

## 论著·技术与方法

## 基于深度学习的结直肠息肉内镜图像分割和分类方法比较

陈 健<sup>1</sup>, 王珍妮<sup>1</sup>, 夏开建<sup>2</sup>, 王甘红<sup>3</sup>, 刘罗杰<sup>1#</sup>, 徐晓丹<sup>1#</sup>

1. 江苏省常熟市第一人民医院 (苏州大学附属常熟医院) 消化内科, 常熟 215500; 2. 江苏省常熟市医学人工智能与大数据重点实验室, 常熟 215500; 3. 江苏省常熟市中医院消化内科, 常熟 215500

**[摘要]** **目的**·比较不同深度学习方法在结直肠息肉内镜图像分割和分类任务中的性能, 以确定最优方法。**方法**·从3家医院采集4个结肠息肉数据集, 涵盖1 534个静态图像及15个肠镜视频。所有样本均经病理学验证, 分为锯齿状病变和腺瘤性息肉2类。使用LabelMe工具进行多边形标注, 将标注结果转换为整数掩膜格式。数据用于训练不同架构的深度神经网络, 包括卷积神经网络、Transformer以及这2种技术的融合, 建立有效的语义分割模型。对比不同架构模型自动诊断结肠息肉的多项性能指标, 包括平均交并比 (mIoU)、整体准确率 (aAcc)、平均准确率 (mAcc)、平均Dice系数 (mDice)、平均F分数 (mFscore)、平均精确率 (mPrecision) 和平均召回率 (mRecall)。**结果**·开发了4种不同架构的语义分割模型, 包括2种深度卷积神经网络架构 (Fast-SCNN 和 DeepLabV3plus)、1种Transformer架构 (Segformer) 以及1种混合架构 (KNet)。在对291张测试图像进行综合性能评估中, KNet最高mIoU为84.59%, 显著优于Fast-SCNN (75.32%)、DeepLabV3plus (78.63%) 和 Segformer (80.17%)。在“背景”“锯齿状病变”和“腺瘤性息肉”3个类别上, KNet的交并比 (IoU) 分别为98.91%、74.12% 和 80.73%, 均超越其他模型。KNet在关键性能指标上也表现优异, 其中aAcc、mAcc、mDice、mFscore 和 mRecall 分别达到98.59%、91.24%、91.31%、91.31% 和 91.24%, 均优于其他模型。尽管在mPrecision上, 91.46%并非最突出, 但KNet的整体性能仍领先。在80张外部测试图像的推理测试中, KNet保持了81.53%的mIoU, 展现出良好的泛化能力。**结论**·利用基于KNet混合架构的深度神经网络构建的结直肠息肉内镜图像语义分割模型表现出优异的预测性能, 具有成为检测结直肠息肉高效工具的潜力。

**[关键词]** 深度学习; 结直肠息肉; 卷积神经网络; Transformer; 图像分割**[DOI]** 10.3969/j.issn.1674-8115.2024.06.012 **[中图分类号]** R574.6; TP181 **[文献标志码]** A

## Comparative study on methods for colon polyp endoscopic image segmentation and classification based on deep learning

CHEN Jian<sup>1</sup>, WANG Zhenni<sup>1</sup>, XIA Kaijian<sup>2</sup>, WANG Ganhong<sup>3</sup>, LIU Luojie<sup>1#</sup>, XU Xiaodan<sup>1#</sup>

1. Department of Gastroenterology, Changshu No. 1 People's Hospital (Changshu Hospital Affiliated to Soochow University), Jiangsu Province, Changshu 215500, China; 2. Changshu Key Laboratory of Medical Artificial Intelligence and Big Data, Jiangsu Province, Changshu 215500, China; 3. Department of Gastroenterology, Changshu Traditional Chinese Medicine Hospital, Jiangsu Province, Changshu 215500, China

**[Abstract]** **Objective**·To compare the performance of various deep learning methods in the segmentation and classification of colorectal polyp endoscopic images, and identify the most effective approach. **Methods**·Four colorectal polyp datasets were collected from three hospitals, encompassing 1 534 static images and 15 videos. All samples were pathologically validated and categorized into two types: serrated lesions and adenomatous polyps. Polygonal annotations were performed by using the LabelMe tool, and the annotated results were converted into integer mask formats. These data were utilized to train various architectures of deep neural networks, including convolutional neural network (CNN), Transformers, and their fusion, aiming to develop an effective semantic segmentation model. Multiple performance indicators for automatic diagnosis of colon polyps by different architecture models were compared, including mIoU, aAcc, mAcc, mDice, mFscore, mPrecision and mRecall. **Results**·Four different architectures of semantic segmentation models were developed, including two deep CNN architectures (Fast-SCNN and

**[基金项目]** 苏州市第二十三批科技发展计划项目 (SLT2023006); 常熟市医学人工智能与大数据重点实验室能力提升项目 (CYZ202301); 苏州市护理学会科研项目 (SZHL-B-202407)。**[作者简介]** 陈 健 (1987—), 男, 副主任医师, 硕士; 电子信箱: szcsdoctor@gmail.com。**[通信作者]** 徐晓丹, 电子信箱: xxddoctor@gmail.com。刘罗杰, 电子信箱: Luojie13542@163.com。<sup>#</sup>为共同通信作者。**[Funding Information]** Suzhou City's 23rd Science and Technology Development Plan (SLT2023006); Changshu City Key Laboratory of Medical Artificial Intelligence and Big Data Capability Enhancement Project (CYZ202301); Scientific Research Project of Suzhou Nursing Association (SZHL-B-202407)。**[Corresponding Author]** XU Xiaodan, E-mail: xxddoctor@gmail.com. LIU Luojie, E-mail: Luojie13542@163.com. <sup>#</sup>Corresponding authors.

DeepLabV3plus), one Transformer architecture (Segformer), and one hybrid architecture (KNet). In a comprehensive performance evaluation of 291 test images, KNet achieved the highest mIoU of 84.59%, significantly surpassing Fast-SCNN (75.32%), DeepLabV3plus (78.63%), and Segformer (80.17%). Across the categories of “background”, “serrated lesions” and “adenomatous polyps”, KNet’s intersection over union (IoU) were 98.91%, 74.12%, and 80.73%, respectively, all exceeding other models. Additionally, KNet performed excellently in key performance metrics, with aAcc, mAcc, mDice, mFscore, and mRecall reaching 98.59%, 91.24%, 91.31%, 91.31%, and 91.24%, respectively, all superior to other models. Although its mPrecision of 91.46% was not the most outstanding, KNet’s overall performance remained leading. In inference testing on 80 external test images, KNet maintained an mIoU of 81.53%, demonstrating strong generalization capabilities. **Conclusion** The semantic segmentation model of colorectal polyp endoscopic images constructed by deep neural network based on KNet hybrid architecture, exhibits superior predictive performance, demonstrating its potential as an efficient tool for detecting colorectal polyps.

