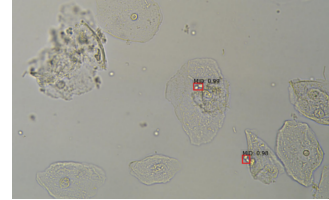


DOI: 10.12086/oe.2022.210361

## 采用注意力机制的显微图像智能检测方法



郝如茜, 王祥舟, 张静, 刘娟秀\*, 杜晓辉, 刘霖

电子科技大学光电科学与工程学院, 四川成都 611731

**摘要:** 显微图像普遍具有背景复杂、细胞多重叠的特点, 传统的图像处理方法由于其技术局限性, 无法实时准确地完成识别任务。针对上述问题, 本文提出一种采用注意力机制的显微图像智能检测方法, 该方法对目标检测模型 DETR 进行改进, 利用分组卷积机制对输入特征进行降维并分别训练不同卷积核实现特征提取, 从而提高模型对于显微图像中目标物体的注意力, 并增加了检测准确率。本文以阴道炎致病微生物检测为应用场景, 实验结果表明, 改进的模型平均查准率均值 mAP 为 96.3%, 比原模型提高约 10%。同时, 本方法在细胞重叠、粘连以及背景杂质繁多等场景下, 仍具有高鲁棒性和强检测能力。每张显微图像检测耗时约为 88.8 ms, 可以满足实时检测的需求。因此, 本文提出的模型可实时准确地对常见的阴道炎致病微生物进行实时高效自动检测, 可满足显微镜临床使用的实际需求。

**关键词:** 显微图像; 注意力机制; 深度学习; 阴道炎; 智能检测

**中图分类号:** TP391

**文献标志码:** A

郝如茜, 王祥舟, 张静, 等. 采用注意力机制的显微图像智能检测方法 [J]. 光电工程, 2022, 49(3): 210361

Hao R Q, Wang X Z, Zhang J, et al. An automatic object detection method for microscopic images based on attention mechanism[J]. *Opto-Electron Eng*, 2022, 49(3): 210361

## An automatic object detection method for microscopic images based on attention mechanism

Hao Ruqian, Wang Xiangzhou, Zhang Jing, Liu Juanxiu\*, Du Xiaohui, Liu Lin

School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

**Abstract:** The microscopic image has the characteristics of complex background and overlapping cells. Due to the technical limitations, traditional image processing methods cannot accurately complete the real-time recognition task. To address the above-mentioned problems, we propose an automatic detection method for microscopic images using attention mechanism. This method improves the original DETR architecture by introducing a split-transform-merge mechanism, which reduces the dimensionality of input features and trains multiple groups of convolution kernels for feature extraction, thereby effectively improving the model's feature extraction ability for the targets and increasing the accuracy of model detection rate. The experimental results show that the mAP of the improved model was 96.3%, which is 10% higher than that of the original model DETR. Meanwhile, the proposed method has superior detection capabilities for scenarios such as cell overlap, adhesion, and complex background.

收稿日期: 2021-11-13; 收到修改稿日期: 2022-01-11

基金项目: 国家自然科学基金资助项目 (61905036); 中央高校基本科研基金资助项目 (ZYGX2021YGCX020)

\*通信作者: 刘娟秀, E-mail: juanxiul@qq.com。

版权所有©2022 中国科学院光电技术研究所

Moreover, the detection time for each leucorrhea image was about 88.8 ms, which can satisfy the requirement of real-time microscopy examination.

**Keywords:** microscopic image; attention mechanism; deep learning; vaginitis; automatic detection

## 1 引言

显微镜检是一种应用广泛的疾病筛查方式, 由于显微图像中有形成数量繁多, 且背景充满杂质, 加之门诊患者数量庞大, 医生在长时间工作下极易产生视觉疲劳, 误诊或漏诊的可能性大大增加。而阴道炎是一种常见的妇科疾病, 并且该疾病难以一次性治愈, 有研究数据表明 40% 至 45% 的女性首次感染阴道炎后, 在一段时间内出现复发性感染<sup>[1]</sup>。常见的感染性阴道炎有三种, 分别为霉菌性阴道炎, 滴虫性阴道炎和细菌性阴道炎<sup>[2]</sup>。白带常规人工镜检是目前大多数医院对于阴道炎的筛查方式, 该方法可以简单且有效地对疾病进行早期诊断<sup>[3]</sup>。但是, 传统镜检需要有经验的医务人员在显微镜下对白带样本中的各个有形成分进行细致地观测, 并通过病理形态学的知识对霉菌、滴虫和线索细胞这些常见致病菌进行确认。因此, 本文提出一种采用注意力机制的显微图像智能检测方法, 并以阴道炎致病微生物检测作为应用场景, 为这一问题提供切实可行的解决方案, 具有重要的研究价值和临床意义。

随着光学成像技术的飞速发展, 在显微医学设备的前端图像获取领域已有很多学者做出相关研究成果。比如田鹏等人<sup>[4]</sup>提出利用均匀球面波数字同轴全息显微成像方式, 获得宽视场且无脏点和杂光干扰的显微图像。Zhang 等人<sup>[5]</sup>利用非线性光学技术进行生物样本的快速、无标记和化学特异性成像, 从而加快对于生物样本的研究分析进程。顾鑫等人<sup>[6]</sup>提出一种基于微流体芯片的衍射相位成像系统实现对水源性寄生虫的实时监测和无标记定量测量, 为单个微生物研究工作提供了切实可行的解决方案。崔光芒等人<sup>[7]</sup>结合多尺度分析算法和梯度绝对值算子, 提出一种适用于显微图像清晰度评价算法, 该算法有较好鲁棒性和抗噪性, 具有很好的实际应用价值。在现有的显微医学辅助诊断设备的后端图像处理部分, 仍然主要依赖于传统的图像处理方法和机器学习技术, 通过设计复杂的算法提取细胞的形态特征后, 人为设置阈值或者训练分类器进行分类识别。比如杜晓辉等人<sup>[8]</sup>先通过大

