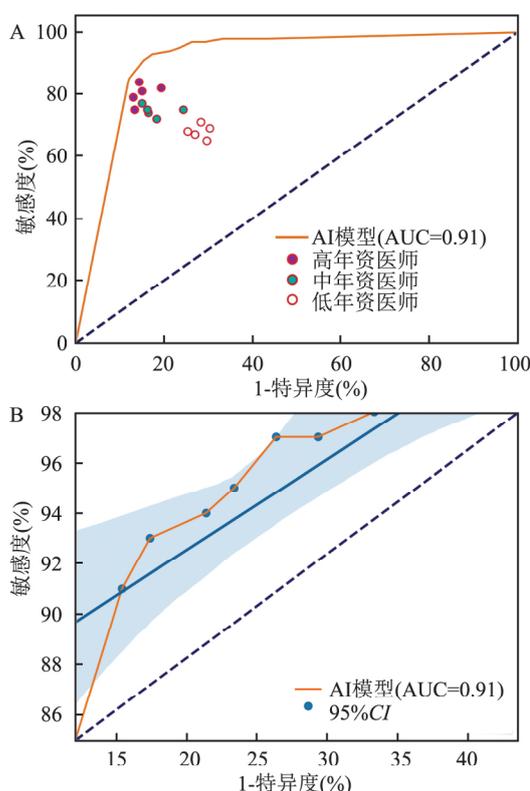


图 7 Mask R-CNN 在对照实验的 ROC 曲线

A: Mask R-CNN 在对照实验的整体 ROC 曲线; B: Mask R-CNN 在对照实验的局部 ROC 曲线

2.3 Mask R-CNN 系统与内镜医师对照实验 对照实验中增加了正常黏膜图像,删减了 AGC 图像。在 70% 的阈值下(图 7),Mask R-CNN 系统的各项指标较为平衡,准确率、敏感度和特异度分别为 85.25%、93% 和 82.67%,NPV 为 97.25%,AUC 为 0.91。

实验共邀请了 15 位内镜医师,分为高年资医师组、中年资医师组及低年资医师组,每组各 5 名。高年资医师组识别 EGC 的准确率(83.80%)、敏感度(80.20%)、特异度(85.00%) 高于中年资医师组(准确率 78.00%、敏感度 74.60%、特异度 79.13%) 和低年资医师组(准确率 70.90%、敏感度 68.00%、特异度 71.87%)。

Mask R-CNN 系统的敏感度明显优于高年资医师组(93.00% vs 80.20%, $\chi^2 = 7.059$, $P < 0.001$),

表 2 Mask R-CNN 系统与内镜医师对照

诊断者	准确率(%)	敏感度(%)	特异度(%)
Mask R-CNN	85.25	93.00	82.67
(n/总数)	(341/400)	(93/100)	(248/300)
高年资医师	83.80	80.20	85.00
(95% CI)	(81.70 85.90)	(75.95 84.45)*	(81.84 88.17)
χ^2 值	0.32	7.06	0.603
中年资医师	78.00	74.60	79.14
(95% CI)	(75.88 80.12)*	(72.34 76.86)*	(76.07 82.15)
χ^2 值	7.10	12.47	1.21
低年资医师	70.90	68.00	71.87
(95% CI)	(68.87 72.93)*	(65.22 70.78)*	(69.36 74.38)*
χ^2 值	24.06	19.91	9.96

与 Mask R-CNN 比较: * $P < 0.05$

3 讨论

胃癌是全球癌症相关死亡的第四大原因^[8],胃

癌的早期发现十分重要。内镜检查是公认发现 EGC 的标准方式,然而,WLI 下区分黏膜的微小变化颇有难度。放大内镜、NBI 等几种较先进的技术对于内镜设备和内镜医师均有一定要求,不适于常规筛查。

