

# 基于多尺度卷积神经网络的胶囊内窥镜出血点快速识别\*

谢雪娇<sup>1</sup>, 陆 枫<sup>2</sup>, 李书展<sup>2</sup>, 周 到<sup>3</sup>

(1. 华中科技大学同济医学院附属同济医院, 湖北 武汉 430030;

(2. 华中科技大学计算机学院大数据技术与系统国家工程研究中心, 湖北 武汉 430074;

3. 中南民族大学生物医学工程学院认知科学国家民委重点实验室, 湖北 武汉 430074)

**摘 要:**无线胶囊内窥镜技术 WCE 已广泛应用于胃肠道疾病辨识中,然而随之产生的海量影像学数据为医生阅片带来了沉重负担。针对 WCE 图像出血点自动识别中存在的颜色和纹理特征不明显、易与正常器官混淆,细节特征模糊与病灶尺寸大小不一,以及含有较多杂质等问题,提出残差多尺度全卷积神经网络对含出血点的 WCE 图像进行快速分类辨识。通过引入残差学习网络中跳跃连接以及 Inception 网络中多尺度卷积核的思想,使简洁的网络结构能够有效提取图像的各类病灶细节特征。从实验结果看,网络的灵敏度达到 98.05%,特异度达到 97.67%,准确率达到 97.84%,优于 ResNet50 和 Inception-v4 网络,识别效率高,且收敛速度更快,计算性能有所提升。总之,该网络兼顾出血点识别效率和性能,实用性较强。

**关键词:**深度学习;无线胶囊内窥镜;卷积神经网络;残差网络;多卷积核

**中图分类号:**TP319.56

**文献标志码:**A

**doi:**10.3969/j.issn.1007-130X.2020.10.015

## Quick bleeding point detection in WCE image based on multi-scale convolutional neural network

XIE Xue-jiao<sup>1</sup>, LU Feng<sup>2</sup>, LI Shu-zhan<sup>2</sup>, ZHOU Dao<sup>3</sup>

(1. Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030;

2. National Engineering Research Center for Big Data Technology and System,

School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074;

3. Key Laboratory of Cognitive Science, School of Biomedical Engineering,

South-Central University for Nationalities, Wuhan 430074, China)

**Abstract:** With the full application of Wireless Capsule Endoscopy (WCE) in the detection of gastrointestinal diseases, screening out a small number of lesion images from the massive imaging data brings a heavy burden to doctors. To solve the problems existing in the automatic detection of WCE images, such as inconspicuous colour and texture features, ease of being confused with healthy organs, fuzzy detail features and different sizes of lesions, and high impurities, we propose a residual-based multi-scale fully convolutional neural network to classify and detect lesions in WCE image. By introducing the concepts of skip connection in residual learning network and multi-scale convolution kernel in the inception network, the model can effectively extract the detailed features of various lesions in the image. The experimental results show that the sensitivity of the model reaches 98.05%, the specificity reaches 97.67%, and the accuracy reaches 97.84%. It is better than the classical deep residual network ResNet50 and the standard width multi-scale Inception-v4 algorithm. The model has high recognition rate, fast

\* 收稿日期:2020-04-30;修回日期:2020-07-15

基金项目:中南民族大学中央高校科研基本业务费(CZY20039,CZQ18013)

通信作者:周到(zhoudao@mail.scuec.edu.cn)

通信地址:430074 湖北省武汉市洪山区民族大道中南民族大学生物医学工程学院

Address: School of Biomedical Engineering, South-Central University for Nationalities, Minzu Avenue, Hongshan District, Wuhan 430074, Hubei, P. R. China

convergence speed, and improved computing performance. In short, the algorithm model takes into account the efficiency and performance of bleeding point detection, and has strong practicability.

