

基于改进 LRCN 的鱼群摄食强度分类模型

徐立鸿¹ 黄 薪¹ 刘世晶²

(1. 同济大学电子与信息工程学院, 上海 201800; 2. 中国水产科学研究院渔业机械仪器研究所, 上海 200092)

摘要: 实现饵料的自动投喂是自动化水产养殖的重点,对鱼群的摄食强度进行识别能够为精准投饵提供参考。目前大多数关于鱼群摄食强度的研究都是基于循环养殖池或者自制鱼缸中,并不适用于开放式养殖池塘。基于实际环境,采用水上观测方式建立了鱼群摄食强度视频数据集,并提出了一种基于改进长期卷积循环网络(LRCN)的鱼群摄食强度分类模型,将注意力机制 SE 模块嵌入卷积神经网络中,通过 SE-CNN 网络提取视频帧的特征,输入至双层 GRU 网络中,最后通过全连接分类层得出视频类别。提出的 SE-LRCN 模型实现了对鱼群摄食视频的强度三分类。试验结果表明,本文提出的模型分类准确率达到 97%,F1 值达到 94.8%,与改进前的 LRCN 模型相比,准确率提高 12 个百分点,F1 值提高 12.4 个百分点。研究模型可以更精细地识别鱼群的摄食强度,为自动化精准投饵提供参考。

关键词: 水产养殖; 鱼群摄食强度; 长期卷积循环网络; 视频分类; 挤压和激励模块

中图分类号: S951.2 **文献标识码:** A **文章编号:** 1000-1298(2022)10-0236-06

OSID:



Recognition of Fish Feeding Intensity Based on Improved LRCN

XU Lihong¹ HUANG Xin¹ LIU Shijing²

(1. College of Electronics and Information Engineering, Tongji University, Shanghai 201800, China

2. Fishery Machinery and Instrument Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200092, China)

Abstract: The realization of automatic feeding of bait has always been the focus and difficulty of automatic aquaculture. Recognition of the fish feeding intensity can provide a reference for accurate feeding. At present, many laboratories have researches on the fish feeding intensity, but most of the researches on the fish feeding intensity are based on circulating farming ponds or self-made fish tanks, which are not suitable for open farming ponds. Aiming at the actual environmental background and difficulties, the water observation method was used to build a data acquisition system and produce a video data set of fish feeding intensity. Then a fish school feeding intensity classification model was proposed based on improved long-term recurrent convolutional networks(LRCN), which embeded the attention mechanism squeeze-and-excitation block(SE-Block) into the convolutional neural networks. The SE-CNN networks was used to extract the features of the video frames, then input the features into the double-layer gate recurrent unit networks. Finally, the video classification results were obtained through the fully connected classification layer. At the end, the proposed SE-LRCN model realized the intensity three classification of the fish school feeding intensity video. The test results showed that the classification accuracy of the proposed model reached 97%, and the F1 score reached 94.8%. Compared with the long-term recurrent convolutional networks before the improvement, the accuracy was increased by 12 percentage points, and the F1 score was increased by 12.4 percentage points. The research model can more finely recognize the fish feeding intensity, and provide a reference for automatic accurate feeding.

Key words: aquaculture; fish feeding intensity; long-term recurrent convolutional networks; video classification; squeeze-and-excitation block

收稿日期: 2021-11-21 修回日期: 2021-12-15

基金项目: 国家重点研发计划项目(2017YFD0701700)和上海市科技兴农重点项目(沪农科推字(2019)第3-2号)

作者简介: 徐立鸿(1960—),男,教授,博士生导师,主要从事智能控制和水产养殖机器视觉技术等研究,E-mail: xulihong@tongji.edu.cn

0 引言

水产养殖业在中国渔业发展中具有重要的地位^[1-2],同时水产养殖是我国渔业主要生产方式,其中开放式池塘养殖产量占全国水产养殖总量的 80%^[3]。随着水产养殖业的不断发展和扩大,工业化的鱼类养殖技术也在不断地更新和适应不同的养殖环境,实现饵料自动投喂是水产养殖实现自动化、工业化的关键^[4-5]。饵料是鱼类生长的必需品,同时也直接影响水产养殖的经济收益,饵料投喂不足会影响鱼群的生长发育,饵料投喂过量会造成经济浪费和水体富营养化。科学控制水产养殖的投饵量对自动化水产养殖有重要意义。

