

DOI:10.13205/j.hjgc.202506019

张宇航,苑明哲,王文洪,等.改进YOLOv5的活性污泥微生物镜检方法[J].环境工程,2025,43(6):188-196.

ZHANG Y H, YUAN M Z, WANG W H, et al. An improved YOLOv5-based microscopic examination method for activated sludge microorganisms[J]. Environmental Engineering, 2025, 43(6): 188-196.

改进YOLOv5的活性污泥微生物镜检方法

张宇航^{1,2,3} 苑明哲^{2,3} 王文洪³ 张 佳¹ 肖金超^{2,3} 曹飞道³

(1. 沈阳工业大学 信息科学与工程学院, 沈阳 110870; 2. 中国科学院沈阳自动化研究所, 沈阳 110016;
3. 广州工业智能研究院, 广州 511458)

摘 要:针对现有活性污泥微生物检测算法精度低,漏检率较高的问题,提出了一种改进YOLOv5的微生物检测方法。利用K-Means++算法生成最适合本数据集的锚框组,在YOLOv5主干网络中用C3GC模块代替原有C3模块加强特征信息提取,颈部网络的特征金字塔融合坐标注意力机制与全局到空间聚合模块加强特征信息的融合。基于实际采集的活性污泥微生物数据,通过数据增强方法构建了训练和测试数据集。改进算法在测试数据集上召回率达到97.4%,平均精准度达到99.2%,模型大小仅为23.5 MB,与原始YOLOv5模型和其他主流检测模型及变种相比,准确率与回归率均取得了较大提升,并能够满足在移动端的快速部署,证明了改进算法的有效性。

关键词:污水处理;微生物检测;YOLOv5;特征金字塔;坐标注意力机制;全局到空间聚合模块

An improved YOLOv5-based microscopic examination method for activated sludge microorganisms

ZHANG Yuhang^{1,2,3}, YUAN Mingzhe^{2,3}, WANG Wenhong³, ZHANG Jia¹, XIAO Jinchao^{2,3}, CAO Feidao³

(1. School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China;
2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
3. Guangzhou Industrial Intelligence Research Institute, Guangzhou 511458, China)

Abstract: To address the research shortages, as well as the low accuracy and high missed detection rate of existing activated sludge microbial detection algorithms, an improved YOLOv5-based microbial detection method was proposed. The K-Means++ algorithm was used to generate the most suitable anchor frame group for this dataset. The original C3 module in the YOLOv5 backbone network was replaced with the C3GC module to enhance feature extraction. In addition, the feature pyramid fusion coordinate attention mechanism in the neck network and the global-to-spatial aggregation module were employed to strengthen the fusion of feature information. Based on the actual collected microbial data of activated sludge, the training and test datasets were constructed using a data augmentation method. The improved algorithm achieved a recall rate of 97.4%, an average accuracy of 99.2%, and a model size of only 23.5MB on the test dataset, which demonstrated a certain improvement in accuracy and regression rate compared to the original YOLOv5 model and other mainstream detection models and variants, while meeting the requirements for rapid deployment on mobile terminals. The experimental data showed that the improved model sacrificed a small amount of detection speed in exchange for a significant improvement in accuracy and regression rate, proving the effectiveness of the improvement. This study provides a valuable reference for the development of

收稿日期:2024-09-28; 修改日期:2024-11-08; 接收日期:2024-12-10

基金项目:国家自然科学基金面上项目(62273332);国家自然科学基金面上项目(62273337);中科院科技服务网络计划(STS)-东莞专项(20211600200072)

第一作者:张宇航(2000—),硕士研究生,主要研究方向为机器视觉、深度学习、智能制造。zhangyuhang@gz.sia.cn

*通信作者:苑明哲(1971—),博士,研究员,主要研究方向为工业物联网、分布式控制系统和工业过程控制领域。mzyuan@sia.cn

artificial intelligence in the field of environmental governance and serves as an example for the advancement of intelligent equipment.

Keywords: sewage treatment; microbial detection; YOLOv5; characteristic pyramids; coordinate attention mechanism; global-to-spatial aggregation module

0 引言

随着社会经济的快速发展,居民经济收入不断提高,人均生活用水量和污水排放量也在增加。统计数据显示,2021—2022年广州市全市生活污水排放量由75598.17万t达到了148258.10万t^[1]。

活性污泥法^[2]及其衍生改良工艺是处理城市污水最广泛使用的方法。该工艺在人工充氧的条件下,将污水与多种微生物混合培养,形成活性污泥。通过活性污泥的生物凝聚、吸附和氧化作用,有效去除污水中有机污染物。活性污泥中的微生物种群与数量变化反映了污水处理系统运行的质量和状态。相关工作人员取样后利用显微镜观察统计污水处理系统中生物相和活性污泥数量的变化,评估污水处理系统的状态,作出及时调整,确保处理系统能够继续正常运行,整个检测流程称之为“镜检”^[3]。传统的镜检方式过度依赖于人工,效率低下。人眼长时间保持同一姿势观测样本容易产生视觉疲劳,并且无法对已观测区域进行标定,容易出现漏检误检和重复计数的情况。

随着机器视觉和深度学习的不断发展,目标检测越来越多地应用到各个现实场景。研究者利用相机替换显微镜中的目镜,采用电动平台代替载物台完成显微镜对样本的遍历,结合目标检测算法对活性污泥中的微生物进行识别统计,辅助人工监测污水处理系统的运行状况。

