

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Wu Yijian, Liu Tie, Chen Nan, Ding Hui, Shao Zhuhong. DGSeg-YOLO: A Collaborative Multi-Task Model for Cell Detection and Instance Segmentation in Esophageal Pathological Images[J/OL]. Journal of Image and Graphics, XXXX: 1-16. DOI: 10.11834/jig.260263. (吴一键, 刘铁, 陈楠, 丁辉, 邵珠宏. DGSeg-YOLO: 一种面向食道病理图像细胞检测与实例分割的协同多任务模型[J/OL]. 中国图象图形学报, XXXX: 1-16. DOI: 10.11834/jig.260263.) [DOI: 10.11834/jig.260263]

DGSeg-YOLO: 一种面向食道病理图像细胞检测与实例分割的协同多任务模型

吴一键, 刘铁*, 陈楠, 丁辉, 邵珠宏

首都师范大学信息工程学院, 北京 邮编 100048

摘要: 目的 针对食道病理图像中细胞密集分布、重叠严重及形态复杂等特点, 现有检测与实例分割方法难以兼顾目标定位精度与边界表达能力, 多任务学习过程中检测与分割分支之间的特征协同不足, 影响复杂病理场景下的识别性能。方法 提出一种面向食道病理图像细胞检测与实例分割的协同多任务模型 DGSeg-YOLO。在统一网络框架下联合完成细胞检测与实例分割任务, 设计双向特征交互模块(bidirectional feature interaction module, BFIM), 实现检测与分割分支间的跨任务特征融合; 设计自适应检测感知分割模块(adaptive detection-aware segmentation module, ADGS), 利用BFIM输出的跨任务交互特征生成目标感知权重, 并对分割特征进行自适应加权增强; 结合难度感知联合损失函数(difficulty-aware joint loss, DAL), 动态协调检测与分割任务优化过程, 提高模型对复杂样本的学习能力。结果 在自建食道病理细胞数据集上的实验结果表明, DGSeg-YOLO 在边界框平均精度均值(box mean average precision, Box mAP)@0.5、掩码平均精度均值(mask mean average precision, Mask mAP)@0.5 和 Dice 指标上分别达到 0.900、0.909 和 0.8351, 在细胞密集分布及重叠区域具有较好的检测与实例分割性能。消融实验验证了 BFIM、ADGS 及 DAL 各模块对模型性能提升的有效性。同时, 所提出模型已集成至食道病理智能分析系统, 可实现病理细胞的自动检测、实例分割及统计分析, 验证了方法在实际应用中的可行性。结论 DGSeg-YOLO 通过检测与分割任务间的协同特征学习, 增强了复杂病理图像中细胞目标的结构表达与边界刻画能力, 在保持模型规模适中的同时兼顾检测性能与工程应用需求, 可为食道病理细胞自动分析提供一种有效的技术方案。

关键词: 深度学习; 食道病理图像; 目标检测; 图像分割; 多任务学习

DGSeg-YOLO: A Collaborative Multi-Task Model for Cell Detection and Instance Segmentation in Esophageal Pathological Images

Wu Yijian, Liu Tie*, Chen Nan, Ding Hui, Shao Zhuhong

Information Engineering College, Capital Normal University, Beijing 100048, China

Abstract: Objective Esophageal cancer is one of the most prevalent malignant tumors of the digestive system, characterized by inconspicuous early-stage symptoms and a relatively low survival rate. Early detection and accurate diagnosis are therefore crucial for improving clinical outcomes. Histopathological image analysis serves as the gold standard for esophageal cancer diagnosis. Method We propose a collaborative multi-task model DGSeg-YOLO. Under a unified network framework, we jointly complete cell detection and instance segmentation tasks. We design a bidirectional feature interaction module (BFIM) to achieve cross-task feature fusion between detection and segmentation branches. We design an adaptive detection-aware segmentation module (ADGS) to utilize the cross-task interaction features output by BFIM to generate target-aware weights, and perform adaptive weighted enhancement on the segmentation features. We combine a difficulty-aware joint loss function (DAL) to dynamically coordinate the optimization process of detection and segmentation tasks, improving the model's learning ability for complex samples. Result Experimental results on the自建食道病理细胞数据集 (自建食道病理细胞数据集) show that DGSeg-YOLO achieves box mean average precision (Box mAP)@0.5, mask mean average precision (Mask mAP)@0.5, and Dice index of 0.900, 0.909, and 0.8351, respectively, demonstrating good performance in cell detection and instance segmentation in areas with dense distribution and overlap. Ablation experiments verify the effectiveness of BFIM, ADGS, and DAL modules in improving model performance. Meanwhile, the proposed model has been integrated into an esophageal pathology intelligent analysis system, realizing automatic detection, instance segmentation, and statistical analysis of pathological cells, verifying the feasibility of the method in practical applications. Conclusion DGSeg-YOLO enhances the structural expression and boundary刻画能力 of cell targets in complex pathological images through collaborative feature learning between detection and segmentation tasks, maintaining a suitable model scale while兼顾检测性能与工程应用需求 (兼顾检测性能与工程应用需求), providing an effective technical solution for automatic analysis of esophageal pathological cells.

收稿日期: 2026-05-13; 修回日期: 2026-09-17

* 通信作者: 刘铁 tliu@cnu.edu.cn

基金项目: 西安交通大学混合增强智能全国重点实验室开放课题(项目编号: HMHAI202407), 国家自然科学基金面上项目(项目编号: 62476178); 北京市自然科学基金项目(项目编号: 4242034, L252009)

Supported by: National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University (HMHAI 202407), National Natural Science Foundation of China (62476178), Beijing Natural Science Foundation (4242034, L252009)

© 中国图象图形学报版权所有

geal cancer diagnosis, where accurate cell detection and instance segmentation provide essential quantitative information for identifying abnormal cellular structures. In recent years, deep learning-based methods have significantly advanced pathological image analysis. However, automatic analysis of esophageal pathological images remains challenging due to dense cell distribution, large morphological variations, severe cell overlap, and blurred cell boundaries. Conventional single-task models perform detection and segmentation independently, limiting their ability to exploit the complementary relationship between object localization and structural representation. Although multi-task learning has shown promising potential, insufficient feature collaboration and task interference often restrict the overall performance. Therefore, this study aims to develop a collaborative multi-task framework to enhance feature interaction between detection and segmentation while improving the robustness of cell analysis in complex pathological scenarios. **Method** To address the above challenges, a collaborative multi-task model, termed **DGSeg-YOLO**, is proposed for joint cell detection and instance segmentation within a unified framework. The model adopts a shared backbone network to extract multi-scale feature representations from pathological images, providing rich semantic and spatial information for downstream tasks. To strengthen cross-task collaboration, a **bidirectional feature interaction module (BFIM)** is introduced between the detection and segmentation branches. BFIM enables bidirectional feature interaction and adaptive information fusion while preserving task-specific representations, thereby enhancing complementary feature learning and alleviating feature interference between the two tasks. Furthermore, an **adaptive detection-aware segmentation (ADGS) module** is designed to refine segmentation features using detection-aware feature representations generated through task collaboration. By enhancing discriminative information in regions containing densely distributed and overlapping cells, the proposed module improves the representation of complex pathological structures and boundary delineation. To alleviate the influence of imbalanced sample distributions during training, a **difficulty-aware joint loss (DAL)** is proposed to dynamically coordinate the optimization of detection and segmentation tasks according to the distribution of complex samples, thereby improving the robustness of multi-task learning. The proposed DGSeg-YOLO is evaluated on a self-constructed esophageal pathological cell dataset containing both isolated and overlapping cells with diverse morphological characteristics. The dataset is annotated with both bounding boxes and pixel-level instance masks to support joint optimization of detection and segmentation. Comprehensive experiments, including quantitative comparisons, ablation studies, and qualitative visualizations, are conducted to evaluate the effectiveness of the proposed framework. **Result** Experimental results demonstrate that DGSeg-YOLO achieves a **Box mAP@0.5 of 0.900**, a **Mask mAP@0.5 of 0.909**, and a **Dice coefficient of 0.8351** on the self-constructed esophageal pathological cell dataset. The proposed model exhibits stable detection and segmentation performance, particularly in pathological regions with densely distributed and overlapping cells. Ablation studies further verify that BFIM effectively enhances cross-task feature interaction, ADGS improves feature representation for complex pathological structures, and DAL contributes to more balanced optimization during multi-task training. Compared with representative lightweight detection and segmentation models, DGSeg-YOLO provides a favorable balance between detection accuracy, segmentation quality, and model complexity while maintaining robust performance in challenging pathological scenarios. In addition, the proposed algorithm has been integrated into an intelligent pathological cell analysis system for automated cell detection, instance segmentation, and quantitative pathological analysis, demonstrating its feasibility for practical pathological image analysis. **Conclusion** This paper proposes DGSeg-YOLO, a collaborative multi-task framework for cell detection and instance segmentation in esophageal pathological images. By enhancing cross-task feature interaction and collaborative learning between detection and segmentation branches, the proposed method improves the structural representation and boundary delineation of cells in complex pathological scenarios, especially in densely distributed and overlapping regions. Experimental results demonstrate that DGSeg-YOLO achieves competitive performance while maintaining a lightweight architecture suitable for engineering deployment. The proposed framework provides an effective technical solution for automated pathological image analysis and offers a useful reference for the application of collaborative multi-task learning in medical image analysis and computer-aided diagnosis systems.

