

doi:10.6041/j.issn.1000-1298.2023.10.040

基于计算机视觉的鱼类低氧胁迫行为检测与跟踪算法

李道亮^{1,2} 姜国旗^{1,2} 杨建安^{1,2} 白羽^{1,2} 谢琰^{1,2} 王承国³

(1. 中国农业大学国家数字渔业创新中心, 北京 100083; 2. 中国农业大学信息与电气工程学院, 北京 100083;
3. 中国农业大学烟台研究院, 烟台 264670)

摘要: 为了准确检测、跟踪加州鲈鱼因水中溶解氧含量低产生的胁迫行为, 本文构建了一种改进的 YOLO v5 与 DeepSORT 组合网络算法。在算法方面提出 2 个改进方案: 在原 YOLO v5 的 Backbone 和 Neck 中分别加入 2 个基于移位窗口的自注意力 Swin Transformer 模块, 提升了网络对目标特征信息的提取能力, 以此提升原模型的检测效果; 采用 Warmup 和 Cosine Annealing 结合的学习率策略, 使多目标跟踪算法 DeepSORT 前期收敛速度更快、更稳定。实验结果表明, 在目标检测方面, 相对于原 YOLO v5, 改进的 YOLO v5 的 mAP@0.5、mAP@0.5:0.95 和召回率分别提升 1.9、1.3、0.8 个百分点, 在不完全遮挡情况下, 改进的算法表现出更好的检测效果。在目标跟踪方面, DeepSORT 算法的 MOTA、MOTP 和 IDF1 分别提升 4.0、0.7、10.7 个百分点, 并且加州鲈鱼在遮挡前后的 ID 切换频率得到明显抑制。改进的 YOLO v5 与 DeepSORT 跟踪算法更适合于检测、跟踪加州鲈鱼的低氧胁迫行为, 能够为加州鲈鱼的养殖提供技术支持。

关键词: 加州鲈鱼; 胁迫行为; 目标检测; 多目标跟踪

中图分类号: TP391 文献标识码: A 文章编号: 1000-1298(2023)10-0399-08 OSID: 

Detection and Tracking Algorithm of Fish Hypoxia Stress Behavior Based on Computer Vision

LI Daoliang^{1,2} JIANG Guoqi^{1,2} YANG Jian'an^{1,2} BAI Yu^{1,2} XIE Yan^{1,2} WANG Chengguo³

(1. National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China
2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
3. Yantai Institute, China Agricultural University, Yantai 264670, China)

Abstract: In order to accurately detect and track the stress behavior of micropterus salmoides due to low dissolved oxygen content in water, an improved YOLO v5 and DeepSORT combined network algorithm was constructed. In terms of algorithm, two improvement schemes were proposed: two self-attention Swin Transformer modules based on shifted windows were added to the Backbone and Neck of the original YOLO v5, which improved the network's ability to extract target feature information, thereby improving the detection effect of the original model; the learning rate strategy combined with Warmup and Cosine Annealing made the convergence speed of the multi-target tracking algorithm DeepSORT faster and more stable in the early stage. The experimental results showed that in terms of target detection, compared with the original YOLO v5, the mAP@0.5, mAP@0.5:0.95 and recall rate of the improved YOLO v5 were increased by 1.9, 1.3 and 0.8 percentage points, respectively. In the case of incomplete occlusion, the improved algorithm could show better detection results. In terms of target tracking, the MOTA, MOTP, and IDF1 of the DeepSORT algorithm were increased by 4.0, 0.7 and 10.7 percentage points respectively, and the ID switching frequency of micropterus salmoides before and after occlusion was significantly suppressed. The improved YOLO v5 and DeepSORT tracking algorithms were more suitable for detecting and tracking the hypoxic stress behavior of micropterus salmoides, and can provide technical support for the breeding of micropterus salmoides.

Key words: micropterus salmoides; stress behavior; target detection; multi-target tracking

0 引言

在水产养殖中,鱼类福利对于鱼的生长有着至关重要的影响^[1-3],其中溶解氧是影响鱼类福利水平的重要指标之一。在工厂化水产养殖中,溶解氧含量与鱼类养殖密度密切相关,一旦水池长时间处于低氧状态下,鱼将会大面积的出现生长放缓、生殖能力下降、免疫功能低下和发病率、死亡率升高等现象^[4-5]。传统水产养殖方式主要依赖养殖人员的经验和肉眼观察来判断鱼是否受低氧环境胁迫,但其实时性差、可靠性低、准确性低。为了工厂化水产养殖能够尽可能减少人工干预、增加鱼的福利以取得最大化的收益,有必要提出一种对鱼的低氧胁迫行为可实时、可靠、准确判断的方法。

