论著。

# 基于深度学习的人工智能技术在早期胃癌诊断中的应用

王智杰 高杰 孟茜茜 杨婷 王则远 陈兴春 王东 李兆申

【摘要】目的 构建和验证一个用于早期胃癌自动识别的深度学习模型,旨在提高早期胃癌的识别和诊断水平。方法 从长海医院消化内镜中心数据库选取 2014 年 5 月至 2016 年 12 月期间 5 159张胃镜图像,其中包括早期胃癌1 000张,良性病变及正常图像4 159张。首先选取4 449张图像(其中早期胃癌图像768 张,其他良性病变及正常图像3 681张)用于深度学习模型的训练。然后将剩余的710张图像用于模型的验证,同时再交给4 名内镜医师进行诊断。最后统计相关结果。结果深度学习模型用于早期胃癌诊断的准确率89.4%(635/710)、敏感度88.8%(206/232)、特异度89.7%(429/478),每张图像的诊断时间为(0.30±0.02)s,均优于相比较的4 名内镜医师。结论 本研究构建的深度学习模型用于早期胃癌的诊断具有较高的准确率、特异度和敏感度,可在胃镜检查中辅助内镜医师进行实时诊断。

【关键词】 诊断; 早期胃癌; 人工智能; 深度学习基金项目:国家自然科学基金项目(81370589)

Artificial intelligence based on deep learning for automatic detection of early gastric cancer Wang Zhijie\*, Gao Jie, Meng Qianqian, Yang Ting, Wang Zeyuan, Chen Xingchun, Wang Dong, Li Zhaoshen.
\*Department of Gastroenterology, Changhai Hospital, Naval Medical University, Shanghai 200433, China Corresponding author: Wang Dong, Email:dr.dongwang@qq.com

[Abstract] Objective To develop and validate a model based on deep learning for automatic diagnosis of early gastric cancer (EGC) to improve detection and diagnosis of EGC. Methods A total of 5 159 images (including 1 000 images of EGC and 4 159 images of other benign lesions or normal patients) obtained from May 2014 to December 2016 were collected from endoscopic database in changhai Hospital. Then 4 449 images were selected randomly for a deep convolutional neural network (CNN) training, of which 768 were diagnosed as EGC and 3 681 diagnosed as other benign lesions or normal. The remaining 710 images were used to test the model by comparing with diagnostic results of four endoscopists. Results The deep learning model showed accuracy of 89.4% (635/710), sensitivity of 88.8% (206/232) and specificity of 89.7% (429/478) for EGC. The mean time required for diagnosis was  $0.30\pm0.02$  s. The performance of the model was superior to that of four endoscopists. Conclusion The model based on deep learning has high accuracy, sensitivity and specificity for detecting EGC, which could assist endoscopists in real-time diagnosis.

[Key words] Diagnosis; Early gastric cancer; Artificial intelligence; Deep learning Fund program; National Natural Science Foundation of China (81370589)

据统计,目前胃癌是我国癌症患者死亡的第二大病因,2015年我国约有49.8万人死于胃癌,平均

DOI: 10.3760/cma.j.issn.1007-5232.2018.08.004

作者单位:200433 上海,海军军医大学附属长海医院消化内科(王智杰、高杰、孟茜茜、杨婷、王东、李兆申);北京诺道认知医学科技有限公司(王则远、陈兴春)

通信作者:王东,Email:dr.dongwang@qq.com

每天1 364人<sup>[1]</sup>。尽管我国胃癌发病率及死亡率近年来已呈现显著下降趋势,但由于人口增长和严重老龄化,新发胃癌的患者数量依然庞大(每年约 68 万例)并且仍在不断增加。有研究显示,我国胃癌人群 5 年相对生存率仅为 27. 4%<sup>[2]</sup>,其重要原因在于大多数患者在初次确诊时已处于肿瘤晚期,这给个人和家庭造成了极大的负担,也对国计民生造成