在相关研究过程中,主要存在以下问题:1)在实际样本中,部分微生物混合在污泥杂质中,造成前景与背景的混乱,进而形成微生物边界不清晰,无法准确定位识别微生物。2)在制作活性污泥微生物数据集时,发现不同种类微生物以及同种类微生物的不同个体都存在一定的体积差异,特征信息分散复杂,难以有效利用。3)在活性污泥样本中,部分微生物在图像中所占像素较少,检测过程中极易遗漏,通常会造成检测过程出现很高的漏检率。

基于以上问题,本文建立了1个用于活性污泥微生物检测的数据集,包括7种重要的用于监测污泥处理系统的运行状态的微生物,如钟虫、楯纤虫、表壳

虫等;并设计了1个新YOLOv5模型,使用K-Means++算法生成最佳先验框,主干网络使用C3GC模块加强特征提取,颈部网络中融合GLSA模块与坐标注意力机制加强特征信息融合,并通过消融实验与对比试验验证了改进算法的有效性。

1 研究现状

1.1 目标检测算法

目标检测技术^[4]主要分为2种:第1种是基于候选区域的两阶段的目标检测模型,主要分为2个过程,第1步提取可能包含目标的候选区域;第2步在候选区域上进行分类检测,得到每个区域的对象类别。常见的两阶段算法有基于区域的快速卷积神经网络(Fast R-CNN)^[5]和基于区域的掩码卷积神经网络(Mask R-CNN)^[6]。第2种是基于回归的一阶段目标检测模型,直接分离特定的类别,并对边界进行回归,YOLO(You Only Look Once)^[7]系列模型是一阶段目标检测最具代表性的模型。与两阶段模型相比,一阶段模型具备更快的速度,但精度略低。随着研究的深入,更加先进的网络架构与即插即用的功能模块层出不穷,研究者们逐渐在YOLO系列模型上实现了精度与速度的平衡,且由于YOLO模型的结构相对简单、利于部署,两阶段检测模型也已逐渐淡出研究者的视野。

1.2 YOLOv5架构

YOLOv5是由UltralyticsLLC公司发布的一种单阶段目标检测算法,有YOLOv5s、YOLOv5m、YOLOv5l和YOLOv5x共4种版本,模型的计算复杂度递增,检测精度递增,检测速度递减,模型大小递增。原始YOLOv5架构分为输入端、主干网络(Backbone)、颈部网络(Neck)和检测头(Detect)。具体包括输入端的Mosaic数据增强、自适应锚框计算与自适应图片缩放;主干网络端的Conv模块,C3模块与SPPF模块;颈部网络端的双向特征金字塔结构;输出端的损失函数以及预测框筛选的NMS函数。其架构如图1所示。

1.3 特征提取

在YOLO中,特征提取通常采用深度卷积神经网络

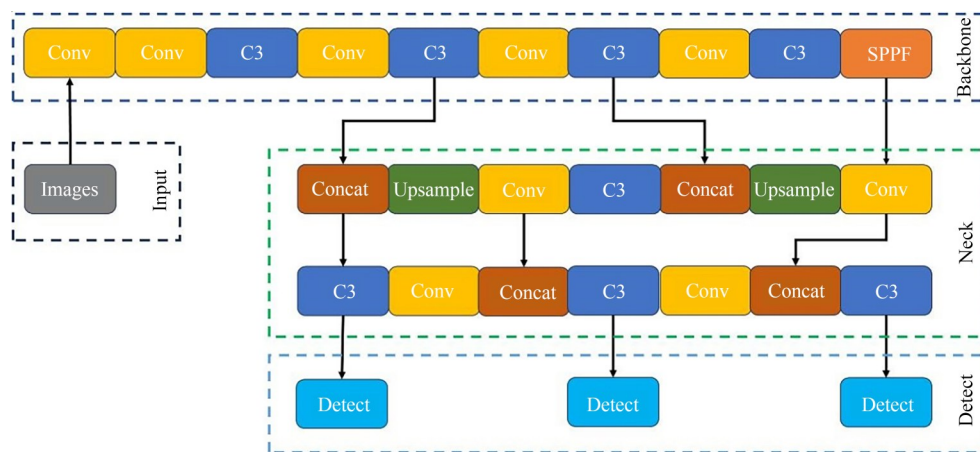


图 1 YOLOv5 原始架构

Figure 1 Original architecture of YOLOv5

络(CNN),通过多个卷积层和池化层进行图像处理,逐步提取出不同层次的特征信息。低层特征捕捉边缘和纹理,高层特征识别更复杂的形状和对象。当原有骨干网络特征提取效果不满足要求时,通常选择对其替换或改进,如文献[8]利用 MobileNetV3 代替原有主干网络,减少网络参数的同时保证了模型的检测精度。文献[9]通过将 SKNet 融入 YOLOv5 主干网络,提高对模糊目标的特征提取能力。

1.4 注意力机制

注意力机制(attention mechanism)是一种在深度学习中广泛应用的技术,目的在于帮助模型聚焦于输入数据中的重要部分,提高其处理效率和表现。在深度学习系统中,注意力机制允许模型根据任务需要动态选择输入的某些部分进行重点处理,而不是均匀地对待所有输入。文献[10-12]利用坐标注意力机制提高小目标在检测任务中的权重,实现了对苹果花朵、小麦和细胞这类小目标的准确检测。

1.5 多尺度特征融合

特征融合是目标检测任务中的一个重要难题。在主干网络中,由于不同检测对象的大小不同,在持续的采样过程中网络提取到不同尺度的特征。直接使用这些特征进行预测会出现一些问题,例如下采样的倍数越大,小目标的信息容易丢失,并且单一尺度的特征利用也不利于网络的学习。