计算机视觉技术作为一种非入侵式且稳定性较好的方法,逐渐成为鱼类物种行为检测的首选^[6]。计算机视觉常见的任务包括图像分类、目标检测^[7-8]、目标跟踪^[9-10],在水产养殖行业已经得到了较广泛的应用。针对浑浊的水下环境,HUSSAM等^[11]将彩色多尺度 Retinex 颜色增强技术与 YOLO 算法结合以实现较高的鱼类检测精度。DU 等^[12]提出了一种 ResNet50-LSTM 算法来识别鱼育种过程中的基本行为,该算法在光照强度低、养殖密度高、养殖环境复杂的养殖车间中检测效果显著,有很好的鲁棒性和有效性。CAI 等^[13]将 YOLO v3 和 MobileNet 进行结合,优化了 MobileNet 骨干网络中特征图的确定策略,相对于原模型其检测精度提升 3.06 个百分点,且推理速度也有明显提升。HU 等^[14]提出了一种改进的 YOLO v4 算法用于检测残余饲料颗粒,其对密集连接和去冗余操作进行特征映射改进,AP50 增长 27.21 个百分点,计算量减少 30%,但其更适用于极小目标的检测。

随着深度学习的迅速发展和目标检测算法精度不断提升,基于目标检测的多目标跟踪算法已成为主流,根据目标跟踪类型分为单目标跟踪和多目标跟踪。WANG 等^[15]提出了一种检测、跟踪鱼类异常行为的 YOLO v5 和 SiamRPN++ 组合网络,其主要对检测头、上采样和下采样进行改进,检测精度和速度效果较好,但 SiamRPN++ 采用迭代的方式实现多目标跟踪,原理上还是单目标追踪,当迭代次数过多时,会消耗大量的 GPU 资源。SI 等^[16]采用 YOLO v5 和多目标跟踪技术 DeepSORT 开发了一套准确可靠的多斑马鱼幼虫跟踪系统,其检测精度和跟踪精度的鲁棒性相对较好,但对于检测器和跟踪器还有很大的提升空间。

因此,本文构建一种改进的 YOLO v5 和

DeepSORT 组合网络算法用于检测和跟踪鱼类低氧胁迫行为,为鱼类高效养殖提供技术支持,其主要包含检测器 YOLO v5 和跟踪器 DeepSORT 两部分,在检测器原基础模型上增加基于移位窗口的自注意力机制 Swin Transformer 用于提升 YOLO v5 中 Backbone 和 Neck 对特征信息的提取能力,进而提高 YOLO v5 对鱼类低氧胁迫行为检测精度,在跟踪器方面采用 Warmup 和 Cosine Annealing 结合的策略对学习率进行调整以加速模型收敛,以提升跟踪模型整体性能。

1 数据采集与数据处理

1.1 数据采集

鱼类低氧胁迫行为识别所用到的实验数据取自于重庆农业科学院鱼菜共生 AI 工厂,选取对象为 50 条亚成年加州鲈鱼,采用海康威视变焦摄像头(4.0×10⁶像素)以速度 25 f/s 拍摄视频数据片段共 90 个,拍摄时间为 2022 年 6—8 月,期间拍摄时间段为 09:00—11:30、13:30—17:30。整个视频数据采集的过程中,除实验池中溶解氧含量会出现大幅度变化外,其余水质参数指标均控制在亚成年加州鲈鱼正常生活范围内,实验场地(1.5 m×2 m)如图 1 所示。



图 1 实验场地
Fig. 1 Experimental site

1.2 数据处理

从拍摄的视频中抽取图像序列并通过 Labelimg 工具进行标注、Python 图像截取,其中 1 439 幅图像用于检测器,9 184 幅图像用于跟踪器。对于检测器图像,为了防止过拟合、避免样本不均衡,并为了提升模型的鲁棒性及泛化能力,分别对其中的 200、600、170 幅图像添加高斯噪声、椒盐噪声及亮度调节,并对数据集进行随机划分,训练集与验证集图像数量比为 8:2。对于跟踪器图像,将截取图像按照 Market1501 数据集格式进行分配,训练集与验证集比为 9:1。

2 目标检测和目标跟踪算法原理

采用检测和目标跟踪算法 YOLO v5 与

DeepSORT。YOLO v5 作为检测器对视频帧中的目标进行检测,检测结果作为跟踪器 DeepSORT 的输入,更高的检测精度能提升 DeepSORT 的跟踪性能。DeepSORT 提取检测器检测的目标特征,计算前后两帧目标的匹配程度,为每个所追踪到的目标分配对应 ID。

2.1 目标检测算法

目前,主流的目标检测算法主要分为 One - Stage 和 Two - Stage。One - Stage 直接通过主干网络给出类别和位置信息,其算法包括 YOLO^[17-18]系列、SSD^[19]和 EfficientDet^[20]。Two - Stage 先由算法生成一系列样本的候选框,再通过卷积神经网络提取相关特征进行目标的分类和定位,其代表算法为 R - CNN^[21]系列。通常情况下,前者相对后者检测速度较快,在某些情况下检测精度略高。

2.1.1 Swin Transformer 注意力机制

Swin Transformer^[22]作为一种通用的计算机视觉主干网络,主要解决 Transformer 从 NLP 领域调整到 CV 领域所带来的实体尺寸、分辨率等差异。其作为一种分层级的 Transformer,通过移位窗口将自

