

丁寅, 陈明, 栗征, 等. 基于全局图推理与改进三维动态卷积的鱼类摄食行为分析[J]. 江苏农业学报, 2024, 40(10): 1863-1874.
doi:10.3969/j.issn.1000-4440.2024.10.011

基于全局图推理与改进三维动态卷积的鱼类摄食行为分析

丁寅^{1,2}, 陈明^{1,2}, 栗征³, 薛江浩^{1,2}

(1.上海海洋大学信息学院, 上海 201306; 2.农业农村部渔业信息重点实验室, 上海 201306; 3.上海时邻信息科技有限公司, 上海 201314)

摘要: 本研究提出一种基于时间动作检测的轻量化视频分类网络, 旨在解决水产智能化养殖中饵料的投喂不均和水体污染等问题, 提高投喂准确性和效率。该网络以 ResNet 3D 为基础, 引入深度可分离卷积模块和三维动态卷积模块, 以降低模型规模和参数量; 同时采用图卷积全局推理模块和稠密卷积模块构建区域和全局关系, 增强网络深层特征的表达, 提高网络分类准确率。经试验验证, 该模型检测准确率可达 96.70%, 相较变分自动编码器卷积网络和 3D ResNet-GloRe 网络, 其准确率分别提高 7.7 个百分点和 4.4 个百分点; 同时, 该模型的参数量和计算量也明显降低, 分别为 1.10 M 和 3.87 G。研究结果表明, 该基于时间动作检测的轻量化视频分类网络可以有效提高水产养殖中饵料的智能化投喂的准确性和效率, 减少饵料投喂不均以及水体污染等问题, 具有较高的应用价值。

关键词: 鱼类行为; 机器视觉; 视频分类; 全局图推理; 动态卷积

中图分类号: S951.2 **文献标识码:** A **文章编号:** 1000-4440(2024)10-1863-12

Fish feeding behavior analysis based on global graph reasoning and improved three-dimensional dynamic convolution

DING Yin^{1,2}, CHEN Ming^{1,2}, LI Zheng³, XUE Jianghao^{1,2}

(1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China; 3. Shanghai Shilin Information Technology Co., Ltd., Shanghai 201314, China)

Abstract: In this study, a lightweight video classification network based on temporal action detection was proposed to solve the problems of uneven feeding of bait and water pollution in intelligent aquaculture, and improve the accuracy and efficiency of feeding. Based on ResNet 3D, a deep separable convolution module and a three-dimensional dynamic convolution module were introduced into the network to reduce the model size and parameter quantity. The graph convolution global inference module and DenseBlock module were used to construct the regional and global relationships, which could enhance the expression of deep features of the network and improve the classification accuracy of network. Experimental results showed that the detection accuracy of the model could reach 96.70%, which was 7.7 percentage points and 4.4 percentage points higher than that of the variational autoencoder convolutional network and 3D ResNet-GloRe network, respectively. The number of parameters and calculation amount of the model were also significantly reduced, which were 1.10 M and

3.87 G, respectively. Therefore, the lightweight video classification network based on time motion detection could effectively improve the accuracy and efficiency of intelligent feeding in aquaculture, reduce the problems of uneven feeding of bait and water pollution, and had high application value.

收稿日期: 2023-09-07

基金项目: 广东省重点领域研发计划项目(2021B0202070001)

作者简介: 丁寅(1998-), 男, 江苏泰州人, 硕士研究生, 主要从事农业信息化、机器视觉与视频分类研究。(E-mail) 2794873588@qq.com

通讯作者: 陈明, (E-mail) mchen@shou.edu.cn

Key words: fish behavior; machine vision; video classification; global graph inference; dynamic convolution

据世界粮农组织统计,2022 年水产品产量达 $1.191\ 8\times 10^8$ t,其中人工养殖的水产品产量占比 52%^[1]。现有人工养殖产业中,饵料成本支出约占总成本的 40%^[2-3],控制饵料精准投喂有助于降低养殖成本。此外,饵料投喂不足会导致鱼类生长减缓,养殖周期变长;而投喂过量会导致水体污染,增加鱼类患病概率^[4]。准确识别鱼类行为是精准投喂的关键。现阶段饵料投喂量多由养殖人员人工判断,主观性强,准确性差^[5]。

现有对鱼类行为的研究主要采用水质检测、声学标签及计算机视觉等技术^[6]。其中,计算机视觉技术因具有对观测事物无侵害、运行效率高等特点被广泛运用^[7-8]。基于计算机视觉技术的鱼类行为分析方法分为通过观察鱼群行为状态进行直接检测和通过水面饵料残留量进行间接检测。