**Key words:** deep learning; wireless capsule endoscopy; convolutional neural network; residual network; multiple convolution kernels

## 1 引言

全球最新的癌症调查报告中显示,癌症在胃部、直肠、食道的发病率分别高达 5.7%,3.9%,3.2%,在所有癌症发病率中分别排第 6,8,9 位,3 种癌症的致死率加起来达到 16.7%<sup>[1]</sup>。显而易见,对消化道疾病的早期发现与及时治疗是有效防止病变恶化的关键。近年来,无线胶囊内窥镜 WCE(Wireless Capsule Endoscopy)已被广泛应用于消化道检查,相对于传统的侵袭性方法 PE(Push Enteroscopy)和利用放射学检查的 CT、钡餐等方法,WCE 不仅对人体的伤害小,还具有更高的病变检出率,在消化道疾病的诊疗方面做出了巨大贡献<sup>[2]</sup>。医生可以通过内窥镜胶囊拍摄的 WCE 图像进行分析和诊断,以判断受检者的消化道内是否存在病变。一般胶囊内镜每秒拍摄 2 幅图像,一个受检者的整个检测过程需要 6~8 h,共产生 50 000~70 000 幅 WCE 图像数据。这些海量 WCE 图像为医生诊断带来了沉重负担。按照一秒读一幅图像来算,读完 60 000 幅图像大约需要 15 h 左右。即便是专业经验丰富的医生也至少需要 2~3 h 的时间来完成阅片。此外,包含病变的图像通常只占有 WCE 图像中的小部分,很容易被忽略。且有些病症由于微小或被遮挡而难以识别,使得漏诊、误诊的情况依然存在。可见,快速、准确地辨识 WCE 图像病症是受检患者与医护人员的迫切需求。

目前已有利用人工智能方法分析 WCE 图像实施辅助影像病灶识别的研究<sup>[3]</sup>,主要分为 2 类,一类是基于传统机器学习方法的 WCE 图像病灶识别,另一类是基于深度学习方法的识别。

传统的机器学习方法可以大致分为 2 个步骤<sup>[4-7]</sup>,先是使用传统的特征提取方式提取 WCE 图像的颜色、纹理或形状等特征<sup>[7]</sup>;然后依据提取的特征采用线性回归或者支持向量机 SVM(Support Vector Machine)等传统机器学习方法构造相应的分类器进行分类。研究的重点主要集中在使用何种方法提取图像的何种特征以及分类器的选择上。这些实验通常涉及的样本数量较少,参与实

验的医学影像数据大多经过去噪、正则化和图像增强等预处理,以及对感兴趣的区域 ROI(Region of Interest)进行手工标注。这类方法在处理原始质量不高的大规模数据时,普适性往往受限。

以卷积神经网络 CNN(Convolutional Neural Network)为基础的 WCE 图像病灶识别研究目前也已经陆续展开<sup>[8-11]</sup>。Li 等人<sup>[8]</sup>针对出血图像数量与正常图像数量不平衡及 WCE 图像的颜色、纹理和亮度的变化等问题,提出了一种基于卷积神经网络的肠道出血识别模型。针对大约 1 300 幅出血图像和 40 000 幅正常图像识别,其 F1 值达到了 98.87%。Coelho 等人<sup>[9]</sup>基于 U-Net 网络对小肠出血点进行识别。Fan 等人<sup>[10]</sup>采用 AlexNet 进行 WCE 图像小肠溃疡和糜烂识别,准确率高达 95%。Aoki 等人<sup>[11]</sup>使用 5 360 幅含溃疡病症的 WCE 图像,训练了一个基于 SSD(Single Shot multibox Detector)模型的神经网络,经过包含 10 440 幅小肠图像(440 幅图像含溃疡)的独立测试集的测试,灵敏度、特异度和准确率分别为 88.2%,90.9%和 90.8%。这些方法大多使用了现成的网络结构,未对网络结构进行个性化修改。

然而,作为医学影像的 WCE 图像具有一些严重影响识别分类效果的特征,比如,阴性与阳性样本的颜色和纹理特征非常不明显,消化道残留物对识别影响较大等<sup>[12]</sup>;又如,针对同一种病灶,有的病灶尺寸非常微小,只占几个像素点;而有的病灶,如溃疡或者糜烂,尺寸又很大。现成的网络不能很好地处理这些问题,致使在实际应用时准确率大幅度下降。为了提高判断与预测的准确性,现有神经网络不断以几乎恒定的计算开销增加网络的深度和宽度<sup>[13,14]</sup>。为了使 WCE 出血点的识别能够安装到可移动设备上实现边检查边识别,还需要在保证准确率的同时,降低网络结构的复杂度,确保 WCE 图像的快速识别,以便及时发现病灶点,减轻医生负担。

