DOI: 10. 19650/j. cnki. cjsi. J2210094

# 基于 Transformer 结构的深度学习模型 用于外周血白细胞检测\*

冷 冰<sup>1,2</sup>,冷 敏<sup>3</sup>,常智敏<sup>2</sup>,葛明锋<sup>1,2</sup>,董文飞<sup>1,2</sup>

(1.中国科学技术大学生物医学工程学院(苏州),生命科学与医学部 苏州 215163;2.中国科学院苏州生物医学 工程技术研究所 苏州 215163;3.聊城市心脏病医院 聊城 252200)

摘 要:虽然血细胞分析仪已广泛应用于医院中,但人工镜检仍是白细胞检测的"金标准"。本文提出了一种基于 DETR 的 Transformer 结构的深度学习模型 T-DETR 用于外周血白细胞的检测,旨在缓解人工镜检的压力。首先,使用 PVTv2 作为 DETR 的骨干提取多尺度特征图来提高检测精度。然后,将可变形注意力模块引入到 DETR 模型中,减少计算复杂度以加快模型收 敛。最后,为了得到最优权重,在筛选后的公共白细胞数据集上使用了迁移学习的训练方式。实验结果表明,T-DETR 在 COCO 数据集上 mAP 为 0.476,在白细胞数据集上的 mAP 为 0.954,优于 DETR 和经典 CNN 模型,验证了 Transformer 结构的模型在医 学图像检测中应用的可行性。

关键词: 白细胞;目标检测;深度学习;Transformer;DETR

中图分类号: TH776 文献标识码: A 国家标准学科分类代码: 520. 2040

## Deep learning model based on Transformer architecture for peripheral blood leukocyte detection

Leng Bing<sup>1,2</sup>, Leng Min<sup>3</sup>, Chang Zhimin<sup>2</sup>, Ge Mingfeng<sup>1,2</sup>, Dong Wenfei<sup>1,2</sup>

(1. School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Suzhou 215163, China; 2. Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, Suzhou 215163, China; 3. Liaocheng Cardiac Hospital, Liaocheng 252200, China)

**Abstract**: Although blood cell analyzers have been widely used in hospitals, the manual microscopy is still the "gold standard" for leukocyte detection. In this article, T-DETR, a DETR-based deep learning model with Transformer architecture is proposed for the detection of peripheral blood leukocytes, which aims to relieve the pressure of manual microscopy. First, PVTv2 is used as the backbone of DETR to extract multi-scale feature maps to improve detection accuracy. Then, the deformable attention module is introduced into the DETR model to reduce the computational complexity to accelerate the model convergence. Finally, to obtain the optimal weights, the training method of transfer learning is used on the filtered public leukocyte dataset. Experimental results show that T-DETR has an mAP of 0. 476 on the COCO dataset and 0. 954 on the leukocyte dataset, which is better than DETR and the classical CNN model. Results verify the feasibility of the Transformer structured model for medical image detection applications. **Keywords**; leukocyte; object detection; deep learning; Transformer; DETR

0 引 言

白细胞作为人体与疾病抗争的"卫士",可以抵抗细

收稿日期:2022-07-07 Received Date: 2022-07-07

菌、病毒、寄生虫等外来病原体的攻击。白细胞分为 5类:中性粒细胞、嗜酸性粒细胞、嗜碱性粒细胞、单核细 胞、淋巴细胞。白细胞的增多或减少都是身体可能发生 病变的表现。因此,白细胞的检测成为医生进行诊断疾 病和治疗观察的重要手段<sup>[1]</sup>。

<sup>\*</sup>基金项目:国家重点研发计划(2021YFB3602200)项目资助

目前医院广泛使用的血细胞分析仪是基于电阻抗原 理,常出现检测异常的情况。人工镜检作为血细胞分析 仪的补充,是白细胞检测的"金标准"。受到实验室政策 和资源情况的限制,人工血涂片复查的比率往往比血液 学审查国际共识小组制定的"规则"要求低,但工作量依 然很大<sup>[2]</sup>。此外,人工镜检对人员的专业性要求较高,且 容易出现观察者之间的差异。

基于传统机器学习的方法,文献[3]提出了动态信 号提取算法,结合支持向量机实现了血细胞与肿瘤细胞 的高精度分类。文献[4]提出基于回归树的特征选择方 法,将显著特征输入到支持向量机中进行分类。传统方 法需要预处理、分割、特征提取和选择等繁琐步骤,结果 容易受到染色条件、光线强弱等的影响<sup>[5]</sup>。