在智能化投饵领域内,目前基于机器视觉投饵控制研究背景主要是分为室内循环养殖池和开放式养殖池塘。如 ZHOU 等^[6]采用近红外摄像头从养殖池上方获取鱼群状态,通过图像处理等技术提出了 SIFFB 和 FIFFB 共 2 个指标,实现了对鱼群摄食强度的评估。之后他们收集鱼群摄食过程中的图像训练卷积神经网络(Convolutional neural networks, CNN),实现对鱼类食欲水平的分级^[7]。陈明等^[8]提出了一种基于特征加权融合鱼类摄食活动强度评估方法。张佳林等^[9]采集了大西洋鲑水下摄食行为视频,提出了一种基于帧间关系的鱼类摄食行为视频分类算法,唐宸等^[10]在此基础上进行改进,实现了基于光流法的鱼群摄食行为视频三分类。YANG 等^[11]提出了 Efficient Net-B2 的双重注意力网络用于鱼群的细粒度短期摄食行为分析。但是这些方案只适用于水质清晰、光照均匀的室内养殖环境,并不适用于环境复杂开放式养殖池塘。

与循环养殖池相比,开放式养殖池塘受水质混浊、光照不均等环境因素影响,智能化难度较大,成为研究的重点。如胡利永等^[12]以投饵过程中拍摄的水面摄食图像中有鱼部分以及鱼群摄食引起的水花部分作为特征区域,提出了一种投饵量计算模型。王勇平等^[13]通过直方图均衡、帧间差分法等对鱼群摄食期间的图像进行处理,利用椭圆拟合得到目标鱼群面积来进行投饵量的控制。覃学标等^[14]借助水下图像增强技术提高鱼类检测的准确度。

视频识别技术已经应用于各个领域,有许多视频分类深度网络取得了较好的效果,其中包括基于双流法的视频分类^[15-17]和基于 3D 卷积^[18-20]的分类算法。MÅLØY 等^[21]提出了双流循环神经网络(Dual stream recurrent network, DSRN)实现了鱼类摄食行为的两分类,但是网络同时引用了 VggNet^[22]和 ResNet^[23],参数多且计算量大。UBINA 等^[24]提

取视频的光流特征输入 3D 卷积神经网络^[20]实现摄食强度的四分类,但是需要无人机拍摄摄食图像,成本过高,不适用于大部分的池塘养殖。

长期卷积循环网络^[25](Long-term recurrent convolutional networks, LRCN)是一种结合了传统卷积网络和长短期记忆网络的新型网络结构,具有处理时序信息的能力。SENet^[26](Squeeze-and-excitation networks)是一种典型的注意力模型,它的子结构 SE Block 可以随意插入到任何网络中,能够有效提高网络的准确率,且计算量小。

针对实际生产环境问题,本文提出一种基于改进 LRCN 的鱼群摄食强度分类模型,将注意力机制应用于鱼群摄食视频的分类模型中,以通过较少的计算量实现对视频的分类,使模型更具有鲁棒性。

1 数据采集与数据集制作

1.1 试验材料与环境

为证明算法的有效性,试验以草鱼、鲫鱼混合养殖摄食强度分类为例,这些鱼类具有较强的夺食性,可以在养殖池塘水面上观察到较为明显的强度差异,此算法同样适用于其他具有夺食性的池塘养殖淡水鱼摄食强度的分类。试验数据采集于上海瑞钵水产有限公司,该水产基地位于上海市崇明区,数据采集选定 3 号池塘,该池塘水深约 2 m,长约 50 m、宽约 25 m,面积约为 1 333 m²,一个池塘配有一个 ZY150 型颗粒饲料投饵机(佛山市中渔科技有限公司),安装在岸边的延伸台,如图 1 所示,采用扇形喷撒式投饵方案,投饵半径约 6 m。