了很大的不良影响。得益于较高的早期诊治率,同为胃癌高发地区的韩国和日本患者 5 年生存率则分别高达 64.6%和 71.5% [3-4]。

据相关资料显示,我国早期胃癌的诊治率约为10%<sup>[5-7]</sup>,远低于韩国(50%)和日本(70%),有着巨大的进步空间。因此,提升早期诊治能力无疑是提高我国胃癌患者生存率的有效策略。然而患者数量巨大、内镜医师数量短缺、医师早癌诊断技术和经验不足、先进内镜设备缺乏等问题,严重影响了我国早期胃癌诊断水平的提高,这些问题在基层医疗机构显得尤为突出。

因此,我们对人工智能(artificial intelligence, AI)技术在内镜图像识别中的应用进行了初步探索。本研究构建和验证了一个用于早期胃癌自动识别的深度学习<sup>[8]</sup>模型,旨在提高早期胃癌的识别和诊断水平。

# 资料与方法

#### 一、病例资料

5 159张胃镜图像全部从长海医院消化内镜中心数据库选取,时间范围为 2014 年 5 月至 2016 年 12 月,其中包括早期胃癌图像1 000张,良性病变及正常图像4 159张,良性病变包括胃溃疡、胃息肉、胃黏膜下隆起和糜烂性胃炎 4 种。所有纳入研究的病变均经过活检或手术病理确诊同时病灶范围明确。其中,早期胃癌的胃镜图像均从我院行内镜下黏膜剥离术 (endoscopic submucosal dissection, ESD)的 330 名患者资料中选取,术后病理诊断包括低级别上皮内瘤变、高级别上皮内瘤变以及局限于黏膜或黏膜下层的癌。考虑算法及临床实际,本研究将上述 3 种 ESD 术后病理结果均归为"早期胃癌",因此本文所称"早期胃癌"并不同于目前公认的早期胃癌定义[9]。

另外,所有图像均在白光、非放大模式下拍摄, 所使用的胃镜有 Olympus 公司的 260 和 290 系列以 及 Fujinon 公司的 580 和 590 系列。

## 二、模型构建和验证

1.图像标注:内镜医师以病理为标准将所有图像分成早期胃癌和其他(包括良性病变和正常)两个大类,同时对病变图像中的病灶范围进行精准标注。

2.图像预处理:算法工程师将标注好的图像进 行裁剪、颜色空间变换、去噪、图像形态学操作以及 归一化等系列处理,从而消除人为和环境干扰,更好地展示图像特征,增强算法的鲁棒性。另外将所有图像数据分成10个子样本,8个子样本用于训练(训练集),2个子样本保留用于验证模型(测试集)。其中,训练集中共有4449张图像(早期胃癌图像768张,其他良性病变及正常图像3681张,均从各自的图像库中随机选取),其余为测试集。

#### 3.构建和训练模型。

算法工程师采用自主研发的并行计算平台搭载 19 层卷 积神经 网络(convolutional neural networks, CNN)构建多个深度学习模型,分别将训练集投入模型进行训练,通过观察和比较确定一个最优模型。

其中 CNN 包括卷积层、池化层和全连接层,其工作流程如下:首先,卷积层进行卷积操作时,将卷积核在输入图像上以 stride 步长进行滑动,按照 wx+b 计算,w 对应卷积核权值,x 对应不同的数据窗口,b 表示各卷积层偏置。激活函数可对上层传来的数据做处理,公式为:f(x)= max(0,x)。然后,通过卷积提取特征之后,池化层实现对特征的进一步汇总,这里使用最大池化,见图 1。最后,全连接层将上层输入的多维矩阵变换成二维矩阵,并与该层权重做矩阵相乘运算。Softmax 层是神经网络的输出层,负责输出每一张图像属于每种分类的概率。CNN 模型对胃镜图像进行分类识别过程见图 2。

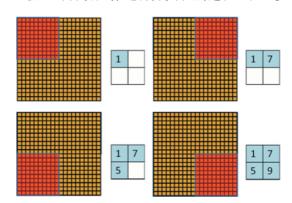


图 1 池化原理 其中红色区域代表 2×2 的池化核,以步长为 2 进行池化,整体的黄色区域代表传统胃镜的原始图像,数字代表红色区域中的最大值,整个过程是将4×4 的矩阵图转化成 2×2 的矩阵

