

引用格式:

袁红春, 陈香枝. 基于多层记忆增强和残差时空变换器的鱼类异常运动行为检测[J]. 湖南农业大学学报(自然科学版), 2024, 50(5): 119–124.

YUAN H C, CHEN X Z. Abnormal fish movement behavior detection based on multilayer memory enhancement and residual spatio-temporal transformer[J]. Journal of Hunan Agricultural University(Natural Sciences), 2024, 50(5): 119–124.

投稿网址: <http://xb.hunau.edu.cn>



基于多层记忆增强和残差时空变换器的 鱼类异常运动行为检测

袁红春, 陈香枝

(上海海洋大学信息学院, 上海 201306)

摘要: 由于人工提取抽象特征方法捕获视频异常存在特征学习不足、特征选择困难和泛化性差的问题, 笔者将计算机视觉技术引入鱼群异常运动行为检测研究中, 采用无监督的学习方式, 提出一种结合多层记忆增强和残差时空变换器的鱼类异常运动行为检测方法。该方法以 U-Net 网络为基础, 利用其编码器和解码器对视频帧编码和解码, 并根据预测帧和真实帧之间的差异实现异常行为检测。为了加强连续视频帧之间的时空信息特征联系, 提出残差时间变换器模块和残差空间变换器模块以提升网络对时间信息和空间信息的建模能力。由于卷积神经网络具有一定的泛化能力, 使用记忆增强模块代替 U-Net 网络中的跳跃连接, 降低编码器对异常帧的表示能力。此外, 采用生成对抗网络(GAN)技术生成更加真实的预测帧, 从而提升网络的检测精度。结果表明: 该方法能有效提取鱼群的运动特性和外观特性, 在自制的两类鱼群数据集上的 AUC(曲线下面积)分别达 0.916 和 0.921, 实现了鱼群异常运动行为检测。

关键词: 鱼类异常行为检测; 计算机视觉; 无监督学习; U-Net; 时空变换器

中图分类号: S951.2; TP391.41

文献标志码: A

文章编号: 1007-1032(2024)05-0119-06

Abnormal fish movement behavior detection based on multilayer memory enhancement and residual spatio-temporal transformer

YUAN Hongchun, CHEN Xiangzhi

(School of Information, Shanghai Ocean University, Shanghai 201306, China)

Abstract: The manual method of extracting abstract features to capture video anomalies was no longer applicable to large-scale aquaculture due to the problems of insufficient feature learning, difficult feature selection and poor generalization. In this study, computer vision technology was introduced into the study of fish movement behavior anomaly detection, and an unsupervised learning approach was used to propose a fish movement behavior anomaly detection method that combined multilayer memory enhancement and residual spatio-temporal transformer to effectively extract the motion correlation and appearance characteristics of fish. Firstly, based on U-Net network, its encoder and decoder were used to implement encoding and decoding of video frames, and behavior anomaly detection was achieved based on the difference between predicted and real frames. In order to strengthen the connection of spatio-temporal information features between consecutive video frames, the residual temporal transformer module and the residual spatial transformer module were proposed to enhance the network's ability to model temporal and spatial information. Since the convolutional neural network had certain generalization ability, the memory enhancement module was used instead of the jump connection in the U-Net network to alleviate the ability of the encoder to represent the anomalous frames. In addition, Generative Adversarial Networks was used to generate more realistic prediction frames, thereby improving the

收稿日期: 2023-03-26

修回日期: 2024-08-23

基金项目: 国家自然科学基金项目(41776142)

作者简介: 袁红春(1971—), 男, 江苏海门人, 博士, 教授, 主要从事人工智能应用研究, hcyuan@shou.edu.cn

detection accuracy of the network. The results indicated that this method could effectively extract the motion and appearance characteristics of fish. On two self-made fish datasets, the area under the curve(AUC) reached 0.916 and 0.921, respectively, achieving fish movement behavior anomaly detection.