Key words: deep learning; esophageal pathological images; object detection; image segmentation; multi-task learning

论文引用格式: Wu Yijian, Liu Tie, Chen Nan, Ding Hui, Shao Zhuhong. DGSeg-YOLO: A Collabora-

© 中国图象图形学报版权所有

tive Multi-Task Model for Cell Detection and Instance Segmentation in Esophageal Pathological Images [J/OL]. Journal of Image and Graphics. DOI:10.11834/jig.260263. (吴一键, 刘铁, 陈楠, 丁辉, 邵珠宏. DGSeg-YOLO: 一种面向食道病理图像细胞检测与实例分割的协同多任务模型[J/OL]. 中国图象图形学报. DOI:10.11834/jig.260263.)

0 引言

随着人口结构老龄化加剧及生活方式转变,恶性肿瘤已成为威胁人类健康的重要疾病之一。食道癌作为我国常见的消化系统恶性肿瘤,具有发病率和死亡率较高且早期症状隐匿等特点(Li等,2021; Abnet等,2018; Sung等,2021)。开展早期筛查与精准诊断对提升患者生存率具有重要价值(Pennathur等,2013)。病理诊断作为肿瘤确诊的关键依据(He等,2025),在食道癌的筛查与分型过程中发挥着核心作用。细胞层面的形态信息是病理分析的重要基础,因此,对细胞进行准确的检测与分割对于获取量化特征具有重要意义。传统病理分析主要依赖医生对显微图像的人工判读(Madabhushi等,2016),不仅效率较低、工作负担较重,而且在细胞分布密集、重叠严重及背景复杂的情况下,结果容易受到经验因素影响,存在一定的不确定性。随着数字病理技术和深度学习的发展,基于计算机视觉的自动化分析方法逐渐兴起,为提升食道病理细胞识别的准确性与效率提供了新的解决思路。然而,由于食道病理图像中细胞形态差异显著、染色不均且常呈密集堆叠状态,仍对检测与分割算法的精度和鲁棒性提出了更高要求。

近年来,随着深度学习技术的发展,基于卷积神经网络(convolutional neural networks, CNN)的病理图像细胞分析方法取得了显著进展。细胞检测方面,学者将目标检测模型引入医学图像领域,如基于快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)的方法被用于实现细胞的自动检测与分类(Kutlu等,2020)。YOLO(You Only Look Once)系列模型因其高效性被广泛应用于细胞检测任务中, Huang等(2021)基于YOLOv5提升了细胞检测与定位性能; Shi等(2022)通过引入注意力机制增强模型对密集小目标的表

能力; Xia等(2022)和Gu等(2024)分别从多尺度特征建模与特征注意力角度对检测模型进行改进。此外,针对细胞密集与重叠问题, Pan等(2018)通过密度图回归方法增强模型对重叠细胞的检测能力。细胞分割方法中,以U-Net为代表的编码器-解码器结构被广泛应用于医学图像分割任务(Ronneberger等,2015),并在此基础上不断发展。Chen等(2021)将Transformer引入U-Net结构以增强全局特征建模能力; Zhu等(2022)通过注意力机制提升模型对关键区域的关注能力。针对细胞粘连与重叠问题, Graham等(2019)提出HoVer-Net模型实现聚集细胞核的有效分离; He等(2021)通过方向特征建模像素间关系以提升实例分割能力。尽管上述方法分别在细胞检测或实例分割任务中取得了较好的效果,但大多针对单一任务进行优化,未能充分利用检测与分割之间的互补信息。

为进一步提高检测与分割性能,多任务学习逐渐成为医学图像分析的重要研究方向。Li等(2023)提出Mask DINO,在统一框架下实现目标检测与分割任务的协同建模,验证了多任务学习的有效性。针对检测与分割任务间的不平衡问题, Nan等(2024)进一步通过任务平衡机制提升协同性能。此外, Heuer等(2021)从结构设计角度实现多任务协同学习, YOLO-Med(2024)则在医学图像中实现了检测与分割的端到端联合建模。Ramesh等(2018)尝试通过多任务学习联合优化检测与分割任务,以提升对重叠细胞的分离能力。总体来看,现有多任务方法主要采用共享特征或简单特征融合策略实现任务协同,对于检测与分割任务间的差异性建模及跨任务信息交互关注相对不足。在细胞密集分布、重叠严重的复杂病理场景下,如何充分发挥检测与分割任务的互补优势,进一步提升细胞结构表达与边界刻画能力,仍然是值得研究的问题。

尽管现有方法在细胞检测、实例分割及多任务学习方面取得了一定进展,但在复杂食道病理图像中仍面临诸多挑战。首先,现有检测方法主要以目标定位为目标,通常输出边界框信息,难以准确刻画细胞的精细轮廓和形态特征,限制了其在病理定量分析中的应用。其次,多任务学习模型大多采用共享特征或简单特征融合策略,检测与分割任务之间缺乏针对性的协同建模。由于两类任务在特征关注重点和优化目标上存在差异,直接共享特征容易引

起任务间干扰,影响整体性能。再次,食道病理图像中细胞密集分布且重叠严重,细胞边界模糊、实例间粘连明显,使目标定位与实例边界分离更加困难。此外,训练数据中单个细胞与重叠细胞样本分布不均衡,复杂样本学习不足,进一步影响模型的检测与分割性能。因此,如何增强检测与分割任务间的特征协同,提高模型对密集重叠细胞的特征表达能力,并缓解复杂样本带来的训练偏差,仍是食道病理细胞智能分析中亟待解决的问题。

面向食道病理图像细胞检测与实例分割的协同多任务模型 DGSeg-YOLO,在统一框架下联合优化检测与实例分割任务,通过增强跨任务特征交互,提高复杂病理场景下细胞目标的检测与分割性能。本文的主要工作如下:(1)提出双向特征交互模块(Bidirectional Feature Interaction Module, BFIM)。针对检测与分割任务特征需求存在差异的问题,通过双向特征交互实现跨任务信息融合,在保留任务特异性表示的同时增强共享特征表达,减轻多任务学习中的特征竞争与信息冗余。(2)设计自适应检测感知分割模块(Adaptive Detection-Aware Segmentation, ADGS)。利用跨任务交互特征生成目标感知权重,并对分割特征进行自适应加权增强,增强模型对密集分布及重叠细胞区域的特征表达能力,提高复杂病理场景下的实例边界刻画能力。(3)构建难度感知联合损失函数(Difficulty-Aware Joint Loss, DAL)。根据训练样本中重叠细胞的分布动态调整检测与分割任务的优化权重,缓解复杂样本带来的训练偏差,提高模型在复杂病理场景下的学习稳定性。实验结果表明,DGSeg-YOLO在自建食道病理细胞数据集上取得了较好的检测与实例分割性能,并已集成至食道病理智能分析系统,实现了细胞检测、实例分割及统计分析等功能,验证了该方法在实际病理辅助分析中的应用潜力。

1 相关工作

1.1 病理图像细胞检测

近年来,随着深度学习技术的发展,基于卷积神经网络的方法逐渐成为病理图像细胞检测研究的主流。早期研究多采用CNN结合人工特征实现细胞定位与识别。Zhao等(2017)利用细胞核颜色与形态特征进行初步检测,并通过CNN提取深层特征完

成白细胞分类,有效提升了检测与识别性能。随后,研究者开始关注复杂病理场景中的特征表达能力,Shanthi等(2019)通过构建深度卷积网络,实现了对宫颈癌细胞复杂形态特征的自动学习与识别。针对病理图像中类别不平衡问题,Hagos等(2021)通过引入细胞权重图并改进损失函数,提高了模型对稀有细胞的检测能力。在此基础上,随着目标检测框架的发展,研究逐渐向多尺度特征建模与检测性能优化方向演进。Zhou等(2025)提出MSRT-DETR模型,通过多尺度特征序列融合与小目标检测机制,有效提升了细胞检测精度。Ferreira等(2025)结合主成分分析与YOLO框架,并通过多数据集融合与低质量图像模拟训练,增强了模型在复杂数据条件下的鲁棒性。Wen等(2023)提出CentroidNet,通过改进锚点匹配策略与间距设计,提高了乳腺癌Ki67细胞核中心点检测的精度与效率。上述方法主要通过优化检测框架与特征建模策略,不断提升模型在复杂场景下的检测性能。另一方面,为降低对精细标注数据的依赖,弱监督学习方法逐渐受到关注。Geijs等(2024)基于多示例学习,在无需像素级标注的情况下实现了全视野病理图像中基底细胞癌的检测与分型。同时,针对不同检测对象的研究也不断拓展,Kuo等(2024)提出ChromosomeNet模型,通过融合一阶段与二阶段检测策略,实现了对染色体的高精度检测。总体来看,现有方法在标准数据集上已取得较好效果,但在细胞密集、尺度变化显著及重叠严重的复杂场景下,仍存在检测精度与鲁棒性不足的问题。

1.2 病理图像细胞分割

在病理图像细胞分割研究中,基于深度学习的方法已成为主流,相关工作主要围绕特征表达、多尺度建模以及监督方式展开。早期研究多从网络结构改进入手,以提升对复杂细胞形态的刻画能力。例如,Raza等(2019)提出Micro-Net,通过引入多分辨率输入及中间层特征融合机制,实现对多种细胞结构的鲁棒分割。进一步地,Srivastava等(2021)设计多尺度残差融合网络MSRF-Net,通过跨尺度特征交互增强模型对不同尺寸细胞的表征能力。朱智勤等(2026)提出FAFTNet,通过融合频率自适应与特征变换机制,优化高低频信息利用,有效提升了医学图像分割精度。刘威等(2025)提出DEMS-GIF网络,引入Transformer特征融合与注意力增强策略,强化