#### 4.验证模型。

利用测试集来测试训练得到的模型,以此来做为评价分类器的性能指标。一般情况下,训练集迭代次数30次或50次,每次迭代后用测试集验证,随

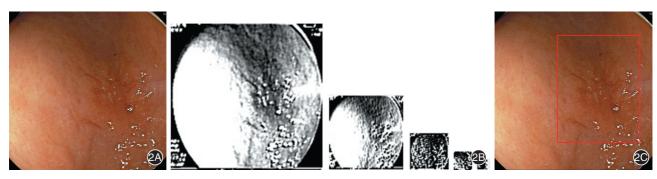


图 2 卷积操作降维过程 2A:传统胃镜拍摄的视频截图原图;2B:通过卷积操作后生成的特征图,共进行了四次卷积操作;2C:卷积层的输出,实现病灶的智能识别

着迭代次数的增加模型精度趋于稳定。

5.模型和内镜医师的比较。

选择 4 名本院内镜医师,分高年资和低年资两组,其中高年资内镜医师的选择标准为从事内镜工作大于 5 年,胃镜检查超过5 000例,低年资内镜医师标准为从事内镜工作小于 3 年,胃镜检查例数为1 000到2 000例。4 名医师分别独立地对电脑屏幕所展示的测试集图像进行识别和判断,同时记录其诊断每张图像所需的时间(仅分类病种,不包括识别病灶范围的时间)。同时模型对相同的测试集进行识别,记录结果。

## 三、统计分析

采用 SPSS 21 统计软件处理数据,计量资料采用  $\bar{x}\pm s$  表示,计数资料采用卡方检验,P<0.05 认为 差异有统计学意义。

#### 结 果

测试集共有710 张图像,其中早期胃癌图像232 张,其他良性病变及正常图像478 张。232 张早期胃癌图像中,深度学习模型正确识别206 张(敏感度为88.8%),478 张其他病变图像正确识别429张(特异度为89.7%),总的准确率为89.4%。另外,每张图像的诊断时间为(0.30±0.02)s。

而 2 名低年资内镜医师和 2 名高年资内镜医师

的敏感度分别为 66.8%、79.3%和 80.6%、84.5%,特异度分别为 71.5%、57.5%和 79.5%、85.3%,识别准确率分别为 70.0%、64.6%和 79.8%、85.1%,平均诊断时间分别为(2.52±1.32)s、(1.93±1.51)s和(2.15±1.41)s、(3.10±2.12)s。卡方检验表明,模型组与内镜医师相比有较高的敏感度、特异度和准确率,详细结果见表 1。

# 讨 论

本研究对人工智能图像识别技术在消化内镜 领域内的应用进行了初步探索。通过回顾性收集的5 159张胃镜图像,整理后投入深度学习模型进行训练和验证,测试结果显示其诊断早期胃癌的敏感度为88.8%,特异度为89.7%,准确率为89.4%,每张图像的诊断时间为(0.30±0.02)s。卡方检验结果表明,模型组的敏感度、特异度、准确率均高于内镜医师组,总体表现令人满意。

早期发现、早期诊断和早期治疗是降低癌症死亡率并提高生存率的主要策略。鉴于我国胃癌诊治的严峻形势,2014年由中华医学会消化内镜学分会联合中国抗癌协会肿瘤内镜学专业委员会牵头,组织发布了中国早期胃癌筛查及内镜诊治共识意见<sup>[10]</sup>,早期胃癌的诊治得到了医学界前所未有的重视。此背景下,以ESD为代表的内镜治疗技术近年

表 1	一 一 知 紀 上 4	名医师诊断早期胃瘍比较
	- ハ I 谷 IV ニ I 4	

诊断者	敏感度	特异度	准确率	诊断时间(s,x±s)
计算机模型	88. 8% ( 206/232 )	89. 7% (429/478)	89. 4% (635/710)	0.30±0.02
低年资医师1	66. 8% ( 155/232) <sup>a</sup>	71. 5% ( 342/478 ) <sup>a</sup>	70.0% (497/710) a	2. 52±1. 32
低年资医师 2	79. 3% ( 184/232) <sup>b</sup>	57. 5% ( 275/478 ) <sup>a</sup>	64. 6% (459/710) <sup>a</sup>	1.93±1.51
高年资医师 1	80. 6% ( 187/232 ) °	79. 5% ( 380/478 ) <sup>a</sup>	79. 8% ( 567/710) <sup>a</sup>	2. 15±1. 41
高年资医师 2	84. 5% ( 196/232 ) <sup>d</sup>	85. 3% (408/478) e	85. 1% (604/710) °	3. 10±2. 12