**[Key words]** deep learning; colorectal polyp; convolutional neural network; Transformer; image segmentation

结直肠癌是全球第二大癌症死因<sup>[1]</sup>。中国结直肠癌的发病和死亡人数持续增加。根据2020年的癌症统计数据,在所有恶性肿瘤中,结直肠癌的发病率和死亡率分别排在第二位和第五位<sup>[2]</sup>。该病可能始于结肠中的非癌症息肉,如果不及时治疗,可能诱发癌症<sup>[3]</sup>。通常经“腺瘤-癌”模式进展,由腺瘤样息肉发展到结直肠癌的病例约占所有结直肠癌的75%<sup>[4]</sup>。通常,息肉恶性转化前有一个持续数年、可检测并可治疗的临床前阶段<sup>[5]</sup>。因此,早期筛查、早期检出和及时切除肠息肉对降低结直肠癌发生率至关重要。

结肠镜是临床上主要的诊疗方法之一<sup>[6]</sup>。然而,在当前的结直肠息肉内镜诊疗实践中,存在两大挑战:息肉的漏诊和对其性质的准确鉴别。息肉漏诊普遍存在,漏诊率为2%~26%<sup>[7-8]</sup>。结肠镜检查后结直肠癌(post-colonoscopy colorectal cancer, PCCRC)的发生率正逐年上升,约占所有结直肠癌病例的9%<sup>[9]</sup>。这是指在结肠镜检查未发现异常的情况下,随后确诊为结直肠癌的情况。

近年来,人工智能技术迅速发展,已成功应用于各个领域。在医学领域中,深度学习(deep learning)取得了巨大的研究进展,尤其是在消化内镜领域。卷积神经网络(convolutional neural network, CNN)的引入,已经实现了胃肠镜检查质量控制、胃早期癌识别、胃早期癌深度预测以及肠息肉监控等多项功能。Transformer技术在医学领域的热度和关注度正迅速超越CNN,成为该领域的新焦点。

深度学习技术在区分息肉性质方面取得了重大进展,多项研究已证实其有效性<sup>[10-12]</sup>。本研究聚焦于CNN、Transformer以及两者的融合技术在结肠息肉图像分割中的应用,对比不同架构模型自动诊断结肠息肉的多项性能指标,以全面评估和比较不同深度学

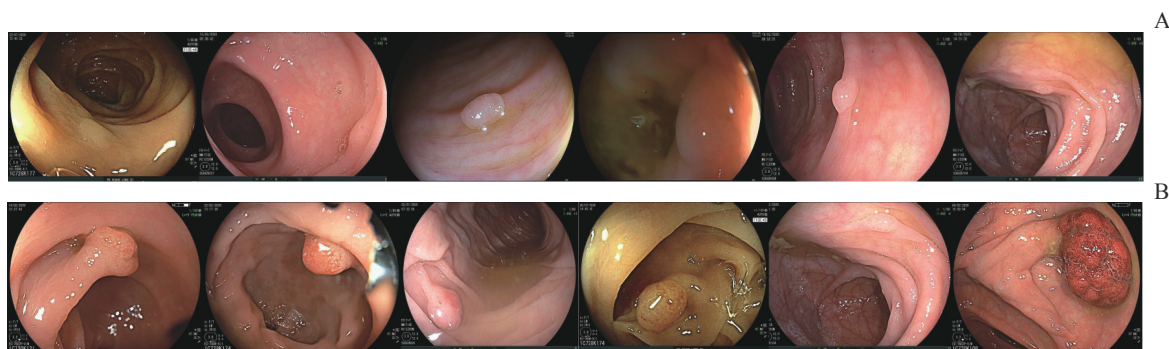
习模型在结肠息肉图像分割方面的表现,以期开发一款能够辅助消化内镜医师检测和鉴别结肠息肉的人工智能模型。

## 1 材料与方法

### 1.1 数据集

采用了4个数据集,时间跨度为2018年1月至2023年6月。数据集1和数据集2分别来自苏州大学附属常熟医院和常熟市中医院,包括1 454张肠息肉图像,用于模型的开发和测试。数据集3由常熟市辛庄人民医院提供,包含80张肠息肉图像,作为模型的外部静态图像测试集。数据集4也来自常熟市辛庄人民医院,包括15个肠镜视频,内含21个息肉,用于模型的外部视频测试。外部测试集仅用于评估模型性能,未参与训练或调参,以保证其独立性。结肠镜检查采用多个品牌的内窥镜系统,共13台电子内窥镜,包括开立(SonoScape HD-550)、奥林巴斯(OLYMPUS CV-V1)、宾得(Pentax EPK-i7000)和富士(ELUXEO 7000)。图像包括腺瘤性息肉和锯齿状病变图像(图1)。为了增强模型的泛化能力和鲁棒性,选择的内镜图像覆盖了不同清洁度的肠道情况。所有检出息肉均经切除并进行组织病理学检查;排除炎症性肠病、活动性结肠炎、凝血障碍、家族性息肉病和急诊结肠镜检查患者。

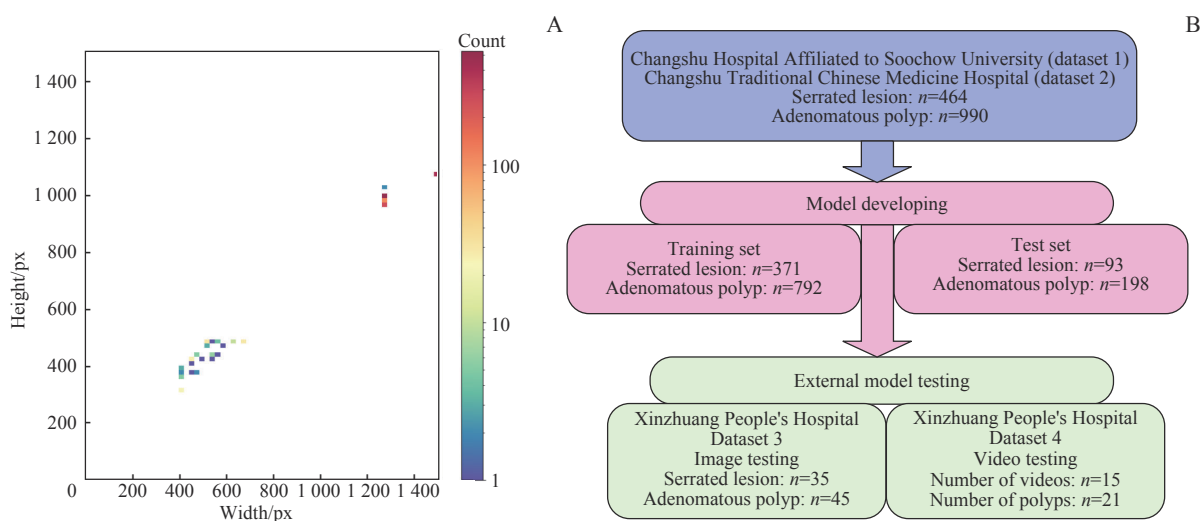
用于模型构建的数据集1和数据集2中的图像尺寸分布情况见图2A。其中,红色区域表示图像尺寸较为集中的区域,蓝色区域则表示图像尺寸较为稀少的区域。数据集包括不同尺寸的图像,尤其是1 280×995和1 504×1 080尺寸的图像比较常见,分别占据总数的25.3%和23.6%。训练集、测试集和外部测试集中不同类型图像的分布情况见图2B。