图 1 自动投饵机

Fig. 1 Automatic feeding machine

试验池塘中草鱼体长约 70 cm,平均鱼体质量 2.5 kg;鲫鱼体长约 35 cm,平均鱼体质量 1 kg。投饵方案为定时定量投饵,每日 09:00—16:00 间隔 1 h 投饵,投饵机抛撒出定量的饵料后会暂停 1 min,等待鱼群摄食完成后再继续下一次投饵动作,每次投饵完全结束需要约 30 min。

1.2 视频采集方案

试验的视频采集采用水上视角,如图 2 所示。摄像头通过立杆安置在池塘投饵机的上方,距离水面约 4 m,摄像头斜向下进行拍摄,拍摄区域中心为

投饵机的饵料抛撒区域。视频采集帧率为 30 Hz,视频帧的原始尺寸为 1 920 像素 × 1 080 像素。摄像头采集不同时间段、不同光照条件下投饵前中后期的鱼群摄食视频,通过 Python 语言的 OpenCV 图像处理库对原始长视频进行处理。视频采集的具体方案如表 1 所示,在光照为中等或较弱时,每次视频采样时间为 10 min;当光照较强时,由于可能出现视频曝光过曝的问题,因此将采集时间设置为 25 min 以保证能采集到足够的数据集。

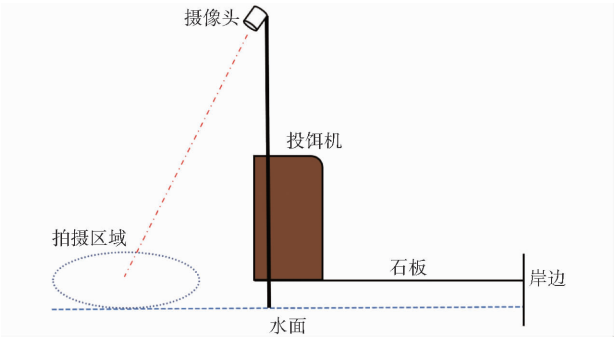


图 2 视频采集示意图
Fig. 2 Structure diagram of video capture

表 1 摄食期间视频采集方案

Tab. 1 Video capture plan during feeding			
天气	时间段	光照强度	采集时间/min
晴	09:00—10:00	中	10
	12:00—13:00	强	25
	15:00—16:00	中	10
阴	09:00—10:00	弱	10
	12:00—13:00	弱	10
	15:00—16:00	弱	10

1.3 数据集制作

将原始采集的水上视频按照未摄食和摄食 2 种状态进行划分,之后结合相关文献及当地养殖专家经验对鱼群摄食期间的视频进行强度分类,分成强和弱两种状态。未摄食状态:鱼群对投喂的饵料无反应,不浮至水面,水面平稳;弱摄食状态:少量鱼群上浮至水面,对投喂饵料主动进食;强摄食状态:大量鱼群上浮至水面抢夺饵料,激起水花。

视频样本的时长设置为 2 s(60 帧),尺寸裁剪为 600 像素 × 500 像素,视频帧图像如图 3 所示。时间长度为 2 s 的视频可以较好地表现当前时间鱼群的摄食强度,同时对设备要求也更低。在实际生产环境中,可以根据现场情况适当调整视频时长,以适应具体的养殖环境。

视频的标注工作由 3 名研究人员同时展开,每一视频样本必须由 2 名或 2 名以上人员确认为同一类别才可完成标注,否则视为无效样本。数据集共



图 3 视频帧
Fig. 3 Video frame

包含 1 071 个时长 2 s 的视频,其中未摄食视频样本 355 个,弱摄食样本 335 个,强摄食样本 381 个,训练集、验证集和测试集按照 7:2:1 的比例划分,具体数量如表 2 所示。

表 2 数据集信息
Tab. 2 Dataset information

摄食强度	训练集	验证集	测试集
未摄食	248	71	36
弱摄食	235	67	33
强摄食	267	76	38
总计	750	214	107

2 基于改进 LRCN 的鱼群摄食强度分类模型

2.1 长期卷积循环网络

长期卷积循环网络 LRCN 用于具有空间输入的序列预测问题,结构如图 4 所示(图中 $y_1, y_2, \dots, y_i$ 表示不同时间下视频帧的分类结果)。用于视频分类时,卷积神经网络提取输入视频帧的特征,这些具有时序关联的特征送入后续的长短期记忆网络(Long short-term memory, LSTM)^[26]进行处理,并得到时序预测结果的输出。