注:与计算机模型组比较, aP<0.001, bP=0.005, cP=0.014, dP=0.172, eP=0.040

来在我国发展迅速[11],相比之下早期胃癌的诊断水平却仍然低下且提高缓慢。造成这一现象的原因很多。首先是先进内镜设备资源的不足和内镜医师的缺乏,据 2015 年的一项全国性调查研究显示,我国消化内镜人力资源和设备资源总体不足且地方分布不均,优质医疗资源集中在少数大医院,这一问题在基层医疗机构显得尤为突出[12]。另外一个重要原因,我们认为在于内镜医师本身。胃镜工作量大、内镜操作不规范、早癌意识薄弱以及诊断经验欠缺等因素,使得胃镜检查过程中内镜医师对早期胃癌"视而不见、见而不识、识而不辩"的现象较为普遍。而由于医疗资源的矛盾涉及诸多因素,短期内难以得到解决。因此,如何能充分利用有限的医疗资源,快速、有效地提高我国的早期胃癌诊断能力,正是我们现在所思考的。

AI的出现为解决这一问题提供了新的思路。 在医疗方面,尽管现在绝大部分 AI 尚未真正进入临 床实践.但 AI 已经展现了巨大的潜力和广阔的前 景。尤其在医学图像识别方面,目前 AI 取得的研究 成果令人欢欣鼓舞。2017年 JAMA 刊文,有研究团 队将深度学习算法应用于乳腺癌淋巴转移的自动 诊断,对270 张淋巴病理切片(其中110 张有癌细胞 转移,160 张无转移)进行学习后建立了算法,然后 与一个由 11 名病理医师组成的专家小组进行对比 测试,结果显示部分算法的诊断水平优于病理医 师[13]。同年,Nature 以封面形式报道了 AI 在皮肤 癌诊断方面的研究成果,斯坦福大学的研究人员利 用深度神经网络对2000多种不同类型,近13万张 的皮肤病变图像进行学习后,开发的算法在皮肤癌 诊断上的表现与皮肤科医师相当,综合准确率达 91%[14]。在消化内镜图像识别领域,我国台湾学者 Chen 等[15]报道了利用深度神经网络对结肠息肉的 性质进行计算机辅助诊断,研究收集了1476张肿瘤 性息肉和681 张增生性息肉图像用于算法的训练和 验证。最终,其识别敏感度和特异度分别达到 96.3%及78.1%。

近年来,我科对 AI 在内镜图像识别领域的应用进行了积极探索<sup>[16-17]</sup>。本研究采用的深度学习不同于传统的图像识别模式。传统的图像识别方法,如 Support Vector Machine (SVM), Decision Trees (DTs)等<sup>[18-19]</sup>,需要专家对图像特征进行手工设计,然后算法再从中提取特征,这导致其只能识别特定病变,而且一旦给予的特征不够,将得不到满意的

识别结果。同时手工设计主要依靠设计者的先验知识,很难利用大数据的优势。而目前关于医学图像识别最新的研究<sup>[20-22]</sup>往往采用基于 CNN 的深度学习算法。深度学习<sup>[8]</sup>无需先验知识,可直接从图像大数据中对最有预测性的特征进行自动学习并对图像进行分类识别,给予它的数据量越大,深度学习的优势更能够发挥,识别精度也越高。已有研究对比了深度学习和非深度学习方法用于前列腺癌磁共振图像的识别,结果表明深度学习识别效果更优<sup>[23]</sup>。显然,深度学习对于处理数据的方式决定了它在如今的大数据时代将有更为广阔的发展空间。