注意的计算限制在非重叠的局部窗口上,并允许跨窗口连接。

Swin Transformer 整体架构如图 2a 所示,它首先是通过面片分割模块将输入的 RGB 图像分割成不重叠的面片,每个面片都被看作一个“token”,其特征被设置为原始像素 RGB 的连接。本文使用的是 4 × 4 的面片尺寸,其特征维数为 4 × 4 × 3 = 48,并将该原始值特征上应用线性嵌入层,将其投影到任意维度 C。Swin Transformer 共包含 4 个阶段,每个阶段都包含 2 部分,分别是 Patch Merging (第 1 阶段为 Linear Embedding) 和连续的 Swin Transformer Block,目的是用来计算相关性和采样,其第 1 阶段不会改变特征图尺寸,其余 3 个阶段在改变特征图的同时改变了通道数。Swin Transformer Block 是在标准的 Transformer 基础上,替换多头注意力计算部分 MSA (Multi-head self - attention) 为基于移位窗口的模块 W - MSA (Windows multi-head self - attention) 和 SW - MSA (Shifted windows multi-head self - attention),其余结构不变,如图 2b 所示。

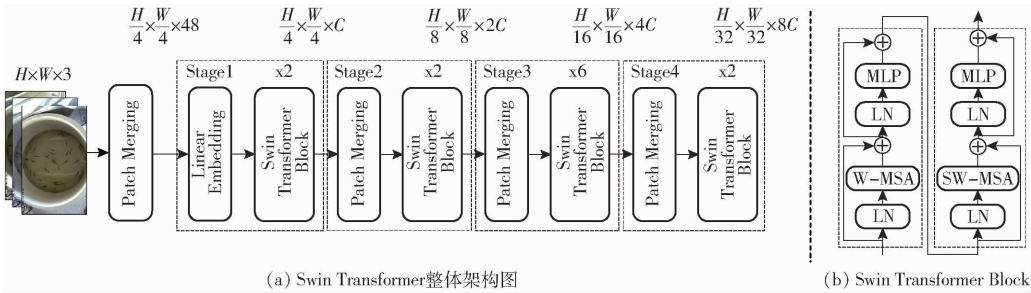


图 2 Swin Transformer 注意力机制

Fig. 2 Swin Transformer attention mechanism

2.1.2 YOLO v5 算法优化

本文构建了一种新的目标检测算法 ST - YOLO v5,网络结构如图 3 所示,其主要由 4 部分组成,分别为输入端、Backbone、Neck 和 Prediction。输入端主要对输入数据进行 Mosaic 数据增强、自适应锚框计算以及自适应图像缩放。Backbone 主要由 Focus、C3 及 SPP 等模块组成,主要目的是目标特征提取。Neck 采用 FPN + PAN 结构,其中 FPN 和 PAN 的走向相反,前者从上往下传递强语义特征,后者从下往上传递强定位信息,让模型的参数在不同的层次进行聚合。Prediction 的主体部分由 3 个 Detect 组成,其利用基于网格的 anchor 在不同尺度的特征图上进行目标检测,进行最终预测的输出。

在 ST - YOLO v5 中,首先将原 YOLO v5 的 C3 模块与 Swin Transformer 模块结合成 C3STR 模块,并分别对 Backbone 中最后 2 个 C3 模块、Neck 中前 2 个 C3 模块进行 C3STR 模块的改进替换。经过实

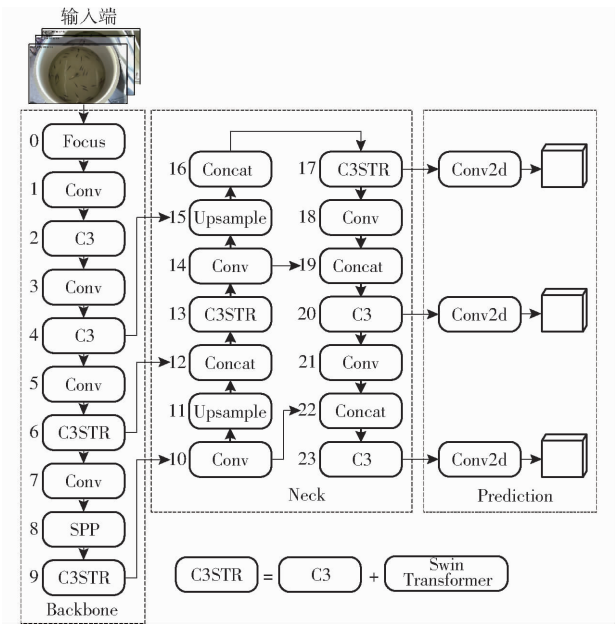


图 3 ST - YOLO v5 网络结构图

Fig. 3 ST - YOLO v5 network structure

验对比发现,ST-YOLO v5 的检测效果明显优于原 YOLO v5。因此,ST-YOLO v5 更适合作为检测器。

2.2 目标跟踪算法

SORT 算法是 DeepSORT 的前身^[23-24],其跟踪过程是首先通过检测器(如 Faster R-CNN、YOLO 等)把每一帧的物体检测出来,随后通过卡尔曼滤波算法^[25-26]和匈牙利算法^[27]分别进行帧预测、匹配。SORT 检测速度较快,但是当物体发生遮挡时会出现频繁 ID 切换,进而导致跟踪效果不佳。因此,DeepSORT 在 SORT 模型的基础上引入级联匹配(Matching cascade)和新轨迹确认(Confirmed),可以有效地改善遮挡情况下目标跟踪的效果。