Keywords: fish behavior anomaly detection; computer vision; unsupervised learning; U-Net; spatio-temporal transformer

智慧渔业的发展是中国水产养殖技术中不可或缺的一部分,努力实现渔业资源、远洋渔业及渔船监管等信息平台一体化发展,并与现代信息技术相结合,是实现渔业科学化管理和智能化管理^[1-2]的重要手段。其中,对于鱼类异常行为的监管是非常重要的环节。由于外部环境的刺激,鱼群会随之产生一系列的应激反应^[3-4],如回避、无规律游动以及死亡等。实时监控鱼群的行为有助于快速辨别水体的质量,以免造成巨大的经济损失。传统的人工检测方法不仅耗时费力,难以实现持续监控,而且效果不佳,无法满足大规模的工业化水产养殖的需求。为了实现鱼群异常行为检测的智能化和自动化,越来越多的研究^[5-14]将其与计算机视觉技术结合起来,降低投入成本的同时,有效地提高了检测效率。

有学者^[5-6]运用计算机视觉、机器学习等技术监测水质异常、评估鱼群摄食强度以及检测鱼群异常行为。LIU等^[7]提出了一种基于计算机视觉的方法测量大西洋鲑鱼的摄食活动,通过定义重叠系数矫正图像中鱼体重叠引起的计算误差,成功测量了鲑鱼在任意给定时间内的摄食活动。汪靖阳等^[8]提出联合高斯混合模型和卷积神经网络模型对鱼体图像进行语义分割,利用鱼体像素计算上溯轨迹、游泳速度和摆尾幅度,平均像素准确率和平均交并比分别达95.9%和93.3%。孙龙清等^[9]通过鱼类行为实现水质检测,即构建出计算量较少、推理速度快、适合水质监测的Ghost操作,通过将三路卷积归并为一层卷积降低模型的参数数量,准确率提高了1.42%,但降低了图像的处理速度。黄志涛等^[10]采用自适应背景差分及光流法得到鱼体的运动特征,包括运动鱼体的速度、转角等,同时通过灰度共生矩阵提取能量、熵、对比度、相关性和逆差距等5个图像纹理特征值,实现了对鱼类摄食行为的识别和检测,该方法相比于单一特征检测法的准确率提高了6.42%。ZHOU等^[11]先运用深度学习框架将音频信号转换为声学特征,再通过卷积神经网络

对鱼的摄食强度进行分类,其分类的准确率达90%。ANAS等^[12]先采用图像增强技术消除水体环境带来的影响,再采用基于YOLO的目标检测算法获取鱼群的行为轨迹,最后将获取的轨迹应用于朴素贝叶斯算法、 K 近邻算法和随机森林算法检测鱼群异常行为,其检测的准确率分别为90%、89%和86%。李星辉^[13]同样使用鱼群的轨迹判定异常,通过目标检测及数据增强等视觉技术提升模型的性能,利用时间上下文信息优化结果,与基准模型相比,优化模型准确率提高了2.6%。袁红春等^[14]通过灰度共生矩阵得到灰度的空间相关性,描述出斑马鱼的纹理特征,采用光流估计算法Lucas-Kanade计算鱼群的运动矢量,构建轻量化梯度促进模型对鱼群异常行为进行检测,其准确率对比模型的提高了0.5%。

现有的方法对于鱼类行为识别都取得了一定的效果,但在特征提取和泛化能力上仍然存在不足,尤其在捕捉连续视频帧的时空信息时表现不佳。针对这些问题,笔者提出一种无监督的深度学习方法用于鱼类异常行为检测。该方法以U-Net网络为基础,引入残差时间变换器和残差空间变换器的结构,以便学习连续帧之间更加丰富的时空信息;通过同时学习参与注意力机制的信息与未参与注意力机制的信息,既增强前后视频帧的交互性,又在加深网络深度的同时缓解注意力机制的收敛性。为了缓解U-Net中卷积神经网络带来的泛化性,引入记忆增强模块,以扩大异常帧与正常帧之间的差异。最后,运用生成对抗网络(GAN)学习正常模式下的视频帧以生成更加真实的预测帧。