另外,本研究所用胃镜图像均为白光、非放大 模式下拍摄所得。原因在于我国早期胃癌诊断的 主要问题在于筛查,即非特异性症状及无症状人群 的胃镜检查(我们最大的内镜检查人群),术前不知 道患者有无病变,主要目的是通过胃镜去寻找病 变。精查即使不能准确分型,也可通过内镜活检弥 补,而筛查则没有弥补手段,漏诊后果严重。不同 于食管癌筛查采用的 NBI、蓝激光成像等模式,胃腔 由于空间巨大导致特殊光的亮度相对不足,且拍摄 角度也差别较大,因此白光更适合于胃癌的筛查。 同时上文中,我国台湾学者的研究[15]采用的图像均 在 NBI 和最大放大倍数下拍摄,而实际上电子染色 及放大内镜并未在多数医院得到普及,且很多内镜 医师操作和识别经验不足。此外还有几项研究表 明,尽管利用 NBI 对息肉进行内镜光学诊断可以取 得类似病理学诊断的准确率,但操作者仅限于专家 级内镜医师且需要一定时间的学习,所以除了部分 高水平医院外不推荐其他普通医院将 NBI 作为息 肉性质诊断的常规手段[24-25]。因此我们认为对白 光、非放大图像进行自动识别的实际适用范围更 大,更具临床意义。同时我们选择最难筛查的胃癌 作为研究对象,去解决难题,也更符合中国目前的 薄弱点。

研究结果还显示了 AI 图像诊断的另一个重要优势即实时性。算法对一张图像做出诊断仅需 (0.30±0.02)s,这是日后胃镜检查过程中协助内镜 医师进行实时辅助诊断的基础。

最后,本研究的不足之处在于用于训练的图像数据(包括早期胃癌和用作干扰的其他病种数量)相对较少,模型的初始构建和训练仍需人工进行标注,精度有待进一步提升。其次,由于是回顾性地

收集图像数据,病灶拍摄的角度相对单一,同时对图像质量进行了筛选,剔除了部分不典型或干扰多的图像,因此影响了模型的临床适用性。目前该模型只能对初步整理后的内镜图像进行识别,实际胃镜检查过程中则会受术前准备、病灶拍摄的角度、光源等诸多干扰因素影响,诊断水平可能降低。接下来的研究中,我们将充分利用本中心海量数据的优势,增加投入训练的数据量,同时对图像进行前瞻性收集,拍摄远近、正侧等多个不同角度的病灶图像,尽量丰富训练集,不断提升算法的诊断效能。同时还将对相关配套软件进行研发,加快使其投入临床实践。

本研究构建的深度学习模型用于早期胃癌的诊断具有较高的准确率、特异性和敏感性,将来可在胃镜检查中辅助内镜医师进行实时诊断。AI 让基层内镜医师获得跟资深专家同质化的诊断效果成为可能,而目前存在的医疗资源分配不均、基层内镜医师技术水平参差不齐等问题,正是推进 AI 发展的动力所在。总之,我们认为 AI 技术在内镜图像识别领域大有可为。

#### 参考文献

- [ 1 ] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China,
   2015 [ J ]. CA Cancer J Clin, 2016, 66 (2): 115-132. DOI:
   10.3322/caac.21338.
- [2] Zeng H, Zheng R, Guo Y, et al. Cancer survival in China, 2003-2005; a population-based study [J]. Int J Cancer, 2015, 136(8):1921-1930. DOI: 10.1002/ijc.29227.
- [ 3 ] Matsuda T, Ajiki W, Marugame T, et al. Population-based survival of cancer patients diagnosed between 1993 and 1999 in Japan: a chronological and international comparative study[J]. Jpn J Clin Oncol, 2011,41(1):40-51. DOI: 10.1093/jjco/hyq167.
- [4] Oh CM, Won YJ, Jung KW, et al. Cancer statistics in Korea; incidence, mortality, survival, and prevalence in 2013 [J]. Cancer Res Treat, 2016, 48 (2): 436-450. DOI: 10.4143/crt.2016.089.
- [5] Baptista V, Singh A, Wassef W. Early gastric cancer; an update on endoscopic management[J]. Curr Opin Gastroenterol, 2012, 28(6):629-635. DOI: 10.1097/MOG.0b013e328358e5b5.
- [ 6 ] Bu Z, Ji J. Controversies in the diagnosis and management of early gastric cancer [J]. Chin J Cancer Res, 2013,25(3):263-266. DOI: 10.3978/j.issn.1000-9604.2013.06.15.
- [7] 上海市早期胃癌临床协作组. 上海不同等级 10 个医疗机构早期胃癌的筛选结果比较[J].中华消化内镜杂志,2007,24(1):19-22. DOI: 10.3760/cma.j.issn.1007-5232.2007.01.005.
- [8] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015,521(7553):436-444. DOI: 10.1038/nature14539.