2.2.1 学习率策略

在训练过程初期因距离目标太远,一般采用较大的学习率,但较大的学习率容易导致不稳定性。DeepSORT 的学习率策略为每过 20 轮迭代,学习率变为原学习率的 1/10。而这种学习率策略会使学习率出现阶梯现象,导致模型不能快速、稳定收敛。为了解决这种慢且不稳定的收敛情况,本文采用 Warmup 策略^[28]对前 5 个 epoch 进行预热学习率调整。该策略的核心在于逐步预热,它将学习率从一个小值逐渐增加到一个大值,可有效地解决模型前期收敛慢、不稳定问题。

通常在模型训练过程中,随着模型训练的进行,会逐渐降低学习率,这样才有助于找到最优解。因此,当采用 Warmup 策略逐步预热学习率到所设预值时,则采用 Cosine Annealing 策略^[29-30]对学习率进行调整。在整个训练过程中学习率的变化曲线如图 4 所示。

由图 4 可知,采用 Warmup 和 Cosine Annealing 结合的学习率调整策略能够更好地减少学习率阶梯现象。通过大量测验来验证这种学习率结合调整策略的可行性,验证结果如图 5 所示。从收敛角度看,相对于原学习率调整策略, Warmup 和 Cosine Annealing 结合的学习率调整策略能够加快跟踪模型收敛速度,减少了模型前期在收敛过程中阶梯现

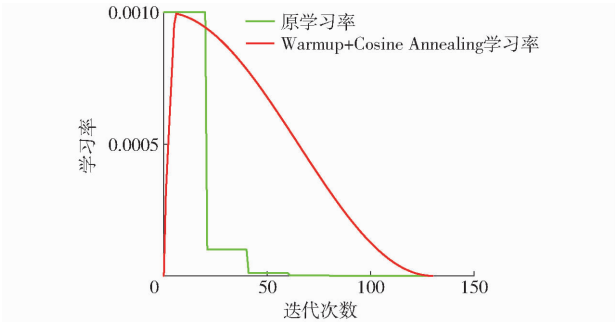


图 4 学习率变化曲线

Fig.4 Learning rate change curves

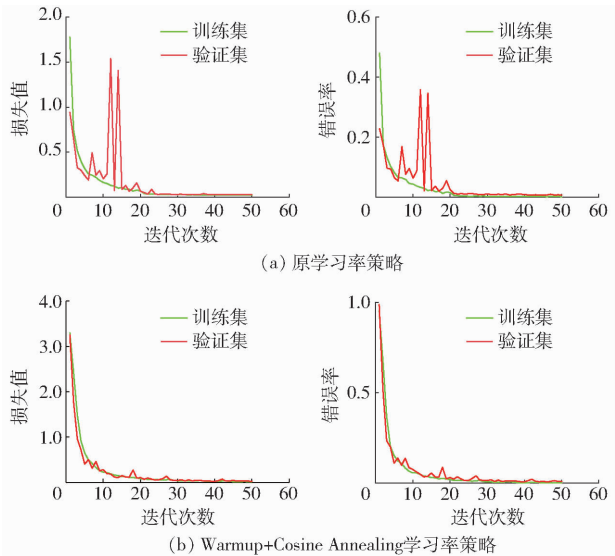


图 5 不同学习率策略下损失值和错误率的变化曲线

Fig.5 Change curves of loss value and error rate at different learning rates strategy

象。因此,在跟踪模型训练过程中,采用 Warmup 和 Cosine Annealing 结合的学习率策略比原有的学习率策略更适用于 DeepSORT 模型训练。

2.2.2 DeepSORT 算法

DeepSORT 核心思想是采用传统的单假设追踪方法,使用递归的卡尔曼滤波和逐帧的数据关联。DeepSORT 首先采用卡尔曼滤波预测轨迹,随后利用匈牙利算法将检测框和预测框进行级联匹配和 IOU 匹配,最后进行卡尔曼滤波更新,算法流程如图 6 所示。

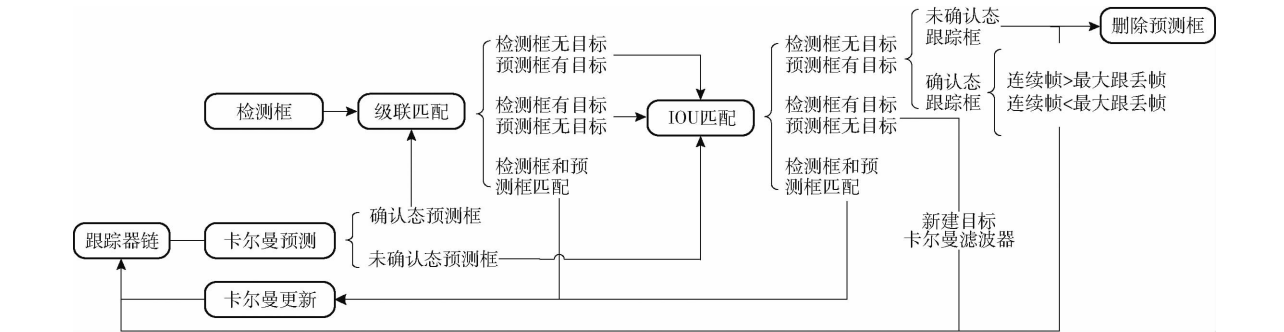


图 6 DeepSORT 跟踪流程

Fig.6 DeepSORT tracking process

在轨迹状态预测方面,相对于 SORT 算法中的 7 个维度信息,DeepSORT 则使用 $(u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$ 8 个维度信息对运动状态进行描述,分别为检测框的中心坐标 (u, v) ,长宽比 γ ,高 h ,以及这 4 个参数在图像坐标中各自的速度。并且采用带有恒定运动速度和线性观测模型的标准卡尔曼滤波器进行轨迹的预测,其观测值为 (u, v, γ, h) 。在卡尔曼滤波器预测期间,使用计数器来对关联帧进行计数,如果在未超过预定值前轨迹关联则计数器重置 0,否则删除该轨迹。