津阈值法和连通域标记分割前景图像, 再对每个前景区域提取其局部二值纹理特征, 最后训练支持向量机对上皮细胞进行分类。该方法的不足之处在于其前景图像提取步骤需要人为设置阈值, 阈值过高容易漏选上皮细胞, 而阈值过低则容易选取过多杂质影响整体模型分类速度。由于白带样本是液体, 具有一定厚度和流动性, 所以里面的有形成分很容易堆叠, 图像通常存在部分失焦模糊的现象, 该前景提取方法鲁棒性低, 导致准确率不高。自 2012 年 Krizhevsky 等人<sup>[9]</sup>提出 AlexNet 并赢得当年的 ILSVRC 图像识别大赛冠军, 深度学习方法在很多领域取得了突破性的成就, 其中包括目标识别领域。常用的基于深度学习的目标检测模型架构主要分为两类: 双阶段检测模型和单阶段检测模型。双阶段检测模型以 Faster R-CNN 为代表, 在第一阶段先通过区域推荐网络获得所有感兴趣区域 ROI (region of interest), 之后第二阶段再同时训练分类器和位置回归器对目标区域进行分类和定位。Wen 等人<sup>[10]</sup>对 Faster R-CNN 的区域推荐网络进行改进, 采用不规则比例的锚点, 在白细胞、红细胞和血小板的识别中取得了较好的效果。双阶段目标检测模型大部分时间浪费在选取 ROI 区域上, 整体耗时较长。为了满足实时检测的需求, 以 YOLO、SSD 为代表单阶段目标检测模型, 通过人为预先设置一系列的锚点或者可能的对象中心的网格, 以此来代替区域推荐阶段, 将整体的目标检测视为一个简单的回归问题, 对预测框坐标点和对应的目标类别概率进行计算。Jia 等人<sup>[11]</sup>使用 SSD 模型对宫颈癌细胞检测, 他们增加了中心损失函数以解决类内差距比类间差距大的问题, 提高了检测准确率。Jiang 等人<sup>[12]</sup>将通道注意力机制和空间注意力机制加入原始的 YOLO 网络中, 提高了 YOLO 网络对于图像特征的提取能力, 完成了血细胞计数任务。单阶段目标检测模型虽然速度较快, 能完成实时检测的需求, 但是普遍精度不高。本文中检测的白带样本普遍背景复杂且致病微生物体积较小, 所以这类检测模型并不适用于本文的应用场景。

Transformer 是一种基于自注意力机制的新颖的模型框架, 它最先被应用在自然语言处理领域, 并获

得了傲人的成绩<sup>[13]</sup>。最近学者们将 Transformer 应用于图像分类、目标检测等图像处理领域, 也取得了优异的检测指标。来自美国科技公司 Facebook 的 Carion 等人<sup>[13]</sup>提出了一种基于 Transformer 的目标检测模型 DETR, 该框架利用了 Transformer 编码器-解码器结构中新颖的基于局部和整体的注意力机制, 从而有效解决重叠目标检测问题, 提高了整体目标检测精度。并且, 利用匈牙利二分图分配算法匹配检测结果和实际目标, 代替了手动设置锚点和非最大抑制的过程, 将目标检测转化为一个集合预测问题, 从而极大简化了整体检测流程并大大减少了检测时间。DETR 也被应用于显微图像处理领域, 比如 Prangemeier 等人<sup>[14]</sup>提出了一种基于 DETR 的酵母菌分割图像算法, 证明 DETR 的注意力机制使其对于重叠的细胞有着优越的目标分割性能。

由于白带镜检作为白带常规的一项重要检验项目, 在医院中需求很大, 患者数量众多, 因此阴道炎致病微生物算法需要处理速度快, 能满足实时检测的需求。同时白带显微图像具有细胞数量多、背景复杂、细胞易堆叠和致病微生物体积较小等特点, 所以对于算法的性能和准确率也要求极高。本文选择基于注意力机制的 DETR 作为显微图像智能检测模型, 对 DETR 模型进行改进, 为其中特征提取部分使用的 ResNet50 模型<sup>[15]</sup>加入分组卷积机制, 从而提高模型对于目标物体特征的注意力, 改进后的模型可实现高精度且高效率地对显微图像中微生物自动识别检测。

## 2 显微图像中微生物自动检测理论

### 2.1 自注意力机制

DETR 中的 Transformer 是一种由多个编码器和多个解码器堆叠的模型架构, 它通过特有的多头注意力机制计算输出和输入的全局对应关系, 从而不仅提高了对于目标区域的注意力, 而且通过关注整体的输入信息降低了背景的干扰<sup>[13]</sup>。多头注意力机制对于本文中的白带图像尤为重要, 因为它计算了整个输入图像中所有对象的注意力分数权重, 提高了待检测的致病微生物注意力权重, 同时降低其余有形成分和杂质的权重, 因此在细胞堆叠和杂质较多的白带显微图像检测中占据较大优势。

单头注意力机制的工作原理是将输入的特征先通过与不同加权矩阵相乘获得查询矩阵、键矩阵和值矩

阵, 再通过查询矩阵和键矩阵计算权重矩阵, 最后将值矩阵用所得的权重矩阵相乘求得加权和, 所得结果即为注意力矩阵。假设  $Q$  是查询矩阵,  $K$  是键矩阵,  $V$  是值矩阵,  $d_k$  是  $K$  的维度, 单头注意力的计算公式如下所示:

$$A_{\text{Attention}}(Q, K, V) = S_{\text{softmax}}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (1)$$

本文所用的 DETR 中每层 Transformer 层都由 8 个单头注意力模块组成。白带图像输入模型的特征提取部分的主干 CNN 后, 将获得的特征图输入  $1 \times 1$  的卷积中, 降维成一维的特征向量。之后, 用所得的特征向量与不同的权重矩阵进行线性变换, 得到  $Q$ 、 $K$  和  $V$  矩阵, 根据式 (1) 计算每个单头注意力模块的注意力矩阵, 之后把多个单头注意力模块的输出连接起来, 最后再进行线性变换以获得最终的注意力结果。