本文分析了 WCE 图像中现有卷积神经网络对出血点进行识别可能存在的问题,提出了残差多尺度全卷积神经网络,结合残差的思想提取和保留图像的浅层特征,采用多尺度卷积核对不同尺度出血点特征进行有效提取。实验表明,该网络可以对

出血点细节特征进行像素级的有效识别,准确率、灵敏度、特异度分别达到 97.87%, 98.05%, 97.67%, 优于经典深度残差网络 ResNet50<sup>[13]</sup> 和经典多尺度 Inception-v4 网络<sup>[14]</sup>, 且计算复杂度更小, 收敛更快。

## 2 WCE 图像出血点识别分析

WCE 图像与自然景物图像有很大的不同, 首先, 它来源于胶囊内窥镜在体内的拍摄, 需要通过无线发射装置向体外传输。考虑到能耗等问题, 其分辨率只有  $480 \times 480$  像素。此外, WCE 图像通常还受到如下 3 种情况影响。

首先, 如图 1 所示, 由于无线胶囊内窥镜是在人体器官腔内部近距离拍摄, 并且拍摄的对象是人体的器官组织, 所以图像的色彩变化很小, 图像的颜色和纹理变化也十分不明显。这需要网络对细节和易混淆的特征进行灵敏度很高的捕捉, 本文称之为像素级特征的提取。

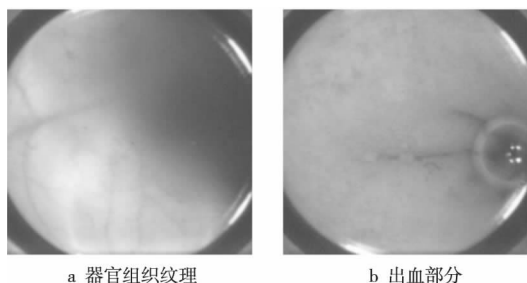


Figure 1 WCE images of the bleeding parts not much different from the texture of the organ tissues

图 1 出血部分与器官组织纹理区别不大的 WCE 图像

其次, WCE 图像中病灶的尺寸差别会非常大。其原因有 2:

(1) 很多出血点的尺寸非常微小, 只占几个像素点, 而有的出血点又很大, 比如溃疡或者糜烂, 如图 2 所示;