**Note:** A. Serrated lesions. B. Adenomatous polyps.

图1 数据集中息肉图像示例

Fig 1 Examples of polyp images from the datasets



**Note:** A. Distribution of image sizes in the dataset. B. Distribution of images across various categories.

图2 数据集图像分布情况

Fig 2 Distribution of images in the datasets

## 1.2 图像标注

所有图像和视频展示的息肉均经病理学验证。如遇到活检和切除标本的病理结果不一致, 优先采用切除标本的结果。两位具有10年以上工作经验的资深内镜医师根据病理结果对肠镜图像进行分类。将结肠息肉分为腺瘤性息肉(管状腺瘤、绒毛状腺瘤、绒毛管状腺瘤)和锯齿状病变(增生性息肉、无蒂锯齿状病变、锯齿状腺瘤), 基于3个关键考量: 首先, 大多数结肠息肉可以归类为腺瘤性或锯齿状病变; 其次, 病理学上, 在区分腺瘤性息肉和锯齿状病变方面表现出极高的一致性, 而在区分增生性息肉和无蒂锯齿状病变方面则一致性较差; 最后, 本研究数据集中其他类型息肉的数量不足, 无法用于开发一个稳健的“多类别”模型。为保证图像质量, 采用定期维护的高分辨率内窥镜, 按照肠道准备指南实行结肠镜检查前的肠道准备<sup>[13]</sup>。

使用Python3.7版本LabelMe(v5.3.1)图形图像

标注工具<sup>[14]</sup>对结肠镜图像中的息肉进行了多边形(Polygons)标注, 并根据病理学结果将其分为3类: 正常肠道、腺瘤性息肉和锯齿状病变。为了适应深度学习模型的训练需求, 开发一个专用的Python脚本, 用以将LabelMe格式的JSON文件转换成整数掩膜格式。在这种格式中, 每个类别被赋予一个唯一的整数标识: 正常背景标为0, 腺瘤性息肉标为1, 锯齿状病变标为2。这样的转换使模型能够直接处理像素级的分类信息, 显著提高了训练效率和图像分割的准确性。

制定规范化的图像标注程序, 并对标注人员进行专业训练, 以确保他们能够精准地识别息肉并严格按照既定程序操作。采用双重标注系统, 允许不同的标注人员独立地标注同一图像, 并对选定结果进行随机核对, 以提高标注的一致性和精确率。定期邀请经验丰富的专家复审标注结果, 并根据其反馈进行调整, 以持续提高标注的品质。



### 1.3 图像预处理

本研究进行一系列系统的图像预处理和增强步骤。采用在线数据增强策略<sup>[15-16]</sup>,所有数据增强操作都在模型训练过程中实时执行,不生成新的图像文件。这保证了在每个训练周期(epoch)中,模型都能够接触到略有差异的图像版本。这些操作包括随机上下和左右翻转、多角度旋转、仿射变换等,旨在增加训练数据量,防止模型过拟合,同时提升模型的泛化能力。在预处理阶段,使用RandomResize策略随机调整图像大小,此策略设定2 048×1 024为基准尺寸,并保持图像宽高比,通过应用0.5~2.0的缩放比例,动态调整图像尺寸,使其相对于基准尺寸缩小至50%或放大至200%。此外,还实施了50%概率的随机水平翻转,以及512×512尺寸的RandomCrop进行随机裁剪,以帮助模型从部分信息中识别目标。采用PhotoMetricDistortion进行光度失真处理,增强模型对于不同光照和颜色变化的适应性。

### 1.4 模型训练配置

优化深度学习分割任务,对4种预训练模型采用迁移学习策略<sup>[17]</sup>。这包括2个基于CNN的模型:Fast-SCNN,预训练于Cityscapes数据集;DeepLabV3plus,预训练于ADE20K数据集。此外,还使用了基于Transformer架构的Segformer模型,以及融合了CNN和Transformer架构的KNet模型,两者均在ADE20K数据集上进行了预训练。通过迁移学习策略,在收集并标注的息肉分割数据集上对这些预训练模型进行微调。这一策略不仅能够充分利用在大规模数据集上获得的先进特征,还大幅降低了从头开始训练所需的数据量和计算资源。

在训练过程中,采用随机梯度下降(stochastic gradient descent, SGD)作为优化器,并配合预设的多项式衰减(PolyLR)学习率调度器优化学习率,从而提高模型训练的效率。模型训练共设定为20 000次迭代,每批次处理5个样本,以保证学习过程的有效性。以Fast-SCNN的迁移学习过程为例,全面更新了Fast-SCNN骨干网络、2个辅助头和DepthwiseSeparableFCNHead解码头的参数。利用OpenMMLab(resnet101\_v1c)上的预训练权重,这种策略不仅限于网络的特定层,而是涵盖了整个网络结构。通过SGD优化器和多项式衰减学习率调度,在40 000次迭代期间,学习率逐步降低以适应新数据。

### 1.5 模型性能评估

为全面评估结肠息肉图像分割模型的性能,采用多元化的评估指标:整体准确率(aAcc),用于反映模型在所有预测中的整体准确性;平均交并比(mIoU),作为图像分割领域的核心指标,评估预测与实际区域的重叠度;平均准确率(mAcc)和平均Dice系数(mDice),提供类别级别的性能评估和衡量预测与真实区域的重叠程度;平均F分数(mFscore),评价查准率(Precision)与查全率(Recall)的平衡;平均精确率(mPrecision)与平均召回率(mRecall),关注正类别的识别与覆盖能力;此外,通过混淆矩阵定量分析,实现对模型在像素级分类精度上的全面评估。

### 1.6 实验平台

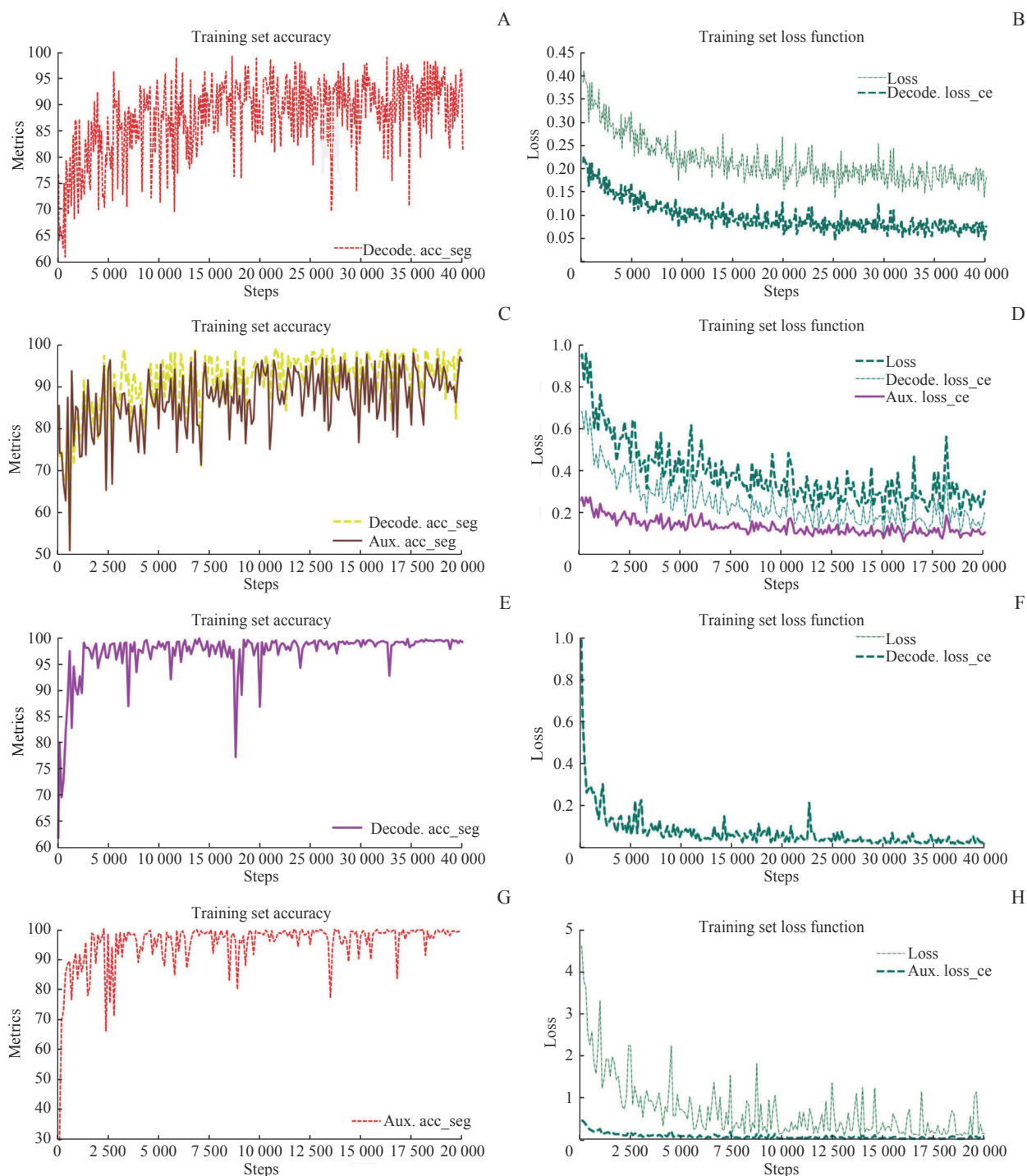
利用PyTorch和其他Python库,构建、训练深度学习模型并处理图像。数据的整理、分析和可视化通过Pandas(1.3.4)、NumPy(1.21.4)、Matplotlib(3.5.0)和Plotly(5.4.0)实现。模型优化使用PyTorch(1.10.0+cu113),而模型的保存和加载则依赖于H5py(3.6.0)。另外,使用Matplotlib、tqdm和opencv-python进行数据可视化和图像处理。