### 2.2 基于匈牙利二分图分配的损失函数

DETR 模型的输出由两部分组成, 目标物体的预测框和目标物体的种类。该输出的数量  $N$  是由人为预先设定,  $N$  应远大于输入图像中目标物体的数目。本文结合白带图像的实际情况, 将  $N$  设置为 100。这就意味着, 并非所有的预测框都是表示待检测致病微生物, 有一些预测框应指示背景  $\emptyset$ 。为了确定哪些预测框表示目标, 哪些表示背景, DETR 模型采用了基于最小化损失的匈牙利二分图分配算法<sup>[16]</sup>。同时, 通过该算法在预测和真实值之间寻找二分匹配, 本模型减去了手动设置锚点和非最大抑制的过程, 提高了检测效率和准确率。假设  $y$  是实际集合, 包含着目标致病微生物和背景,  $\hat{y}$  是预测集合,  $P_N$  代表预测集合和实际集合中的所有可能匹配的排列组合情况, 通过对所有排列组合情况计算预测和真实值间的损失大小, 用匈牙利二分图分配算法求得损失最小的最佳组合情况  $\hat{\sigma}$ , 公式如下所示:

$$\hat{\sigma} = \operatorname{argmin}_{\sigma \in P_N} \sum_i^N \mathcal{L}_{\text{match}}(y_i, \hat{y}_{\sigma_i}). \quad (2)$$

模型的整体损失函数即为在最佳匹配下的损失。由于 DETR 模型的输出分为预测框和种类, 则模型的损失函数也由分类损失和边界框损失组合而成。分类损失采用交叉熵损失, 如下所示:

$$L_{\text{class}} = - \sum_i^N \log \hat{p}_{\sigma_i}(c_i). \quad (3)$$

边界框损失则采用 LI 损失和 GIoU 损失的线性组合, 如下所示:

$$L_{\text{box}} = \sum_i^N \lambda_{L_i} \|b_i - \hat{b}_{\sigma_i}\| + \lambda_{\text{giou}} L_{\text{giou}}(b_i - \hat{b}_{\sigma_i}). \quad (4)$$

而当实际标签为背景 $\emptyset$ 时, 背景并没有实际的目标坐标, 所以边界框损失为 0, 此时仅有分类损失, 背景和背景之间的损失为常数, 根据文献 [13] 所述, 当该常数设置为预测得到的概率时效果最好, 本文的模型遵从了这一参数设定。

### 2.3 改进特征提取部分的模型架构

原始 DETR 中用于图像特征提取的模型为残差网络 ResNet50<sup>[15]</sup>, 该模型的基本残差单元如图 1(a) 所示, 它将输入  $X$  与卷积模块的输出  $F(X)$  相加进行身份映射, 残差单元的输出结果  $X+F(X)$  则作为输入传入下一残差单元, 实现整个模型结构的跳跃连接, 实验证明该模型可有效缓解层数较多的神经网络中梯度消失的问题。

由于白带显微图像背景复杂, 细胞特征多变, 所

以本文对原始 ResNet 模型<sup>[15]</sup> 进行改进, 使用具有分组卷积结构的 ResNeXt50\_32×4d 模型<sup>[17]</sup>, 该模型的基本单元结构如图 1(b) 所示, 每组卷积的具体操作为先将输入  $X$  通过  $1\times 1$  卷积进行降维, 之后再使用  $3\times 3$  卷积核进行卷积操作, 最后再将其恢复为输入  $X$  的维度大小。

每组卷积的操作都相同, 在 ResNeXt50\_32×4d 中基数设置为 32, 即一共分为 32 组, 而 4d 则表示输入特征图被  $1\times 1$  卷积将维度减少成 4。这种分组卷积类似于注意力机制, 对于输入  $X$  通过使用和训练不同的卷积核进行特征提取, 可以让神经网络学习输入的不同特征。同时,  $1\times 1$  的卷积核降维操作也降低了参数的数量, 大大减少了计算复杂度。

### 2.4 显微图像智能检测模型总体流程图

本文所使用的显微图像智能检测模型总体流程图如图 2 所示, 首先将训练集中的白带图像直接送入改

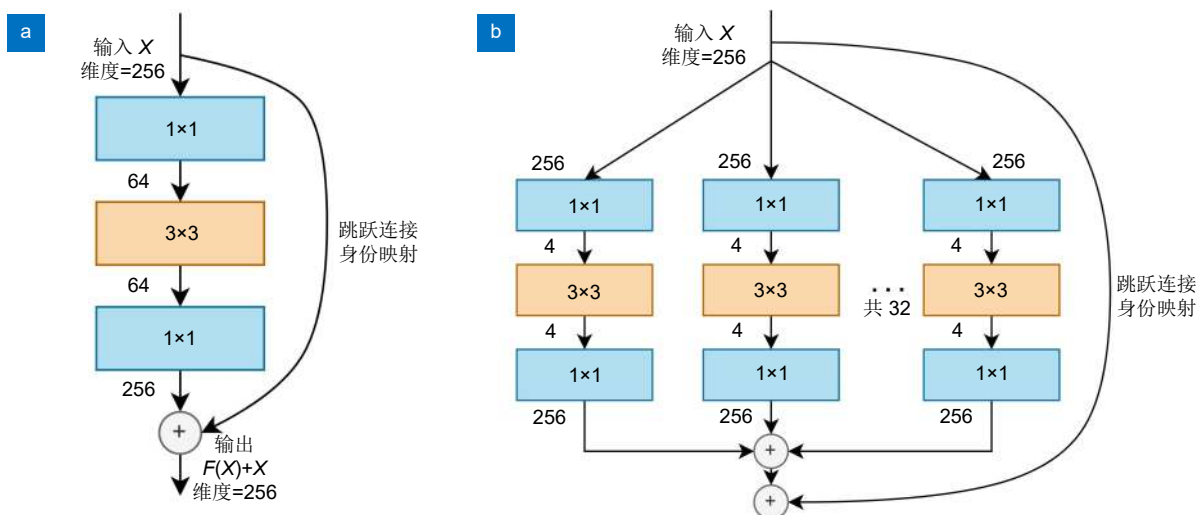


图 1 ResNet 和 ResNeXt 模型的基本单元。

(a) ResNet 模型基本单元; (b) ResNeXt 模型基本单元

Fig. 1 The basic blocks of ResNet and ResNeXt.