多层特征表达,在多个数据集上取得了较优的分割性能。上述方法主要通过改进网络结构与特征表达能力,不断提升分割模型在复杂场景下的性能表现。在此基础上,研究者开始关注复杂场景下的细胞分割问题。Ramesh与Tasdizen(2018)将检测与分割任务进行联合建模,利用多任务学习策略实现对重叠细胞的有效分割,提高了模型在密集细胞区域的表现。随着标注成本问题的凸显,弱监督方法逐渐成为研究热点。Zhang等(2024)提出MTCSNet,仅依赖细胞质心与边界点等稀疏标注信息即可完成分割任务,在降低标注成本的同时保持了较好的分割性能。近年来,Transformer结构被引入医学图像分割任务中,Hörst等(2024)提出CellViT模型,通过大规模组织学图像预训练,有效提升了细胞核实例分割的精度。此外,为增强模型在复杂数据分布下的适应能力,Shen等(2025)提出多模态融合方法,通过引入多源信息提升模型的泛化性能。总体来看,现有方法在标准数据集上已取得较好效果,但在细胞密集、边界模糊及重叠严重的复杂病理场景中,仍存在分割精度下降与泛化能力不足的问题。

2 方 法

在模型架构设计上,本文以前期细胞检测模型LightCellNet(Chen等,2025)为基础构建DGSeg-YOLO。LightCellNet主要面向食道病理细胞识别与检测任务,通过在YOLOv11框架中引入倒置残差模块和通道注意力机制,提高模型对单个细胞与重叠细胞的检测能力。与LightCellNet侧重单一检测任务不同,本文在其轻量化特征提取与多尺度融合结构基础上进一步引入实例分割分支,并设计跨任务特征交互与目标感知特征调制模块,使模型能够在统一框架下同时输出细胞边界框、类别信息和实例掩码。

鉴于食道病理图像中的细胞目标尺度较小、分布密集且重叠现象普遍,检测模型需同时具备较强的特征提取能力、较高的推理效率以及网络轻量化特性。近年来,基于变换器(transformer)和检测变换器(detection transformer, DETR)的目标检测模型在检测精度方面取得了显著进展,但此类模型通常具有较大的参数规模和较高的计算开销,对训练资源和部署环境要求较高。相比之下,YOLO系列模型

在检测精度与推理效率之间保持了良好的平衡,已成为实际应用中广泛采用的目标检测框架。相较于YOLOv8等轻量级检测模型,YOLOv11在特征提取效率、网络结构设计以及推理性能方面均有所提升,同时具备良好的训练稳定性,更适合作为多任务协同框架的基础模型。此外,与尚处于快速迭代阶段的YOLOv12等新模型相比,YOLOv11具有更成熟稳定的工程实现和较高的部署可靠性。考虑到本文研究重点在于检测与分割任务间的协同优化机制,而非单纯提升检测骨干网络性能,同时兼顾实际设备对端侧推理效率和部署成本的要求,因此本文选择YOLOv11作为基础框架,并在此基础上开展面向检测与实例分割任务的协同建模研究。

DGSeg-YOLO在统一框架下联合建模检测与实例分割任务,并在高层语义特征层引入跨任务信息交互机制,以实现检测与分割的协同优化。具体而言,通过双向特征交互模块(BFIM)实现检测与分割特征的解耦与融合,在此基础上引入自适应检测感知分割模块(ADGS),利用目标感知响应对分割特征进行调制与增强,同时结合难度感知联合损失函数缓解类别不平衡问题。所提出方法在保持模型结构简洁的同时,有效提升了复杂细胞场景下的检测与分割性能。

2.1 多任务协同模型整体架构

DGSeg-YOLO的整体结构如图1所示。模型以病理图像作为输入,首先通过Backbone网络提取多尺度特征。骨干网络采用改进的YOLOv11结构(Stem与C3k2模块),输出不同分辨率的特征层 P_3 、 P_4 和 P_5 。在Neck部分,沿用LightCellNet中的特征融合结构,在特征金字塔网络(feature pyramid network, FPN)的基础上引入倒置残差模块(inverted residual module, iRMB)以增强特征表达能力,并结合通道注意力机制对多尺度特征进行加权融合。

在此基础上,增强后的特征被输入至多任务协同学习模块,其结构如图2所示。该模块以高层语义特征 P_5 为核心,通过双向特征交互模块(BFIM)对检测与分割特征进行解耦与融合,并生成用于检测分支的增强特征。检测分支结合 P_3 、 P_4 及增强后的 P_5 特征进行目标预测,输出细胞的边界框及类别信息。与此同时,分割分支在特征交互的基础上,引入自适应检测感知分割模块(ADGS),利用跨任务交互特征生成目标感知权重,并对分割特征进行自适应

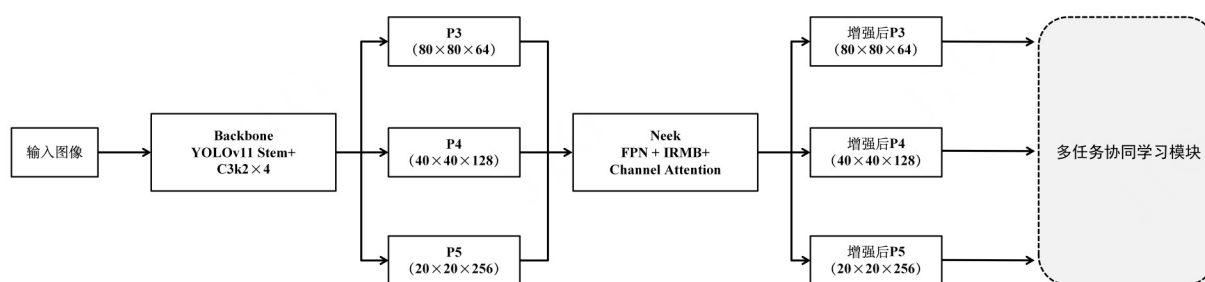


图1 DGSeg-YOLO网络模型结构

Fig. 1 Structure Diagram of DGSeg-YOLO Network Model

加权增强,从而获得更加精细的特征表示,最终通过分割头生成细胞实例掩码。

通过上述结构设计,DGSeg-YOLO在统一框架下实现了检测与分割任务的协同建模,并通过跨任务特征交互与检测感知机制有效缓解任务间的干扰问题,从而提升模型在细胞密集与重叠场景下的整体识别性能。

与前期 LightCellNet 相比,DGSeg-YOLO 的任务目标由细胞识别与检测扩展为细胞检测与实例分割协同建模。LightCellNet 主要输出细胞类别与边界框;DGSeg-YOLO 在轻量化特征融合结构基础上引入 BFIM、ADGS、DAL 及实例分割分支,可同时输出细胞类别、边界框和实例掩码,为细胞计数、轮廓分析和形态定量提供结构化结果。

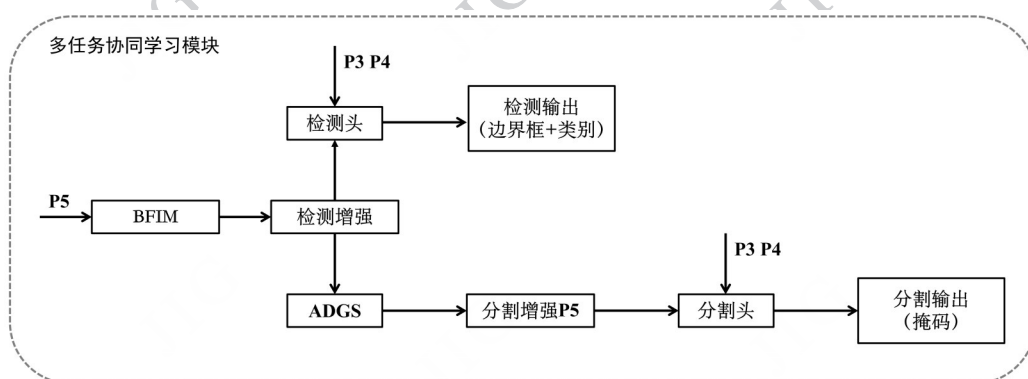


图2 多任务协同学习模块

Fig. 2 Multi-task collaborative learning module

2.2 双向特征交互模块

在传统的检测与分割联合框架中,两类任务通常直接共享 Neck 输出的特征表示。然而,检测与分割在特征侧重点上存在着明显的差异。检测任务更关注目标的空间位置信息,依赖边缘和关键点等局部结构特征;而分割任务则更强调区域一致性与语义完整性,需要对像素级边界进行精细建模。若二者完全共享同一特征表达,容易在特征学习过程中产生相互干扰,进而影响检测定位精度与分割结果的边界质量(Crawshaw, 2020)。在食道病理图像中,由于细胞分布密集且重叠严重,这种矛盾更加突出,尤其是在堆叠细胞区域,不同任务对同一位置特征的关注重点存在冲突,导致特征表达不稳定。为

避免不同激活函数符号混用,本文统一采用 SiLU ($\cdot$) 表示 Sigmoid 加权线性单元激活函数,采用 $\sigma(\cdot)$ 表示 Sigmoid 函数,采用 $\text{ReLU}(\cdot)$ 表示修正线性单元 (rectified linear unit, ReLU) 激活函数。后续公式与文字说明均遵循该符号约定。

针对以上问题,本文设计了双向特征交互模块 BFIM (Bidirectional Feature Interaction Module), 其结构如图 3 所示。该模块以高层语义特征 P_5 为输入,通过构建相对独立的检测分支与分割分支,实现任务特征的分离学习。同时,在两个分支之间引入信息交互机制,使分割分支中蕴含的语义信息能够对检测特征进行补充,从而提升检测特征的表达能力。