近年来,深度学习极大地提高了目标检测领域的 先进水平。其中,卷积神经网络(convolutional neural network, CNN)在视觉任务上的成功依赖于卷积层所编 码的强大的归纳偏置<sup>[6]</sup>。最近,视觉转换器 (Transformer)使用自注意力来聚合不同位置的信息,表 现出比 CNN 更高的性能<sup>[7]</sup>。根据模型结构的不同,目 标检测模型可分为基于 CNN 和 Transformer 的两种 类型。

基于 CNN 的一阶段目标检测网络如你只需看一次 网络(you only look once, YOLO)<sup>[8]</sup>,直接对候选框分类 并对位置进行回归,检测速度快但精度低。二阶段网络 如更快的区域卷积神经网络(faster region based CNN, Faster R-CNN)<sup>[9]</sup>,先在提议框上进行有无目标的分类和 回归,然后进行目标多分类和框回归,检测精度高但速度 较慢。文献[10]整合了 YOLOv3 用于宫颈细胞的检测, 开发了一个完整的宫颈液涂片诊断系统。文献[11]利 用软注意力机制改进 Faster R-CNN 用于白细胞的检测, 最佳 mAP 可达 0.744。文献[12]使用 YOLOv4 进行急性 淋巴白血病患者血液中的母细胞检测,在公共数据集上的最佳 mAP@[0.5]可达 0.96。

基于 Transformer 的 检测 Transformer (detection transformer, DETR)<sup>[13]</sup>是一种端到端的目标检测模型,采用集合预测的方法直接给出预测结果,省去了人工设计阶段和后处理过程。文献[14]提出的可变形注意力模块(deformable attention module, DAM)可以融合多尺度特征,增强了 DETR 对于小目标的检测能力。文献[15]提出的空间调制的共同注意力机制引入了空间先验和多尺度,加快了 DETR 模型的收敛。在细胞检测应用方面,文献[16]提出了一种基于注意力的实例分割的模型 Cell-DETR,达到了与掩码区域卷积神经网络(mask region based CNN, Mask R-CNN)相同的性能。

基于深度学习的白细胞检测方法尚有以下不足:1) 基于 CNN 的检测模型受到人工设计的步骤及后处理等 因素的影响,性能有待提高。2)基于 Transformer 的 DETR 作为一种新型目标检测网络,其收敛速度和小目 标检测能力有待提升。

本文面向实验室的白细胞检测真实场景,将流程简 洁的 DETR 作为基础模型进行改进,提出了一种 Transformer 架构的端到端目标检测网络基于转换器的 DETR(transformer based DETR, T-DETR)。

## 1 T-DETR 的原理

T-DETR 模型的流水线框架分为三部分,如图 1 所示。第1部分使用金字塔视觉转换器第二版(pyramid vision Transformer version 2, PVTv2)<sup>[17]</sup>提取多尺度的特征图,以提高检测精度。第2部分将多尺度特征图输入到编码器和解码器模块中,该模块中的 DAM 自然聚合多尺度特征并提高模型收敛速度。第3部分将处理后的结果与真值进行二部图匹配,输出预测集合。



图 1 所提出的模型流水线框架



## 1.1 DETR 模型

原始的 DETR 模型以 ResNet 为骨干网络提取单尺 度特征,编码器接收展平的特征做自注意力处理,解码器 则接收编码器的输出和一组可学习的目标查询集,对目 标和全局图像上下文的关系进行推理,输出预测集合。 该模型消除了人工设计的组件及后处理过程,简化了检 测流程。DETR 的模型结构及流程如图 2 所示。DETR 模型采用匈牙利算法定义损失函数使集合匹配的损失最小,包含分类损失和边界框损失<sup>[13]</sup>:

$$\mathcal{L}_{\text{Hungarian}}(y, \hat{y}) = \sum_{i=1}^{N} \left[ -\log \hat{p}_{\hat{\sigma}(i)}(c_i) + 1_{|c_i \neq \emptyset|} \mathcal{L}_{\text{box}}(b_i, \hat{b}_{\hat{\sigma}}(i)) \right]$$
(1)

其中,  $\hat{p}$  是物体被预测为  $c_i$  类的概率。  $\mathcal{L}_{\text{box}}(b_i, \hat{b}_a(i))$  是边界框损失,  $b_i$  是真值,  $\hat{b}_a(i)$  是预测边界框。