(a) Basic block of ResNet; (b) Basic block of ResNeXt

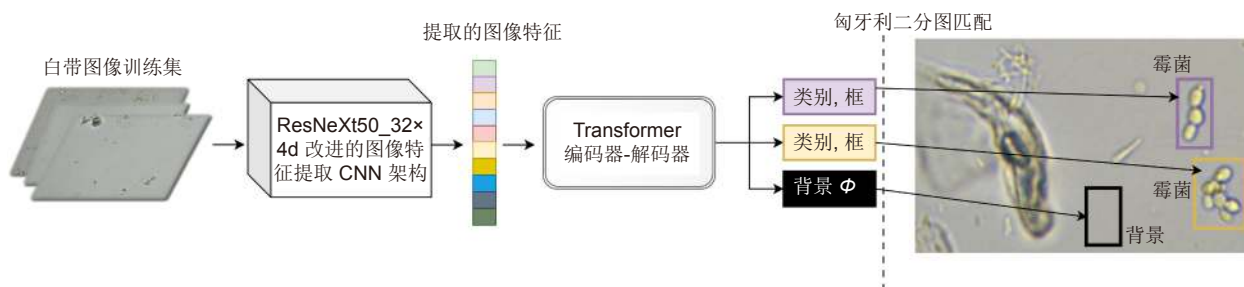


图 2 算法总体流程图

Fig. 2 Workflow of the proposed algorithm

进的图像特征提取 CNN 神经网络 ResNeXt50\_32×4d 中, 该改进的 CNN 通过分组卷积机制进行白带图像的特征提取, 此机制采用特征降维和不同组卷积核分离训练的方式, 大幅提高了神经网络对于特征的提取能力。

其次, 提取到的图像特征传入具有多头注意力机制的 Transformer 编码器-解码器中, 该模块的注意力机制从复杂的背景中将较大权重放置于待识别的目标物体上, 可有效提升模型对于重叠细胞的识别能力。

最后, 输出的预测集合包括预测的目标类别和定位框。本检测模型采用匈牙利二分图匹配算法将预测值和真实值进行匹配, 找到使类损失和边界框损失最小的一种匹配方式。该过程将预测值中的目标和背景进行区分, 代替了手动设置锚点和非最大抑制的过程, 提高了整体检测速率和准确率。本模型利用找到的最佳匹配方式计算整体网络损失和进行反向传播训练, 完成整体的参数更新。

本文提出的端对端的显微图像智能检测模型, 结合分组卷积和注意力机制, 可高准确率地完成对三种常见的阴道炎致病微生物霉菌、滴虫和线索细胞实时检测的任务。

### 3 实验结果与分析

#### 3.1 数据集采集

显微白带图像光学采集系统由奥林巴斯 CX31 的

生物显微镜和图谱 EXCCD01400KMA CCD 相机组成。显微镜搭载着 40 倍数值孔径为 0.65 的物镜。相机的曝光时间为每帧 40 ms。该光学采集系统的视场约为  $0.41 \text{ mm} \times 0.26 \text{ mm}$ 。

通过与四川省成都市第六人民医院合作, 一共采集了来自 112 个病人的 1089 张白带显微图像, 每张图像的分辨率为  $1920 \text{ pixels} \times 1200 \text{ pixels}$ 。由医生对每张图像里的滴虫、霉菌和线索细胞进行标注。这三种常见阴道炎致病菌形态各不相同, 图 3(a)~3(c) 分别展示了本数据集中的霉菌、滴虫和线索细胞的显微图像。

本文数据集中一共包含 3057 个标注细胞, 具体为 2901 个霉菌, 114 个滴虫, 42 个线索细胞。本文将数据集中的图像按照 6:2:2 的比例划分成训练集, 验证集和测试集, 具体的划分情况见表 1。

#### 3.2 评价指标

为了对模型对致病微生物的检测效果定量分析, 本文采用了目标检测任务中的常见评价指标, 它们分别为查准率 (precision,  $P$ ), 召回率 (recall,  $R$ ), 平均查准率 (average precision,  $AP$ ) 和平均查准率均值 (mean average precision,  $mAP$ )。它们的公式如下所示:

$$P = \frac{TP}{TP + FP}, \quad (5)$$

$$R = \frac{TP}{TP + FN}, \quad (6)$$



图 3 三种常见的阴道炎致病菌显微图像。

(a) 霉菌; (b) 滴虫; (c) 线索细胞

Fig. 3 Microscopic images of the three common pathogenic cells that cause vaginitis.

(a) Mildew; (b) Trichomonas; (c) Clue cell

表 1 数据集具体划分情况

Table 1 The details of dataset split

数据集类别	图像张数	标注细胞总数	霉菌个数	滴虫个数	线索细胞个数
训练集	653	1844	1757	63	24
验证集	218	593	562	21	10
测试集	218	620	582	30	8

$$AP = \sum_n (R_n - R_{n-1}) P_n, \quad (7)$$

$$mAP = \frac{1}{N_{\text{class}}} \sum_{i=1}^{N_{\text{class}}} AP_i, \quad (8)$$

其中:  $TP$  指实际为致病微生物, 且正确被识别为致病微生物的数目;  $FP$  指实际为背景或其他细胞, 但错误地被识别为致病微生物的数目;  $FN$  为指实际为致病微生物, 但错误地被识别为背景或其他细胞的数目。

本文设定交并比 IoU(intersection over union) 阈值为 0.5, 即当预测的边界框和实际目标框的交集, 与这两者的并集的比例大于 0.5 时, 且类别判定正确时, 此次预测为正确预测。

### 3.3 模型训练参数

本文中的所有实验均在型号为 NVIDIA GeForce RTX 2070 显卡上运行, 使用的软件版本为 Python 3.8, 和开源框架 PyTorch 1.5.1。本文设置训练周期 epochs 数目为 200, 每个训练周期按批给模型输入白带图片, 批大小 Batch size 设置为 1, 模型的 CNN 特征提取器和 Transformer 的初始学习率均设置为 0.00001, 并且每 100 个 epochs 衰减为原来的十分之一。同时, 使用 AdamW 作为模型训练优化器, 其权重衰减设置为 0.0001。

### 3.4 实验结果分析和对比

本文使用训练集不断训练更新阴道炎致病微生物检测模型, 并在每个 epoch 结束后观测所得的模型在验证集上的平均查准率均值 mAP。为了说明改进的

模型可有效提升神经网络对于目标细胞的检测能力, 本文描绘了原始 DETR 和改进的 DETR 在验证集上对每个 epoch 进行评估的 mAP 曲线图。如图 4 所示, 横轴为训练周期数, 纵轴为 mAP 指标, 红色和蓝色分别代表原始的模型和改进后的模型。可以看出最开始 0 至 40 epoch 时原始模型和改进后的模型差距并不大, 但在 40 epoch 至 200 epoch 时, 改进后的模型的表现稳定地超过原始模型, 改进的模型具有较好的检测性能。