利用OpenMMLab(<https://openmmlab.com>)的开源语义分割算法库MMSegmentation,进行结直肠息肉分割模型的训练、测试。MMSegmentation通过模块化设计,提供了易于构建和定制的分割模型。

## 2 结果

### 2.1 模型训练和预测结果

图3A、C、E和G展示了Fast-SCNN、DeepLabV3plus、Segformer和KNet这4个深度学习分割模型在训练过程中准确率的持续提升,表明模型能够稳定地识别训练数据中的关键特征,从而实现对数据的有效拟合。图3B、D、F和H展现了这4个模型在损失函数上的下降趋势。随着训练的深入,每个模型的损失函数都呈现出明显的下降趋势,并逐步稳定化,这显示模型正在有效地向收敛状态进展。图4显示4个模型在测试集上的多项性能指标(aAcc、mIoU、mAcc、mDice、mFscore、mPrecision、mRecall)随训练过程的变化。虽然这些模型在特定指标的变化趋势上存在差异,但最终都显示出了随着训练进程的稳定性,没有出现明显的性能反弹或波动。



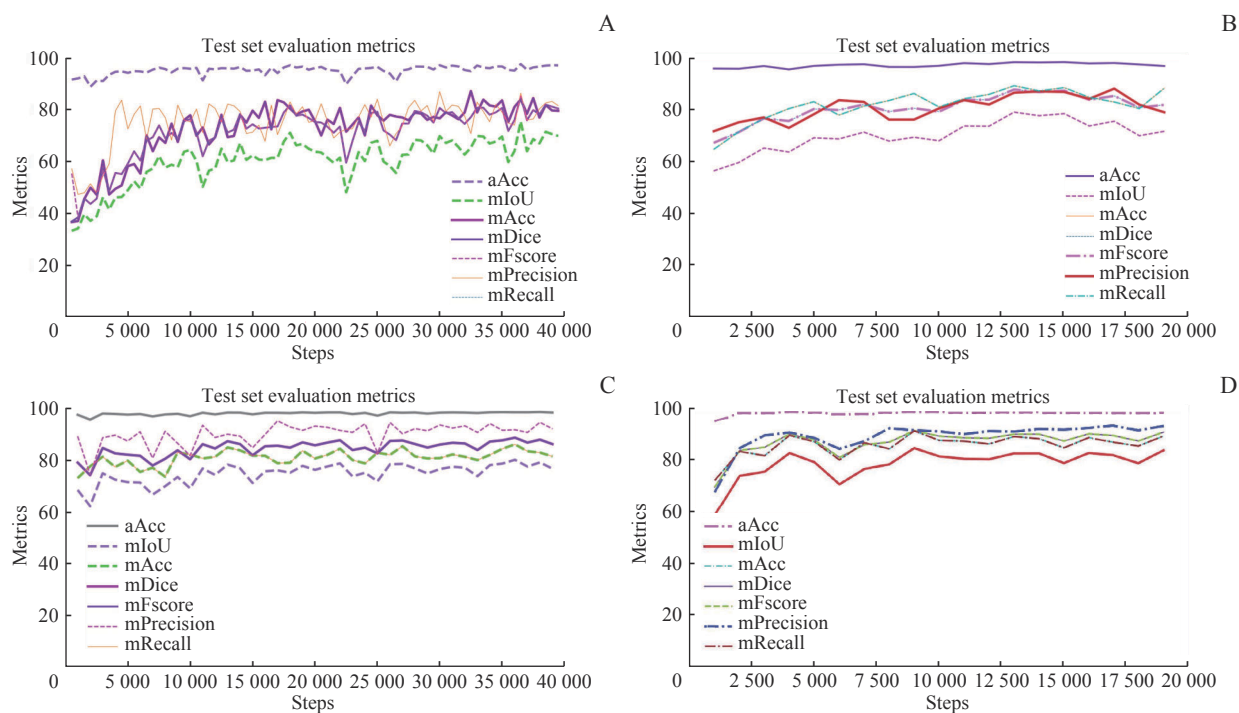
**Note:** A/B. Accuracy (A) and loss function (B) fluctuation of the Fast-SCNN model. C/D. Variations in accuracy (C) and loss function (D) in the DeepLabV3plus model. E/F. Segformer model's accuracy (E) and loss function (F) changes. G/H. KNet model's accuracy evolution (G) and loss function (H) alterations. All figures encapsulate the decoder segmentation accuracy (decode.acc\_seg) and auxiliary classifier segmentation accuracy (aux.acc\_seg).