为解决上述问题,随着计算机技术发展,AI 辅助内镜诊疗应运而生。在胃癌筛查方面,Tang et al^[9]开发了一种用于检测 EGC 的实时 AI 系统,能以每张图片 15 ms 的速度进行诊断。该研究为回顾性多中心研究,不能反映 AI 的临床实用效果。Wu et al^[10]开发了一个名为 ENDOANGEL-LD 的系统,在视频实验中,准确率为 72.0% (72/100),敏感度为 100.0% (38/38),高于 4 位专家的水平(准确率 68.0%,敏感度 85.5%)。Wu et al^[11]还提出一项随机对照实验,评价 AI 系统提示盲点、动态监测 EGC 的性能。该系统输出预测的平均值 [标准差 (SD)] 总时长即为 230 (SD 60) ms,因此被设置为仅以每秒 2 帧的速度实时处理视频。该研究首次将视频按照相应帧数转化为图像,但仍未给出 AI 在该情况下的性能评价。

目前常采用的基于深度学习的目标检测算法,主要分为两大类。一类是李素琴等^[12]采用的以 YOLO 为代表的单阶段算法,具有较高的检测速度,但在精度上有待提升。另一类是 Tang et al^[13]采用的以 DCNN、Fast R-CNN 等为代表的两阶段算法,其相对于单阶段算法有着更高的检测精度,但存在训练时间过长、计算耗费时间、实验设备要求高等不足。同时,上述算法无法将特征图与原像素精准对齐,仅仅获得了粗糙的表述^[14]。

针对现有胃癌筛查研究的不足,笔者开发了基于 Mask R-CNN 的内镜下辅助诊断 EGC 新技术。Mask R-CNN 综合此前优秀的深度学习研究成果,在同一网络中,不仅可以又快又准地进行多目标检测,得到高准确率,还能精准地完成语义分割,获取目标轮廓^[15]。

本研究中的 Mask R-CNN 系统在 WLI 静态图片中的表现十分优异,准确率为 90.25%,敏感度为 91.06%,特异度为 89.01%,与病理学金标准诊断相比,差异无统计意义。在前瞻视频实验中,Mask R-CNN 系统能识别出所有视频中的 EGC 病灶 (100%)。与既往不同的是,为了进一步评估其性能,笔者将视频流转化为相应帧数的图片进行检测,并在加入了检测结果进行投票筛选的模块,有效减少误判。Mask R-CNN 系统诊断 EGC 的准确率、敏

感和特异度分别为 90.27%、84.86% 和 91.87%,采集数据并输出结果的延迟仅为 80 ms,最高可处理 35 帧/s 的实时视频。Mask R-CNN 系统在达到上述优秀性能的同时,可以满足既往研究^[16]中 EGC 病灶边界精准描绘的要求。在对照实验中,Mask R-CNN 系统的总体诊断能力不低于高年资医师。本研究表明,Mask R-CNN 系统能从正常、炎症、AGC 等多种不同程度的黏膜背景中识别诊断出 EGC。

本研究仍存在局限性:① 本研究收集的数据均来自本院内镜室,不能排除选择偏倚。② 本研究中的训练集均来自人工标记,标记结果可能受到医师水平影响。③ 本研究纳入胃部疾病种类有限,只针对 EGC 的识别诊断进行一般设计,并不能更好地区分黏膜病变的具体类型。在接下来的研究中,本课题组将进一步加入多种病变,以提高 Mask R-CNN 系统的临床适用性。④ 本研究没有将测试集加入训练集中再次学习,在自主学习方面有所欠缺。

综上所述,本课题组所开发的 Mask R-CNN 系统具有在 WLI 下实时检测 EGC 的优秀性能,有望在不发达地区内镜医师的培训和临床实践诊断 EGC 中发挥作用。

参考文献

- [1] 冯振游,胡孔旺,丁慧明,等. 全反式维甲酸与胃癌预后相关性及其机制研究[J]. 安徽医科大学学报,2022,57(8):1216-21.
- [2] Katai H, Ishikawa T, Akazawa K, et al. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001-2007) [J]. *Gastric Cancer*, 2018, 21(1): 144-54.
- [3] Zhao Y, Xue D, Wang Y, et al. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy [J]. *Endoscopy*, 2019, 51(4): 333-41.
- [4] Yang R, Yu Y. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis [J]. *Front Oncol* 2021, 11:638182.
- [5] Yao K, Anagnostopoulos G K, Ragunath K. Magnifying endoscopy for diagnosing and delineating early gastric cancer [J]. *Endoscopy* 2009, 41(5): 462-7.
- [6] Dixon M F. Gastrointestinal epithelial neoplasia: Vienna revisited [J]. *Gut* 2002, 51(1): 130-1.
- [7] Horiuchi Y, Aoyama K, Tokai Y, et al. Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging [J]. *Dig Dis Sci*, 2020, 65(5): 1355-63.