在经历 200 个 epochs 后, 本文根据 mAP 的数值选取在验证集上表现最好的模型作为最终得到的模型, 之后将其在测试集上使用相关评价指标进行量化分析。原始模型和改进的模型在测试集上分别计算所得的平均查准率均值 mAP 和不同阴道炎致病微生物的平均查准率 AP 可见表 2。本文改进模型的 mAP 为 96.3%, 比原始 DETR 所得的 87.5% mAP 提高约 10%。改进的模型对于霉菌、滴虫和线索细胞的 AP 值分别为 93%, 95.8% 和 100%, 均比原始模型的 AP 值更高。其中, 线索细胞的 AP 值达到 100%, 原因可能与测试集中线索细胞样本较少有关, 测试集中仅有 8 个线索细胞, 所以出现了全部线索细胞检测成功的情况。

为了进一步直观地表现改进的模型和原始模型在各个阴道炎致病菌检测上的性能差异, 本文将两种模型的查准率—召回率 PR (precision-recall) 曲线进行了对比, 如图 5 所示, 橙色代表使用 ResNeXt 作为特征提取器的改进的模型, 蓝色代表使用 ResNet 作为主干 CNN 的原始模型。对比橙、蓝两色曲线, 可以看

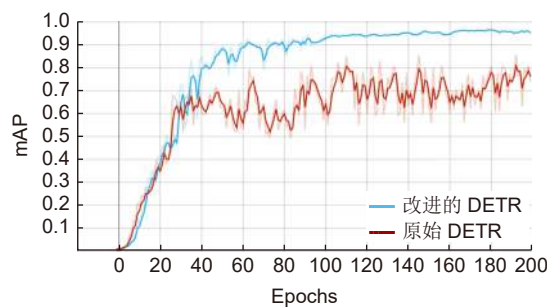


图 4 验证集表现结果对比

Fig. 4 The performance of the original DETR and the improved DETR on validation dataset

表 2 原始模型和改进模型的 mAP 和 AP 指标对比

Table 2 The comparison of mAP and AP results of the original model and the proposed model

模型	mAP/%	霉菌AP/%	滴虫AP/%	线索细胞AP/%
原始DETR	87.5	87.0	85.9	89.7
本文改进模型	96.3	93.0	95.8	100.0

出几乎每个子图的橙色曲线都比蓝色曲线更越靠近右上角, 且橙色曲线的下方面积更大, 说明改进后的模型的 mAP 值和 AP 值均比原始模型更高, 即改进后的模型检测性能优于原始模型。

为了更加详尽地展示原始模型和改进的模型的性能对比效果, 本文分别计算了在最置信度阈值下对于三种阴道致病菌的查准率 (precision) 和召回率 (recall) 指标, 如表 3 所示, 可见本文改进的模型所表现的查准率和召回率均等于或高出原始模型, 可见改进后的模型性能有了明显提升。

为了说明本文的采用注意力机制的显微图像智能检测方法对于细胞重叠和复杂背景的情形均保持较高的准确率和鲁棒性, 图 6 展示了部分白带显微图像的检测结果图。

如图 6(a) 所示, 霉菌的检测结果用红色矩形框显示, 该图有两个霉菌目标, 一个在图的中间位置, 完全重叠在上皮细胞上, 另一个霉菌在该图靠近右下角位置, 与一个上皮细胞紧密相连。如图 6(b) 所示, 滴虫的检测结果用蓝色矩形框表示, 图中有三个滴虫, 但它们因为在白带液体的不同层上, 对焦导致左下角的滴虫有边界模糊的情形。并且, 图中中间的细胞为白细胞, 与滴虫大小类似, 但形态有细微的差别, 本方法成功地将它们区分。如图 6(c) 所示, 线索细胞的检测结果用黄色矩形框表示, 该样本的充满了球菌、杆菌等阴道菌群, 并伴有数量众多的上皮细胞和白细胞, 该白带图像的背景极其复杂, 线索细胞难以从背景中区分。因此, 由图 6 的检测结果可见, 本文提出的模型对于难度较高的检测场景 (比如细胞重叠、粘连和

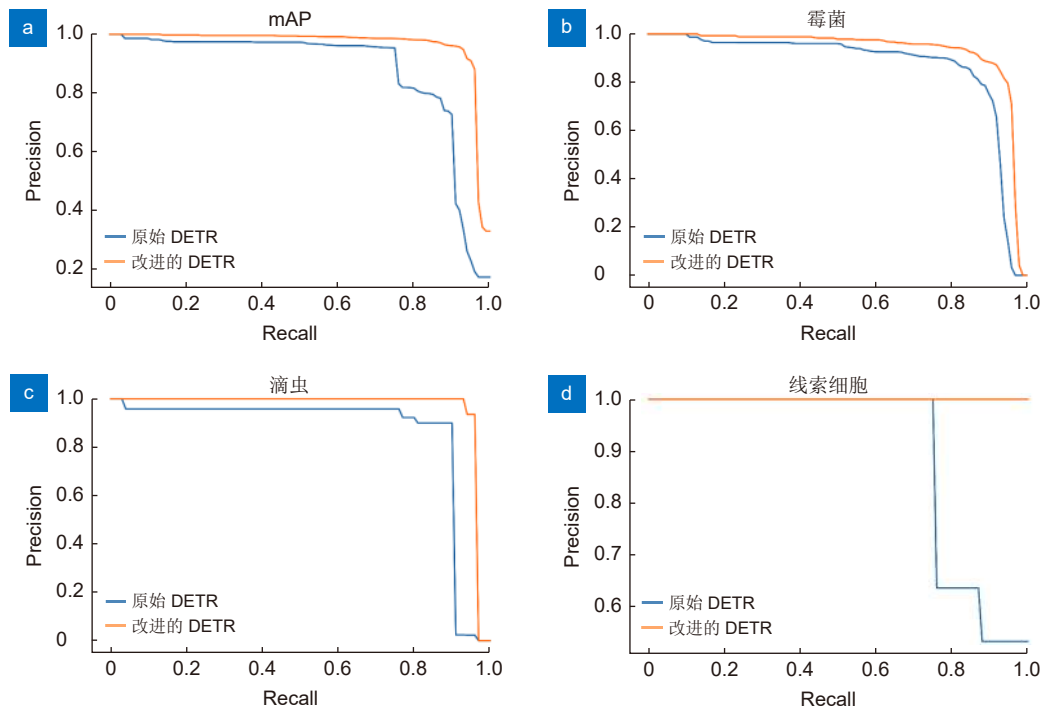


图 5 原始模型和改进的模型的 PR 曲线对比。

(a) mAP 的 PR 曲线; (b) 霉菌的 PR 曲线; (c) 滴虫的 PR 曲线; (d) 线索细胞的 PR 曲线

Fig. 5 The comparison of PR curves computed from the original model and the improved model.