图3 4种分割模型在训练集上的性能指标随训练进程的变化

Fig 3 Changes in performance metrics for four segmentation models in the training set throughout the training process

图5显示4种不同架构的息肉语义分割模型 (Fast-SCNN、DeepLabV3plus、Segformer、KNet) 在外部测试集上的预测效果, 通过4张随机选取的结肠息肉静态图像进行展示。该图包含6列, 依次展示原始息肉图像、真实标注 (ground truth, GT), 以及上述4种分割

模型的预测图。在模型预测图中, 腺瘤性息肉的识别区域以绿色标识, 而锯齿状病变的识别区域以红色标识, 背景则未着色。KNet正确预测了所有息肉类型, 并且KNet对图5A、B、C的预测区域与GT的一致性较高。然而, 在图5D中, KNet漏诊了一枚微小锯齿状息肉。



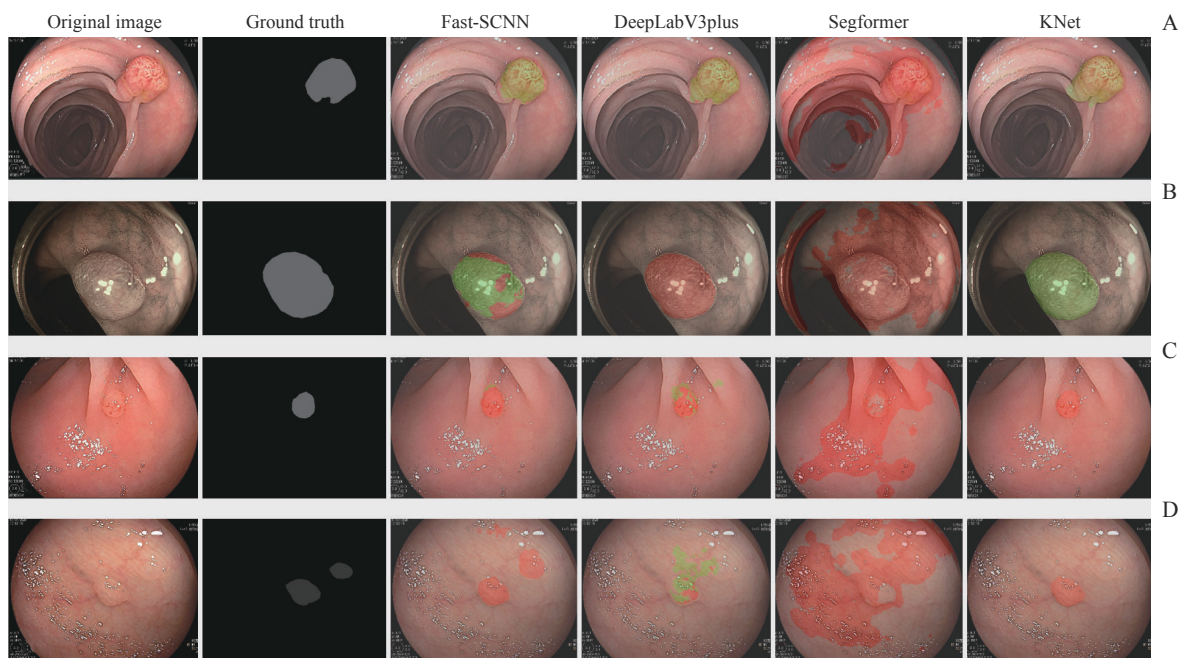
**Note:** A. Fast-SCNN model. B. DeepLabV3plus model. C. Segformer model. D. KNet model.

图4 4种分割模型在测试集上的性能指标随训练进程的变化

Fig 4 Changes in performance metrics for four segmentation models in the test set throughout the training process

Fast-SCNN模型也准确预测了所有息肉类型,并成功识别了KNet漏诊的那枚微小锯齿状息肉。尽管Fast-SCNN的预测结果在主要区域的颜色标记大体准确,但在图5B和5D中存在一定比例的颜色标记误差。Segformer在随

机选取的4张外部测试集图像上展现出较差的分割性能,表明模型的泛化能力有限。DeepLabV3plus在预测区域与GT的一致性方面表现较好,但在图5B和5D中错误地分类了息肉类型。



**Note:** A/B. Adenomatous polyps. C/D. Serrated lesions. Green represents the predicted areas for adenomatous polyps, and red represents the predicted areas for serrated lesions.

图5 不同模型在外部测试集上的预测结果对比

Fig 5 Comparison of prediction results from different models in an external test set



## 2.2 模型性能评价

**2.2.1 模型在测试集上的平均性能指标** 对比不同架构深度学习模型在291张测试图像上的平均分割性能指标。Fast-SCNN作为CNN类模型之一,具有97.42%的aAcc,但其在mIoU和mAcc上的表现稍显逊色,分别为75.32%和83.80%。而DeepLabV3plus模型在所有性能指标上均显著提升,尤其是在mIoU和mAcc上,分别达到了78.63%和88.78%。Transformer架构的Segformer模型,在维持较高aAcc的同时呈现出均衡的性能,其mIoU和mAcc分别为80.17%和84.71%,并在mPrecision上达到了92.97%,但mRecall略低(84.71%)。然而,在所有模型中,融合了CNN和Transformer优势的混合架构(Hybrid)模型KNet表现最为突出。KNet在几乎所有评价指标上都展现出了高性能:aAcc达到了98.59%,mIoU和mAcc分别为84.59%和91.24%,证明了其在识别息肉区域上的高准确性;KNet的mDice和mFscore均超过91%,显示了其在分割精确率和一致性上的优越表现。同时,该模型在mPrecision和mRecall上也保持了高水平,分别为91.46%和91.24%。

在图6所示的雷达图中,KNet模型(命名为KNet-ColoSeg)展现了显著的性能优势。KNet-ColoSeg在图表上的指标几乎完全超越其他比较模型,形成了一种明显的包裹现象。KNet-ColoSeg在aAcc、mPrecision和mRecall等核心指标上表现尤为优异。

表1 4种模型在测试集上的分类预测性能指标(%)

Tab 1 Performance metrics for classification predictions of the four models in the test dataset (%)

Model	Category	IoU	Acc	Dice	Fscore	Precision	Recall
Fast-SCNN	Background	97.74	99.02	98.86	98.86	98.70	99.02
	Serrated lesion	59.08	71.13	74.27	74.27	77.70	71.13
	Adenomatous polyp	69.15	81.24	81.76	81.76	82.30	81.24
DeepLabV3plus	Background	98.26	98.96	99.12	99.12	99.28	98.96
	Serrated lesion	63.63	83.38	77.78	77.78	72.88	83.38
	Adenomatous polyp	74.00	84.02	85.06	85.06	86.12	84.02
Segformer	Background	98.20	99.6	99.09	99.09	98.60	99.60
	Serrated lesion	67.22	71.95	80.39	80.39	91.09	71.95
	Adenomatous polyp	75.09	82.57	85.77	85.77	89.23	82.57
KNet	Background	98.91	99.50	99.45	99.45	99.40	99.50
	Serrated lesion	74.12	87.33	85.14	85.14	83.05	87.33
	Adenomatous polyp	80.73	86.88	89.34	89.34	91.94	86.88

KNet模型(命名为KNet-ColoSeg)具有最高的交并比(IoU),属性能最优的模型。KNet-ColoSeg在背景类别的分割中表现卓越,其IoU达到98.91%,

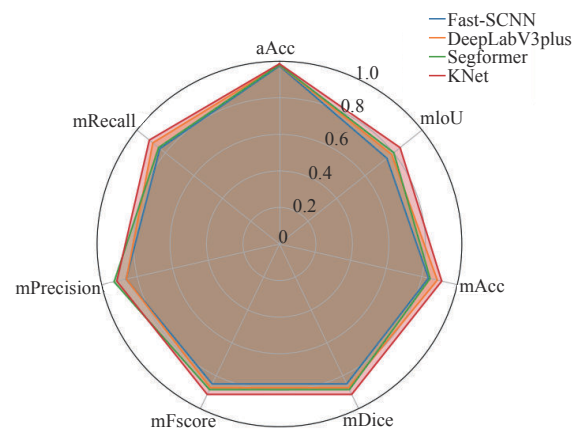


图6 测试集上4种结肠息肉分割模型的平均分割性能指标对比  
Fig 6 Comparative metrics of the average performance of four colon polyp segmentation models in the test set