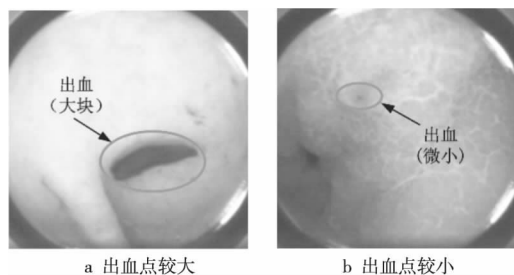


Figure 2 WCE images with different bleeding sizes

图 2 含尺寸大小不一出血点的 WCE 图像

(2) 由于 WCE 图像是近距离拍摄的, 同样的病灶从不同的角度拍摄会引起较大尺寸变化。此

时, 需要网络既能对像素级特征进行判别, 又能对大块特征进行抽象和提取。

最后, 如图 3 所示, WCE 图像中不可避免地会存在各类气泡或者食物残渣等消化道杂质, 并且这些杂质还极易附着在镜头上, 影响神经网络对特征的抽取和识别。

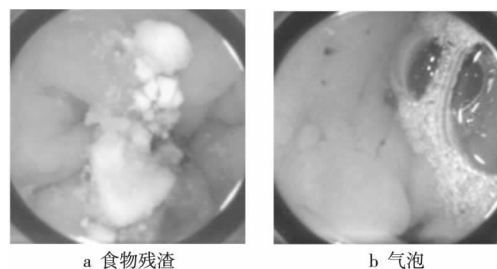


Figure 3 WCE images containing food debris and air bubbles

图 3 包含食物残渣和气泡的 WCE 图像

## 3 基于残差的多尺度全卷积网络

一般而言, 随着卷积层次的加深<sup>[13,15]</sup>, 得到的特征抽象程度越高。浅层特征属于细粒度特征, 语义信息较少, 上下文信息丰富, 不容易丢失特征且目标位置更加准确。对于微小出血病灶而言, 浅层特征是非常重要的特征。与之相反, 深层特征属于粗粒度特征, 抽象程度更高, 语义信息丰富, 上下文信息较少, 适于识别大病灶区域, 然而不利于识别目标的精确定位。如果加宽网络对多尺度特征进行辨识<sup>[14]</sup>, 则往往会带来计算量的急剧增加, 同时降低对细粒度特征的辨识度。据此, 利用 CNN 辨识 WCE 出血点需要在一定计算参数训练复杂度的约束下考虑模型细粒度和粗粒度等不同尺寸特征的有效提取。对此, 本文采用 Long 等人<sup>[15]</sup>提出的全卷积网络 FCN (Fully Convolutional Network) 为主干网, 以确保像素级特征的提取效果。FCN 擅于从抽象的特征中恢复出每个像素所属的类别, 即从图像级别的分类延伸到像素级别的分类。进而, 将残差思想运用至 FCN, 使得网络可以提取和保留 WCE 图像的浅层特征, 在与多尺度卷积得到的高级特征级联后, 输入至网络下层。最后, 采用了类似于 Inception 网络的多尺度卷积核, 加强网络对不同尺度出血点特征的有效提取。

### 3.1 基于残差思想的全卷积神经网络

本文以全卷积神经网络为主干, 设计了一种基于残差思想的多尺度全卷积神经网络, 整个网络的结构如图 4 所示。网络主要由 5 个多尺度卷积-跳

跃连接模块组成,网络的最后一层使用了 Softmax 函数,用作输出。

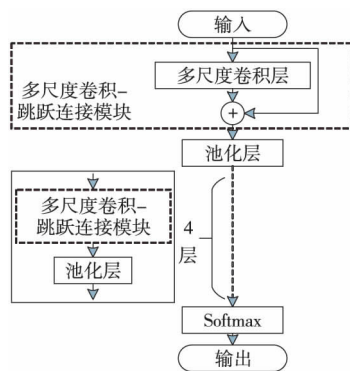


Figure 4 A multi-scale fully convolutional neural network structure based on residuals

图4 基于残差思想的多尺度全卷积神经网络结构

### 3.2 多尺度卷积-跳跃连接模块

多尺度卷积-跳跃连接模块主要由多尺度卷积模块与跳跃连接模块组成。多尺度卷积模块是一个包含5层卷积运算的小型卷积神经网络。

如图5所示,输入首先经过 $1 \times 1$ 的卷积运算(192个滤波器)处理后输入至下面3个并排的卷积层。这3个卷积层的滤波器数目都为64个,分别执行 $2 \times 2$ 、 $4 \times 4$ 和 $8 \times 8$ 的卷积运算,能捕获小、中、大3类不同尺度的特征。3个并行卷积层的输出特征图(Feature Map)连接在一起,经过 $1 \times 1$ 的卷积(192个)过滤后得到多尺度卷积模块的输出。跳跃连接模块将输入经过 $1 \times 1$ 卷积后,与多尺度卷积模块的输出叠加,再执行池化操作。该池化操作由过滤器大小为 $1 \times 1$ 、步长为2的卷积层执行。这里使用卷积层而不是传统的最大池化层,是为了让网络从训练数据中自动学习出最合适的降采样方法。跳跃连接模块采用与多尺度卷积模块相同的 $1 \times 1$ 卷积操作有2个作用:(1)有助于减少网络的参数总数;(2)能保证跳跃连接的输出维度与多尺度卷积模块的输出维度相同,可以执行叠加操作。