设输入特征为 $F \in R^{B \times C \times H \times W}$, 其中 B 表示批大
© 中国图象图形学报版权所有

小, C 表示通道数, H 与 W 分别表示特征图的空间尺寸。在检测分支中, 通过卷积、批归一化(batch normalization, BN)以及 SiLU($\cdot$) 激活函数对输入特征进行变换, 得到检测特征:

$$F_{det} = \text{SiLU}(\text{BN}(W_d * P_5 + b_d)) \quad (1)$$

式中 W_d 和 b_d 分别表示卷积层的权重和偏置参数, $*$ 表示卷积操作, $\text{SiLU}(\cdot)$ 表示 SiLU 激活函数。在分割分支中, 采用结构相同但参数独立的卷积变换提取区域语义特征:

$$F_{seg} = \text{SiLU}(\text{BN}(W_s * F + b_s)) \quad (2)$$

式中 W_s 与 b_s 为分割分支对应的参数。通过参数解耦, 使两个分支分别关注定位信息与语义信息, 从而减轻特征冲突。为将分割分支的语义信息引入检测特征, 首先通过 1×1 卷积生成语义权重图:

$$W = \sigma(W_{1 \times 1} * F_{seg} + b_{1 \times 1}) \quad (3)$$

式中左端量表示语义权重图, 用于刻画各空间位置的重要程度; $W_{1 \times 1}$ 表示卷积核参数, $\sigma(\cdot)$ 表示 Sigmoid 函数, 用于将权重归一化至 $[0, 1]$ 。随后将语义权重作用于检测特征, 实现特征增强:

$$F'_{det} = F_{det} \odot W \quad (4)$$

式中 $\odot$ 表示逐元素乘法。最后通过残差连接融合原始检测特征与增强特征, 得到模块输出:

$$F_{out} = F_{det} + F'_{det} \quad (5)$$

式中左端量表示模块最终输出特征, 右端第二项为式(4)得到的加权增强检测特征。

通过上述过程, BFIM 在保持检测特征空间结构的同时, 引入分割分支的语义信息, 实现跨任务特征的有效交互与融合。该模块不仅缓解了检测与分割之间的特征冲突问题, 还为后续检测感知分割增强模块提供了更具判别性的特征表示, 从而提升模型在复杂细胞场景下的整体性能。

2.3 自适应检测感知分割模块

尽管 BFIM 模块通过特征解耦与跨分支信息交互缓解了检测与分割之间的特征冲突问题, 但在复杂病理场景中, 分割任务仍面临较大挑战。实例分割不仅需要识别目标类别, 还需在像素级别精确划分实例边界。对于食道病理图像中的堆叠细胞, 由于细胞之间存在遮挡与粘连, 其交界区域的像素归属具有较高不确定性, 仅依赖分割分支自身特征往往难以获得稳定且准确的边界表达。检测与分割分支经 BFIM 交互后形成的跨任务交互特征同时包含检测定位响应与分割语义信息。基于该特征生成目

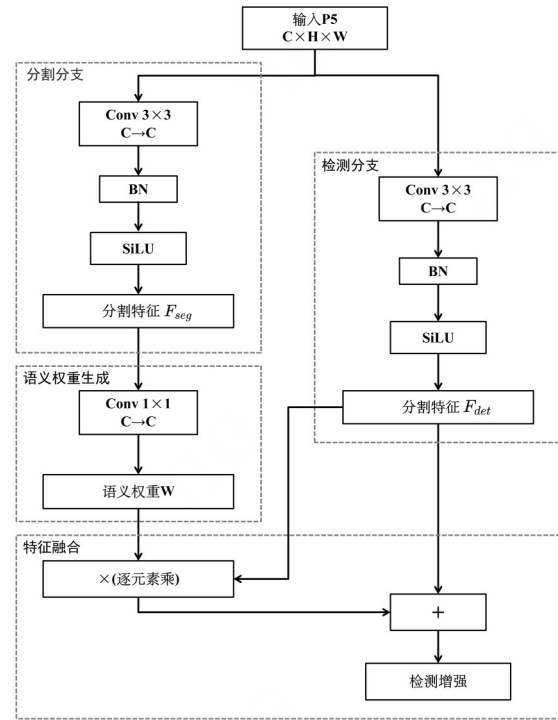


图3 BFIM 模块结构图

Fig. 3 BFIM Module Structure Diagram

标感知权重, 并对分割特征进行自适应加权增强, 可在一定程度上突出细胞相关区域、抑制背景干扰并提升边界刻画能力。

基于此, 本文提出自适应检测感知分割模块 ADGS (Adaptive Detection-Aware Segmentation), 其结构如图4所示。ADGS 利用 BFIM 输出的跨任务交互特征生成目标感知权重, 并对分割特征进行自适应加权增强。该模块在不改变整体模型结构的基础上强化细胞相关区域响应, 从而提升复杂病理场景下的分割特征表达能力。

设 BFIM 输出的检测增强特征为 F_{out} , 同时输入经通道增强的特征 P_5 作为分割基础特征。首先, 从 F_{out} 中提取空间信息以生成目标感知权重图。具体而言, 通过 1×1 卷积进行通道压缩, 再经 3×3 卷积恢复特征表达, 并结合非线性激活函数生成归一化权重, 其计算过程如下:

$$A = \sigma(W_2 * \text{ReLU}(W_1 * F_{out} + b_1) + b_2) \quad (6)$$

式中 W_1 与 W_2 分别表示 1×1 与 3×3 卷积核参数, b_1 与 b_2 为偏置项, $\text{ReLU}(\cdot)$ 表示 ReLU 激活函数, $\sigma(\cdot)$ 表示 Sigmoid 函数。所生成的注意力图 A 用于表征各空间位置的重要程度。随后, 利用目标感知权重对基础特征进行加权, 得到增强特征:

$$F_{\text{enh}} = P5 \odot A \quad (7)$$

式中左端量表示目标感知加权后的增强特征； $\odot$ 表示逐元素乘法。该过程能够强化目标区域特征响应，同时抑制无关背景信息，使分割特征更加集中于潜在目标区域。为避免在目标感知加权过程中损失原有语义信息，进一步采用特征融合策略，将原始特

征与增强特征在通道维度进行拼接，并通过卷积进行融合，得到最终分割特征：

$$F_{\text{seg}} = W_f * (P5 \parallel F_{\text{enh}}) + b_f \quad (8)$$

式中双竖线表示通道拼接操作， W_f 与 b_f 分别为融合卷积的权重与偏置参数。最终得到的 F_{seg} 被送入分割头，与 P_3 、 P_4 特征共同完成实例分割预测。

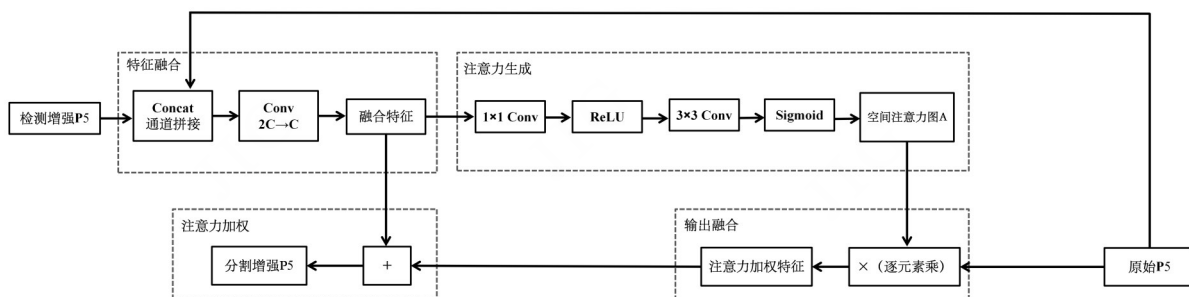


图4 ADGS模块结构图

Fig. 4 ADGS Module Structure Diagram

在整体框架中，BFIM与ADGS形成协同关系。前者侧重于检测与分割特征的解耦与交互，生成跨任务交互特征 F_{out} ；后者利用该特征生成目标感知权重，并对分割特征进行自适应加权增强，从而获得更具判别力的分割表示。其信息流关系可表示为：

$$P5 \xrightarrow{\text{BFIM}} F_{\text{out}} \rightarrow \text{Detection Head} \quad (9)$$

$$(F_{\text{out}}, P5) \xrightarrow{\text{ADGS}} F_{\text{seg}} \rightarrow \text{Segmentation Head} \quad (10)$$

与传统注意力机制不同，ADGS中的目标感知权重由BFIM输出的跨任务交互特征生成，属于一种跨任务交互特征驱动的检测感知分割增强方式。该设计能够有效利用检测与分割协同后的目标响应，增强分割特征对细胞相关区域的表达，从而在细胞密集与重叠场景下改善分割结果的边界表达。需要说明的是，ADGS使用的是BFIM输出后的跨任务交互特征，因此本文并不将其定义为单向的纯检测先验引导，而是将其视为一种检测感知的分割特征加权增强机制。

2.4 难度感知联合损失函数

在检测与分割的联合训练过程中，损失函数的构建对模型的学习效果具有重要影响。针对本文所构建的食道病理细胞数据集，训练中存在两方面问题：一是类别分布不均衡，二是多任务损失之间的协调难度较大。数据集中单个细胞数量远多于堆叠细胞，在常规损失函数作用下，模型参数更新更易受到多数类样本主导，从而削弱对堆叠细胞的识别

与分割能力。同时，在多任务学习框架中，模型需要同时优化边界框回归、类别分类及掩码分割等多个目标，不同损失项在数值尺度与收敛速度上存在差异，若采用固定权重进行组合，容易造成训练过程中的不平衡现象。

为缓解上述问题，本文在YOLOv11-seg标准损失函数基础上，引入一种难度感知联合损失函数（Difficulty-Aware Joint Loss, DAL）。该方法通过利用训练批次中的类别分布信息，对关键损失项进行自适应加权，从而在不引入额外复杂参数的前提下提升模型对少数类样本的学习能力。