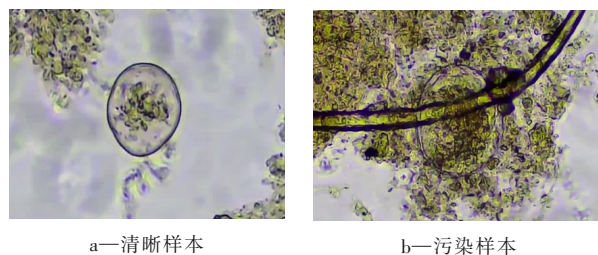
为解决多尺度检测问题,文献[13]提出特征图金字塔网络(feature pyramid networks, FPN),将特征提取过程中自上而下的高层语义信息与低层语义信息结合,增强了网络特征提取能力。然而 FPN 结构只增强了语义信息,忽略了定位信息的传递。因此,

文献[14]设计了路径聚合网络(path aggregation network, PAN),自下而上地连接低层不同尺度特征图对 FPN 结构进行补充,实现语义信息与定位信息的结合,这也是目前最常用的特征融合金字塔结构。

2 模型改进与优化

2.1 全局上下文网络模块

活性污泥中的微生物在显微镜下一般具有规则且光滑的边缘,这是辨别其种类的重要判据。在真实的活性污泥样本中,部分微生物与杂质混合重叠,导致光无法正常透射过微生物内部,使原本具有一定透明度的微生物内部变得杂乱,边缘信息遭到干扰。对图 2 中 2 幅图片进行比较,污染样本的目标微生物有效信息被杂质分隔、遮挡,从而分散在有限的视野范围中,导致常用的卷积(如 3×3 卷积)感受野有限,无法捕捉到有效特征。



a—清晰样本

b—污染样本

图 2 清晰样本与污染样本对比

Figure 2 Comparison of clean and contaminated samples

受全局上下文网络(global context network, GCNet)^[15]的启发,在主干网络中引入 C3GC 模块解决上述问题。通过对特征通道进行调整,使用非本地网络(non-local networks)建立全局的特征感知,并使用压缩激励的方法平衡权重,突出有效特征。

GCNet结合了非本地网络(non-local networks)的全局信息聚合能力和压缩激励网络(squeeze-and-excitation networks, SENet)的计算效率。非本地网络通过捕捉图像内各个位置之间的长距离依赖关系,实现了对整体图像特征的全局感知,从而突破了传统卷积操作局部感受野的限制。与此同时,SENet通过动态地重新加权通道特征,增强了模型对有用信息的捕捉能力和对不相关信息的抑制能力。

GCNet实现了群居信息的有效整合,融合模型原有的C3模块形成了改进的C3GC模块。这一结构不仅继承了Non-Local Networks的能力,通过非局部操作建立起图像各个部分之间的长距离关系,实现全局范围内的特征关注;还融合了SENet的通道注意力机制,优化了特征通道间的权重分配,从而更加高效地聚焦于图像中的关键信息。这种综合方法使得模型能够在不显著增加计算负担的情况下,有效捕获并利用全局上下文信息,极大地提升了目标检测任务中的性能表现。Non-Local block输出可以表示为式(1):

$$Z_i = x_i + W_z \sum_{j=1}^{N_p} \frac{f(x_i, x_j)}{C(x)} (W_v \bullet x_j) \quad (1)$$

式中: i 为查询位置的索引; j 为枚举所有可能的位置; $f(x_i, x_j)$ 表示位置 i 和 j 之间的关系,并且具有归一化因子 $C(x)$ 。 W_z 和 W_v 表示线性变换矩阵(如 1×1 卷积)。为了简化,将 $f(x_i, x_j)/C(x)$ 表示为位置 i 和 j 之间的归一化成对关系。

但是Non-local结构对每个像素点均计算其注意力分布,造成了巨大的计算资源浪费。因此GCNet利用其优势并对其结构进行简化,简化后可表示为式(2):

$$Z_i = x_i + W_z \sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{\sum_{m=1}^{N_p} e^{W_k x_m}} (W_v \bullet x_j) \quad (2)$$

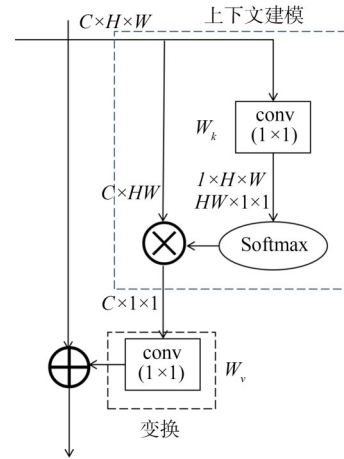
简化后的Non-Local block结构如图3所示。

整个GCNet block可以简化表示为式(3):

$$Z_i = F(x_i, \delta(\sum_{j=1}^{N_p} a_j x_j)) \quad (3)$$

式中: $\sum_{j=1}^{N_p} a_j x_j$ 表示权重模块; F 表示聚合全局特征到每个位置的特征上。

GCNet吸取了SENet结构的优点,优化了模型的计算量,在维持模型运算速度的同时提高模型的精



注: $C \times H \times W$ 为图片输入; C 为通道数; H 为高度; W 为宽度; HW 为先后取 W 和 H 方向数据,下同。

图3 简化Non-local模块

Figure 3 The simplified Non-local module

准度,SE模块的架构如图4所示。最终,全局上下文(GC)块的详细架构如图5所示,公式见式(4):

$$Z_i = x_i + W_{v2} \text{RELU}(\text{LN}(W_{v1} \sum_{i=1}^{N_p} \frac{e^{W_k x_j}}{\sum_{m=1}^{N_p} e^{W_k x_m}} x_j)) \quad (4)$$