- [ 9 ] Association JGC. Japanese classification of gastric carcinoma-2nd English edition-[J]. Gastric Cancer, 1998,1(1):10-24. DOI: 10.1007/s101209800016.
- [10] 中华医学会消化内镜学分会,中国抗癌协会肿瘤内镜专业委员会.中国早期胃癌筛查及内镜诊治共识意见(2014年,长沙)[J].中华消化内镜杂志,2014,31(7):361-377. DOI: 10.3760/cma.j.issn.1007-5232.2014.07.001.
- [11] Cai MY, Zhou PH, Yao LQ. Current status of endoscopic resection in China[J]. Dig Endosc, 2012, 24 (Suppl 1): 166-171.
  DOI: 10.1111/j.1443-1661.2012.01268.x.
- [12] 王洛伟,辛磊,林寒,等.中国消化内镜技术发展现状[J].中华消化内镜杂志,2015,32(8):501-515. DOI: 10.3760/cma.j.issn.1007-5232.2015.08.001.
- [13] Ehteshami BB, Veta M, van Diest P J, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer [J]. JAMA, 2017, 318 (22);2199-2210. DOI: 10.1001/jama.2017.14585.
- [ 14 ] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks [ J ]. Nature, 2017,542(7639):115-118. DOI: 10.1038/nature21056.
- [15] Chen PJ, Lin MC, Lai MJ, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis [J]. Gastroenterology, 2018, 154 (3): 568-575. DOI: 10.1053/j.gastro.2017. 10.010.
- [16] 张敏敏, 金震东, 蔡哲元, 等. 计算机辅助判别超声内镜图像诊断胰腺癌的实验研究[J]. 中华消化内镜杂志,2009,26(04): 180-183. DOI: 10.3760/cma,i,issn.1007-5232, 2009. 04.004.
- [17] 朱建伟, 王雷, 储怡宁, 等. 图像分析技术在自身免疫性胰腺炎与慢性胰腺炎超声内镜鉴别诊断中的应用[J]. 中华消化内镜杂志, 2015, 32(4): 225-228. DOI: 10.3760/cma.j. issn.1007-5232.2015.04.002.
- [18] Erickson BJ, Korfiatis P, Akkus Z, et al. Machine learning for medical imaging [J]. Radiographics, 2017, 37 (2): 505-515. DOI: 10.1148/rg.2017160130.
- [19] Orrù G, Pettersson-Yeo W, Marquand AF, et al. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review [J]. Neurosci Biobehav Rev, 2012,36(4):1140-1152. DOI: 10.1016/j.neubiorev.201-2.01.004.
- [20] Lekadir K, Galimzianova A, Betriu À, et al. A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound[J]. IEEE J Biomed Health Inform, 2017,21(1):48-55. DOI: 10.1109/JBHI.2016.2631401.
- [21] Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network [J]. IEEE Trans Med Imaging, 2016,35(5):1207-1216. DOI: 10.1109/TMI.2016.2535865.
- [22] Rajkomar A, Lingam S, Taylor AG, et al. High-throughput classification of radiographs using deep convolutional neural networks
  [J]. J Digit Imaging, 2017, 30(1): 95-101. DOI: 10.1007/s10278-016-9914-9.

- [23] Wang X, Yang W, Weinreb J, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning [J]. Sci Rep, 2017,7(1):15415. DOI: 10.1038/s41598-017-15720-y.
- [24] Rex DK. Narrow-band imaging without optical magnification for histologic analysis of colorectal polyps [J]. Gastroenterology, 2009, 136(4):1174-1181. DOI: 10.1053/j.gastro.2008. 12.009.
- [25] Rees CJ, Rajasekhar PT, Wilson A, et al. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study [J]. Gut, 2017,66(5):887-895. DOI: 10.1136/gutjnl-2015-310584.

(收稿日期:2018-02-09) (本文编辑:唐涌进)