YOLOv11-seg的标准多任务损失函数由多个子项构成，其表达形式为：

$$L_{\text{standard}} = \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{seg}} L_{\text{seg}} + \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{dfl}} L_{\text{dfl}} \quad (11)$$

式中 L_{box} 表示边界框回归损失，采用完全交并比（complete intersection over union, CIoU）形式：

$$L_{\text{box}} = 1 - \text{CIoU}(b, b^*) \quad (12)$$

式中 b 与 b^* 分别表示预测框与真实框。分类损失 L_{cls} 采用二元交叉熵（Binary Cross Entropy, BCE）形式：

$$L_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^N (p_i^* \log p_i + (1 - p_i^*) \log (1 - p_i)) \quad (13)$$

式中 N 为样本数量， p_i 为预测概率， p_i^* 为真实标签。分割损失 L_{seg} 同样采用BCE损失，在像素级别度量预测掩码与真实掩码之间的差异；分布焦点损失 L_{dfl} 用于优化边界框坐标分布。损失项权重系数设

置为:

$$\lambda_{\text{box}} = 7.5, \lambda_{\text{seg}} = 1.0, \lambda_{\text{cls}} = 0.5, \lambda_{\text{dfl}} = 1.5 \quad (14)$$

为增强模型对堆叠细胞的学习能力, 本文根据每个训练批次中的类别分布, 引入难度感知权重因子。设当前批次中目标总数为 N_{total} , 其中堆叠细胞数量为 N_{overlap} , 则其占比可表示为:

$$r_{\text{overlap}} = \frac{N_{\text{overlap}}}{N_{\text{total}}}, r_{\text{overlap}} \in [0, 1] \quad (15)$$

基于该比例定义权重函数:

$$w = 1 + (\alpha - 1)r_{\text{overlap}} \quad (16)$$

式中重叠细胞比例用于刻画当前训练批次中复杂样本的分布情况。由于重叠细胞通常伴随边界模糊、实例粘连和像素归属不确定等问题, DAL 根据该比例对分类损失和分割损失进行动态加权, 以增强模型对复杂样本的学习能力。参数 α 用于控制权重调节幅度, 本文将其设置为 1.5, 以在增强复杂样本学习的同时保持训练稳定性。具体而言, 本文所称“难度感知”并非对单个样本视觉难度的精确估计, 而是基于训练批次中重叠细胞比例构建的轻量级启发式动态加权策略。

引入该权重后, 联合损失函数形式为:

$$L_{\text{total}} = \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{seg}} w L_{\text{seg}} + \lambda_{\text{cls}} w L_{\text{cls}} + \lambda_{\text{dfl}} L_{\text{dfl}} \quad (17)$$

式中 λ_{box} 、 λ_{seg} 、 λ_{cls} 与 λ_{dfl} 分别表示边界框损失、分割损失、分类损失以及分布焦点损失的权重系数。相关参数主要参考 YOLO 系列模型的默认配置, 并结合验证集实验进行调整。本文最终设置 $\lambda_{\text{box}}=7.5$, $\lambda_{\text{seg}}=1.0$, $\lambda_{\text{cls}}=0.5$, $\lambda_{\text{dfl}}=1.5$ 。实验结果表明, 在一定参数范围内模型性能变化较为稳定, 说明所提出方法具有较好的鲁棒性。权重 w 仅作用于分类损失与分割损失, 而不作用于边界框回归损失与分布焦点损失。这是因为类别不平衡主要影响分类判别与像素级分割, 而边界框回归属于几何优化问题, 与类别比例关系较弱。

在实现过程中, 每个训练批次前向传播时首先统计堆叠细胞比例 r_{overlap} , 并计算对应权重 w , 随后在分类与分割损失计算中进行动态加权。该方法仅涉及简单的计数操作, 计算开销较小, 对训练效率影响可以忽略。最终得到的损失 L_{total} 用于反向传播以更新模型参数。

3 实验与结果分析

3.1 数据集

本文实验基于自建食道病理细胞数据集开展。数据来源于临床病理切片, 经数字病理扫描设备获取高分辨率图像, 并在预处理后构建用于细胞检测与分割的研究数据集。为满足多任务学习需求, 数据集同时提供目标检测框标注与像素级分割掩码标注。其中, 目标检测标注采用 LabelImg 工具完成, 实例分割标注基于开源工具 Labelme 实现。

数据构建时根据细胞形态及空间分布特征, 将样本划分为单细胞与堆叠细胞两类。本文对所有图像进行精细化标注, 确保检测框与分割掩码在空间位置上的一致性, 为模型训练提供可靠监督信息。由于实际样本中堆叠细胞样本数量明显少于单个细胞, 存在一定程度的类别不平衡问题。为增强模型对少数类样本的学习能力, 对包含堆叠细胞的样本进行数据增强处理, 具体包括水平翻转、垂直翻转及组合变换。通过引入几何变换, 有效扩充了样本数量并提升了数据多样性。在数据划分过程中, 采用分层随机划分策略, 使训练集与测试集中不同类别的样本比例保持一致, 从而避免类别分布偏差。同时, 保证同一图像仅出现在一个子集中, 以防止数据泄露, 提高实验结果的客观性与可靠性。

在上述数据构建与划分策略下, 本文最终形成的自建食道病理细胞数据集共包含 668 张图像, 约按 7:1.5:1.5 的比例划分为训练集、验证集和测试集, 分别包含 467 张、100 张和 101 张图像。数据集共标注 3581 个细胞实例, 其中单个细胞 3241 个, 堆叠细胞 340 个, 单个细胞与堆叠细胞的实例数量比例约为 9.5:1, 表明数据集中存在一定程度的类别不平衡。部分食道病理细胞样本图像如图 5 所示。

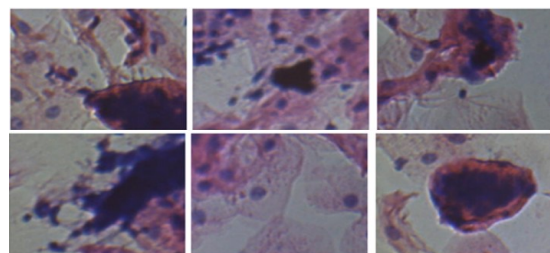


图5 食道病理细胞数据集示例

Fig. 5 Example of esophageal pathological cell dataset

3.2 实验细节

本文基于 PyTorch 深度学习框架完成模型训练与测试,实验在 Windows 操作系统环境下运行,硬件平台为 NVIDIA GeForce RTX 4050 Laptop GPU。训练过程中采用 AdamW 优化器进行参数更新,并结合余弦退火学习率调度策略动态调整学习率,初始学习率设为 0.01,最终衰减至其 0.01 倍。模型共训练 200 个 epoch,批次大小为 8,权重衰减系数为 0.0005,其余优化器参数采用默认设置。为提升模型泛化能力并缓解过拟合,引入多种数据增强策略,包括随机水平翻转、随机垂直翻转、HSV 颜色空间扰动及 Mosaic 增强等。在训练后期关闭 Mosaic 增强,以降低数据分布波动对模型收敛的影响。所有对比实验均在统一配置下进行,以保证结果的公平性与可比性。

此外,为评价模型在实际硬件平台上的运行效率,本文在 NVIDIA GeForce RTX 4050 Laptop GPU 环境下开展推理性能测试,端到端处理时间包括图像输入、模型推理和结果后处理。DGSeg-YOLO 的参数量为 5.4 M,计算量为 12.5 GFLOPs;在上述硬件平台下,平均推理速度为 68.7 FPS,单幅图像平均端到端处理延时为 15.5 ms/image。

3.3 评价指标

本文从检测精度与分割质量两方面定量评估 DGSeg-YOLO 模型在食道病理细胞检测与实例分割中的性能。检测任务采用精确率(Precision)、召回率(Recall)及平均精度均值(mean average precision, mAP)作为评价指标。

精确率表示模型检出的目标中正确检出目标所占的比例,反映检测结果的准确性;召回率表示全部真实目标中被模型正确检出的比例,反映模型检出目标的完整性。两者分别用于评价模型的误检与漏检情况。

交并比(intersection over union, IoU)衡量预测框与真实框的重叠程度。平均精度(average precision, AP)综合反映某一类别在不同置信度阈值下的精确率与召回率表现,mAP 为各类别 AP 的平均值。本文使用 Box mAP 评价检测性能:Box mAP@0.5 表示 IoU 阈值为 0.5 时各类别 AP 的均值;Box mAP@0.5:0.95 表示在 IoU 阈值从 0.5 至 0.95、步长为 0.05 的 10 个阈值下,对各类别 AP 进一步取平均,用于综合评价模型在不同定位精度要求下的检测性能。

对于实例分割任务,采用预测掩码与真实掩码之间的像素级 IoU 进行实例匹配,并使用 Mask mAP 评价整体分割性能。Mask mAP@0.5 与 Mask mAP@0.5:0.95 的阈值设置及平均方式与 Box mAP 一致,区别在于重叠程度由实例掩码计算。此外,引入 Dice 系数衡量预测掩码与真实掩码的像素重叠程度,其取值范围为 0~1,数值越大表示分割结果与真实标注越一致。本文在 IoU 阈值为 0.5 的条件下进行实例匹配,并统计不同场景下的 Dice 均值。

3.4 消融实验

为验证所提出方法中各关键模块的有效性,本文基于基线模型 LightCellNet 开展消融实验,并从检测与分割两个任务层面对模型性能进行分析。实验结果如表 1、表 2 所示。