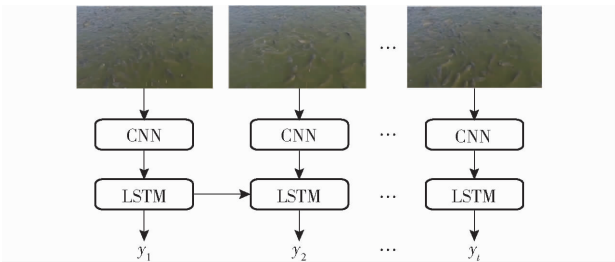


图 4 LRCN 结构图

Fig. 4 Structure diagram of LRCN

LRCN 将 CNN 应用于每个输入的视频帧,并将每个输入视频帧的输出作为单个时间步骤传递给 LSTM 网络。本文选取长期卷积循环网络作为基准模型,融合 SE 模块,并对 LSTM 网络部分进行改进,实现网络性能的提升。

2.2 挤压和激励模块

SENet 是一种新型网络结构,引入了通道注意力机制。SE 模块建模特征通道之间的相互依赖关

系的核心在于网络会自动获取到每个特征通道的重要程度,根据重要程度去放大有用的特征及抑制无用的特征,示意图如图 5 所示。

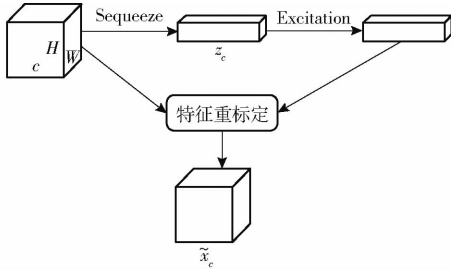


图 5 SE 模块示意图

Fig. 5 Structure diagram of SE block

SE 模块首先对特征图进行 Squeeze 操作,顺着空间维度来进行特征压缩,将每个大小为 $W \times H$ 的二维特征通道压缩至 1×1 ,数学描述为

$$z_c = F_{sq}(x_c) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

式中 x_c ——原始特征图,尺寸为 $H \times W \times c$

z_c ——Squeeze 操作后 $1 \times 1 \times c$ 的特征图

c ——通道数

F_{sq} ——Squeeze 操作函数

之后进行 Excitation 操作,通过参数来为每个特征通道生成权重,其中参数被学习用来显式地建模特征通道间的相关性,数学描述为

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (2)$$

式中 W_1, W_2 ——可训练的参数,其中 W_1 的维度为 $(c/r) \times c$, W_2 的维度为 $c \times c/r$

r ——缩放参数

F_{ex} ——Excitation 操作函数

$\sigma()$ ——Sigmoid 函数

$\delta()$ ——脉冲函数

s ——通道权重

最后进行特征重标定,将 Excitation 操作后的输出权重通过乘法逐通道加权到先前的特征上,数学描述为

$$\hat{x}_c = F_{scale}(x_c, s_c) = s_c x_c \quad (3)$$

式中 $\hat{x}_c$ ——SE 模块处理后特征图,尺寸与 x_c 相同

F_{scale} ——加权操作函数

s_c ——不同通道的特征权重

利用 SE 模块可嵌入到其他模型中的特点,将 SE 模块嵌入 LRCN 模型的卷积神经网络中,提高网络的特征提取能力。

2.3 改进的 LRCN 模型

门控递归单元^[27] (Gate recurrent unit, GRU) 是循环神经网络 (Recurrent neural network, RNN) 的一种,具有结构简单、参数量少、效果优等特点。LSTM

中含有 3 个门函数来控制输入值、记忆值和输出值,在 GRU 中只有更新门和重置门,所以 GRU 需要更少的数据来泛化。

本文提出了 SE-LRCN 视频分类模型,对鱼群的摄食视频进行强度分类,该模型融合了 SE 模块和 LRCN 网络结构,用双层 GRU 代替原始的单层 LSTM 网络,模型的整体结构如图 6 所示(图中 c_{i1} 代表鱼群处于未摄食状态, c_{i2} 代表鱼群处于弱摄食状态, c_{i3} 代表鱼群处于强摄食状态)。