**2.2.2 模型在测试集不同类别息肉上的性能指标** 对4个已构建的结直肠息肉语义分割模型进行测试集上的性能评估(表1)。评估结果显示,在背景类别的分割中,这些模型呈现出优异的性能,平均IoU达到98.28%,Acc为99.27%,Dice系数和F分数均为99.13%,Precision为98.995%,Recall为99.27%。在腺瘤性息肉类别中,所有模型的平均IoU为74.74%,Acc为83.68%,Dice系数为85.48%,F分数为85.48%,Precision为87.40%,Recall为83.68%。而在处理锯齿状病变类别时,所有模型的平均IoU为66.01%,Acc为78.45%,Dice系数为79.40%,F分数为79.40%,Precision为81.18%,Recall为78.45%。

Acc高达99.50%,同时Dice系数和F分数均保持在99.45%,Precision和Recall分别为99.40%和99.50%。在处理锯齿状病变时,该模型展现出了74.12%的IoU

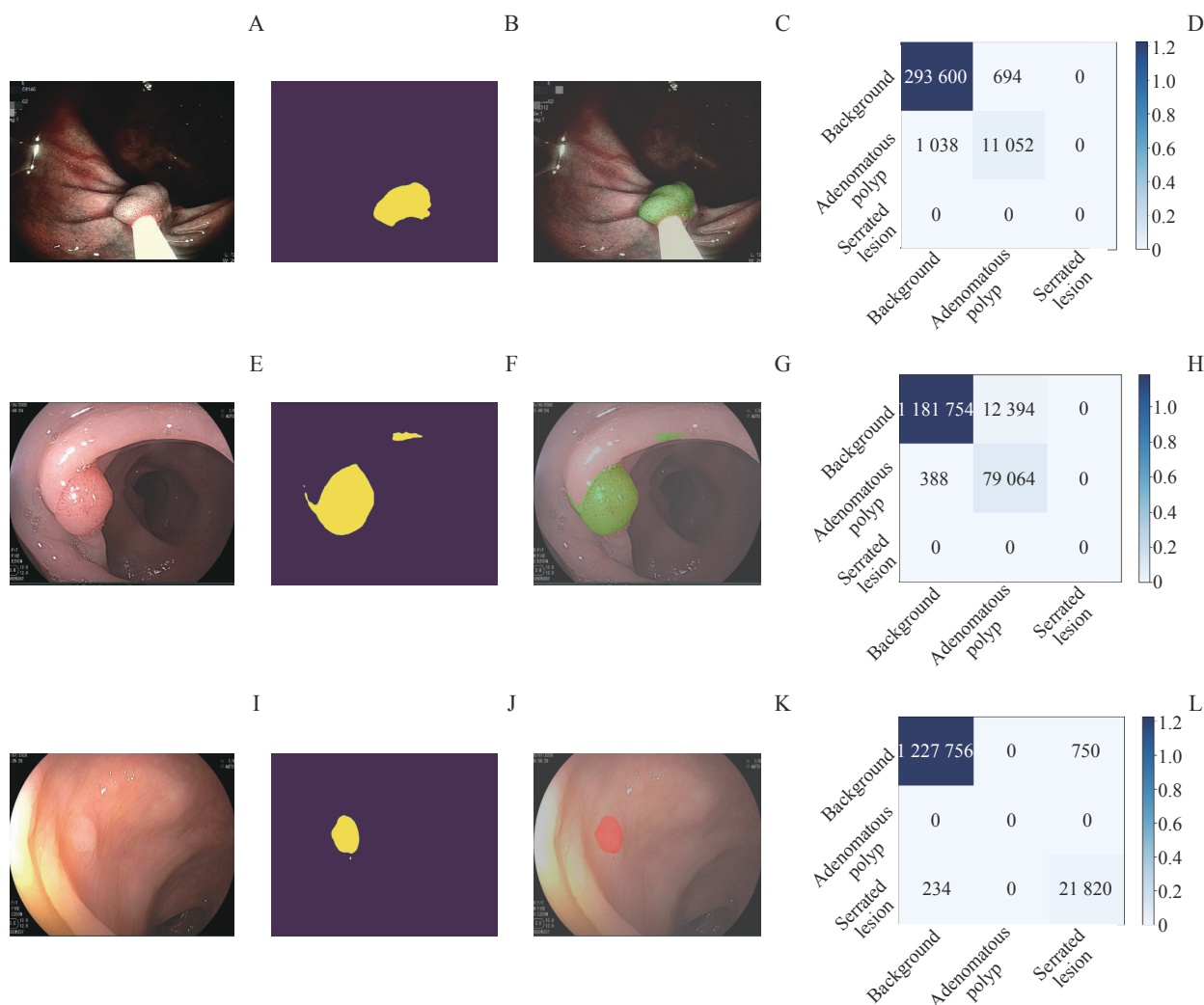
和 87.33% 的 Acc, Dice 系数和 F 分数均为 85.14%, Precision 为 83.05%, Recall 为 87.33%。对于腺瘤性息肉类别, KNet-ColoSeg 的 IoU 为 80.73%, Acc 为 86.88%, Dice 系数和 F 分数均达到 89.34%, Precision 高达 91.94%, Recall 为 86.88%。

### 2.3 人工智能模型预测新的图像和视频

KNet-ColoSeg 模型在外部测试集上的综合性能评估中, aAcc、mAcc 和 mIoU 分别为 97.72%、90.13% 和 81.53%。模型在处理背景类别时具有优势, 达到了 98.81% 的 IoU、99.49% 的 Acc、99.4% 的 Dice 系数和 F 分数, 以及 99.31% 的 Precision 和 99.49% 的 Recall。针对锯齿状病变类别, 模型展现了 63.53% 的

IoU、92.72% 的 Acc、77.70% 的 Dice 系数和 F 分数、66.87% 的 Precision 以及 92.72% 的 Recall。在腺瘤性息肉类别中, 模型实现了 82.25% 的 IoU、84.87% 的 Acc、90.26% 的 Dice 系数和 F 分数、96.38% 的 Precision 和 84.87% 的 Recall。

KNet-ColoSeg 模型在外部测试集上的 3 张随机选取的结直肠息肉图像上的推理预测结果 (图 7), 分别展示了原始息肉图像、模型预测输出、预测结果覆盖在原始图像上的视图, 以及针对单张图像的混淆矩阵。从 3 张原始图像的分析中观察到, 模型成功识别了所有息肉类别。尽管混淆矩阵揭示了少量像素的误分类, 但总体来看, 模型的预测与原始图像高度一致, 显示了其在图像分割准确性方面的出色性能。



**Note:** This figure consists of four columns, including the original polyp images (A, E and I), the model's prediction output (B, F and J), a view of the model's predictions superimposed on the original images (C, G and K), and the confusion matrix for the prediction of individual images (D, H and L).