实验结果表明,不同模块对两类任务的影响存在差异。当仅引入 BFIM 模块时,检测性能出现一定波动,说明仅依赖特征解耦与交互难以充分提升模型对复杂目标的表征能力;而在分割任务上,相关指标变化较小,表明模型对已检测目标的像素级刻画能力基本保持稳定。相比之下,引入 ADGS 模块后,检测与分割性能均得到一定提升。分割任务提升更为明显,说明该模块能够利用跨任务交互特征生成目标感知权重,并对分割特征进行自适应加权增强,从而改善复杂结构区域的分割质量,这一效果在堆叠细胞场景中表现更加突出。在同时引入 BFIM、ADGS 以及难度感知联合损失函数后,模型在当前模块组合下取得了较优的综合性能。检测指

标的提升表明模型对目标的定位与识别能力进一步增强,而分割指标的提升则反映出模型在边界刻画与区域一致性方面的改进。尤其在堆叠细胞场景下,模型表现出更稳定的分割效果,说明该方法能够有效缓解样本不平衡问题,并增强模型对复杂目标的建模能力。

3.5 对比实验

为全面评估模型性能,本文在相同实验设置下选取 YOLOv8n-seg、YOLOv11n-seg 以及 Mask R-CNN 作为对比方法进行对比实验。实验结果如表 3 所示,单阶段模型在部分指标上具有一定优势,这主要得益于其专注于单一任务优化,能够更充分利用模型容量。而本文方法由于采用多任务结构,在检测与分割之间进行特征共享,检测性能略有波动,但整体仍保持较高水平。

在分割性能及综合表现方面,本文方法与轻量化单阶段模型基本持平,同时在模型复杂度与综合性能之间保持了较好的平衡。与 Mask R-CNN 相比,本文方法在 mAP 指标上表现更优,体现出较好的检测与分割综合性能。尽管 Mask R-CNN 在 Dice 指标上略占优势,但其较大的模型复杂度限制了实际应用效率。

DGSeg-YOLO 的研究重点是通过检测感知分割增强机制增强模型在复杂病理场景下的特征表达能力。从实验结果来看,虽然本文方法在整体指标上与 YOLOv8n-seg 和 YOLOv11n-seg 相比未表现出明显优势,但所提出机制能够在一定程度上增强模型对目标区域的关注能力,在细胞重叠度较高、边缘模

糊等复杂病理场景中的表现更具鲁棒性,有助于改善复杂区域的分割质量。这说明多任务协同框架在复杂病理图像分析任务中具有一定的研究价值与应用潜力。

DGSeg-YOLO 引入 BFIM 和 ADGS 模块后参数量有所增加,但检测与实例分割任务共用主干特征提取过程,能够减少重复特征计算。模型参数量为 5.4 M,计算量为 12.5 GFLOPs。在 NVIDIA GeForce RTX 4050 Laptop GPU 平台下,模型平均推理速度为 68.7 FPS,单幅图像端到端处理延时为 15.5 ms/image,表明新增模块所带来的计算开销仍处于实际工程部署可接受范围内。

表 1 细胞检测任务的消融实验结果
Table 1 Ablation experiment results of cell detection task

Class	BFIM	ADGS	DAL	P	R	mAP@0.5	mAP@0.5:0.95
All	\	\	\	0.853	0.791	0.870	0.573
	✓	\	\	0.743	0.777	0.774	0.478
	\	✓	\	0.799	0.832	0.873	0.582
	✓	✓	✓	0.908	0.842	0.900	0.602
Individual cell	\	\	\	0.898	0.909	0.946	0.630
	✓	\	\	0.895	0.822	0.946	0.623
	\	✓	\	0.886	0.933	0.953	0.634
	✓	✓	✓	0.911	0.901	0.949	0.631
Overlapping cell	\	\	\	0.808	0.672	0.793	0.517
	✓	\	\	0.591	0.672	0.602	0.333
	\	✓	\	0.713	0.732	0.792	0.530
	✓	✓	✓	0.905	0.783	0.851	0.572

注:✓表示引入对应模块,\表示未引入对应模块;加粗结果表示该指标在对应类别实验中的最优值。

表 3 给出了不同模型在自建食道病理细胞数据集上的对比结果。本文实验重点在于评估模型在细胞检测、实例分割和结构化结果输出方面的综合能力,而非单独追求某一检测或分割指标的最优值。LightCellNet 主要面向细胞检测任务,本文已在方法部分说明其与 DGSeg-YOLO 在任务目标和输出形式上的差异;在对比模型选择上,本文主要遵循任务一致性、模型复杂度可比性和工程部署适配性原则。由于本文目标是面向食道病理细胞图像构建检测与实例分割一体化的轻量化协同模型,因此实验中选取 Mask R-CNN、YOLOv8n-seg 和 YOLOv11n-seg 作

为代表性对比方法。其中,Mask R-CNN 代表经典两阶段实例分割框架,YOLOv8n-seg 和 YOLOv11n-seg 代表轻量化单阶段实例分割框架,能够从不同模型范式下评价本文方法的综合性能。HoVer-Net、CDNet、MaskDINO、DiMaskDINO 和 YOLO-Med 等方法虽在相关任务中具有代表性,但其模型结构、训练目标、标注依赖和计算复杂度与本文轻量化工程部署场景存在差异。因此,本文未将所有大规模或专用化 SOTA 模型纳入主对比表,而是将实验重点放在与本文任务形式和部署目标更一致的代表性模型上。

表 2 细胞分割任务的消融实验结果

Table 2 Ablation experiment results of cell segmentation task								
Class	BFIM	ADGS	DAL	P	R	mAP@0.5	mAP@0.5:0.95	Dice 均值
All	\	\	\	0.834	0.789	0.870	0.569	0.8325
	✓	\	\	0.753	0.788	0.779	0.486	0.8354
	\	✓	\	0.841	0.830	0.884	0.585	0.8343
	✓	✓	✓	0.926	0.856	0.909	0.614	0.8351
Individual cell	\	\	\	0.896	0.922	0.953	0.611	0.8286
	✓	\	\	0.900	0.888	0.948	0.606	0.8329
	\	✓	\	0.902	0.923	0.958	0.621	0.8298
	✓	✓	✓	0.920	0.909	0.959	0.618	0.8306
Overlapping cell	\	\	\	0.772	0.656	0.787	0.528	0.8762
	✓	\	\	0.606	0.689	0.611	0.367	0.8612
	\	✓	\	0.780	0.738	0.811	0.550	0.8820
	✓	✓	✓	0.932	0.803	0.859	0.610	0.8785

注:✓表示引入对应模块,\表示未引入对应模块;加粗结果表示该指标在对应类别实验中的最优值。

表 3 对比实验结果

Table 3 Results of comparative experiments						
Model	Params	Box		Mask		Dice 均值
		mAP @0.5	mAP @0.5:0.95	mAP @0.5	mAP @0.5:0.95	
Mask R-CNN	44.2M	0.820	—	0.877	—	0.8933
YOLOv8n-seg	3.4M	0.920	0.642	0.915	0.635	0.8340
YOLOv11n-seg	2.8M	0.910	0.650	0.912	0.634	0.8382
DGSeg-YOLO	5.4M	0.900	0.602	0.909	0.614	0.8351

注:—表示该模型未报告对应指标或该指标不适用。

从结果可以看出,DGSeg-YOLO 在部分单项指标上并未取得最高值,参数量也有所增加。这与模型引入 BFIM、ADGS 和 DAL 等跨任务协同模块有关。上述模块并非单纯用于提升某一项 mAP 指标,而是用于增强检测与分割分支之间的特征交互,使模型能够在统一框架下同时输出细胞类别、边界框和实例掩码,为后续细胞计数、轮廓分析和形态定量提供结构化结果。

此外,数据集中单个细胞实例数量多于堆叠细胞,而堆叠、粘连和边界模糊区域对实例边界分离和掩码质量影响更明显,因此完整模型在部分 Dice 指标上可能存在波动。Dice 主要反映预测掩码与真实掩码的区域重叠程度,不能完全代表检测定位、实例分离和结构化输出能力。综合 Box mAP、Mask

mAP、Dice 及应用需求来看,DGSeg-YOLO 在检测与实例分割一体化任务中取得了稳定的综合表现。

3.6 工程应用场景分析

为进一步说明所提出模型的工程应用价值,本文将 DGSeg-YOLO 用于食道病理智能分析流程,实现病理细胞图像的自动检测、实例分割与统计分析。图 6 左为病理图像采集平台,图 6 右为典型食道病理细胞图像。可以看出,实际病理图像中细胞形态差异较大,细胞边界模糊,且存在局部重叠与粘连现象。对于此类图像,仅依赖目标检测结果难以提供完整的细胞轮廓信息,而单独进行实例分割又难以充分利用目标定位信息。

因此,在工程应用场景中,模型不仅需要完成细胞目标定位,还需要进一步输出实例掩码,以支持细

胞计数、轮廓提取和形态定量分析。DGSeg-YOLO 通过统一框架完成检测与实例分割任务,能够减少检测模型与分割模型分离部署带来的流程复杂度。虽然该设计会引入一定额外参数,但其输出形式更符合病理智能分析系统对结构化结果的需求,体现了本文方法在实际应用流程中的综合价值。

在实际工程应用中,病理图像分析系统支持 6 路和 12 路图像处理模式,每路对应的拨片可采集约 1200 张以上高清细胞图像。结合病理辅助分析流程的时效性需求,系统的性能设计目标是在单卡 NVIDIA GeForce RTX 4050 Laptop GPU 硬件平台下