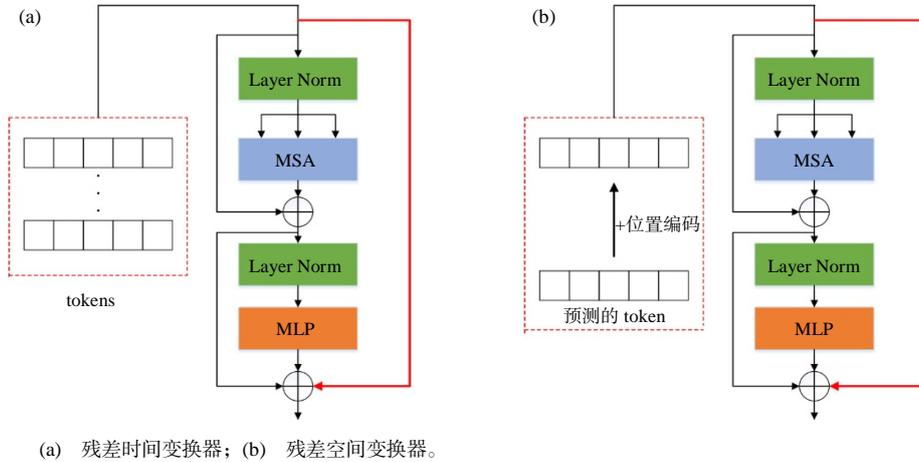
1 异常检测模型

1.1 残差时空变换器

残差时空变换器结构如图1所示。在时间变换器的结构中,将经编码器处理得到的特征图线性投影后输入时序变换器中,按照视频视觉变换器(ViViT)法^[15],将得到的特征图分解成 N 个不重叠

的区块，并经过线性投影转化成多组标记(token)，再对其进行块编码处理；最后，运用变换器模块计算 token 的全局上下文关系。为缓解注意力的收敛性，在多头自注意力层的输入与其输出之间添加残

差连接，这种连接跳过了多层感知器(MLP)，即相对的“非注意力”特征与“注意力”特征一起被集成到一个矩阵中，避免输出的剧烈收敛。



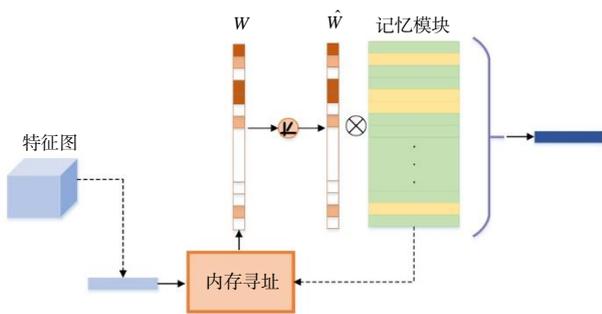
(a) 残差时间变换器；(b) 残差空间变换器。

图 1 残差时空变换器结构

Fig.1 The structure of residual spatio-temporal transformer

1.2 记忆增强模块

为缓解卷积神经网络的过度泛化能力，引入记忆增强模块^[6]记录并存储正常行为模式下的原型特征。在训练过程中，记忆增强模块会对记忆项进行更新和存储；而在测试阶段，记忆项的内容会被固定，如果输入的是异常行为模式下的数据则会得到更大的误差，更容易区分正常模式和异常模式。其主要结构如图 2 所示。



W 为注意力权重； \hat{w} 为经过稀疏处理的注意力权重。

图 2 记忆增强模块结构

Fig.2 The structure of memory enhancement module

1.3 基于生成对抗网络的模型结构

在给定的连续视频帧 I_1, I_2, \dots, I_t 上，使用这些连续的帧预测未来帧，真实的未来帧记作 I_{t+1} ，预测出的未来帧记作 \hat{I}_{t+1} 。图 3 为基于生成对抗网络的异常检测模型的整体框架图，经过预测之后， I_{t+1} 和 \hat{I}_{t+1} 之间的差异被用来计算规则性分数并用于异常检测。主要采用无监督的学习方式，即在训练时全部使用正常的样本进行训练，使得模型更好地学习正常事件；在测试阶段，通过训练好的模型来对测试集中的数据进行测试，实现异常检测。