在匹配问题上,DeepSORT 采用匈牙利算法来解决检测框和预测框之间关联性的问题。其中对于运动和外观信息的整合采用马氏距离来描述,其计算式为

$$d^{(1)}(i, j) = (\mathbf{d}_j - \mathbf{y}_i)^T \mathbf{S}_i^{-1} (\mathbf{d}_j - \mathbf{y}_i) \tag{1}$$

式中 $\mathbf{d}_j$ ——第 j 个缺氧鲈鱼检测框的状态
 $\mathbf{y}_i$ ——第 i 个目标跟踪器对缺氧鲈鱼经过卡尔曼滤波后预测框的状态
 $\mathbf{S}_i$ ——检测框和预测框之间的协方差矩阵

当处在缺氧状态下,加州鲈鱼由于自身产生的应激反应会产生激烈游动,并且在密集型养殖场景中,较长时间遮挡现象普遍存在。这些现象会使卡尔曼预测的位置不准确,导致马氏距离的度量不准确,进而导致 ID 切换现象频繁发生。对于这种不确定性的现象,DeepSORT 引入余弦距离作为辅助,对于每个检测框 $\mathbf{d}_j$ 求一个外观描述符 $\mathbf{r}_i$,其中 $\|\mathbf{r}_i\| = 1$ 。此外,为每个跟踪器构建一个存储序列用来存放最近关联成功的 100 帧的 R_i ,其中 $R_i = \{\mathbf{r}_k^{(i)}\}_{k=1}^{L_k}$ 。第 i 个跟踪器的 R_i 与当前第 j 个检测结果之间的最小余弦距离 $d^{(2)}(i, j)$ 的计算式为

$$d^{(2)}(i, j) = \min \{1 - \mathbf{r}_j^T \mathbf{r}_k^{(i)} \mid \mathbf{r}_k^{(i)} \in R_i\} \tag{2}$$

DeepSORT 分别从运动物体可能位置的信息和外观信息两个方面考虑分配问题。为了构建关联问题,DeepSORT 通过加权和将两个指标结合,其计算式为

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j) \tag{3}$$

3 实验结果与分析

3.1 检测器和跟踪器评估指标

3.1.1 目标检测器评估指标

对 ST-YOLO v5 性能评价主要指标是精确率、召回率、平均精度、帧率以及权重。在指标计算方面,在 ST-YOLO v5 中采用混淆矩阵,真阳性(True positive, TP)、真阴性(True negative, TN)、假阳性(False positive, FP)、假阴性(False negative, FN)。

3.1.2 目标跟踪器评估指标

在多目标跟踪评估方面,采用 TrackEval 工具对 DeepSORT 进行整体性能评估,主要评估指标分别为跟踪准确度(MOTA)、跟踪精度(MOTP)、ID 切换数(IDSW)和综合指标(IDF1)。其中 MOTA 是跟踪准确度,MOTA 是对漏报数(FN)、误报数(FP)、ID 切换数(IDSW)的一个整体性评估,是衡量多目标跟踪模型 DeepSORT 在检测和跟踪方面一个非常直观的性能参数。MOTA 的计算式为

$$MOTA = 1 - \frac{\sum (FN + FP + IDSW)}{N} \tag{4}$$

式中 N ——真实轨迹总数量

MOTP 为跟踪精度,主要考虑的是检测框和目标真实边框的重叠情况,其计算式为

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \tag{5}$$

式中 $d_{t,i}$ ——当前帧位置匹配误差
 c_t ——当前帧匹配成功个数

IDF1 为正确识别的检测数与平均真实数和计算检测数之比,主要评估 DeepSORT 算法的稳定性,值越大表示越稳定,其计算式为

$$IDF1 = \frac{IDTP}{IDTP + 0.5IDFP + 0.5IDFN} \tag{6}$$

式中 $IDTP$ ——正确识别的目标数量
 $IDFP$ ——错误识别为目标数量
 $IDFN$ ——未能识别为目标数量

3.2 目标检测器选型

针对于不同场景和目标物,One-Stage 和 Two-Stage 的算法检测效果不同。为了对目标检测模型进行准确选型,分别对 One-Stage 中的 SSD、YOLO v4、YOLO v5、YOLO v6、YOLO v7、EfficientDet 算法和 Two-Stage 中的 Faster R-CNN 进行对比实验,其实验结果如表 1 所示。表中 mAP@0.5 表示 IOU 阈值为 5 时识别对象的平均精度。