式中: $\sum_{i=1}^{N_p} \frac{e^{W_k x_j}}{\sum_{m=1}^{N_p} e^{W_k x_m}} x_j$ 为网络模块的权值,

$W_{v2} \text{RELU}(\text{LN}(W_{v1} \sum_{i=1}^{N_p} \frac{e^{W_k x_j}}{\sum_{m=1}^{N_p} e^{W_k x_m}} x_j))$ 为通道转换模块;其

中,RELU为非线性激活函数,LN为层标准化。在原来简化的Non-local模块的变换部分,融合了SE模块中Bottleneck结构,并使用层标准化运算解决优化问题,而上下文建模部分保留了简化的Non-local模块的结构,既能得到Non-local适应特征之间长距离依赖的性能,又能像SE模块一般减少计算量,解决提取特征多样性的丢失问题,提高了检测的准确率。

2.2 坐标注意力机制

活性污泥中不同种类的微生物体积不同,并且同一种类的微生物也会具有一定的体积差异。模型在学习过程中注意力往往会集中在中大型目标,而忽略小目标的检测。其原因是小目标在图片中所占像素低,在频繁的下采样过程中会出现特征丢失,无法捕捉到有效特征信息。这对于检测某些外形较小的目标微生物时是非常不利的,所以更希望模型能够在大、中、小3种目标之间的权重达到平衡。

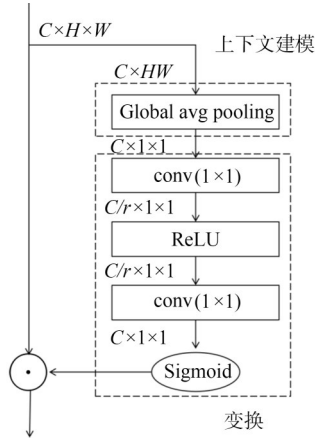


图4 SE模块

Figure 4 SE module

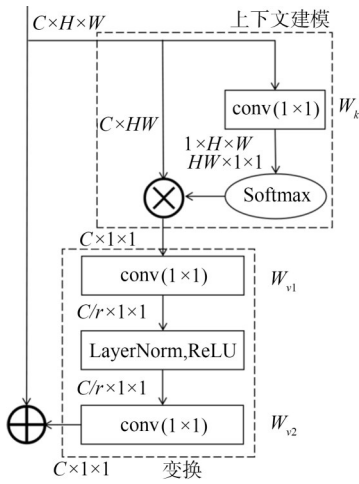


图5 GCNet模块

Figure 5 GCNet module

2021年Hou等^[16]首次提出了坐标注意力机制(coordinate attention, CA),这是针对通道注意力机制提出的一种新的注意力机制,结构如图6所示。

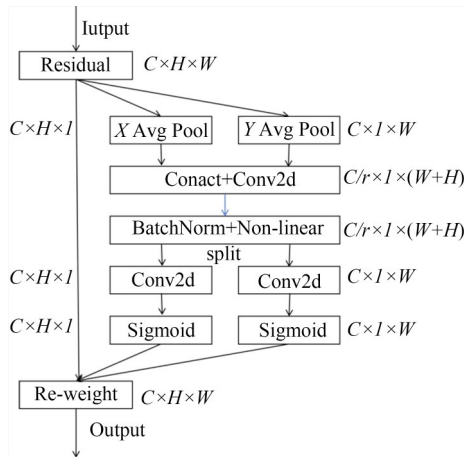


图6 坐标注意力机制结构

Figure 6 Structure diagram of the coordinate attention mechanism

由图6可知:坐标注意力机制通过坐标信息嵌入和坐标生成2个步骤,对通道关系与远程依赖进行编码。坐标注意力模块可以看作计算单元,其目的是增强移动网络学习特征的表达能。它可以取任意中间张量 $X = [x_1, x_2, \dots, x_c] \in R^{C \times W \times H}$ 作为输入并输出一个有增强表示能力的同样尺寸的输出 $Y = [y_1, y_2, \dots, y_c]$ 。其中, C 为通道数; H 和 W 分别为输入图片的高和宽。

对于坐标信息嵌入来说,为了促进注意力模块可以获取精准位置信息的空间长距离依赖关系,CA模块将全局池化分为一对一维特征编码操作。对于输入的特征图 X ,使用大小为 $(H, 1)$ 或 $(1, W)$ 的池化核,分别沿横坐标和纵坐标对每个通道进行编码,也就是高度为 h 的第 c 个通道与宽度为 w 的第 c 个通道的输出,输出公式如式(5)、(6)所示:

$$Z_c^h(h) = \frac{1}{w} \sum_{0 \leq i \leq w} X_c(h, i) \quad (5)$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} X_c(j, w) \quad (6)$$

上述2种变换分别沿着2个方向进行特征聚合,产生一对方向感知映射 Z^h 和 Z^w 。这2种变换允许注意力块沿着一个空间方向捕获远程依赖关系,并沿着另一个空间方向保留精确的位置信息,这有助于网络更准确地定位感兴趣的对象。