以 U-Net 网络、残差时空变换器和多层记忆增强模块作为生成网络，如图 4 所示。编码器的输入是分辨率为 256 像素 \times 256 像素、包含 3 通道的图片，输出是分辨率为 32 像素 \times 32 像素、包含 512 通道的特征图。与堆叠连续的帧不同，这些输入帧分开独立进行编码，因而 t 个相同的编码器共享相同的参数，连续的 t 帧被编码成相应的特征图。以这种方

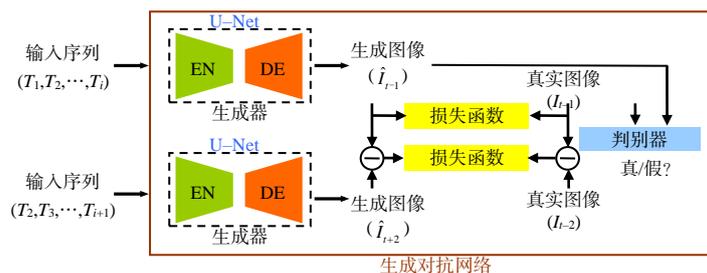


图 3 模型整体框架

Fig.3 The overall framework of the model

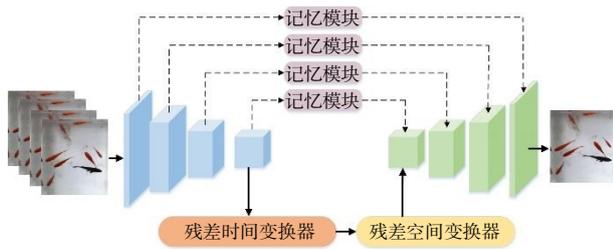


图4 生成器结构

Fig.4 The structure of generator

式, 编码器只专注于提取空间信息。编码器中所有的卷积层使用的激活函数均是 ReLU 函数。

在 U-Net 的结构中, 运用残差时空变换器来提取语义信息, 主要分为时间提取模块和空间提取模块。时间提取模块的变换器接收编码器输出的 t 个特征图, 并输出预测的特征图。随后, 将这些预测的特征图输入空间提取模块变换器中, 以捕获空间上的信息。最后, 将经过空间位置编码得到的预测结果重塑成合适的特征图, 并将其作为变换器模块的最终输出。通过在每层的跳跃连接之间加入记忆增强模块来学习正常样本的各尺度原型特征。在解码的过程中, 解码器通过接受预测特征图 \hat{x}_{t+1} 并输出预测帧 \hat{i}_{t+1} 。解码器的网络结构主要包括使用 ReLU 函数作为激活函数的卷积层和不使用激活函数的反卷积层。本研究中, 利用 patchGAN^[17]来作为判别器。

1.4 损失函数

为了保证生成的预测帧与真实帧相近, 引入外观约束和运动约束^[18]进行优化。在外观约束中, 采用强度损失和梯度损失。由于光流的使用会导致模型极其不稳定^[19], 采用图像差分损失来得到更加稳定的模型。为了使记忆模块中存储的典型特征尽可能地丰富, 引入带有注意力的权重熵损失^[16]来约束模型, 使得记忆增强模块更有效。由于使用了生成对抗网络, 同时采用对抗损失对生成图像和真实图像进行分类。

1.5 异常分数

在测试过程中, 通过生成器生成图像得到预测帧, 真实帧和预测帧之间的异常值可以用来进行异常检测。采用了峰值信噪比(PSNR)来评估异常检测中的图像质量。

2 验证试验及其结果

2.1 试验数据集

鱼群数据集均为上海海洋大学人工智能鱼实验室自制的数据集, 包括红鲫鱼(fish)数据集和斑马鱼(zebrafish)数据集。其中, 红鲫鱼数据集的训练集包含 6 段正常视频图像, 共 2 088 张; 测试集包含 4 段正常视频图像和 2 段异常视频图像, 共 2 410 张。斑马鱼数据集的训练集仅包含正常视频图像, 共 2 000 张; 测试集包含正常行为和异常行为的 2 500 张图像。