(a) PR curve of mAP; (b) PR curve of mildew; (c) PR curve of trichomonas; (d) PR curve of clue cell

表 3 原始模型和改进模型的查准率和召回率指标对比

Table 3 The comparison of precision and recall results of the original DETR and the proposed improved DETR

评价指标	模型	霉菌	滴虫	线索细胞
查准率(precision)/%	原始DETR	85.3	90.0	100
	本文改进模型	88.0	100	100
召回率(recall)/%	原始DETR	85.0	90.0	75.0
	本文改进模型	91.0	93.0	100

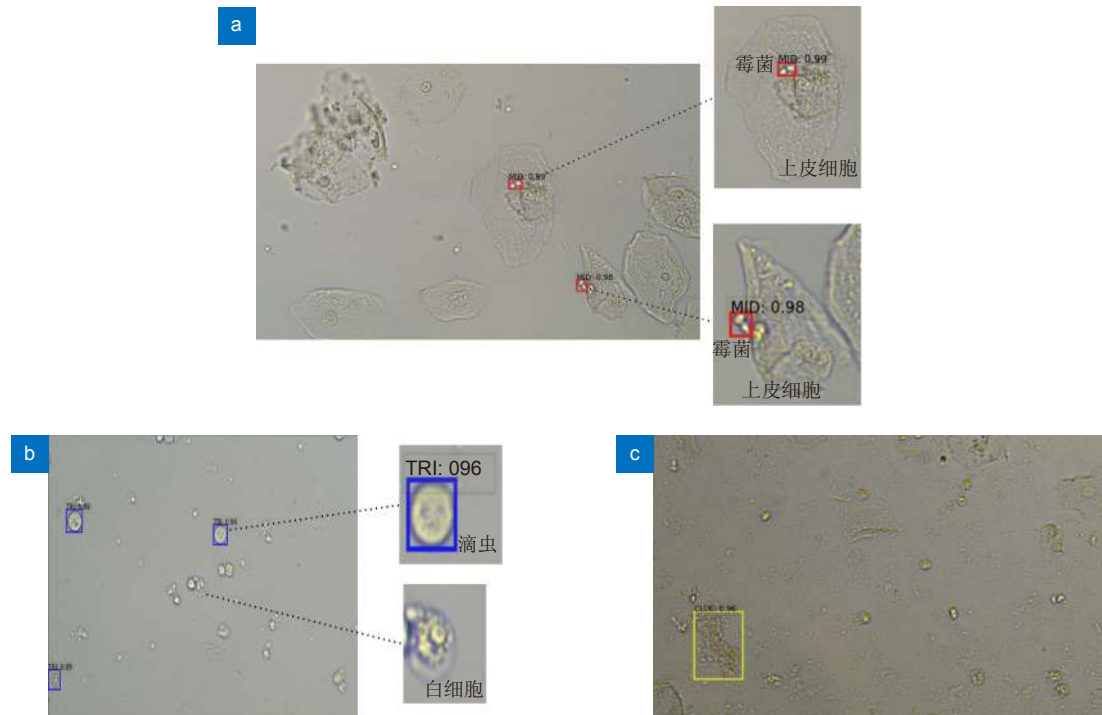


图6 三种常见阴道炎致病微生物检测结果图。

(a) 霉菌检测结果图; (b) 滴虫检测结果图; (c) 线索细胞检测结果图

Fig. 6 The detection results of the three common pathogenic cells.

(a) Detection results of mildew; (b) Detection results of trichomonas; (c) Detection results of clue cell

复杂背景等)均表现出较强的检测性能。同时, 本文将改进的模型对整体测试集的总时长除以测试集的图片张数, 计算得出平均每张白带显微图像的识别时间约为 88.8 ms, 所以本文提出的采用注意力机制的显微图像智能检测方法可满足实时检测的需求。

本文对原始 DETR 模型的特征提取部分的模型架构进行了改进, 通过引入分组卷积机制, 改进后的 DETR 模型相较于原始模型提高了对于复杂场景的特征提取能力, 图 7 对比了原始模型和改进模型对于同一幅白带显微图像的检测结果及解码器最末层的注意力权重可视化结果。

图 7(a) 为原始白带图像, 图 7(b) 为人工标注后的目标图像, 可见图的中间和右下角分别有两个霉菌目标。图 7(c) 为原始 DETR 的检测结果, 可见原始模型对于右下方与上皮细胞紧密相连的霉菌未能准确检测。图 7(d) 为原始模型注意力权重可视化图, 同时为了更好地展示模型的注意力权重, 图 7(e) 将注意力热力图与原始图像融合, 由这两幅图可见原始模型并未关注到右下角的霉菌目标, 将其当成了背景从而导致漏检。作为对比, 图 7(f) 为改进的 DETR 的检测结果, 目标细胞均被成功检测。并且, 通过观察图 7(g)

和图 7(h) 这两幅改进的模型注意力热力图, 可见通过引进分组卷积机制, 模型对于复杂场景的特征提取能力更强, 对于与其他细胞粘连的目标细胞特征也赋予了注意力权重, 从而成功检测出所有待检测目标。