完成检测、实例分割及统计分析等完整处理流程,其中 6 路和 12 路模式下单次任务的整体处理时间均应控制在 5 min 以内。在上述硬件平台和处理流程下进行测试, DGSeg-YOLO 的平均推理速度为 68.7 FPS, 单幅图像平均端到端处理延时为 15.5 ms/image。按每路约 1200 张图像计算, 6 路模式共处理约 7200 张图像, 整体处理时间约为 1.86 min; 12 路模式共处理约 14 400 张图像, 整体处理时间约为 3.72 min, 均低于 5 min 的工程设计要求。结果表明, 在既定硬件资源条件下, DGSeg-YOLO 能够满足病理智能分析系统的处理时效需求。

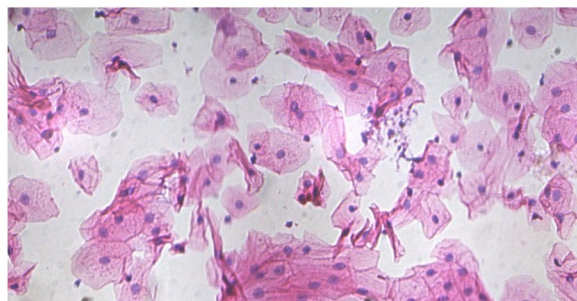
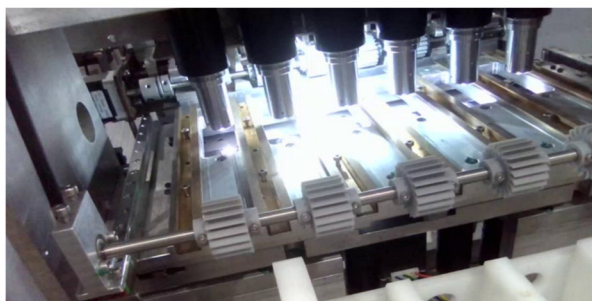


图 6 食道病理智能分析系统应用场景示意图

Fig. 6 Application scenario of the esophageal pathological cell analysis system.

3.7 可视化分析

为直观展示模型在复杂场景下的表现,选取测试集中的典型样本进行可视化对比,如图 7 所示。图 7 主要展示 YOLOv11n-seg 与 DGSeg-YOLO 在相同测试图像上的预测结果,其中第一列为原始图像,第二列为 YOLOv11n-seg 的预测结果,第三列为 DGSeg-YOLO 的预测结果。图中矩形框表示模型预测的细胞检测区域,彩色掩码表示模型预测的实例分割结果,标签中的数值表示预测置信度。需要指出的是,图 7 主要用于定性展示不同模型的预测效果,模型的定量评价已在表 1—表 3 中通过 Precision、Recall、Box mAP、Mask mAP 和 Dice 等指标给出。可以看出, DGSeg-YOLO 在实例定位与掩码生成方面均表现出较好的稳定性。相比 YOLOv11n-seg, 在细胞分布密集、局部重叠或边界模糊的区域, 本文方法能够获得相对完整的实例掩码和更加清晰的实例边界, 对相邻细胞之间的结构区分更加合理, 有助于减少掩码粘连和误分割现象。同时, 在细胞间距较小的情况下, 模型仍能够保持较好的实例分离效果, 体现出对复杂形态结构的适应能力。

对 ADGS 模块注意力分布进行可视化分析, 如图 8 所示。图 8 主要展示 ADGS 模块生成的目标感知注意力响应, 其中第一列为原始图像, 第二列为 ADGS 注意力热力图, 第三列为注意力热力图与模型预测检测结果的叠加显示。热力图中红色和黄色区域表示目标感知响应较强, 蓝色区域表示响应较弱; 叠加图中的检测框和标签为模型预测结果。该图主要用于分析 ADGS 模块对细胞相关区域的关注情况, 而非真实标注与预测结果的对照图。结果表明, 高响应区域主要集中于细胞目标所在位置, 而背景区域响应较弱, 说明该模块能够有效突出关键区域并抑制无关信息干扰。在多目标场景中, 注意力能够形成多个相对独立的响应区域, 对应不同细胞的位置。在重叠或边界模糊的情况下, 仍能够保持对目标区域的良好聚焦能力。该特性有助于提升模型对复杂区域的表达能力, 从而改善掩码的细节质量。

综合来看, 可视化结果从定性角度验证了所提方法在复杂病理图像中的有效性, 尤其在处理细胞密集和重叠场景时表现更加稳定。上述可视化分析

主要用于辅助理解模型预测行为和 ADGS 模块的响应特征,本文方法的定量性能比较以实验表格中的评价指标为依据。

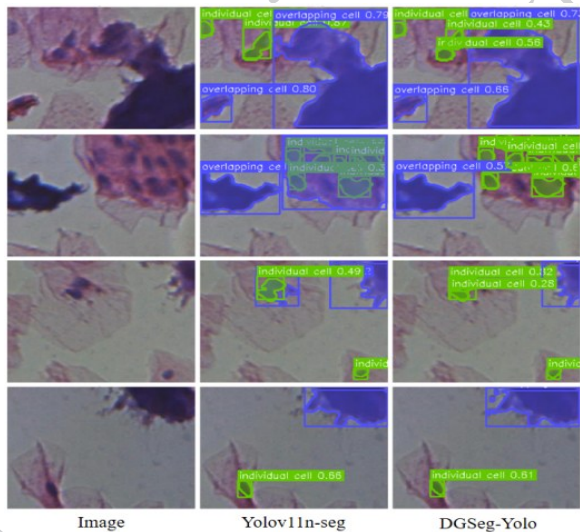


图7 DGSeg-YOLO 与 YOLOv11n-seg 检测与分割结果对比图
Fig. 7 Comparison chart of detection and segmentation results between DGSeg-YOLO and YOLOv11n-seg

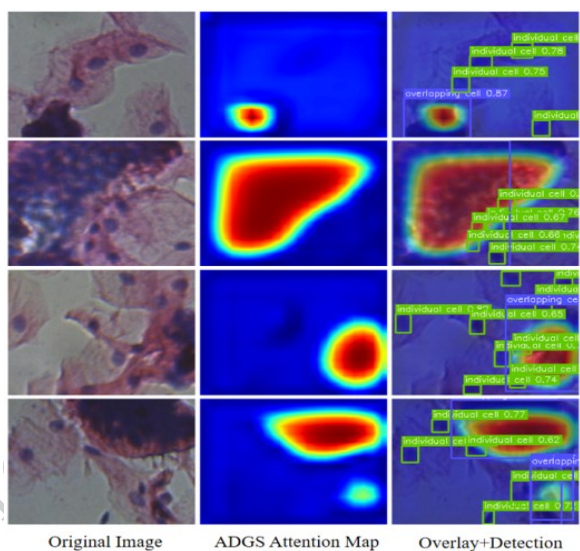


图8 ADGS 注意力图
Fig.8 ADGS attention map

4 结 论

本文围绕食道病理图像中细胞重叠严重与结构复杂等问题,构建了一种基于检测感知的食道病理细胞检测与分割多任务协同模型 DGSeg-YOLO。通过引入双向特征交互机制,实现检测与分割特征的

有效解耦与信息融合;同时结合检测感知分割增强策略,利用跨任务交互特征生成目标感知权重,并对分割特征进行自适应加权增强,从而提升模型对细胞相关区域的表达能力。在此基础上,引入难度感知联合损失函数,对复杂样本进行自适应优化,进一步增强模型对堆叠细胞的表征能力。实验结果表明,该方法在复杂病理细胞场景下取得稳定表现,并能够在统一框架下输出细胞检测、实例分割及统计分析结果,具有一定的工程应用潜力。

未来研究方向将进一步扩展多样化数据资源并优化特征建模方法,以提升模型在复杂细胞场景下的泛化能力与鲁棒性,同时探索其在实际病理辅助诊断中的应用潜力。

参考文献(References)

- Abnet C C, Arnold M and Wei W Q. 2018. Epidemiology of esophageal squamous cell carcinoma. *Gastroenterology*, 154 (2): 360-373 [DOI: 10.1053/j.gastro.2017.08.023]
- Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2021-02-05]. [2026-04-01]. <https://arxiv.org/abs/2102.04306> [DOI: 10.48550/arXiv. 2102.04306]
- Chen N, Liu T, Xin J H, Hou K Y, Ding H and Shao Z H. 2025. An attention mechanism and lightweight structure-based esophageal histopathological cell recognition method // 2025 IEEE 3rd International Conference on Computer, Vision and Intelligent Technology (ICCVIT). Baoding, China: IEEE: 1-6 [DOI: 10.1109/ICCVIT67848.2025.11391422]
- Crawshaw M. 2020. Multi-task learning with deep neural networks: a survey [EB/OL]. [2020-09-10]. [2026-04-01]. <https://arxiv.org/abs/2009.09796> [DOI: 10.48550/arXiv. 2009.09796]
- Ferreira F R T and do Couto L M. 2025. Using deep learning on microscopic images for white blood cell detection and segmentation to assist in leukemia diagnosis. *The Journal of Supercomputing*, 81 (2): 410 [DOI: 10.1007/s11227-024-06903-2]
- Geijs D J, Dooper S, Aswolinskiy W, Hillen L M, Amir A L and Litjens G. 2024. Detection and subtyping of basal cell carcinoma in whole-slide histopathology using weakly-supervised learning. *Medical Image Analysis*, 93: 103063 [DOI: 10.1016/j.media.2023.103063]
- Graham S, Vu Q D, Raza S E A, Azam A, Tsang Y W, Kwak J T, et al. 2019. HoVer-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 54: 101571 [DOI: 10.1016/j.media.2019.101571]