直接检测法运用较广。陈彩文等^[9]利用灰度共生矩阵提取鱼群纹理特征并使用主成分分析法进行降维,实现对鱼群摄食和非摄食状态的识别;陈明等^[10]利用背景减除法等提取图像特征,并结合 Relief 等算法进行加权融合来检测鱼类摄食状态;陈志鹏等^[11]利用光流法提取鱼群运动速度转角等特征来检测鱼群运动的快慢,从而进行摄食状态识别;刘世晶等^[12]利用帧间光流特征和改进递归神经网络来检测草鱼摄食状态;Zhou 等^[13]采用深度学习方法训练 LeNet5 模型,对鱼类摄食强度进行分类,检测鱼类摄食状态;朱明等^[14]使用轻量级神经网络 MobileNetV3-Small 采集投喂后 80~110 s 图片对鲈鱼摄食状态进行检测。

除直接检测鱼类行为数据外,对饵料等进行检测也能间接反映鱼群行为。钱阳^[15]通过计算连续图像中饵料像素差值对鱼类摄食强度进行量化,建立最小二乘支持向量机来检测饵料数量,从而间接检测鱼类摄食状态;Li 等^[16]基于图像掩模局部强度直方图构建自适应阈值算法,检测水下残留饵料数量,从而间接检测鱼类摄食状态;Liu 等^[17]通过自适应大津阈值和线性时间分量标签算法来检测剩余饵料颗粒数量。

以上鱼类行为研究大多基于单个图片特征,此类方法的弊端是单一的图像特征无法完全体现鱼群的运动信息且特征提取对光照和背景等环境条件要

求苛刻。此外,传统机器学习的识别效果依赖特征的选取,当数据类别过多或类别不平衡时,算法容易失效^[18-19]。因此,本研究拟采用基于视频数据和深度学习的视频分类方法对鱼类行为进行研究。

在视频分类领域常见的鱼类摄食行为分类方法有以下几种,Mäløy 等^[20]通过双流循环神经网络对水下鱼类视频进行分类;张佳林等^[21]采用变分自动编码器对视频序列编码预处理,计算特征后再进行卷积分类;Feng 等^[22]采用改进的 3D ResNet-GloRe 网络对鱼类摄食行为进行分类,并根据其摄食行为的强弱来量化鱼类摄食质量的好坏;Shou 等^[23]开发了一种多阶段 3D 卷积网络,通过时间动作检测定位动作标签来准确反映鱼类摄食活动变化。

本研究拟提出一种准确率高、参数量少、计算量小的可用来部署于嵌入式设备中的轻量级神经网络,对鱼类行为视频进行分类,以 ResNet 3D^[24]为基础,引入由 6 层稠密卷积层(DenseLayer)组成的稠密卷积模块(DenseBlock)^[25],通过构建卷积层间的稠密连接来实现特征复用,增强输入特征来提高网络准确率;使用深度可分离卷积模块^[26]进行逐通道卷积和逐点卷积,实现通道和区域分离,减少模型计算量和参数量;引入基于图卷积的全局推理模块^[27],构建全局图获取上下文信息,寻找全局和局部特征关联,加快网络训练速度,提高网络准确率;使用三维动态卷积模块^[28],通过注意力机制根据输入尺寸自适应调整卷积核的大小和形状,更好地提取特征,有效降低参数量和计算量,以期为鱼类摄食行为检测相关研究提供参考,为鱼类养殖饵料精准投喂提供科学依据。

1 材料与方法

1.1 试验材料和数据采集系统

本研究的研究对象为彩鲤和红鲮鱼,试验鱼体长 10~20 cm,平均重量 500 g,饲养于深度 1.5 m、水温 10~15 ℃ 的养殖池中。在试验前,研究对象均已被放养到养殖池中数月,以使其适应环境。投喂时间固定为每天 10:00 和 17:00。试验数据采集系统如图 1 所示,摄像机固定在距养殖池水面高约 1 m 的位置。采集的视频数据帧速率为 1 s 30 帧,分辨率为 1 280×720,格式为 mp4。图像处理语言为 Py-

thon,神经网络搭建工具由 PyTorch 库提供。

1.2 试验数据

本研究将鱼类摄食行为划分为摄食状态、非摄食状态 2 大类,再细分为 4 小类^[29],具体分类见图 2。摄食状态包括摄食行为强和摄食行为弱,非摄食状态包括求食行为强和求食行为弱,分类依据如表 1 所示。以 3~5 s 视频片段为单个样本,用 452 个彩鲤视频片段和 95 个红鲮鱼视频片段作为数据集。选取数据集总量的 20% 用作测试集,剩余部分的 80% 用作训练集,20% 用作验证集。