图 2 DETR 的结构及检测流程<sup>[13]</sup>



原始 DETR 模型在初始化时每个查询对所有位置的 权重相同,导致模型训练时收敛慢;在训练时只采用低分 辨率的单尺度特征图使得小目标的检测性能不佳。因此 本文针对这两方面进行改进。 的 PVTv2 作为 T-DETR 的骨干。它的设计仿照 ResNet, 分为4个阶段,结构如图3所示。每个阶段分为两部分, 其中重叠的图像块嵌入用于对局部连续性信息进行建 模,Transformer 编码器中使用线性空间还原注意力替代 多头注意力层以减少计算和内存成本。本文所使用的版 本为 PVTv2-small,大小与 ResNet-50 模型接近。



为了得到多尺度特征图,我们使用基于 Transformer



图 3 PVTv2 模型的结构<sup>[17]</sup>



#### 1.3 DAM

DAM 只关注参考点周围的几个关键点,可以有效减 少特征映射的复杂度,加速模型的收敛。

将单尺度特征图  $x \in R^{C \times H \times W}$  输入到 DAM 中,则 DAM 的计算公式为<sup>[14]</sup>:

$$DeformAttn(z_q, p_q, x) = \sum_{m=1}^{M} \boldsymbol{W}_m \times \left[\sum_{k=1}^{K} A_{mqk} \cdot \boldsymbol{W}'_m x(p_q + \Delta p_{mqk})\right]$$
(2)

其中, q 是查询元素的索引,  $z_q$  是查询元素的特征表示,  $p_q$  是参考点, M 是注意力头的总数, K 是被采样键的总数。

 $\Delta p_{mqk}$ 和 $A_{mqk}$ 分别表示第m个注意力头中第k个采样点的采样偏移和注意力权重。 $W_m$ 和 $W'_m$ 是可学习的权重。

DAM 还可以通过注意力机制自然聚合 PVTv2 生成 的多尺度特征图,提升检测精度。输入的多尺度特征图 为  $\{x^{l}\}_{l=1}^{L}$ ,其中  $x^{l} \in R^{C \times H_{l} \times W_{l}}$ 。多尺度可变形注意力特 征的计算公式为<sup>[14]</sup>:

$$MSDeformAttn(z_{q}, \hat{p}_{q}, \{x^{l}\}_{l=1}^{L}) = \sum_{m=1}^{M} W_{m} \times$$

$$\sum_{l=1}^{L} \sum_{k=1}^{K} A_{mlqk} \cdot W'_{m} x^{l} (\phi_{l}(\hat{p}_{q}) + \Delta p_{mlqk}) ]$$

$$(3)$$

其中,  $\hat{p}_q \in [0,1]^2$  是每个查询元素参考点的归一化 坐标, *L* 是特征层的总数,  $\Delta p_{mlqk}$ 和  $A_{mlqk}$ 分别表示第 *l* 个特 征层中的第 *m* 个注意力头中第 *k* 个采样点的采样偏移和 注意力权重。 $\phi_l(\hat{p}_q)$  对 $\hat{p}_q$ 坐标进行缩放以对应相应的特 征层。应当注意的是, 当 *L* = 1 时, 式(3)将退化为 式(2)。

## 2 实验验证

#### 2.1 数据集

在大规模自然数据集上进行预训练,并在医学数据 集上进行迁移学习(transfer learning, TL)训练,已成为医 学成像领域的常规做法。模型的 TL 训练比从头开始 (from scratch, FS)的训练表现出更好的性能<sup>[18]</sup>。

本文 T-DETR 模型预训练使用的数据集是 COCO 数据集<sup>[19]</sup>,该数据集包含 118 k 训练图像和 5 k 验证图 像。本文使用的白细胞数据集为 Raabin 公共数据 集<sup>[20]</sup>,该数据集由两名血液学专家标注,但存在大量有 争议和未识别的细胞,且标注框包含较多背景区域。 本文对无争议的细胞图片进行筛选和重新标注,并由 检验科医生进行确认,筛选后的数据集信息如表 1 所 示。筛选后的白细胞主要有 3 类:嗜酸性粒细胞、单核 细胞、中性粒细胞。训练集和验证集按照 8:2比例划 分。该数据集未包含患者信息,原作者未说明数据存 在伦理风险。