所有的卷积层在经过 ReLU 激活函数后都会进行批正则化 BN(Batch Normalization)处理。批正则化使网络的输出均值为0,标准差为1。实验表明,批正则化处理有助于加快网络收敛,也有助于限制过拟合的现象。

## 4 实验与测试

### 4.1 数据集的构成

从志愿者胶囊内镜图像中选取了胃部和小肠

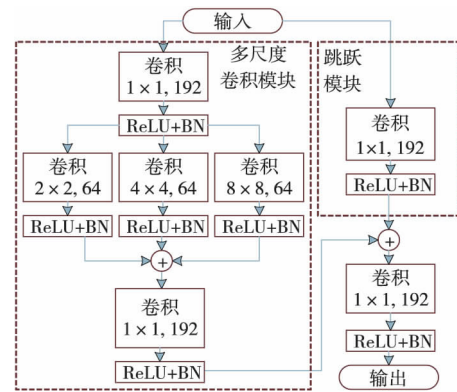


Figure 5 Multi-scale convolution-jump connection module structure

图5 多尺度卷积-跳跃连接模块结构

出血点图像8 624幅,出血点附近的阴性图像23 988幅,共32 612幅。为了平衡阳性样本和阴性样本的数量,对阳性样本进行旋转和镜像翻转处理。再分别随机挑出胃部阳性、阴性样本,以及小肠阳性、阴性样本各2 000幅。最后按照3:1的比例制作了训练集和测试集。

### 4.2 实验测试环境与方案

实验使用的计算机,CPU为Intel酷睿i7 6700HQ,内存大小8 GB,GPU为NVIDIA GeForce GTX 960M。实验使用前向均方根梯度下降算法(Root Mean Square Propagation)进行优化,初始学习速率为0.01。使用TensorFlow与Keras框架实现,CUDA的版本为9.0,cuDNN的版本为6.0。对比网络为应用广泛的深度残差网络ResNet50<sup>[13]</sup>和多尺度Inception-v4<sup>[14]</sup>。

### 4.3 评估指标

实验使用评估分类性能的指标包括准确率ACC(Accuracy)、灵敏度SEN(Sensitivity)和特异度SPC(Specificity),计算方法如式(1)~式(3)所示:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$SEN = \frac{TP}{TP + FN} \quad (2)$$

$$SPC = \frac{TN}{TP + FP} \quad (3)$$

其中,TP表示预测正确的阳性样本数量,FP表示预测错误的阳性样本数量,TN表示预测正确的阴性样本数量,FN表示预测错误的阴性样本数量。本文还使用了ROC曲线以及曲线下面积(AUC)来衡量网络的性能。

最后,为了比较各网络的收敛速率和计算性

能,绘制了迭代次数与损失的关系曲线图,直观地比较各网络的收敛速度。

4.4 实验结果与分析

在同一数据集上,本文提出的网络与 Res-Net50 和 Inception-v4 网络的测试结果对比如图 6 所示。实验结果表明,本文提出网络的准确率 ACC 达到 97.84%,灵敏度 SEN 达到 98.05%,特异度 SPC 达到 97.67%。

从图 6 可以看到,本文提出网络对 WCE 图像出血点分类的各项指标均优于 Inception-v4 和 ResNet50 网络的,兼顾了灵敏度和特异度,在确保漏检率的条件下保证了识别的准确率。与 Inception-v4 网络相比,本文的网络具有跳跃连接的结构,可以更好地保留 WCE 图像的初始特征。因此,在灵敏度上比 Inception-v4 提高了 3.6%。

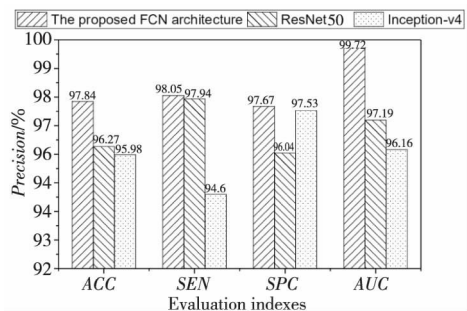


Figure 6 Performance comparison with ResNet50 and Inception-v4

图 6 本文网络与 ResNet50、Inception-v4 网络的性能对比  
本文网络在四重交叉验证下,得到的平均 AUC 值为 99.72%,其 ROC 曲线如图 7 所示。