- 58: 101563 [DOI: 10.1016/j.media.2019.101563]
- Gu W C and Sun K X. 2024. AYOLOv5: improved YOLOv5 based on attention mechanism for blood cell detection. *Biomedical Signal Processing and Control*, 88: 105034 [DOI: 10.1016/j.bspc.2023.105034]
- Hagos Y B, Lecat C S Y, Patel D, Lee L, Tran T A, Justo M R, et al. 2021. Cell abundance aware deep learning for cell detection on highly imbalanced pathological data // 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France: IEEE: 1438-1442 [DOI: 10.1109/ISBI48211.2021.9433994]
- He H L, Huang Z Y, Ding Y, Song G L, Wang L, Ren Q, et al. 2021. CDNet: centripetal direction network for nuclear instance segmentation // Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4026-4035 [DOI: 10.1109/ICCV48922.2021.00399]
- He Z H and Ke Y. 2025. Evolving diagnostic and surveillance strategies in esophageal cancer screening: evidence from population-based studies in China. *Chinese Journal of Cancer Research*, 37(6): 929-936 [DOI: 10.21147/j.issn.1000-9604.2025.06.05]
- Heuer F, Mantowsky S, Bukhari S and Schneider G. 2021. MultiTask-CenterNet (MCN): efficient and diverse multitask learning using an anchor free approach // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 997-1005 [DOI: 10.1109/ICCVW54120.2021.00116]
- Hörst F, Rempe M, Heine L, Seibold C, Keyl J, Baldini G, et al. 2024. CellViT: vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94: 103143 [DOI: 10.1016/j.media.2024.103143]
- Huang D H, Cheng J T, Fan R, Su Z H, Ma Q X and Li J. 2021. Bone marrow cell recognition: training deep object detection with a new loss function // 2021 IEEE International Conference on Imaging Systems and Techniques (IST). New York, USA: IEEE: 1-6 [DOI: 10.1109/IST50367.2021.9651340]
- Huang S Z, Sirejiding S, Lu Y X, Ding Y, Liu L H, Zhou H, et al. 2024. YOLO-Med: multi-task interaction network for biomedical images // ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Seoul, South Korea: IEEE: 2175-2179 [DOI: 10.1109/ICASSP48485.2024.10446165]
- Kuo C E, Li J Z, Tseng J J, Lo F C, Chen M J and Lu C H. 2024. ChromosomeNet: deep learning-based automated chromosome detection in metaphase cell images. *IEEE Open Journal of Engineering in Medicine and Biology*, 6: 227-236 [DOI: 10.1109/OJEMB.2024.3512932]
- Kutlu H, Avci E and Özyurt F. 2020. White blood cells detection and classification based on regional convolutional neural networks. *Medical Hypotheses*, 135: 109472 [DOI: 10.1016/j.mehy.2019.109472]
- Li F, Zhang H, Xu H Z, Liu S L, Zhang L, Ni L M, et al. 2023. Mask DINO: towards a unified transformer-based framework for object detection and segmentation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 3041-3050 [DOI: 10.1109/CVPR52729.2023.00297]
- Li J, Xu J G, Zheng Y D, Gao Y, He S Y, Li H, et al. 2021. Esophageal cancer: epidemiology, risk factors and screening. *Chinese Journal of Cancer Research*, 33(5): 535-547 [DOI: 10.21147/j.issn.1000-9604.2021.05.01]
- Liu W, Zhong M, Liu G W, Wang H N and Ning Q. 2025. Detail-enhanced medical image segmentation focusing on global and intermediate features. *Journal of Image and Graphics*, 30(10): 3392-3410 (刘威, 钟森, 刘光伟, 王浩男, 宁倩. 2025. 聚焦全局与中间层的细节增强医学图像分割. *中国图象图形学报*, 30(10): 3392-3410) [DOI: 10.11834/jig.240463]
- Madabhushi A and Lee G. 2016. Image analysis and machine learning in digital pathology: challenges and opportunities. *Medical Image Analysis*, 33: 170-175 [DOI: 10.1016/j.media.2016.06.037]
- Nan Z X, Li X H, Xiang T and Dai J F. 2024. DI-MaskDINO: a joint object detection and instance segmentation model. *Advances in Neural Information Processing Systems*, 37: 60359-60381 [DOI: 10.52202/079017-1930]
- Pan X P, Yang D X, Li L Q, Liu Z B, Yang H H, Cao Z W, et al. 2018. Cell detection in pathology and microscopy images with multi-scale fully convolutional neural networks. *World Wide Web*, 21(6): 1721-1743 [DOI: 10.1007/s11280-017-0520-7]
- Pennathur A, Gibson M K, Jobe B A and Luketich J D. 2013. Oesophageal carcinoma. *The Lancet*, 381(9864): 400-412 [DOI: 10.1016/S0140-6736(12)60643-6]
- Ramesh N and Tasdizen T. 2018. Cell segmentation using a similarity interface with a multi-task convolutional neural network. *IEEE Journal of Biomedical and Health Informatics*, 23(4): 1457-1468 [DOI: 10.1109/JBHI.2018.2885544]
- Raza S E A, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, et al. 2019. Micro-Net: a unified model for segmentation of various objects in microscopy images. *Medical Image Analysis*, 52: 160-173 [DOI: 10.1016/j.media.2018.12.003]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation // International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Shanthi P B, Faruqi F, Hareesha K S and Kudva R. 2019. Deep convolution neural network for malignancy detection and classification in microscopic uterine cervix cell images. *Asian Pacific Journal of Cancer Prevention: APJCP*, 20(11): 3447-3456 [DOI: 10.31557/APJCP.2019.20.11.3447]
- Shen B, Gu Z, Zhou J L, Yan B Y, Fang Y Q and Wang H F. 2025. A

- deep learning multimodal fusion-based method for cell and nucleus segmentation. *IEEE Transactions on Medical Imaging*, 45(1): 112-124 [DOI: 10.1109/TMI.2025.3592625]
- Shi C, Pan Q and Rehman M. 2022. Cervical cancer cell image detection method based on improved YOLOv4 // 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). Xi'an, China: IEEE: 1996-2000 [DOI: 10.1109/ICSP54964.2022.9778577]
- Srivastava A, Jha D, Chanda S, Pal U, Johansen H D, Johansen D, et al. 2021. MSRF-Net: a multi-scale residual fusion network for biomedical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(5): 2252-2263 [DOI: 10.1109/JBHI.2021.3138024]
- Sung H, Ferlay J, Siegel R L, Laversanne M, Soerjomataram I, Jemal A, et al. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3): 209-249 [DOI: 10.3322/caac.21660]
- Wen K, Jin X, An H, He J and Wang J. 2023. CentroidNet: a lightweight, fast nuclei centroid detection model for breast Ki67 scoring. *Journal of Image and Graphics*, 28(04): 1119-1133 (文可, 金旭, 安虹, 何杰, 王珏. 2023. CentroidNet:轻量快速的乳腺癌Ki67细胞核中心点检测模型. *中国图象图形学报*, 28(04): 1119-1133)[DOI: 10.11834/jig.211207]
- Xia M Y, Zhang G S and Guan B. 2022. Automated Pap smear cervical cancer detection based on multiscale convolutional neural network // 2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS). Emeishan, China: IEEE: 1105-1110 [DOI: 10.1109/DDCLS55054.2022.9858434]
- Zhang B Y, Meng Z, Li H Y, Zhao Z C and Su F. 2024. MTCSNet: one-stage learning and two-point labeling are sufficient for cell segmentation. *IEEE Transactions on Medical Imaging*, 43(11): 3869-3880 [DOI: 10.1109/TMI.2024.3404428]
- Zhao J W, Zhang M S, Zhou Z H, Chu J J and Cao F L. 2017. Automatic detection and classification of leukocytes using convolutional neural networks. *Medical & Biological Engineering & Computing*, 55(8): 1287-1301 [DOI: 10.1007/s11517-016-1590-x]
- Zhou C C, He H Y, Zhou H, Ge F H and Yu P F. 2025. MSRT-DETR: a novel RT-DETR model with multi-scale feature sequence for cell detection. *Biomedical Signal Processing and Control*, 103: 107378 [DOI: 10.1016/j.bspc.2024.107378]
- Zhu N Y, Liu C, Forsyth B, Singer Z S, Laine A F, Danino T, et al. 2022. Segmentation with residual attention U-Net and an edge-enhancement approach preserves cell shape features // 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow, UK: IEEE: 2115-2118 [DOI: 10.1109/EMBC48229.2022.9871026]
- Zhu Z Q, Sun M W, Qi G Q, Li Y Y, Yang M J, Cheng J, et al. 2026. Frequency adaptation and feature transformation network for medical image segmentation. *Journal of Image and Graphics*, 31(1): 0303-0319 (朱智勤, 孙梦薇, 齐观秋, 李媛源, 杨梦杰, 程俊, 等. 2026. 融合频率自适应和特征变换的医学图像分割. *中国图象图形学报*, 31(1): 0303-0319)[DOI: 10.11834/jig.250100]

作者简介

吴一键,男,硕士研究生,首都师范大学信息工程学院,主要研究方向为人工智能与计算机视觉。E-mail:wuyijian@cnu.edu.cn

刘铁,男,副教授,首都师范大学信息工程学院,主要研究方向为具身智能、计算机视觉和智能机器人等。E-mail:tliu@cnu.edu.cn

陈楠,女,硕士研究生,首都师范大学信息工程学院,主要研究方向为计算机视觉。E-mail:2231002094@cnu.edu.cn

丁辉,女,副教授,首都师范大学信息工程学院,主要研究方向为图像处理、计算机视觉,交叉学科相关研究。E-mail:dhui@cnu.edu.cn

邵珠宏,男,副教授,首都师范大学信息工程学院,主要研究方向为计算机视觉、多模态情感分析、医学人工智能。E-mail:zhshao@cnu.edu.cn