表 1 Raabin 数据集的详细信息 Table 1 Details of the Raabin dataset

图片数量	实验设备	成像设备	放大倍数	分辨率
10 323	CX18 显微镜	Samsung Galaxy S5 相机	100	5 312×2 988

#### 2.2 检测性能评估

目标检测模型的性能评估需要同时考虑到模型的准确率和召回率,一般使用平均精度(average precision, AP)来评价。AP值由准确率和召回率曲线下方的面积计算。对于多类别的目标检测,使用平均 AP(mean AP, mAP)来评估检测性能:

$$nAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{4}$$

其中,n是目标类别的总数,AP<sub>i</sub>是第 i 个类别的 AP 值。

本文使用的是 COCO 数据集的常用目标检测评价指标 mAP@[0.5:0.95],该指标为交并比取值 0.5~0.95, 间隔为 0.05 时的 10 个数值所对应 mAP 的均值。该指标能够更全面评价模型检测和定位的准确性。

#### 2.3 模型训练策略

本文基于 pytorch1.8.0 框架搭建模型,使用4个 NVIDIA RTX 3090 GPU 进行分布式训练,系统是 Ubuntu 20.104.1 LTS,CPU为 Intel<sup>®</sup> Xeon<sup>®</sup> Silver 4210R。我 们训练的检测模型有:DETR、T-DETR、Faster R-CNN、 YOLOv3。各模型的训练参数如表 2 所示。

表 2 模型训练的参数设置 Table 2 Parameter settings for model training

						-			
模型	数据集	训练方式	优化器	学习率	动量	权重衰减	轮次	批量大小	用时/h
T-DETR	COCO	FS	AdamW	2×10 <sup>-4</sup>	-	$1 \times 10^{-4}$	70	4	229
DETR	Raabin	TL/FS	AdamW	$1 \times 10^{-4}$	-	$1 \times 10^{-4}$	100	8	11
T-DETR	Raabin	TL/FS	AdamW	$2 \times 10^{-4}$	-	$1 \times 10^{-4}$	100	4	24
Faster R-CNN	Raabin	TL	SGD	2×10 <sup>-2</sup>	0.9	$1 \times 10^{-4}$	100	8	11
YOLOv3	Raabin	TL	SGD	$1 \times 10^{-2}$	0.9	5×10 <sup>-4</sup>	100	8	9

#### 2.4 结果分析

## 本文的实验主要分为3个部分:

1) 实验是在环境中的微软常见物体(Microsoft common objects in context, COCO)数据集上对T-DETR进行FS训练,为T-DETR在白细胞数据集上的TL训练提供预训练权重。总训练周期为70个轮次,在第61个轮次mAP取得最大值0.476,训练过程如图4所示。

2)实验包含两部分内容。首先,在白细胞数据集上 对 T-DETR 和 DETR 进行 TL 训练,两者的预训练权重均 来自于 COCO 数据集。如图 5 所示,T-DETR 在第 40 个 轮次开始收敛,收敛速度较 DETR 快,且 mAP 值高于 DETR。然后,在白细胞数据集上对 T-DETR 和 DETR 进 行 FS 训练。训练时 DETR 不能收敛,且 mAP 值一直为 0,因此结果未在图中展示。虽然 T-DETR 的 FS 训练性 能略低于 TL 训练,但超过了 DETR 的 TL 训练。由此可 见,所提出的 T-DETR 优于 DETR 的性能,且 TL 训练有 助于提升模型性能。

3) 部分实验中, 在白细胞数据集上对 CNN 模型 Faster R-CNN 和 YOLOv3 进行 TL 训练, 得到最优的模型 权重以实现白细胞的检测与识别。如图 6 所示, YOLOv3







图 5 各模型在白细胞数据集上的训练损失和 mAP Fig. 5 Training loss and mAP for each model on the leukocyte dataset

的检测性能明显低于其他模型。虽然 T-DETR 的收敛速 度低于 Faster R-CNN,但随着训练的进行,T-DETR 的检 测性能超过了 Faster R-CNN。





## 2.5 讨 论

本文将 Transformer 结构的目标检测模型应用在医学 图像的领域中,具有一定的创新性。分别从检测性能和 实际检测效果两方面讨论 T-DETR 的优越性:

首先,各模型在白细胞数据集上的检测性能 mAP 如 表 3 所示。经过 TL 训练后的 T-DETR 取得了最优的检 测性能。此外,还对每类白细胞的检测性能 AP 进行了 统计。如图 7 所示,T-DETR(TL)对于三类白细胞的检测 性能几乎均达到了最优,其中对于中性粒细胞的检测性 能最高,可能与它的形态易于识别有关。

表 3 各模型的检测性能 Table 3 Detection performance of each model

	1	
模型	训练方式	mAP
DETR	TL	0.902
T-DETR	TL	0.954
T-DETR	FS	0.943
Faster R-CNN	TL	0.952
YOLOv3	TL	0. 699