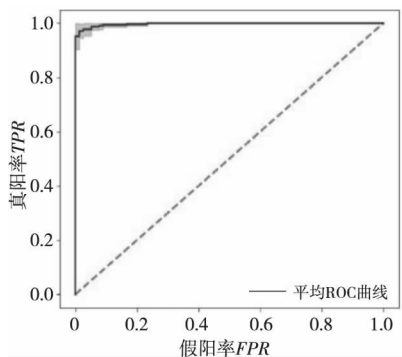


Figure 7 Average ROC curve under the quadruple cross validation set (where the grey part represents the confidence interval)

图 7 四重交叉验证集下平均 ROC 曲线  
(其中灰色部分代表置信区间)

从网络结构上看,如表 1 所示,本文提出的网络层数最少,参数约  $9 \times 10^6$  个;ResNet50 层数居中,参数约  $2.5 \times 10^7$ ;Inception-v4 最深,参数量也

最大,约  $4.3 \times 10^7$  个。

Table 1 Comparison of accuracies and the number of parameters

网络	准确率/%	参数量
本文	97.84	$9 \times 10^6$
ResNet50	96.27	$2.5 \times 10^7$
Inception-v4	95.96	$4.3 \times 10^7$

从计算复杂度上看,本文网络将浮点运算的数量降低了一个数量级。训练的时间也减少了一大半,整个网络训练的过程大约需要 2 h。

迭代次数与损失关系曲线图如图 8 所示,当迭代次数超过 10 000 次时,网络其实已经趋于稳定收敛,测试过程中的损失值小于 0.05,最终的准确率稳定在 0.989。从损失曲线和准确率变化趋势可以看出,本文网络的收敛性能非常好,收敛速度和最终的准确率都优于 ResNet50 和 Inception-v4。

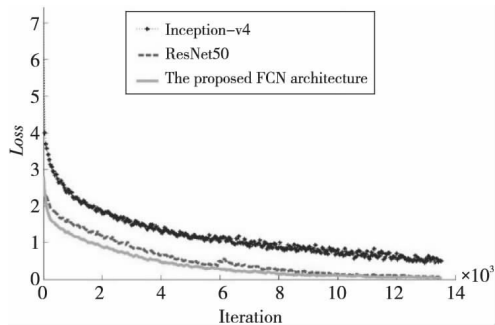


Figure 8 Relationship between the loss and the number of iterations

图 8 损失与迭代次数的关系  
从具体实例来看,ResNet50 对含有多个不同尺度病灶的图像进行辨识时,往往易发生漏检,如图 9a 所示。而本文网络具有多个尺度的卷积核,对含有多个不同尺度病灶的图像辨识具有较明显的优势,如图 9b 所示。

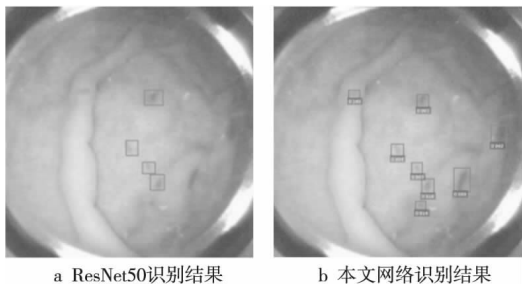


Figure 9 Bleeding parts recognized by ResNet50 and our network

图 9 同一 WCE 图中多个出血点的识别

## 5 结束语

现代人患肠胃疾病的概率日益上升,利用人工智能对 WCE 中的病灶进行有效的自动识别,能够保证肠胃疾病的早发现和早治疗。本文在分析现有神经网络模型的基础上,提出了一种由多尺度卷积-跳跃连接模块构成的简单全卷积神经网络。首先,引入了残差学习网络中跳跃连接的思想,使网络能够有效保存输入图像的初始特征。实验结果表明,与没有这种结构的 Inception-v4 网络相比,本文网络在结构上降低了复杂度,在参数和计算量上有大幅减少,在对 WCE 图像做出血点的分类识别时,灵敏度有所提升;其次,本文网络的收敛速度快,训练时间短,计算性能明显高于 Inception-v4 和 ResNet50,有较强的实用性。