表 1 目标检测算法对比实验结果

Tab.1 Comparative experimental results of target detection algorithms

类型	模型	mAP@ 0.5 / %	模型占用 内存量/MB	帧率/ (f·s ⁻¹)
One-Stage	SSD	75.4	90.61	74.3
	YOLO v4	95.5	244.4	22.4
	YOLO v5	95.1	13.7	47.6
	YOLO v6	94.0	38.8	68.1
	YOLO v7	93.1	142.3	45.8
	EfficientDet	61.3	15.1	14.6
Two-Stage	Faster R-CNN	94.8	315.0	10.4

由表 1 可知,Two-Stage 中最具有代表性的算法 Faster R-CNN 在帧率和权重方面均不如 YOLO 系列,其实时性较差,且不适合在移动端部署。在 One-Stage 中 SSD 和 EfficientDet 的 mAP@0.5 远低于 YOLO 系列。对于 YOLO 系列,YOLO v4 的 mAP@0.5 略高于 YOLO v5,但 YOLO v5 的帧率远高于 YOLO v4,且 YOLO v5 权重远远小于 YOLO v4 的权重。YOLO v6 的帧率高于 YOLO v5,但 mAP@0.5 和权重均不如 YOLO v5。YOLO v7 的 mAP@0.5、权重、帧率均不如 YOLO v5。鉴于本研究对于检测精度和推理速度要求较高,因此采用 YOLO v5 作为基础目标检测模型。

YOLO v5 提供 4 种规模的版本,分别为 YOLO v5s、YOLO v5m、YOLO v5l、YOLO v5x。为了更加准确对目标检测模型版本进行选型,对以上 4 种规模的版本进行了实验对比,对比结果如表 2 所示。表中 mAP@0.5:0.95 表示不同 IOU 阈值(从 0.5 到 0.95,步长 0.05)上的平均精度均值。

表 2 YOLO v5 不同规模版本对比实验结果
Tab.2 Comparative experimental results of YOLO v5 versions of different scales %

模型	mAP@0.5	mAP@0.5:0.95	精确率	召回率
YOLO v5s	95.1	54.2	95.2	93.1
YOLO v5m	94.6	52.9	96.4	90.7
YOLO v5l	94.2	53.4	94.6	92.0
YOLO v5x	93.6	52.2	94.4	91.7

由表 2 可知,YOLO v5s 的 mAP、召回率均高于其余 3 种,且 YOLO v5s 结构相对简单,实时检测性能好。本文将采用 YOLO v5s 作为基础检测器。

3.3 不同注意力机制下检测器性能评估

在深度学习,注意力机制已经成为一个重要的概念^[31],其本质是通过网络自主学习生成一组权重系数,并以动态加权的方式来着重强调所关注的区域,同时忽略不相关信息。在目标检测领域,常见的注意力机制为 SE、ECA、CA、CBAM 等,它们与 Swin Transformer 都能提升目标检测模型对特征信息提取的能力,影响目标检测模型的精度。为了探究 YOLO v5 模型在不同注意力机制下提取加州鲈鱼低氧胁迫行为信息的能力,进行了一系列的对比实验,主要参考指标有 mAP、精确率、召回率,其结果如表 3 所示。当把 SE 和 ECA 添加到 YOLO v5 模型中,其模型在 mAP@0.5、mAP@0.5:0.95 和精确率方面都有所降低。对于注意力机制 CBAM,添加后其各项指标均不如原 YOLO v5 模型。注意力机制 CA 在 mAP@0.5、mAP@0.5:0.95 方面相较于原 YOLO v5 模型有一定的提高,但从整体指标对比

来看提升效果远不如本文所提出的 ST-YOLO v5 模型。ST-YOLO v5 在 mAP@0.5、mAP@0.5:0.95 和召回率方面分别提升 1.9、1.3、0.8 个百分点。因此,本文所提出的 ST-YOLO v5 更适合于检测加州鲈鱼的低氧胁迫行为。

表 3 不同注意力机制对比实验结果
Tab.3 Comparative experimental results of different attention mechanisms %

模型	mAP@0.5	mAP@0.5:0.95	精确率	召回率
YOLO v5	95.1	54.2	95.2	93.1
YOLO v5 + SE	94.8	53.8	94.6	93.6
YOLO v5 + ECA	94.9	54.1	94.8	93.5
YOLO v5 + CBAM	94.9	53.7	93.6	92.5
YOLO v5 + CA	96.3	55.3	94.0	92.3
ST-YOLO v5	97.0	55.5	93.6	93.9

为了进一步验证 ST-YOLO v5 模型检测加州鲈鱼低氧胁迫行为的有效性,对包含加州鲈鱼低氧胁迫行为的随机视频帧图像进行实例测验,视频帧图像包含遮挡和非遮挡两种情况,其实例测验结果如图 7 所示。由图 7 可知,在相同实验环境下,以上 6 种模型均能够识别到加州鲈鱼低氧胁迫行为。但综合视频帧图像的遮挡和非遮挡两种情况来看,对于真实的加州鲈鱼低氧胁迫行为,ST-YOLO v5 模型所提供的置信度相对更高,表明基于移位窗口的自注意力机制 Swin Transformer 模块能够最大程度地提升目标检测模型 YOLO v5 对加州鲈鱼低氧胁迫行为信息的提取能力,进而提升模型对目标行为的检测效果。