2.2 评价指标

采用曲线下面积(AUC)^[20]来评估模型的性能, AUC 的值越大, 模型的性能越好。

2.3 试验环境和参数设置

试验均在以 Pytorch 为框架、NVIDIA GeForce RTX 3090 为 GPU 的环境下完成。为了训练整个网络, 将所有帧的分辨率均调整为 256 像素×256 像素, 并将像素值归一化至[-1, 1]; 同时使用基于 Adam 的随机梯度下降来优化参数。所有的数据集中, 强度损失所占权重、梯度损失所占权重、对抗损失所占权重、记忆交叉熵损失所占权重分别为 1.00、1.00、0.01、0.05, 批尺寸为 4。生成器和判别器的学习率分别设置为 0.000 2 和 0.000 02。

2.4 消融试验

为了验证方法的有效性, 设计 5 组试验(T1、T2、T3、T4、T5)对两类鱼群数据集进行分析, 其中, T1 只使用 U-Net 模块, T2 使用 U-Net 模块和时空变换器模块, T3 使用 U-Net 模块和残差时空变换器模块, T4 使用 U-Net 模块、时空变换器模块和记忆增强模块, T5 使用 U-Net 模块、残差时空变换器模块和记忆增强模块, 模型的 AUC 如表 1 所示。从表 1 可知: 单独采用 U-Net 作为生成器(T1)得到的斑马鱼和红鲫鱼的 AUC 均较低, 引入时空变换器结构之后(T2)AUC 分别提升了 0.015 和 0.012, 说明时空变换器充分学习了时间信息和空间信息, 具有更强的学习和预测能力; 在时空变换器中添加残差结构(T3), 同时学习参与注意力机制信息和未参与注意力机制信息, AUC 又分别提高了 0.009 和 0.011, 缓解了由于网络深度所带来的注意

表 1 两个鱼群数据上的消融试验结果

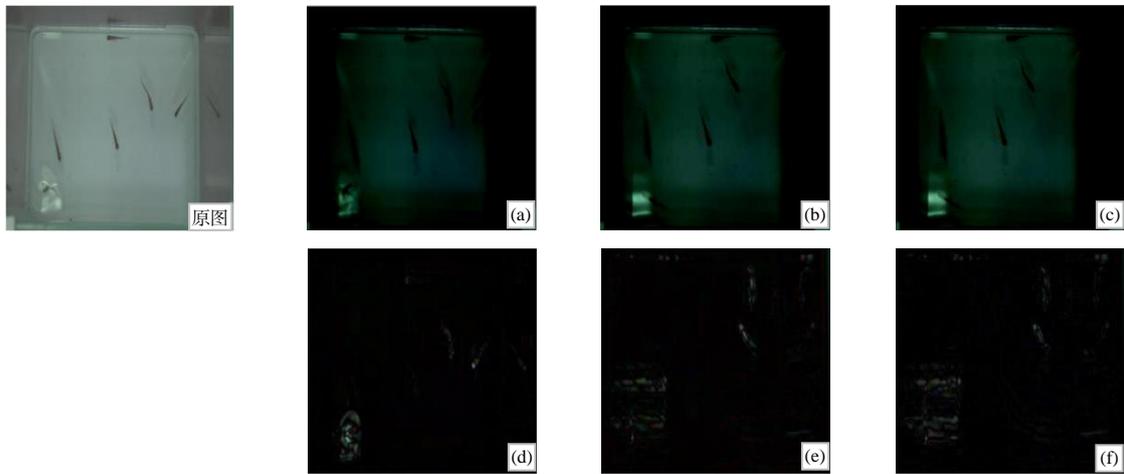
Table 1 Ablation experiments on two fish dataset

组别	使用的模块	AUC	
		斑马鱼	红鲫鱼
T1	U-Net	0.885	0.890
T2	U-Net+时空变换器	0.900	0.902
T3	U-Net+残差时空变换器	0.909	0.913
T4	U-Net+时空变换器+记忆增强模块	0.912	0.919
T5	U-Net+残差时空变换器+记忆增强模块	0.916	0.921

力收敛的情况，提高了模型的检测效果；引入记忆增强模块，T4的AUC较T2的提高了0.012和0.017，T5的AUC较T3的提高了0.007和0.008，对网络性能提升有一定帮助，记忆增强模块能够学习丰富的多尺度原型特征，将其直接输入到解码器后会促