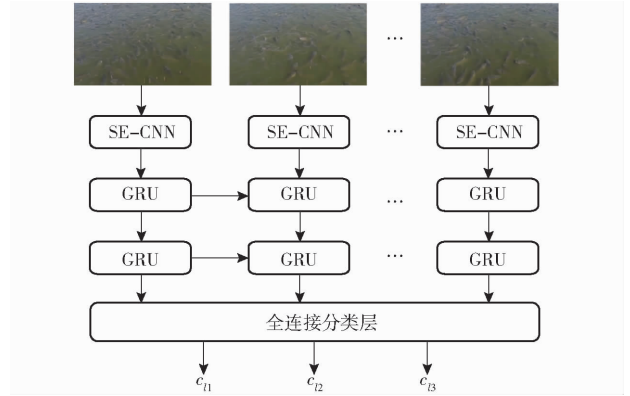


图 6 SE-LRCN 模型示意图

Fig. 6 Structure diagram of SE-LRCN

模型输入层为时长 2 s (60 帧) 的视频样本。SE-CNN 网络部分将 SE 模块嵌入到卷积神经网络中,用于视频帧的特征提取,共包含 5 个卷积层和 1 个平均池化层,中间 3 个卷积层嵌入了 SE 模块 ($r=4$),网络结构如图 7 所示。

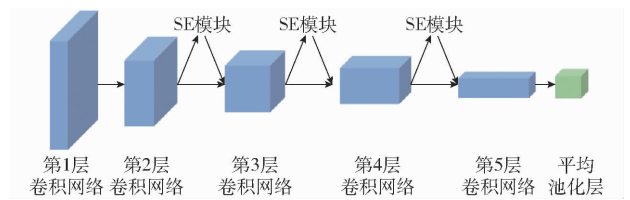


图 7 SE-CNN 网络示意图

Fig. 7 Structure diagram of SE-CNN network

通过 SE-CNN 网络进行视频帧特征提取后,将具有时间分布信息的特征输入至两层的 GRU 网络中,GRU 网络的节点数均为 256,最后通过全连接分类层输出视频样本的类别,输出层采用 Softmax 分类函数预测类别概率,数学描述为

$$p_{c_i} = \frac{e^{a_{c_i}}}{\sum_{k=1}^K e^{a_k}} \quad (c_i = 1, 2, 3) \quad (4)$$

式中 p_{c_i} ——视频样本预测为 c_i 的概率

c_i ——分类类别

K ——类别数量,取 3

a_{c_i} ——全连接分类层预测为 c_i 的原始输出

模型采用交叉熵损失函数 L 进行训练,学习率设置为 0.001。

2.4 评价指标

选取综合准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、F1 值 (F1 score) 4 个评价指标来评估本算法的性能。综合准确率表示预测正确的样本数量占总样本数量的比例,精确率表示预测属于某一类别的样本中预测正确的比例,召回率表示所有属于某一类别的样本被预测正确的比例,F1 值则综合考虑了精确率和召回率,能够避免算法出现保守预测和暴力召回的情况。

3 结果与分析

3.1 循环网络选择

在 SE - LRCN 模型中,GRU 网络的层数会直接影响模型的效果,如表 3 所示。在同样使用 SE - CNN 网络进行视频特征提取的情况下,随着 GRU 网络层数的增加,模型的评价指标均有一定的提升。当 GRU 层数从 1 层上升为 2 层时,模型的准确率提升了 7 个百分点;当层数从 2 层提升为 3 层时,模型准确率只提升了 0.5 个百分点。考虑到 GRU 网络层数提升带来的计算量和参数量,本文最终选择采用两层的 GRU 网络进行视频帧特征的分类。

表 3 不同 GRU 网络层数下模型性能

Tab.3 Model performance at different layers of GRU networks

GRU 网络层数	准确率	精确率	召回率	F1 值
1	90.0	93.6	91.2	92.4
2	97.0	95.7	94.0	94.8
3	97.5	96.0	97.5	96.7