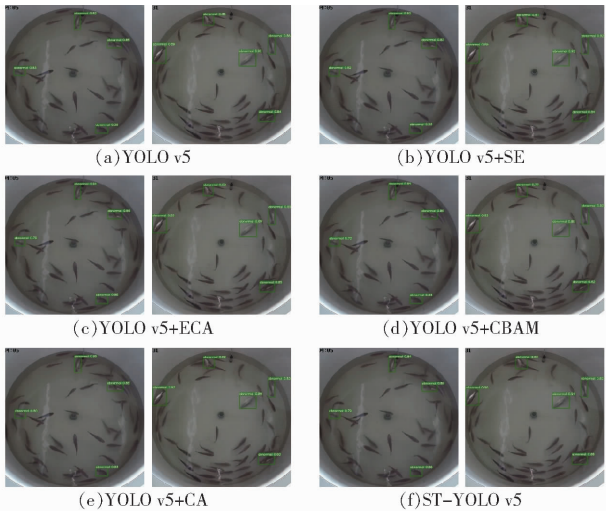


图 7 不同注意力机制下 YOLO 实例检测图像
Fig.7 YOLO instance detection images under different attention mechanisms

3.4 跟踪模型性能评估

为了验证本文所提出的关于加州鲈鱼缺氧行为

跟踪模型的有效性,在所采集的数据集中随机挑选一段视频进行跟踪测验,其测验结果如图 8 所示。根据优化前后 DeepSORT 的实例跟踪结果,在遮挡和非遮挡情况下,优化后的跟踪模型并没有出现漏检、ID 切换、重检的情况,表明本文所优化的跟踪模型相较于原跟踪模型性能优异。

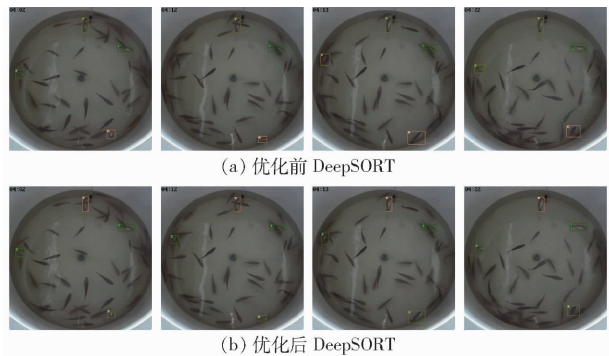


图 8 实例跟踪结果

Fig. 8 Instance tracking results

为了对优化后多目标跟踪模型做一个整体性能评估,采用跟踪性能评估工具 TrackEval,并在采集的加州鲈鱼低氧胁迫行为视频数据集中随机选取一段视频用于跟踪模型整体性能评估,选取的视频数据片段采用 DarkLabel 标注工具进行目标真实路径标注,标注结果和跟踪模型输出结果作为评估工具的输入,优化前多目标跟踪模型与优化后多目标跟踪模型评估结果如表 4 所示。

表 4 跟踪模型评估结果

Tab. 4 Tracking model evaluation results

模型	MOTA/%	MOTP/%	IDF1/%	IDSW
优化前 DeepSORT	83.3	70.9	82.8	2
优化后 DeepSORT	87.3	71.6	93.5	0

由表 4 可知,所提出的改进 YOLO v5 与 DeepSORT 模型相对于优化前跟踪模型在 MOTA、MOTP 和 IDF1 方面分别提升 4.0、0.7、10.7 个百分点。在 IDSW 方面,改进后的算法模型在跟踪缺氧的加州鲈鱼时,并没有因遮挡出现 ID 切换的情况。

综上所述,改进后的算法模型在遮挡和非遮挡情况下性能表现优异,更适用于加州鲈鱼低氧胁迫行为的跟踪。

4 结论

(1)针对工厂化水产养殖中传统水产养殖方式对鱼类低氧胁迫行为判断存在准确率低、可靠性差、实时性差等问题,构建了一种基于 ST-YOLO v5 的多目标跟踪模型 DeepSORT,该模型在鱼类低氧胁迫行为检测、跟踪方面性能表现优异,为鱼类养殖人员提供了一定的技术支持。

(2)对不同类型、不同规模版本的目标检测算法进行对比实验,最终选择 YOLO v5s 作为基础目标检测算法,并构建目标检测算法 ST-YOLO v5。ST-YOLO v5 相对于基础目标检测算法在 mAP@0.5、mAP@0.5:0.95 和召回率方面分别提升 1.9、1.3、0.8 个百分点,模型的整体性能更优异。

(3)在 ST-YOLO v5 的基础上引入多目标跟踪模型 DeepSORT,并采用 Warmup 和 Cosine Annealing 结合的学习率调整策略加速跟踪模型收敛。实验结果表明,基于 ST-YOLO v5 的跟踪模型 DeepSORT 在 MOTA、MOTP 和 IDF1 方面分别提升 4.0、0.7、10.7 个百分点,并且 ID 的切换频率得到抑制。

参 考 文 献

[1] WANG G, AKHTER M, LIU C, et al. Automatic recognition of fish behavior with a fusion of RGB and optical flow data based on deep learning[J]. Multidisciplinary Digital Publishing Institute, 2021,11:2774.