使预测帧更接近真实帧，其与异常帧之间的误差显著增加，从而提高异常检测精度。

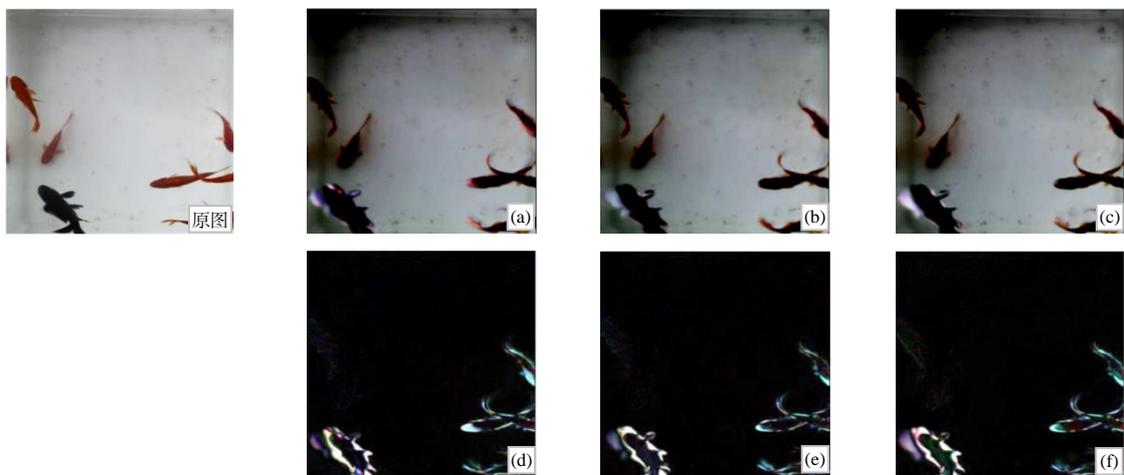
为了更加直观地展示模型的预测效果，在斑马鱼和红鲫鱼数据集上分别展示了试验 T1、T3、T5 的预测图和帧差图。图 5 和图 6 分别展示了斑马鱼第 2 358 帧的检测结果和红鲫鱼第 706 帧的检测结果，异常部分的颜色更加明亮，表示预测出的误差值较大，即说明该帧中的鱼群的行为出现了异常，一定程度上实现了异常定位。可见，该模型在正常数据集上进行训练后，能够在包含异常数据的测试集中表现出良好的异常检测能力及准确的异常定位效果。



(a)、(d) T5 的预测图和帧差图；(b)、(e) T3 的预测图和帧差图；(c)、(f) T1 的预测图和帧差图。

图 5 斑马鱼数据集异常帧的检测结果

Fig.5 Detection results of abnormal frame in zebrafish dataset



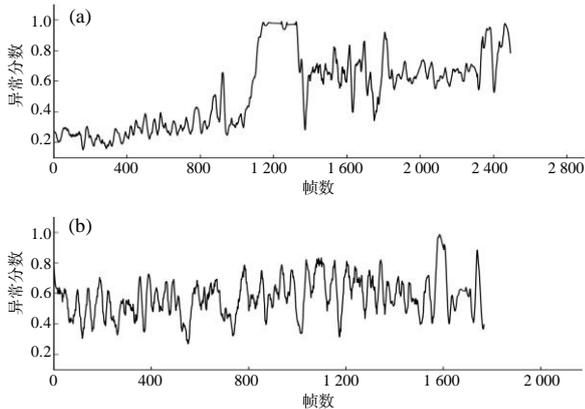
(a)、(d) T5 的预测图和帧差图；(b)、(e) T3 的预测图和帧差图；(c)、(f) T1 的预测图和帧差图。

图 6 红鲫鱼数据集异常帧的检测结果

Fig.6 Detection results of abnormal frame in red crucian carp dataset

2.5 异常检测结果

图7所示为在两类鱼群数据集中的异常分数曲线,其中正常帧的得分较高,异常帧的得分较低。当水体环境发生变化时,鱼群会大量快速朝着没有规律的方向游动,即出现异常行为,异常分数快速发生变化,并迅速降低,说明该方法可以有效地检测鱼类的异常行为。