其次,使用具有一定检测难度的图片输入到具有最 优权重的模型中得到检测与识别结果。如图 8 所示,当 检测目标较多时,YOLOv3(TL)出现了漏检和定位框不 准确的情况,T-DETR(TL)的整体置信度高于其他模型。 另外,在收集到的临床数据(未涉及患者个人信息)上进 行了模型实际检测性能的验证。典型检测结果如图 9 所 示,本文所提出的 T-DETR(TL)在检测的准确率和置信 度方面优于其他模型。



(d) T-DETR 模型FS训练 (d) T-DETR model FS training

(e) Faster R -CNN模型TL训练 (e) Faster R -CNN model TL training

(f) YOLO v3 模型TL训练 (f) YOLO v3 model TL training

图 8 各模型对白细胞检测结果的局部放大图 Fig. 8 Local magnification of leukocyte detection results for each model



(a) 原图 (a) Original image



(d) T-DETR 模型FS训练 (d) T-DETR model FS training



(b) DETR 模型TL训练 (b) DETR model TL training







(c) T-DETR 模型TL训练 (c) T-DETR model TL training



(f) YOLO v3 模型TL训练 (f) YOLO v3 model TL training

图 9 各模型对白细胞临床数据检测结果的局部放大图

Fig. 9 Local magnification of the results of each model on leukocyte clinical data

## 2.6 局限性及展望

以输入图片大小为1200×800为例,对各模型的计算量和参数量进行了计算,在单个 NVIDIA RTX 3090

GPU上对验证集中的 2 000 张图片进行了推理测试,如 表 4 所示。在推理速度上,YOLOv3 最快,而我们提出的 T-DETR 较慢。一个原因是多尺度的特征图增加了计算

量,减慢了推理速度。另一个最重要原因是现有 GPU 加速对 Transformer 模型的优化不足,限制了推理速度<sup>[13]</sup>。

#### 表 4 各模型的计算量、参数量及推理速度

 Table 4 Computational volume, parameters and inference speed of each model

模型	计算量/GFLOPs	参数量/M	推理速度/Fps
DETR	91	41	22. 4
T-DETR	174	37	10.6
Faster R-CNN	206	41	23.5
YOLOv3	193	61	57.3

虽然 T-DETR 的检测速度存在不足,但医学图像的 精确诊断更为重要。最近的研究表明,去掉网络骨干的 DETR 能够在保持良好性能的同时进一步加快推理速 度<sup>[21]</sup>。未来,随着 GPU 加速的优化及模型结构的改进, 基于 Transformer 的模型速度会进一步提升。

此外,本文的检测模型是在正常白细胞数据集上进行的训练,未进行临床验证。未来将收集一些针对特定 疾病(如白血病)的临床数据进行训练和验证,在通过伦 理论证后进行实时测试,扩展模型的适用性。

## 3 结 论

本文借助于 DETR 检测流程简单的优势,将基于 Transformer 的 PVTv2 作为模型的骨干网络,并引入 DAM 降低计算复杂度,构建了一个基于 Transformer 结构的目 标检测模型 T-DETR。该模型对大视野显微图片中白细 胞检测的 mAP 值可达 0.954,超越了经典的 CNN 检测算 法,实现了白细胞的高精度检测。该模型将最近在计算 机视觉领域中表现优异的 Transformer 结构引入到医学图 像领域,是一次有益的尝试。

## 参考文献

- LENG B, LENG M, GE M, et al. Knowledge distillation-based deep learning classification network for peripheral blood leukocytes[J]. Biomedical Signal Processing and Control, Elsevier Ltd, 2022, 75 (December 2021): 103590.
- [2] LA GIOIA A, FIORINI F, FUMI M, et al. A prolonged microscopic observation improves detection of underpopulated cells in peripheral blood smears[J]. Annals of Hematology, 2017, 96(10): 1749-1754.
- [3] 周雨宁,夏华,王晓莉,等.基于谱分解动态散射成像的细胞无标记检测与分类方法[J].电子测量与仪器学报,2022,35(4):1-7.

ZHOU Y N, XIA H, WANG X L, et al. Study on label-free cell detection and classification method by using spectral decomposition-based dynamic scattering imaging[J]. Journal of Electronic Measurement and Instrumentation, 2022, 35(4): 1-7.