从总体看,文中实验选取的数据集相比实际患者产生的 WCE 图像数量要少很多,还需要采用更多数据进行验证和优化。残差多尺度卷积神经网络本身也还存在一些待改进的地方,例如病灶定位等。此外,是否存在更好的网络结构、网络参数如何优化、网络性能能否提升等方面,依然需要不断地探索与尝试。

致谢

本文感谢安翰公司长期以来的合作和支持。

### 参考文献:

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2018, 68(6): 394-424.
- [2] Iddan G J, Merron G, Glukhovskiy A, et al. Wireless capsule endoscopy[J]. Nature, 2000, 405(6785): 417.
- [3] Iakovidis D K, Koulouzidis A. Software for enhanced video capsule endoscopy: Challenges for essential progress[J]. Nature Reviews Gastroenterology & Hepatology, 2015, 12(3): 172-186.
- [4] Charisis V S, Hadjileontiadis L J, Liatsos C N, et al. Capsule endoscopy image analysis using texture information from various colour models[J]. Computer Methods and Programs in Biomedicine, 2012, 107(1): 61-74.
- [5] Zhao Q, Meng M Q H. Polyp detection in wireless capsule endoscopy images using novel color texture features[C]//Proc of the 2011 9th World Congress on Intelligent Control and Automation, 2011: 948-952.
- [6] Yuan Y X, Li B P, Meng M Q H. Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images[J]. IEEE Transactions on Automation Science and Engi-

neering, 2016, 13(2): 529-535.

- [7] Xu Ting-ting, Ji Xiao-dong, Li Wen-hua, et al. Capsule endoscope image classification based on color and texture features[J]. Modern Electronics Technique, 2018, 41(19): 58-62. (in Chinese)
- [8] Li P P, Li Z Y, Gao F. Convolutional neural networks for intestinal hemorrhage detection in wireless capsule endoscopy images[C]//Proc of the 2017 IEEE International Conference on Multimedia and Expo, 2017: 1518-1523.
- [9] Coelho P, Pereira A, Leite A, et al. A deep learning approach for red lesions detection in video capsule endoscopies[C]//Proc of the International Conference Image Analysis and Recognition, 2018: 553-561.
- [10] Fan S H, Xu L M, Fan Y H, et al. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images[J]. Physics in Medicine & Biology, 2018, 63(16): 165001.
- [11] Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network[J]. Gastrointestinal Endoscopy, 2019, 89(2): 357-363.
- [12] Luo Ming-fang, Yong Xin, Zhang Huan, et al. Analysis of related factors affecting the diagnosis rate of unexplained gastrointestinal bleeding in capsule endoscopy[J]. Medical Journal of National Defending Forces in Southwest China, 2019, 29(8): 852-854. (in Chinese)
- [13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//Proc of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//Proc of the 31st AAAI Conference on Artificial Intelligence (AAAI-17), 2016: 4278-4284.
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proc of 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.

### 附中文参考文献:

- [7] 徐婷婷, 吉晓东, 李文华, 等. 基于颜色和纹理特征的胶囊内镜图像分类[J]. 现代电子技术, 2018, 41(19): 58-62.
- [12] 罗明芳, 雍鑫, 张欢, 等. 影响胶囊内镜不明原因消化道出血确诊率的相关因素分析[J]. 西南国防医药, 2019, 29(8): 852-854.

### 作者简介:



谢雪娇(1986-),女,湖北枣阳人,硕士,工程师,研究方向为医疗信息化。E-mail: xuejiao1360@tjh.tjmu.edu.cn

XIE Xue-jiao, born in 1986, MS, engineer, her research interest includes medical informatization.