针对已生成的2项特征聚合映射,首先将两者连接,随后发送到1个共享的 1×1 卷积进行变换 F_1 ,如式(7)所示。

$$f = \delta(F_1([Z^h, Z^w])) \quad (7)$$

式中: δ 表示非线性激活函数。然后,沿空间维度将 f 切分成2个单独的张量,再利用2个 1×1 卷积 F_h 和 F_w 将特征图 f^h 和 f^w 变转化为与输入 X 具有相同通道数的张量,其结果如式(8)、(9)所示。

$$g^h = \sigma(F_h(f^h)) \quad (8)$$

$$g^w = \sigma(F_w(f^w)) \quad (9)$$

最后将 g^h 和 g^w 分别展开作为注意力权重,CA模块的最终输出可用式(10)表示。

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (10)$$

2.3 融合GLSA的特征金字塔结构

在目标检测任务中,融合不同尺度的特征是提高模型分割性能的一个重要手段。对于活性污泥微生物而言,同一份样本中所包含的微生物个体差异往往非常大,如图7所示。其中图7a为钟虫个体,图7b为楯纤虫个体。

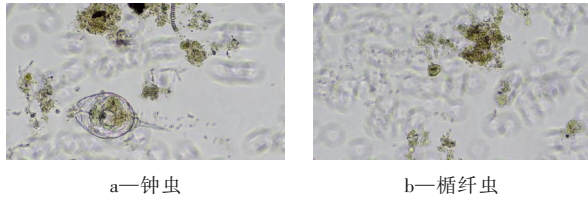


图7 不同种类微生物个体差异

Figure 7 Individual differences among different microbial species

一般来说,随着网络层的深入,下采样倍数越高(即特征图的缩放倍数越高),感受野越大,更容易捕捉大目标的整体特征信息。相应地,浅层网络的感

受野较小,只能捕捉到局部信息和小目标信息。因此,为更好地利用不同尺度的特征信息,保留了YOLOv5原始的FPN和PAN结合的特征金字塔结构,并在颈部网络中融合全局到空间聚合模块^[17](global-to-local spatial aggregation, GLSA)同步提取局部空间细节信息和全局空间语义信息,减少了高层特征中的错误信息,加强特征融合能力。

全局到空间聚合模块由两个独立的局部关注单元与全局关注单元组成,其结构如图8所示。双流结构设计有效地保留了本地和非本地建模功能,并且分离通道的方式很好地平衡了精度和计算资源。

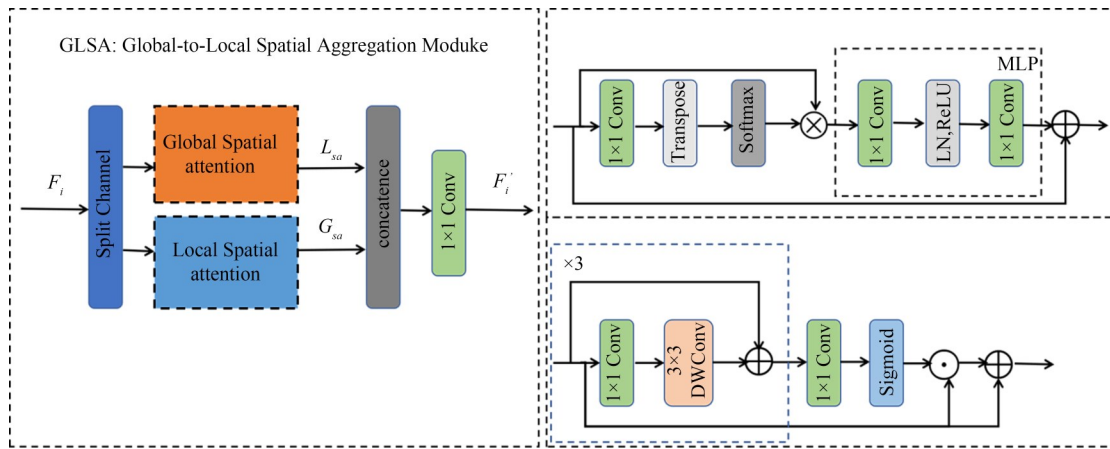


图8 GLSA结构

Figure 8 GLSA structure diagram

由图8可知:具有64通道的特征映射 $\{F_i | i \in (2, 3, 4)\}$ 平均分成2个映射 $F_i^1, F_i^2 [i \in (2, 3, 4)]$,分别输入全局空间关注模块(GSA)与局部控件关注模块(LSA)。这2个注意力模块的输出最后被1个 1×1 的卷积层连接起来,此过程如式(11)、(12)所示。

$$F_i^1, F_i^2 = \text{Split}(F_i) \quad (11)$$

$$F_i' = C_{1 \times 1}(\text{Concat}(G_{sa}(F_i^1), L_{sa}(F_i^2))) \quad (12)$$

式中: G_{sa} 为全局空间注意; L_{sa} 为局部空间注意; F_i 为输出特征。GSA模块强调空间中每个像素的长程关系,通常作为局部空间注意力的补充。将全局空间注意力图 G_{sa} 和 F_i^1 作为输入,计算如式(13)、(14)所示。

$$\text{AttG}(F_i^1) = \text{Soft max}(\text{Transpose}(C_{1 \times 1}(F_i^1))) \quad (13)$$

$$G_{sa}(F_i^1) = \text{MLP}(\text{AttG}(F_i^1) \otimes F_i^1) + F_i^1 \quad (14)$$

式中: $\text{AttG}(\cdot)$ 为全局注意运算; $C_{1 \times 1}$ 为 1×1 卷积; $\otimes$ 为矩阵乘法。 $\text{MLP}(\cdot)$ 由2个全连接层和1个ReLU非线性和规范化层组成。 MLP 的第一层将其输入转换为一个扩展比为2的高维空间,第2层将其恢复为与输入相同的维数。