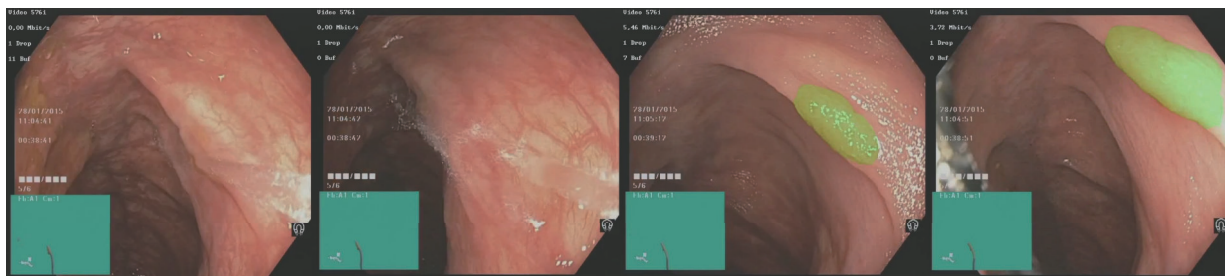
图 7 KNet-ColoSeg 模型在外部测试集上对新息肉图像的预测结果

Fig 7 Prediction results of KNet-ColoSeg model for new polyp images in an external test dataset



为了评估 KNet-ColoSeg 模型在实际应用中的性能, 我们利用该 AI 模型对外部测试集 (数据集 4) 的视频进行实时预测。图 8 展示了 AI 模型在实时定位

和分割息肉的能力, 并演示了其自动化判断息肉性质的过程。



**Note:** The image sequence from left to right demonstrates a serrated lesion case successfully segmented by the AI model after irrigation with physiological saline.

图 8 KNet-ColoSeg 模型对一例视频案例的实时预测过程

Fig 8 KNet-ColoSeg model for real-time prediction of a video case

### 3 讨论

本研究从多个内镜中心收集结肠镜图像数据, 创建了一个专门的数据集, 用于训练和测试针对结直肠息肉分割的人工智能模型。该数据集包括锯齿状病变和腺瘤性息肉图像, 旨在评估不同架构的结肠息肉语义分割模型。通过迁移学习技术, 开发并比较了 Fast-SCNN、DeepLabV3plus、Segformer 和 KNet 深度学习模型。KNet 模型结合了 CNN 和 Transformer 的优势, 在测试集上表现最佳, 其 mIoU 高达 84.59%, 将其命名为 KNet-ColoSeg。在外部测试集上, KNet-ColoSeg 展现出了 81.53% 的 mIoU, 证明了其在临床环境中的广泛适用性和鲁棒性。

结直肠癌的肿瘤发生假说, 根据癌前病变的性质主要分为 3 种途径: 腺瘤-癌变序列 (adenoma-carcinoma sequence)<sup>[18]</sup>、锯齿状病变途径 (serrated neoplasia pathway)<sup>[19]</sup>、de novo 致癌<sup>[20]</sup>。从频度上看, 作为结直肠癌的发展途径, 前两者是结直肠癌发展的主要途径, 但它们在临床病理学和分子异常方面被认为是互斥的<sup>[21-22]</sup>。2019 年 WHO 指南<sup>[23]</sup>提出, 使用“锯齿状病变”定义, 将增生性息肉 (hyperplastic polyps, HP)、传统锯齿状腺瘤 (traditional serrated adenomas, TSA) 和无蒂锯齿状病变 (sessile serrated lesions, SSL) 统称为锯齿状病变。在临床处理和病理诊断中, 区分腺瘤性息肉和锯齿状病变至关重要。因为如果漏诊, 该病可能通过不同途径恶化为癌症。因此, 运用深度学习模型识别和

区分这 2 种病变, 具有重要的临床应用价值。

深度学习技术, 特别是其卓越的特征提取和数据处理能力, 为结肠镜下息肉的智能化识别和诊断提供了有效方法。随着深度神经网络技术的快速发展, 其在息肉图像分割方面的应用也日益增多。ResUNet++ 网络, 通过结合残差块、压缩激发块、空间金字塔池化和注意力机制, 在 CVC-612 数据集上实现了 79.55% 的 Dice 系数和 79.62% 的 mIoU<sup>[24]</sup>; DeepLab 在 PASCAL VOC-2012 息肉语义图像分割任务中达到了 79.7% 的 mIoU<sup>[25]</sup>。尽管, 两者在息肉分割性能上均取得了显著成果, 但未能进一步区分息肉性质。在本研究中, 通过汇集多个内镜中心的结肠镜图像, 构建了一个包含锯齿状病变和腺瘤性息肉的专用数据集。在此数据集上训练的 KNet-ColoSeg 模型获得了 91.31% 的 mDice 系数和 84.59% 的 mIoU, 效果优于上述研究。

在结直肠息肉的研究中, 多种深度学习模型架构已得到广泛应用<sup>[26-28]</sup>。特别是 CNN, 因其在处理图像等固定形状数据方面的优势而备受关注。此外, 最初在自然语言处理领域确立新标准的自注意力机制的 Transformer 模型, 也已成功扩展到计算机视觉领域<sup>[29]</sup>。这些技术的发展不仅提升了分析效率, 还为结直肠息肉的识别和分类提供了新视角。本研究使用 2 种基于 CNN 的模型 (Fast-SCNN 和 DeepLabV3plus)、1 种基于 Transformer 架构的模型 (Segformer) 以及 1 种融合 CNN 和 Transformer 的混合架构模型 (KNet), 并通过在测试集上的比较, 筛选出了最优模型。KNet 在上述 4 种模型中表现出最佳性能, 这得

益于其统一、简洁且高效的框架,利用一组动态可学习的内核,专门生成特定实例或类别的掩码。这种内核更新策略允许 KNet 自适应地根据输入图像中的特征,有效区分不同的实例。结合了 CNN 的空间特征提取能力和 Transformer 的注意力机制, KNet 在处理结直肠息肉图像分割任务时,通过二分匹配而无需非极大值抑制(NMS)的端到端训练和推断过程,显著提升了性能和效率<sup>[30]</sup>。

本研究通过汇集并构建一个涵盖锯齿状病变与腺瘤性息肉的结肠镜图像数据集,所开发的 KNet-ColoSeg 模型在结直肠息肉分割任务上显示出优越的性能,实现了对结直肠息肉图像像素级的分割。利用 CNN 与 Transformer 的结合,这一模型不仅在测试集上实现了 91.31% 的 mDice 系数和 84.59% 的 mIoU,在外部测试集上也表现出 81.53% 的 mIoU,展现了其对于临床诊断支持的潜在应用价值。

尽管本研究在结直肠息肉分割任务中提供了一些见解,仍需注意一些限制。特别是数据集中 SSL 的样本数量有限,限制了在构建能够识别多类别息肉的强健模型方面的进展。在我们的视频案例测试中发现,尽管模型能够在冲洗后识别 SSL 区域,但在确定息肉性质时偶有误分类情况发生。未来工作将着重于数据集的扩增,以便进一步提高模型的准确性和鲁棒性。

#### 利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

#### 伦理批准和知情同意/Ethics Approval and Patient Consent

本研究已通过常熟市第一人民医院(苏州大学附属常熟医院)伦理审查委员会的批准(批准号:L2023045)。由于研究采用非干预性回顾性设计,患者知情同意被豁免。

This study was approved by the Ethics Review Committee of Changshu No. 1 People's Hospital under the approval number L2023045. Due to the non-interventional, and retrospective design of the research, patient consent was waived.

#### 作者贡献/Authors' Contributions

陈健负责研究设计及论文写作;王珍妮、王甘红、刘罗杰参与数据收集和整理;夏开建负责代码解释和报错解决;徐晓丹负责修订论文。所有作者均阅读并同意了最终稿件的提交。

CHEN Jian was responsible for the design and writing of the paper. WANG Zhenni, WANG Ganhong, and LIU Luo jie participated in data collection and organization. XIA Kaijian led the interpretation of codes and the resolution of errors. XU Xiaodan was responsible for revising the manuscript. All authors have read and agreed to the submission of the final manuscript.