循环神经网络的类型同样会影响模型的效果,本文选择 GRU、LSTM 和 RNN^[28] 3 种典型网络结构进行试验,均采用 SE - CNN 网络进行视频特征提取,并使用双层网络结构,对比结果如表 4 所示,双

表 4 不同网络结构下模型性能

Tab.4 Model performance at different types

网络类型	准确率	精确率	召回率	F1 值
GRU	97.0	95.7	94.0	94.8
LSTM	93.0	93.7	94.0	93.8
RNN	92.5	90.0	89.8	89.9

层 GRU 网络结构在各项指标下均优于其他 2 种网络结构。

3.2 SE - LRCN 模型结果分析

SE - LRCN 模型在验证集上的准确率达到 97%,精确率达到 95.7%,与改进前 LRCN 模型相比,准确率提高 12 个百分点(表 5),精确率提高 15 个百分点,召回率提高 9.8 个百分点,F1 值提高 12.4 个百分点,各项指标都有明显的提升。

表 5 SE - LRCN 与 LRCN 模型性能对比

Tab.5 Experiment results comparison between SE - LRCN model and LRCN model

模型	准确率	精确率	召回率	F1 值
SE - LRCN	97.0	95.7	94.0	94.8
LRCN	85.0	80.7	84.2	82.4

由表 3 ~ 5 可知,原始的 LRCN 网络仅依靠卷积神经网络进行视频特征的提取,特征提取能力有限。SE 模块的引入能够整合特征通道的重要程度,有效提升 CNN 的表示能力,极大地提高了网络的性能。GRU 相较于 LSTM 网络具有参数量少等特点,在数据量大的条件下,LSTM 强大表达能力可能会产生更好的结果。但本文数据集的样本数量较少,结构简单、参数少的 GRU 网络能够更好地泛化。双层 GRU 网络结构与单层相比具有更加强大的表达能力和学习能力,能够更好地处理具体有时序信息特征的问题。

4 结论

(1)搭建了开放式养殖池塘水上鱼群摄食视频数据集,为后续视频分类模型训练提供数据支撑。

(2)对传统的 LRCN 视频分类模型进行改进,将注意力机制应用于视频分类中,提出了 SE - LRCN 模型,适用于数据集数量较少的场景。

(3)利用搭建的开放式养殖池塘水上鱼群摄食视频数据集对 SE - LRCN 模型进行训练,最终在验证集上准确率达到 97%,与改进前 LRCN 模型相比有较大的性能提升。

(4)模型实现了针对开放式养殖池塘水上鱼群摄食强度的三分类,能够较为全面地表现鱼群的摄食状态。

参 考 文 献

[1] 郑思宁, 刘强, 郑逸芳. 规模化水产养殖技术效率及其影响因素分析[J]. 农业工程学报, 2016, 32(20): 229 - 235. ZHENG Sining, LIU Qiang, ZHENG Yifang. Analysis of scale aquaculture production efficiency and analysis of its determinants [J]. Transactions of the CSAE, 2016, 32(20): 229 - 235. (in Chinese)

[2] 岳冬冬, 吴反修, 李欣童, 等. 我国水产养殖业生产效率评估及其对渔业统计的启示[J]. 渔业信息与战略, 2021, 36(2): 9.

[3] 苗雷, 田昌凤, 汤涛林, 等. 标准化池塘养殖集中智能控制系统设计与应用[J]. 科学养鱼, 2013(10): 2.

[4] CHANG C M, FANG W, JAN R C, et al. Development of an intelligent feeding controller for indoor intensive culturing of eel