LSA模块模块在给定的特征图的空间维度上有效地提取感兴趣区域的局部特征,主要针对小目标检测。将局部注意力图 L_{sa} 和 F_i^2 作为输入,计算如式(15)、(16)所示。

$$\text{AttL}(F_i^2) = \sigma(C_{1 \times 1}(F_c(F_i^2)) + F_i^2) \quad (15)$$

$$L_{sa} = \text{AttL}(F_i^2) \odot F_i^2 + F_i^2 \quad (16)$$

式中: $F_c(\cdot)$ 表示级联三个 1×1 卷积层和 3×3 深度卷积层; $\text{AttL}(\cdot)$ 为局部注意运算; $\sigma(\cdot)$ 为Sigmoid函数,是逐点乘法。该结构的优势在于可以用较少的参数有效地聚合局部空间信息,改进后的网络结构图如图9所示。

2.4 K-Means++聚类算法

在目标检测任务中,模型不仅需要学习对象的类别,还需要学习对象的位置和大小。每张图片中的对象具有不同的尺度和纵横比,很难学习到不同形状的对象,也导致模型在定位方面表现不足。为了解决这一问题,早期研究者们人为设计了9个不同大小和宽高比的先验框。该方法降低了问题的难度

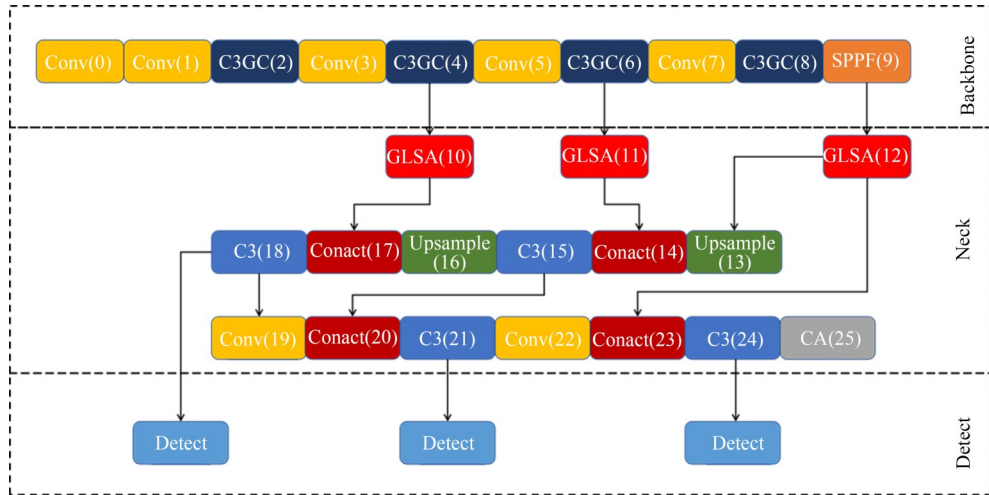


图9 改进模型结构

Figure 9 The improved model structure

和模型学习的难度,被广泛应用于各种目标检测模型。然而,人工设计的高度主观性使得先验框不能很好地拟合数据集。当先验框的尺寸与目标尺寸差异过大时,则会影响模型的检测效果。为此,本文通过分析微生物数据集中所有种类的形状和特征,使用 K -Means++^[18] 算法对训练集进行聚类,得到最贴合数据集 9 个先验框。 K -Means++ 算法的核心是先从数据集随机选取 1 个样本点作为聚类中心,然后计算每个样本与当前已有聚类中心之间的最短距离,然后计算每个样本点被选为下一个聚类中心的概率,最后选择最大概率值(或者概率分布)所对应的样本点作为下一个簇中心;重复上述步骤,直到选出 K 个聚类中心,并自动生成更适合此数据集的先验框,从而降低算法的原始成本,提高准确率与召回率。与传统的 K -Means^[19] 算法相比, K -Means++ 算法优化了簇中心的选择策略。 K -Means 算法聚类首先要完成 K 个聚类中心的初始化,这也使得它的收敛情况严重依赖于簇中心的初始化情况,当在初始化过程中将 K 个或大多数簇中心都初始化到了同一个簇中, K -Means 算法很大程度上不会收敛到全局最小值。

原始的 YOLOv5 模型的先验框为 [10, 13, 16, 30, 33, 23]、[30, 61, 62, 45, 59, 119] 和 [116, 90, 156, 198, 373, 326], 是基于 COCO 数据集生成的 3 种锚框,分别用于检测小、中、大 3 种目标,但并不适用于所使用的数据集。所以通过 K -Means++ 聚类算法重新生成 3 组锚框,分别是 [13, 23, 31, 57, 45, 192], [56, 97, 76, 139, 103, 87], 和 [128, 254, 141, 147, 322, 417]。

3 实验结果与分析

3.1 实验环境

数据采集设备:本项目所使用的数据采集设备为明美 ML31 电动显微镜,搭载 2000 万像素、最大分辨率可达 5480*3648 的 DVP 相机,有效增益 1-16X,带有稳定 LED 光源。载物台采用电动 XYZ 轴平台改造,采用步进电机驱动,最小步长 0.1 μm ,绝对定位精度 $\pm 0.4 \mu\text{m}$ 。设备启动时步进电机以预设步长驱动载物台,使得相机对样本区域进行遍历拍照获得数据。

模型训练平台:本项目使用的模型训练平台采用 Ubuntu 操作系统,处理器为 12th Gen Intel® Core™ i9-12900K 2.6 GHz,搭载 4 张 NVIDIA RTX 3090 GPU,模型训练采用 Pytorch 框架。