## 4 结论

本文针对人工显微镜检耗时长、准确率低和需要大量人力投入等弊端, 提出了一种采用注意力机制的显微图像智能检测方法, 以临床中白带常规检查为应用场景, 该方法对白带图像中的三种阴道炎致病微生物(霉菌、滴虫和线索细胞)进行检测, 并针对显微图像中细胞易重叠以及复杂背景的特点, 对原始的 DETR 中图像特征提取 CNN 进行改进, 采用具有分组卷积机制有效地提升了模型对于目标物体特征图的提取能力。实验结果证明, 本文改进的模型的平均查准率均值 mAP 比原模型提高约 10%, 达到 96.3%。同时, 细胞检测效果图也表明本文模型对于细胞重叠、粘连以及背景有形成成分繁多等场景仍有优越的检测能力。而且, 该模型对于每张显微白带图像的平均检测时间约为 88.8 ms, 可以满足实时检测的需求。因此, 本文提出的方法可基本满足白带常规临床使用的实际



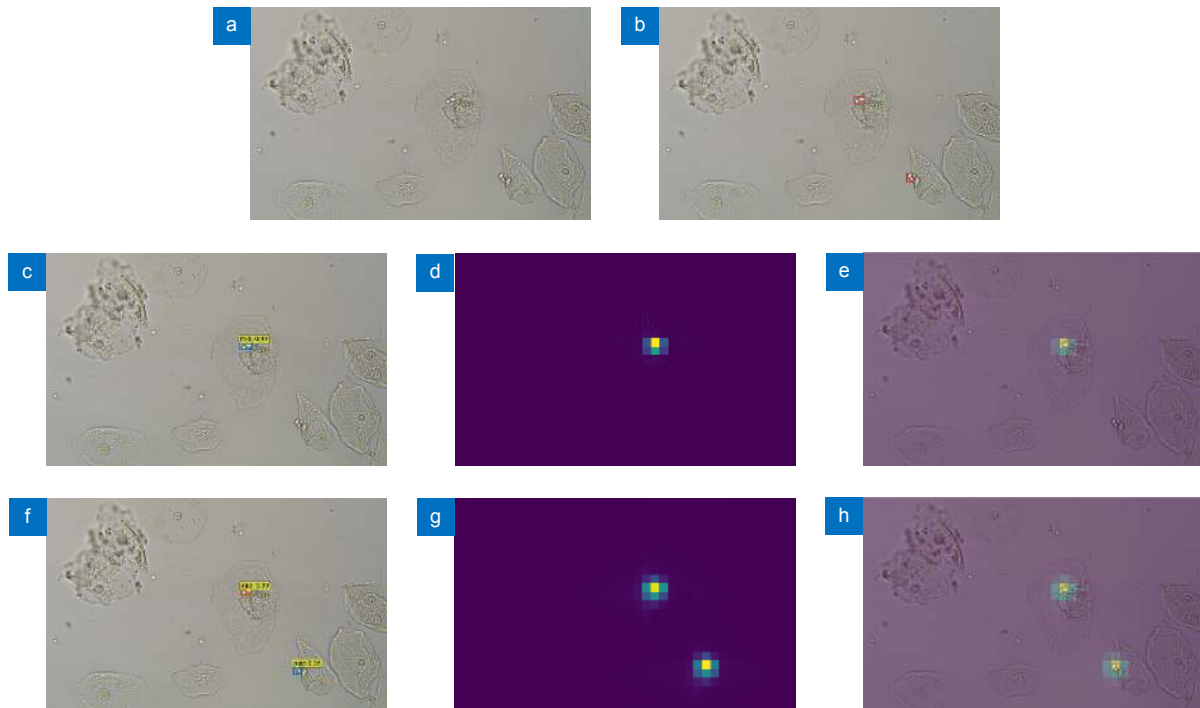


图 7 模型检测结果和注意力权重可视化图对比。

(a) 原始图像; (b) 人工标注的目标图像; (c) 原始模型检测结果图; (d) 原始模型注意力权重可视化图;  
(e) 原始模型的权重可视化图与原图的融合图; (f) 改进后模型检测结果图; (g) 改进后  
模型权重可视化图; (h) 改进后模型的权重可视化图与原图的融合图

Fig. 7 Comparison of detection results and attention weights visualization map.

(a) Original image; (b) Ground truth; (c) Detection results of original DETR; (d) Attention weights visualization of the original DETR;  
(e) Attention weights visualization of the original DETR on the original image; (f) Detection results of the improved DETR; (g) Attention  
weights visualization of the improved DETR; (h) Attention weights visualization of the improved DETR on the original image

需求。并且本方法有较好的拓展性，可以快速迁移到如粪便常规镜检、尿常规检测和血常规检查等显微临床应用。针对线索细胞样本数目较少的问题，本文将持续采集图像、增加标注样本集大小，并考虑引入针对少类别目标加权采样的策略，从而减少样本不均衡现象。同时，对于 DETR 模型训练需要大量样本的问题，本文后续计划进行数据增强、主动学习等缓解样本不足的相关方法研究。

## 参考文献

- [1] Paczkowska M, Chanaj-Kaczmarek J, Romaniuk-Drapala A, et al. Mucoadhesive chitosan delivery system with *Chelidonium herba* lyophilized extract as a promising strategy for vaginitis treatment[J]. *J Clin Med*, 2020, 9(4): 1208.
- [2] Abdul-Aziz M, Mahdy M A K, Abdul-Ghani R, et al. Bacterial vaginosis, vulvovaginal candidiasis and trichomonal vaginitis among reproductive-aged women seeking primary healthcare in Sana'a city, Yemen[J]. *BMC Infect Dis*, 2019, 19(1): 879.
- [3] Mylonas I, Bergauer F. Diagnosis of vaginal discharge by wet mount microscopy: a simple and underrated method[J]. *Obstet Gynecol Surv*, 2011, 66(6): 359-368.
- [4] Tian P, Yan W, Li F X, et al. Biology microscopy using well-distributed sphere digital in-line holography[J]. *Opto-Electron Eng*, 2019, 46(1): 180110.  
田鹏, 严伟, 李凡星, 等. 均匀球面波数字同轴全息生物显微方法[J]. *光电工程*, 2019, 46(1): 180110.
- [5] Zhang S L, Liu L W, Ren S, et al. Recent advances in nonlinear optics for bio-imaging applications[J]. *Opto-Electron Adv*, 2020, 3(10): 200003.
- [6] Gu X, Huang W, Yang L M, et al. Microfluidic diffraction phase microscopy and its application in parasites measurement[J]. *Opto-Electron Eng*, 2019, 46(12): 190046.  
顾鑫, 黄伟, 杨立梅, 等. 微流体衍射相位显微成像及其在寄生虫测量中的应用[J]. *光电工程*, 2019, 46(12): 190046.
- [7] Cui G M, Zhang K Q, Mao L, et al. Micro-image definition evaluation using multi-scale decomposition and gradient absolute value[J]. *Opto-Electron Eng*, 2019, 46(6): 180531.  
崔光芒, 张克奇, 毛磊, 等. 结合多尺度分解和梯度绝对值算子的显微图像清晰度评价方法[J]. *光电工程*, 2019, 46(6): 180531.
- [8] Du X H, Liu L, Zhang J, et al. Detection of epithelial cells in leucorrhea microscopic images based on LBP texture features[J]. *Chin J Liquid Cryst Dis*, 2019, 34(9): 871-878.  
杜晓辉, 刘霖, 张静, 等. 基于LBP纹理特征的白带显微图像中上皮细胞检测方法[J]. *液晶与显示*, 2019, 34(9): 871-878.
- [9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing*

Systems, 2012.