(a) 斑马鱼数据集; (b) 红鲫鱼数据集。

图7 两类鱼群数据集的异常分数曲线

Fig.7 The anomaly score curves of two fish school datasets

3 结论

针对传统特征提取方式存在特征不充分、泛化性较差的问题,提出了一种基于多层记忆增强和残差时空变换器的鱼类异常行为检测方法,充分提取了连续视频帧的时空特征,多层记忆增强的使用进一步降低了网络对异常帧的表示能力,有效地区分了正常帧和异常帧,提高了网络的鲁棒性。试验结果表明,在斑马鱼和红鲫鱼数据集上取得了较高的AUC(0.916、0.921),但真实的水底环境更加复杂,未来将聚焦于尽可能还原水下鱼群的生存环境,以便将该方法更有效地应用于实际场景。

参考文献:

[1] 于宁,徐涛,王庆龙,等.智慧渔业发展现状与对策研究[J].中国渔业经济,2021,39(1):13-21.
 [2] 郑鹏,王欣宇,田宛鑫.碳中和目标下中国海洋渔业转型成效研究[J].中国渔业经济,2021,39(6):1-7.
 [3] 侯晓娇.基于计算机视觉的鱼群行为检测与结构特征研究[D].太原:太原科技大学,2014.
 [4] 付世建,李佳倩.模拟捕食刺激对不同社会性中华倒刺鲃鱼群行为的影响[J].生态学杂志,2022,41(12):2382-2387.
 [5] 杨东海.基于计算机视觉的青鲈鱼游泳跟踪方法研究[D].上海:上海海洋大学,2019.
 [6] 骆桂兰,陈军,王会聪,等.水产养殖过程中计算机

视觉的应用[J].畜牧与饲料科学,2017,38(12):91-92.
 [7] LIU Z Y, LI X, FAN L Z, et al. Measuring feeding activity of fish in RAS using computer vision[J]. Aquacultural Engineering, 2014, 60: 20-27.
 [8] 汪靖阳,吴鑫森,常志强,等.基于GMM-CNN的鱼道内鱼类游泳行为提取方法[J].中国水利水电科学研究院学报(中英文),2023,32(2):194-202.
 [9] 孙龙清,王泊宁,王嘉煜,等.基于G-RepVGG和鱼类运动行为的水质监测方法[J].农业机械学报,2022,53(增刊2):210-218.
 [10] 黄志涛,何佳,宋协法.基于鱼体运动特征和图像纹理特征的鱼类摄食行为识别与量化[J].中国海洋大学学报(自然科学版),2022,52(1):32-41.
 [11] ZHOU C, XU D M, CHEN L, et al. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision[J]. Aquaculture, 2019, 507: 457-465.
 [12] ANAS O, WAGEEH Y, MOHAMED H E D, et al. Detecting abnormal fish behavior using motion trajectories in ubiquitous environments[J]. Procedia Computer Science, 2020, 175: 141-148.
 [13] 李星辉.基于轨迹提取的鱼类异常行为监测[D].杭州:浙江大学,2021.
 [14] 袁红春,王丹,陈冠奇,等.基于LightGBM模型的鱼类异常行为检测[J].渔业现代化,2020,47(1):47-55.
 [15] ARNAB A, DEGHANI M, HEIGOLD G, et al. ViViT: a video vision transformer[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021: 6816-6826.
 [16] GONG D, LIU L Q, LE V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul: IEEE/CVF, 2019: 1705-1714.
 [17] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 5967-5976.
 [18] LIU W, LUO W X, LIAN D Z, et al. Future frame prediction for anomaly detection-A new baseline[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 6536-6545.
 [19] TANG Y, ZHAO L, ZHANG S S, et al. Integrating prediction and reconstruction for anomaly detection[J]. Pattern Recognition Letters, 2020, 129: 123-130.
 [20] RAMACHANDRA B, JONES M J. Street Scene: a new dataset and evaluation protocol for video anomaly detection[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, USA: IEEE, 2020: 2558-2567.

责任编辑:邹慧玲
英文编辑:柳正