[2] LI X, HAO Y, ZHANG P, et al. A novel automatic detection method for abnormal behavior of single fish using image fusion[J]. Computers and Electronics in Agriculture,2022,203:107435.

[3] 何佳,黄志涛,宋协法,等. 基于计算机视觉技术的水产养殖中鱼类行为识别与量化研究进展[J]. 渔业现代化,2019,46(3):7-11.

HE Jia, HUANG Zhitao, SONG Xiefa, et al. Research progress of fish behavior recognition and quantification in aquaculture based on computer vision technology[J]. Fishery Modernization,2019,46(3):7-11. (in Chinese)

[4] PAOLO D, JOHN F S, STEFANO M. The effect of hypoxia on fish schooling[J]. The Royal Society,2017,327:20160236.

[5] NIU B, LI G, PENG F, et al. Survey of fish behavior analysis by computer vision[J]. Journal of Aquaculture,2018,9:5.

[6] 张璐,黄琳,李备备,等. 基于多尺度融合与无锚点 YOLO v3 的鱼群计数方法[J]. 农业机械学报,2021,52(增刊):237-244.

ZHANG Lu, HUANG Lin, LI Beibei, et al. Fish school counting method based on multi-scale fusion and no anchor YOLO v3[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(Supp.):237-244. (in Chinese)

[7] XU A, ZHANG M, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection[J]. Neurocomputing,2023,527:204-232.

[8] KAUR R, SINGH S. A comprehensive review of object detection with deep learning[J]. Digital Signal Processing,2022,132:103812.

[9] MEI Y, SUN B, LI D, et al. Recent advances of target tracking applications in aquaculture with emphasis on fish[J]. Computers and Electronics in Agriculture,2022,201:107335.

[10] DAI Y, HU Z, ZHANG S, et al. A survey of detection-based video multi-object tracking[J]. Displays,2022,75:102317.

[11] HUSSAM E M, ALI F, OMAR A, et al. MSR – YOLO: method to enhance fish detection and tracking in fish farms[J]. Procedia Computer Science,2020,170:539 – 546.

[12] DU L, LU Z, LI D. Broodstock breeding behaviour recognition based on ResNet50 – LSTM with CBAM attention mechanism [J]. Computers and Electronics in Agriculture,2022,202:107404.

[13] CAI K, MIAO X, WANG W, et al. A modified YOLO v3 model for fish detection based on MobileNetv1 as backbone[J]. Aquacultural Engineering,2020,91:102117.

[14] HU X, LIU Y, ZHAO Z, et al. Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO v4 network[J]. Computers and Electronics in Agriculture,2021,185:106135.

[15] WANG H, ZHANG S, ZHAO S, et al. Real-time detection and tracking of fish abnormal behavior based on improved YOLOV5 and SiamRPN + + [J]. Computers and Electronics in Agriculture,2022,192:106512.

[16] SI G, ZHOU F, ZHANG Z, et al. Tracking multiple zebrafish larvae using YOLOv5 and DeepSORT [C] // 2022 8th International Conference on Automation, Robotics and Applications (ICARA). IEEE, 2022: 228 – 232.

[17] JIANG P, DAJI E, LIU F, et al. A review of YOLO algorithm developments[J]. Procedia Computer Science,2022,199: 1066 – 1073.

[18] 邵延华,张铎,楚红宇,等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报,2022,40(10):3698 – 3705.
SHAO Yanhua, ZHANG Duo, CHU Hongyu, et al. Survey of YOLO target detection based on deep learning[J]. Journal of Electronics & Information Technology,2022,40(10):3698 – 3705. (in Chinese)

[19] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21 – 37.

[20] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10781 – 10790.

[21] LIU B, ZHAO W, SUN Q. Study of object detection based on Faster R – CNN [C] // 2017 Chinese Automation Congress (CAC). IEEE, 2017: 6233 – 6236.

[22] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 – 10022.

[23] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking [C] // 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 3464 – 3468.

[24] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C] // 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3645 – 3649.

[25] LING J. Target tracking using Kalman filter based algorithms [C] // Journal of Physics: Conference Series. IOP Publishing, 2021, 2078(1): 012020.

[26] CHEN H, FENG X, HUANG Z, et al. Target tracking based on Kalman filtering techniques [C] // 2022 International Symposium on Control Engineering and Robotics (ISCER). IEEE, 2022: 237 – 245.

[27] HAMUDA E, GINLEY B M, GLAVIN M, et al. Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm [J]. Computers and Electronics in Agriculture,2018,148:37 – 44.

[28] GOYAL P, DOLLÁR P, GIRSHICK R, et al. Accurate, large minibatch sgd: training imagenet in 1 hour [J]. arXiv preprint arXiv:1706.02677, 2017.

[29] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 558 – 567.

[30] LOSHCHILOV I, HUTTER F. Sgdr: stochastic gradient descent with warm restarts [J]. arXiv preprint arXiv:1608.03983, 2016.

[31] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用,2021,57(20):64 – 70.
ZHANG Chenjia, ZHU Lei, YU Lu. A review of attention mechanism in convolution neural networks [J]. Computer Engineering and Application,2021,57(20):64 – 70. (in Chinese)