- [8] 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]. *CA Cancer J Clin*, 2021, 71(3): 209–49.
- [9] Tang D, Wang L, Ling T, et al. Development and validation of a real-time artificial intelligence-assisted system for detecting early gastric cancer: A multicentre retrospective diagnostic study [J]. *EBioMedicine*, 2020, 62: 103146.
- [10] Wu L, Xu M, Jiang X, et al. Real-time artificial intelligence for detecting focal lesions and diagnosing neoplasms of the stomach by white-light endoscopy (with videos) [J]. *Gastrointest Endosc*, 2022, 95(2): 269–80.
- [11] Wu L, He X, Liu M, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial [J]. *Endoscopy*, 2021, 53(12): 1199–207.
- [12] 李素琴, 吴练练, 宫德馨, 等. 基于YOLO算法和ResNet深度卷积神经网络的结直肠息肉检测(含视频) [J]. *中华消化内镜杂志*, 2020, 37(8): 584–90.
- [13] Tang D, 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]. *Front Oncol*, 2021, 11: 622827.
- [14] 温庭栋, 宋文爱, 赵莉, 等. 胃镜下早期胃癌计算机辅助分析研究综述 [J]. *计算机工程与应用*, 2021, 57(10): 39–47.
- [15] 李大军, 何维龙, 郭丙轩, 等. 基于Mask-RCNN的建筑物目标检测算法 [J]. *测绘科学*, 2019, 44(10): 172–80.
- [16] An P, Yang D, Wang J, et al. A deep learning method for delineating early gastric cancer resection margin under chromoendoscopy and white light endoscopy [J]. *Gastric Cancer*, 2020, 23(5): 884–9.

Detection of early gastric cancer in white light imagings based on region-based convolutional neural networks

Jin Jing^{1,2}, Zhang Qianqian^{1,2}, Bill Dong³, Ma Tao³, Wang Xi^{1,2}, Mei Xuecan^{1,2},
Song Shaofang⁴, Peng Jie⁴, Wu Aijiu⁴, Dong Lanfang³, Kong Derun^{1,2}

(¹Dept of Gastroenterology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022;

²Key Laboratory of Digestive Diseases of Anhui Province, Hefei 230022; ³School of Computer Science and Technology, High-tech Campus, University of Science and Technology of China, Hefei 230027;

⁴Hefei Zhongna Medical Instrument Co., Ltd, Hefei 230088)

Abstract Objective To develop an endoscopic automatic detection system in early gastric cancer (EGC) based on a region-based convolutional neural network (Mask R-CNN). **Methods** A total of 3 579 and 892 white light images (WLI) of EGC were obtained from the First Affiliated Hospital of Anhui Medical University for training and testing, respectively. Then, 10 WLI videos were obtained prospectively to test dynamic performance of the RCNN system. In addition, 400 WLI images were randomly selected for comparison with the Mask R-CNN system and endoscopists. Diagnostic ability was assessed by accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). **Results** The accuracy, sensitivity and specificity of the Mask R-CNN system in diagnosing EGC in WLI images were 90.25%, 91.06% and 89.01%, respectively, and there was no significant statistical difference with the results of pathological diagnosis. Among WLI real-time videos, the diagnostic accuracy was 90.27%. The speed of test videos was up to 35 frames/s in real time. In the controlled experiment, the sensitivity of Moks R-CNN system was higher than that of the experts (93.00% vs 80.20%, $\chi^2 = 7.059$, $P < 0.001$), and the specificity was higher than that of the juniors (82.67% vs 71.87%, $\chi^2 = 9.955$, $P < 0.001$), and the overall accuracy rate was higher than that of the seniors (85.25% vs 78.00%, $\chi^2 = 7.009$, $P < 0.001$). **Conclusion** The Mask R-CNN system has excellent performance for detection of EGC under WLI, which has great potential for practical clinical application.

Key words artificial intelligence; region-based convolutional neural network; endoscopy; early gastric cancer