3.2 数据增强

本文所用原始数据集 2000 张,采集自实际污水处理厂,其中包含钟虫、表壳虫、水熊虫、轮虫、楯纤虫、钩虫和吸管虫 7 种为主要的活性污泥指示微生物。在模型训练环节中,为防止模型过早进入过拟合,提高模型的精度与鲁棒性,一般需要采取数据增强手段丰富数据集。因此,为了模拟微生物不同方向上的运动姿态,本文使用旋转处理扩充数据集,以 60° 为步长将图片自 -180° 至 180° 旋转。另外,本文还使用随机高斯噪声添加和亮度调整的方式进一步扩充,并将扩充后的数据集以 8:1:1 的比例划分为训练集、验证集和测试集。在模型的输入端,采用 YOLOv5 自带的 Mosaic 数据增强方式,将 4 张图片随机缩放拼接,放置指定尺寸的大图的左上、右上、左

下、右下位置,并根据每张图片的尺寸变换方式,将映射关系对应到图片标签上。训练时设置训练轮数 epochs=400,批处理大小 batch-size=20,输入图像大小限定 640*640。

3.3 评估参数

本实验主要的评估参数为召回率(Recall)、F1得分(F1-Score)和 IOU 阈值为 0.5 时的平均精度(mAP@.5),其计算分别如式(17)~(20)所示。

$$P = TP / (TP + FP) \quad (17)$$

$$R = TP / (TP + FN) \quad (18)$$

$$F1 = 2TP / (2TP + FP + FN) \quad (19)$$

$$mAP@.5 = \frac{\sum_{i=1}^C AP_i}{C}, \quad IOU = 0.5 \quad (20)$$

式中:TP为真正例;FP为假正例;FN为假反例;C为总类别数目;AP为 Precision-Recall 曲线下面积;mAP为每个类别 AP 的均值;AP_i为第 i 类的 AP 值。

3.4 消融实验

为了分析本文提出的改进方法对 YOLOv5 算法性能的影响,设计了 4 组消融实验对不同的改进部分进行结果分析,实验结果如表 1 所示。在本项目中,更加关注模型的召回率,即所有实际为正类的样本中,有多少被模型正确预测。具体来说,召回率越高,模型能够找出样本中所有微生物的能力越强,漏检率越低,这对于活性污泥微生物检测非常重要。mAP@.5 则体现了模型在数据集下的整体精准度表现,比单一的精准度更加全面。

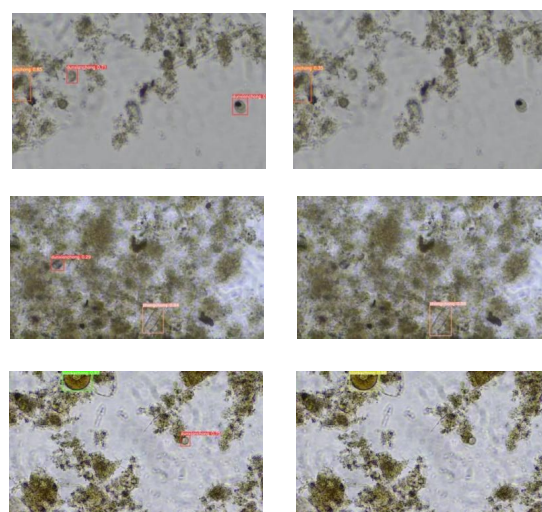
表 1 消融实验实验数据对比

Table 1 Comparison of experimental data from ablation experiments

改进方案	K-Means++	C3GC	CA	GLSA	R	mAP@.5	F1
YOLOv5s	×	×	×	×	0.956	0.986	0.960
改进 1	√	×	×	×	0.967	0.988	0.967
改进 2	√	√	×	×	0.968	0.987	0.972
改进 3	√	√	√	×	0.964	0.991	0.970
改进 4	√	√	√	√	0.974	0.992	0.974

在消融实验过程中,采用 K-Means++ 聚类算法确定锚框相比原始 YOLOv5s 模型召回率提高 1.1 百分点,mAP@.5 提高 0.2 百分点,F1 分数提高 0.7 百分点。在改进 1 基础上应用 C3GC 模块相比 YOLOv5s 召回率提高 1.2 百分点,mAP@.5 提高 0.1 百分点,F1 分数提高 1.2 百分点。在改进 2 的基础上添加坐标

注意力机制相比 YOLOv5s 召回率提高 0.8 百分点,mAP@.5 提高 0.5 百分点,F1 分数提高 1.0 百分点。最后,在改进 3 的基础上颈部网络融合 GLSA 模块相比 YOLOv5s 召回率提高 1.8 百分点,mAP@.5 提高 0.6 百分点,F1 分数提高 0.7 百分点。在添加坐标注意力机制时,模型更加关注精准率出现了召回率的小幅下滑。为验证模型改进的有效性,选取 3 张相同的图片交由原始模型与改进模型进行识别处理。从对比结果看出,改进模型相对于原始模型更能找出样本中的所有待检测目标,说明其具备更高的回归率,大大降低了漏检率,这与模型改进的预期相符。改进后模型的检测效果如图 10 所示。



a—改进模型识别结果 b—YOLOv5s 识别结果

图 10 2 种模型实验结果对比

Figure 10 Comparison of the experimental results of the two models

3.5 对比实验

为了验证本文改进算法的性能,将本文改进算法与目标检测模型 YOLOv5x、YOLOv3-tiny、YOLOv5-transformer 和 CBAM-YOLOv5s 进行对比实验,采用召回率、平均准确率、模型参数总量和模型体积,对比结果如表 2 所示。可知:改进的 YOLOv5 算法召回率达到了 0.974,mAP@.5 达到了 0.992,相比于体量最大的 YOLOv5x 模型,用大约 1/7 的参数总量,回归率提高了 0.3 百分点,mAP@.5 提高了 0.2 百分点。与 YOLOv3-tiny 相比,回归率提高了 6.9 百分点,mAP@.5 提高了 4.4 百分点。与 CBAM-YOLOv5s 相比,回归率提高了 2.6 百分点,mAP@.5 提高了 1.1 百分点。与 YOLOv5-transformer 相比,回归率提高了 1.3 百分点,mAP@.5 提高了 0.5 百分点。