• Received: 2024-01-01

• Accepted: 2024-03-15

• Published online: 2024-06-28

#### 参·考·文·献

- [1] GUNTER M J, ALHOMOU S, ARNOLD M, et al. Meeting report from the joint IARC-NCI international cancer seminar series: a focus on colorectal cancer[J]. *Ann Oncol*, 2019, 30(4): 510-519.
- [2] 中华人民共和国国家卫生健康委员会,中华医学会肿瘤学分会. 中国结直肠癌诊疗规范(2023年版)[J]. *中华外科杂志*, 2023, 61(8): 617-644.  
National Health Commission of the People's Republic of China, Chinese Society of Oncology. Chinese protocol of diagnosis and treatment of colorectal cancer (2023 edition)[J]. *Chinese Journal of Surgery*, 2023, 61(8): 617-644.
- [3] GUPTA S, LIEBERMAN D, ANDERSON J C, et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US multi-society task force on colorectal cancer[J]. *Gastrointest Endosc*, 2020, 91(3): 463-485. e5.
- [4] LIEBERMAN D A, REX D K, WINAWER S J, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer[J]. *Gastroenterology*, 2012, 143(3): 844-857.
- [5] BRETTTHAUER M, KALAGER M, ADAMI H O. Do's and don'ts in evaluation of endoscopic screening for gastrointestinal cancers[J]. *Endoscopy*, 2016, 48(1): 75-80.
- [6] MORI Y, KUDO S E, BERZIN T M, et al. Computer-aided diagnosis for colonoscopy[J]. *Endoscopy*, 2017, 49(8): 813-819.
- [7] LEUFKENS A M, VAN OIJEN M G, VLEGGAAR F P, et al. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study[J]. *Endoscopy*, 2012, 44(5): 470-475.
- [8] KIM N H, JUNG Y S, JEONG W S, et al. Miss rate of colorectal neoplastic polyps and risk factors for missed polyps in consecutive colonoscopies[J]. *Intest Res*, 2017, 15(3): 411-418.
- [9] RUTTER M D, BEINTARIS I, VALORI R, et al. World endoscopy organization consensus statements on post-colonoscopy and post-imaging colorectal cancer[J]. *Gastroenterology*, 2018, 155(3): 909-925. e3.
- [10] GONG E J, BANG C S, LEE J J, et al. No-code platform-based deep-learning models for prediction of colorectal polyp histology from white-light endoscopy images: development and performance verification[J]. *J Pers Med*, 2022, 12(6): 963.
- [11] GARCÍA-RODRÍGUEZ A, TUDELA Y, CÓRDOVA H, et al. *In vivo* computer-aided diagnosis of colorectal polyps using white light endoscopy[J]. *Endosc Int Open*, 2022, 10(9): E1201-E1207.
- [12] WANG P, LIU P, GLISSEN BROWN J R, et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study[J]. *Gastroenterology*, 2020, 159(4): 1252-1261. e5.
- [13] HASSAN C, EAST J, RADAELLI F, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline-Update 2019[J]. *Endoscopy*, 2019, 51(8): 775-794.
- [14] RUSSELL B C, TORRALBA A, MURPHY K P, et al. LabelMe: a database and web-based tool for image annotation[J]. *Int J Comput Vis*, 2008, 77(1): 157-173.



- [15] URBAN G, TRIPATHI P, ALKAYALI T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy[J]. *Gastroenterology*, 2018, 155(4): 1069-1078. e8.
- [16] CHEUNG T H, YEUNG D Y. A survey of automated data augmentation for image classification: learning to compose, mix, and generate[J]. *IEEE Trans Neural Netw Learn Syst*, 2023. DOI: 10.1109/TNNLS.2023.3282258.
- [17] SHIN H C, ROTH H R, GAO M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning[J]. *IEEE Trans Med Imaging*, 2016, 35(5): 1285-1298.
- [18] LESLIE A, CAREY F A, PRATT N R, et al. The colorectal adenoma-carcinoma sequence[J]. *Br J Surg*, 2002, 89(7): 845-860.
- [19] ZHOU Y J, LU X F, CHEN H, et al. Single-cell transcriptomics reveals early molecular and immune alterations underlying the serrated neoplasia pathway toward colorectal cancer[J]. *Cell Mol Gastroenterol Hepatol*, 2023, 15(2): 393-424.
- [20] GOTO H, ODA Y, MURAKAMI Y, et al. Proportion of *de novo* cancers among colorectal cancers in Japan[J]. *Gastroenterology*, 2006, 131(1): 40-46.
- [21] KIM K M, LEE E J, HA S, et al. Molecular features of colorectal hyperplastic polyps and sessile serrated adenoma/polyps from Korea[J]. *Am J Surg Pathol*, 2011, 35(9): 1274-1286.
- [22] LUI R N, SUNG J J Y. Sessile serrated adenoma/polyps: why we should be working flat out to understand more about these flat lesions?[J]. *J Gastroenterol Hepatol*, 2019, 34(10): 1667-1668.
- [23] NAGTEGAAL I D, ODZE R D, KLIMSTRA D, et al. The 2019 WHO classification of tumours of the digestive system[J]. *Histopathology*, 2020, 76(2): 182-188.
- [24] JHA D, SMEDSRUD P, RIEGLER M, et al. ResUNet++: an advanced architecture for medical image segmentation[R]. San Diego: IEEE International Symposium on Multimedia, 2019.
- [25] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Trans Pattern Anal Mach Intell*, 2018, 40(4): 834-848.
- [26] HSU C M, HSU C C, HSU Z M, et al. Colorectal polyp image detection and classification through grayscale images and deep learning[J]. *Sensors (Basel)*, 2021, 21(18): 5995.
- [27] LO C M, YEH Y H, TANG J H, et al. Rapid polyp classification in colonoscopy using textural and convolutional features[J]. *Healthcare (Basel)*, 2022, 10(8): 1494.
- [28] KRENZER A, HEIL S, FITTING D, et al. Automated classification of polyps using deep learning architectures and few-shot learning[J]. *BMC Med Imaging*, 2023, 23(1): 59.
- [29] LIU Z, LV Q, YANG Z, et al. Recent progress in transformer-based medical image analysis[J]. *Comput Biol Med*, 2023, 164: 107268.
- [30] ZHANG W, PANG J, CHEN K, et al. K-Net: towards unified image segmentation [R]. Montreal: Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021), 2021.

[本文编辑] 吴 洋

## 学术快讯

### 上海交通大学基础医学院张思宇团队揭示高级皮层和高级丘脑 差异化调控初级视皮层局部微环路新机制

2024年5月27日,上海交通大学基础医学院解剖学与生理学系张思宇研究员团队在*Nature Communications* 期刊在线发表题为*Organization of corticocortical and thalamocortical top-down inputs in the primary visual cortex*的研究论文。

该研究结合光遗传学和全细胞膜片钳记录系统地解析了来自次级视皮层内侧区、扣带回皮层、眶额叶皮层和丘脑外侧后核的4条自上而下的输入对初级视皮层各亚层兴奋性神经元和抑制性神经元的调控特征和机制,阐明了多条自上而下输入采取的差异化信息处理及相互作用的策略,为研究多样的自上而下调控信号在行为中的相互作用奠定了基础。