- [4] DONG N, ZHAI M, CHANG J, et al. A self-adaptive approach for white blood cell classification towards pointof-care testing[J]. Applied Soft Computing, 2021, 111: 107709.
- [5] BANIK P P, SAHA R, KIM K D. An automatic nucleus segmentation and cnn model based classification method of white blood cell[J]. Expert Systems with Applications, Elsevier Ltd, 2020, 149: 113211.
- [6] RAGHU M, UNTERTHINER T, KORNBLITH S, et al. Do vision transformers see like convolutional neural networks? [J]. Advances in Neural Information Processing Systems (NeurIPS), 2021, 34: 12116-12128.
- [7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. International Conference on Learning Representations (ICLR), 2021.
- [8] REDMON J, FARHADI A. YOLOV3: An Incremental Improvement[J]. ArXiv, 2018, DOI: 10.48550/arXiv. 1804.02767.
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, 2017, 39 (6): 1137-1149.
- [10] ZHU X, LI X, ONG K, et al. Hybrid AI-assistive diagnostic model permits rapid TBS classification of cervical liquid-based thin-layer cell smears [J]. Nature Communications, 2021, 12(1): 1-12.
- [11] WANG D, HWANG M, JIANG W C, et al. A deep learning method for counting white blood cells in bone marrow images [J]. BMC Bioinformatics, BioMed Central, 2021, 22: 1-13.
- [12] KHANDEKAR R, SHASTRY P, JAISHANKAR S, et al. Automated blast cell detection for Acute Lymphoblastic leukemia diagnosis[J]. Biomedical Signal Processing and Control, 2021, 68(May): 102690.
- [13] CARION N, MASSA F, SYNNAEVE G, et al. End-toend object detection with transformers [J]. European Conference on Computer Vision (ECCV), 2020, 12346:

213-229.

- [14] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable transformers for end-to-end object detection[J]. International Conference on Learning Representations (ICLR), 2021(June): 1-14.
- [15] GAO P, ZHENG M, WANG X, et al. Fast convergence of detr with spatially modulated co-attention [C]. IEEE/ CVF International Conference on Computer Vision (ICCV), 2021.
- PRANGEMEIER T, REICH C, KOEPPL H. Attentionbased transformers for instance segmentation of cells in microstructures [J]. Proceedings-2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020, 2020; 700-707.
- [17] WANG W, XIE E, LI X, et al. PVTv2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2021, 8(3): 415-424.
- [18] RAGHU M, ZHANG C, KLEINBERG J, et al. Transfusion: Understanding transfer learning for medical imaging[J]. Advances in Neural Information Processing Systems (NeurIPS), 2019.
- [19] LIN T, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context [J]. European Conference on Computer Vision (ECCV), 2014(June): 740-755.
- [20] KOUZEHKANAN Z M, SAGHARI S, TAVAKOLI S, et al. A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm [J]. Scientific Reports, Nature Publishing Group UK, 2022, 12(1): 1-14.
- [21] LIU F, WEI H, ZHAO W, et al. WB-DETR: Transformer-based detector without backbone [C].
   IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, 1: 2959-2967.

## 作者简介



冷冰,2011年于哈尔滨工业大学获得学 士学位,2015年于哈尔滨工业大学获得硕士 学位。现为中科院苏州医工所助理研究员, 主要研究方向为人工智能与图像识别。 E-mail: lengbing@ sibet. ac. cn

Leng Bing received his B. Sc. degree from Harbin Institute of Technology in 2011, and received his M. Sc. degree from Harbin Institute of Technology in 2015. He is currently an assistant research fellow at Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences. His main research interests include artificial intelligence and image recognition.



冷敏,1994年毕业于聊城市卫生学校, 2009年于潍坊医学院获得学士学位。现为 聊城市心脏病医院副主任技师,主要研究方 向医学检验。

E-mail: yylengmin@ 126. com

Leng Min graduated from Liaocheng Health School in 1994 and received her B. Sc. degree from Weifang Medical University in 2009. She is currently the deputy chief technologist of Liaocheng Cardiac Hospital. Her main research interest is medical laboratory science.



**董文飞**(通信作者),1996年于浙江大 学获得学士学位,2004年于德国波兹坦大学 获得博士学位。现为中科院苏州医工所研 究员,主要研究方向为生物医学光子学。

E-mail: wenfeidong@ sibet. ac. cn

**Dong Wenfei** (Corresponding author) received his B. Sc. degree from Zhejiang University in 1996 and Ph. D. degree from University of Potsdam in 2004. He is now a research fellow at Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences. His main research interest is biomedical photonics.