- [J]. *Aquacultural Engineering*, 2004, 32(2):342–353.
- [5] DU Y, SUN G, ZHENG B, et al. Design and implementation of intelligent gateway system for monitoring livestock and poultry feeding environment based on bluetooth low energy[J]. *Information*, 2021, 12(6):218.
- [6] ZHOU C, LIN K, XU D, et al. Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture[J]. *Computers and Electronics in Agriculture*, 2018, 146:114–124.
- [7] ZHOU C, XU D, CHEN L, et al. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision[J]. *Aquaculture*, 2019, 507(30):457–465.
- [8] 陈明, 张重阳, 冯国富, 等. 基于特征加权融合鱼类摄食活动强度评估方法[J]. *农业机械学报*, 2020, 51(2):245–253.
- CHEN Ming, ZHANG Chongyang, FENG Guofu, et al. Intensity assessment method of fish feeding activities based on feature weighted fusion[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(2):245–253. (in Chinese)
- [9] 张佳林, 徐立鸿, 刘世晶. 基于水下机器视觉的大西洋鲑摄食行为分类[J]. *农业工程学报*, 2020, 36(13):158–164.
- ZHANG Jialin, XU Lihong, LIU Shijing. Classification of Atlantic salmon feeding behavior based on underwater machine vision[J]. *Transactions of the CSAE*, 2020, 36(13):158–164. (in Chinese)
- [10] 唐宸, 徐立鸿, 刘世晶. 基于光流法的鱼群摄食状态细粒度分类算法[J]. *农业工程学报*, 2021, 37(9):238–244.
- TANG Chen, XU Lihong, LIU Shijing. Fine-grained classification algorithm of fish feeding state based on optical flow method[J]. *Transactions of the CSAE*, 2021, 37(9):238–244. (in Chinese)
- [11] YANG L, YU H, CHENG Y, et al. A dual attention network based on Efficient Net – B2 for short-term fish school feeding behavior analysis in aquaculture[J]. *Computers and Electronics in Agriculture*, 2021, 187:106316.
- [12] 胡利永, 魏玉艳, 郑堤, 等. 基于机器视觉技术的智能投饵方法研究[J]. *热带海洋学报*, 2015, 34(4):90–95.
- HU Liyong, WEI Yuyan, ZHENG Di, et al. Research on intelligent bait casting method based on machine vision technology[J]. *Journal of Tropical Oceanography*, 2015, 34(4):90–95. (in Chinese)
- [13] 王勇平, 聂余满, 谢成军, 等. 基于机器视觉的养殖鱼群智能投饵系统设计与研究[J]. *仪表技术*, 2015(1):4.
- [14] 覃学标, 黄冬梅, 宋巍, 等. 模糊水下图像多增强与输出混合鱼类检测方法[J]. *农业机械学报*, 2022, 53(7):243–249.
- QIN Xuebiao, HUANG Dongmei, SONG Wei, et al. Fish detection method of multiple enhanced and outputs blend for blurred underwater images[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(7):243–249. (in Chinese)
- [15] SIMONVAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]// *Advances in Neural Information Processing Systems*. MIT Press, 2014.
- [16] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Convolutional two-stream network fusion for video action recognition[C]// *Computer Vision & Pattern Recognition*. IEEE, 2016.
- [17] NG Y H, HAUSKNECHT M, VIJAYANARASIMHAN S, et al. Beyond short snippets: deep networks for video classification[C]// *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015.
- [18] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks[C]// *IEEE International Conference on Computer Vision*. IEEE, 2015.
- [19] CARREIRA J, ZISSERMAN A. Quo vadis, action recognition? A new model and the kinetics dataset[C]// *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.
- [20] DIBA A, FAYYAZ M, SHARMA V, et al. Temporal 3D convnets: new architecture and transfer learning for video classification[J/OL]. *arXiv e-prints*, 2014. <https://arxiv.org/abs/1711.08200v1>.
- [21] MÅLØY H, AAMODT A, MISIMI E. A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture[J]. *Computers and Electronics in Agriculture*, 2019, 167:105087.
- [22] SIMONVAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]// *International Conference on Learning Representations*, San Diego: ICLR, 2015.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// *IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE, 2016.
- [24] UBINA N, CHENG S C, CHANG C C, et al. Evaluating fish feeding intensity in aquaculture with convolutional neural networks[J]. *Aquacultural Engineering*, 2021, 94:102178.
- [25] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]// *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015.
- [26] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 42(8):2011–2023.
- [27] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting[J/OL]. *arXiv e-prints*, 2015. <https://arxiv.org/abs/1506.04214v2>.
- [28] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J/OL]. *arXiv e-prints*, 2014. <https://arxiv.org/abs/1406.1078v3>.