- [10] Wen T X, Wu H X, Du Y, et al. Faster R-CNN with improved anchor box for cell recognition[J]. *Math Biosci Eng*, 2020, 17(6): 7772–7786.
- [11] Jia D Y, Zhou J L, Zhang C W. Detection of cervical cells based on improved SSD network[J]. *Multimed Tools Appl* (2021). doi: [10.1007/s11042-021-11015-7](https://doi.org/10.1007/s11042-021-11015-7).
- [12] Jiang Z F, Liu X, Yan Z Z, et al. Improved detection performance in blood cell count by an attention-guided deep learning method[J]. *OSA Continuum*, 2021, 4(2): 323–333.
- [13] Carion N, Massa F, Synnaeve G, et al. 2020. End-to-end object detection with transformers[C]//*Proceedings of the 16th European Conference on Computer Science*, 2020: 213–229.
- [14] Prangemeier T, Reich C, Koeppl H. Attention-based transformers for instance segmentation of cells in microstructures[C]//*Proceedings of 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020: 700–707.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.
- [16] Kuhn H W. The Hungarian method for the assignment problem[J]. *Nav Res Logist Quart*, 1955, 2(1-2): 83–97.
- [17] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 5987–5995.

## 作者简介



郝如茜 (1994-), 女, 在读博士研究生, 已于电子科技大学获得硕士学位, 主要从事医学图像处理、计算机视觉、深度学习和机器学习等方面的研究。

E-mail: ruqian\_hao@163.com

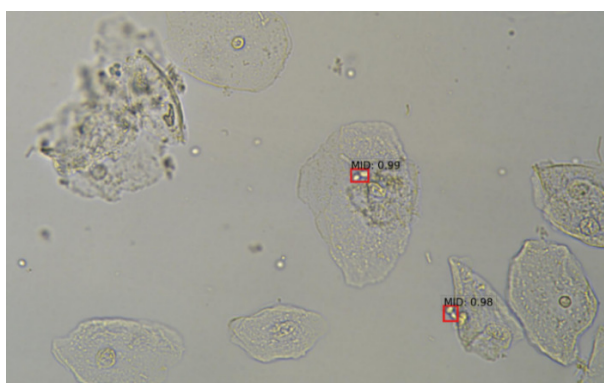


【通信作者】刘娟秀 (1980-), 女, 博士, 副教授, 主要从事机器视觉、光学计算等方面的研究。

E-mail: juanxiul@qq.com

# An automatic object detection method for microscopic images based on attention mechanism

Hao Ruqian, Wang Xiangzhou, Zhang Jing, Liu Juanxiu\*, Du Xiaohui, Liu Lin



The detection results of the common pathogenic cells

**Overview:** The microscopic image has the characteristics of complex background and overlapping cells. The traditional microscopy examination is performed by experienced doctors, and usually takes long time. Vaginitis is a common and potentially harmful gynecological disease. It can significantly increase the risk of women being infected with pelvic inflammatory disease, Human Immunodeficiency Virus (HIV), and premature birth. Routine leucorrhea examination is a common early screening method for vaginitis, but traditional microscopic examination takes a long time, and the test results are largely affected by the subjectivity of the doctor. Due to the technical limitations, traditional image processing methods cannot accurately complete the real-time recognition task. To address the above-mentioned problems, we propose an automatic object detection method for microscopic images based on attention mechanism and perform it on the scenario of vaginitis screen. This method can achieve real-time and efficient detection of three common cells, which are mildews, trichomonas and clue cells. The proposed framework uses DETR architecture based on the self-attention mechanism for detection. Different from the commonly used object detection deep learning models, DETR employs Hungarian bipartite match algorithm to assign the prediction and ground truth based on the least cost, which saves the procedure of generating anchor boxes and region of interest proposal. The self-attention mechanism is also used in DETR to generate the global connection of the targets and the whole image, which enables the model to detect the overlapped targets more accurately. In our scenario of microscopic images with complex background, to achieve better performance in image feature extraction section, we replaced the original CNN backbone with a CNN architecture with a split-transform-merge mechanism. It reduces the dimensionality of input features and trains multiple groups of convolution kernels for feature extraction, thereby effectively improving the model's feature extraction ability for the targets and increasing the accuracy of model detection rate. The experimental results show that the mAP of the improved model was 96.3%, which is 10% higher than that of the original model DETR. Meanwhile, the proposed method has superior detection capabilities for scenarios such as cell overlap, adhesion, and complex background. Moreover, the detection time for each leucorrhea image was about 88.8 ms, which can satisfy the requirement of real-time detection. Therefore, this model can accurately identify pathogenic microorganisms in real time and meet the needs of actual clinical use for leucorrhea routine examination. In addition, this method has high transferability and robustness, which can be quickly migrated to microscopic clinical applications such as routine fecal microscopy, urine composition analysis, blood routine analysis, etc.

Hao R Q, Wang X Z, Zhang J, et al. An automatic object detection method for microscopic images based on attention mechanism[J]. *Opto-Electron Eng*, 2022, 49(3): 210361; DOI: [10.12086/oe.2022.210361](https://doi.org/10.12086/oe.2022.210361)

Foundation item: National Natural Science Foundation of China (61905036) and the Fundamental Research Funds for the Central Universities (ZYGX2021YGCX020)

School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

\* E-mail: [juanxiul@qq.com](mailto:juanxiul@qq.com)