表 2 不同模型实验数据对比

Table 2 Comparison of experimental data of different models

模型	R	mAP@.5	参数	Model_Size
YOLOv5x	0.971	0.990	86749405	173.1 MB
YOLO3-tiny	0.905	0.948	8674496	123.5 MB
CBAM-YOLOv5s	0.948	0.971	7428063	14.8 MB
YOLOv5-transformer	0.961	0.987	7235645	14.4 MB
本研究	0.974	0.992	11700730	23.5 MB

4 结 论

本文针对通用深度学习目标检测算法难以满足活性污泥微生物检测进行高精度检测的问题,提出了一种改进 YOLOv5 的活性污泥微生物检测方法。该方法基于 YOLOv5 网络架构,使用 K-Means++ 算法生成最佳先验框,主干网络使用 C3GC 模块加强特征提取,颈部网络中融合 GLSA 模块与坐标注意力机制加强特征信息融合,提高对小目标的检测精度。实验表明,本文提出的检测方法能够准确地识别活性污泥中的微生物, mAP@.5 达 0.992,召回率达 0.974,实现了高精准度、低漏检率的检测。

参考文献

- [1] 中国环境健康安全网. 广州市 2022 年排放源统计公报 [EB/OL]. ehs.cn, 2023-11-16. <https://www.ehs.cn/article-22489-1.html>
- [2] 原培胜, 张洪荣, 陈现明, 等. 两种改良型活性污泥法在城市污水处理中的比较[J]. 环境工程, 2006, 24(3): 34-35.
- [3] 周仓, 闫攀, 袁帅, 等. 一种用于活性污泥微生物的智能镜检装置及镜检方法: CN201910251070.7[P]. 2024-08-14.
- [4] 吴帅, 徐勇, 赵东宁. 基于深度卷积网络的目标检测综述[J]. 模式识别与人工智能, 2018, 31(4): 335-346.
- [5] Girshick R. Fast R-CNN [J]. Computer Science, 2015. DOI: 10.1109/ICCV.2015.169.
- [6] HE K, GKIOXARI G, PIOTR D, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 202, 42(2): 386-397.
- [7] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-

- GONZÁLEZ J A. A comprehensive review of yolo architectures in computer vision: From YOLOv1 to YOLOv8 and Yolo-Nas [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [8] GUO G, ZHANG Z. Road damage detection algorithm for improved YOLOv5 [J]. Scientific reports, 2022, 12(1): 15523.
- [9] 赵梦, 于红, 李海清, 等. 融合 SKNet 与 YOLOv5 深度学习的养殖鱼群检测[J]. 大连海洋大学学报, 2022, 37(2): 312-319.
- ZHAO M, YU H, LI H Q, et al. Detection of fish stocks by fused with SKNet and YOLOv deep learning [J]. Journal of Dalian Fisheries University, 2022, 37(2): 312-319.
- [10] 杨其晟, 李文宽, 杨晓峰, 等. 改进 YOLOv5 的苹果花生长状态检测方法[J]. 计算机工程与应用, 2022, 58(4): 237-246.
- YANG Q S, LI W K, YANG X F, et al. Improved YOLOv5 method for detecting growth status of apple flowers [J]. Computer Engineering and Applications, 2022, 58(4): 237-246.
- [11] 宋怀波, 王云飞, 段援朝, 等. 基于 YOLOv5-MDC 的重度粘连小麦籽粒检测方法[J]. 农业机械学报, 2022(4): 245-253.
- SONG H B, WANG Y F, DUAN Y C, et al. Detection method of severe adhesive wheat grain based on YOLOv5-MDC Model [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 245-253.
- [12] 王静, 孙紫雲, 郭苹, 等. 改进 YOLOv5 的白细胞检测算法[J]. 计算机工程与应用, 2022, 58(4): 134-142.
- WANG J, SUN Z Y, GUO P, et al. Improved leukocyte detection algorithm of YOLOv5 [J]. Computer Engineering and Applications, 2022, 58(4): 134-142.
- [13] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [J]. IEEE Computer Society, 2017. DOI: 10.1109/CVPR.2017.106.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [J]. IEEE, 2018. DOI: 10.1109/CVPR.2018.00913.
- [15] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE, 2020. DOI: 10.1109/ICCVW.2019.00246.
- [16] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [EB/OL]. arXiv, 2021. DOI: 10.48550/arXiv.2103.02907.
- [17] TANG F, XU Z, HUANG Q, et al. DuAT: Dual-aggregation transformer network for medical image segmentation [C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2023: 343-356.
- [18] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding [C]// Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans, SIAM, 2007: 1027-1035.
- [19] IKOTUN A M, EZUGWU A E, ABUALIGAH L, et al. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data [J]. Information Sciences, 2023, 622: 178-210.