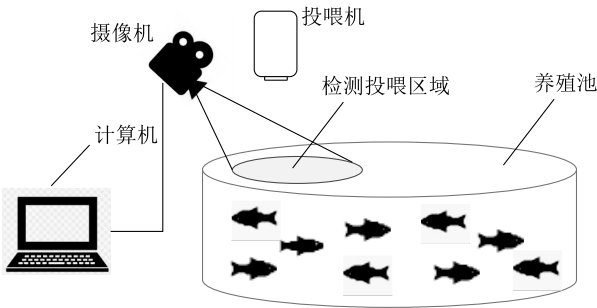


图 1 数据采集系统示意图
Fig.1 Schematic diagram of data acquisition system

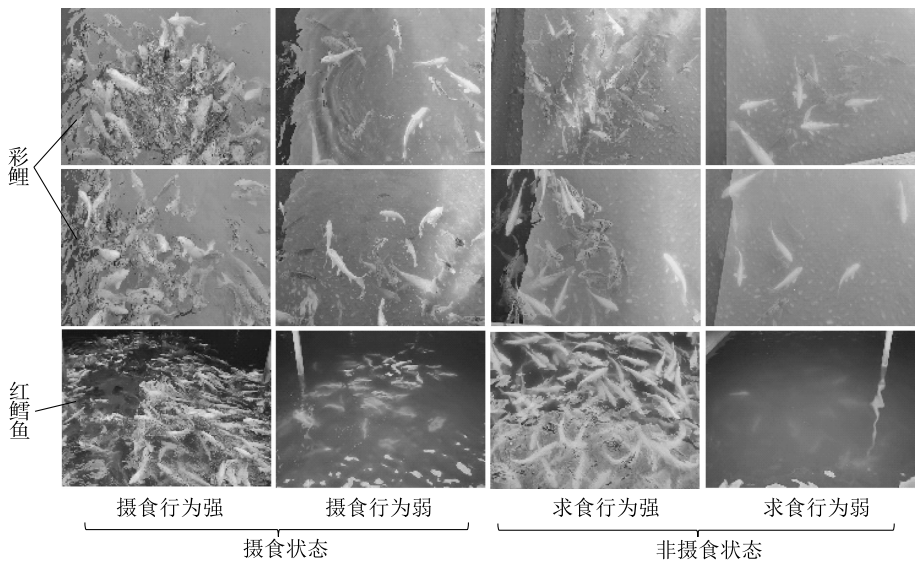


图 2 鱼类摄食行为分类图
Fig.2 Classification diagram of fish feeding behaviors

表 1 鱼类行为分类标准
Table 1 Standard for classification of fish behaviors

摄食行为分类		分类依据
摄食状态	摄食行为强	鱼群游动速度快,对洒下饲料反应激烈,大量上浮水面致使水面溅起大量水花,同一画面中出现的鱼群数量多
	摄食行为弱	鱼群游动速度减慢,少量上浮水面,只对嘴边饲料有反应,水面溅起水花少,同一画面中出现的鱼群较少
非摄食状态	求食行为强	鱼群游动速度快,上浮至水面等待投喂,同一画面出现鱼群多
	求食行为弱	鱼群静止或游动速度慢,同一画面中出现的鱼群数量较少

1.3 试验方法

本研究采用视频分类方法,通过提取输入视频数据的时空特征信息对鱼类摄食行为进行分类。试验流程为:第一,数据收集,采集鱼群摄食视频;第二,数据预处理,将视频大小调整为171×128 像素以加速网络运算,并等间隔提取视频帧消除信息冗余;第三,网络训练,将处理后的数据集投入网络进行训

练;第四,模型测试,用测试集验证和训练网络模型。试验算法应用流程如图 3 所示,将采集好的视频进行数据预处理后投入网络进行分类检测,得出鱼类摄食行为状态,并据此给出指令。
本研究所提算法网络结构如图 4 所示。本研究进行的改进具体如下:(1)引入由 6 层 DenseLayer 连接组成的 DenseBlock 实现特征复用,增强输入特

征,提高网络准确性,通过降维模块(Transition)下采样压缩通道数,减少计算量;(2)引入全局图推理模块(GloRe)构建摄食阶段鱼群和饵料的局部和全体关系,加快网络训练;(3)使用深度可分离卷

积(DCN)通过逐点卷积和逐通道卷积进行区域和通道分离,减少网络计算量;(4)使用动态卷积根据输入特征动态调整卷积核来减少网络规模,降低计算量和参数量。

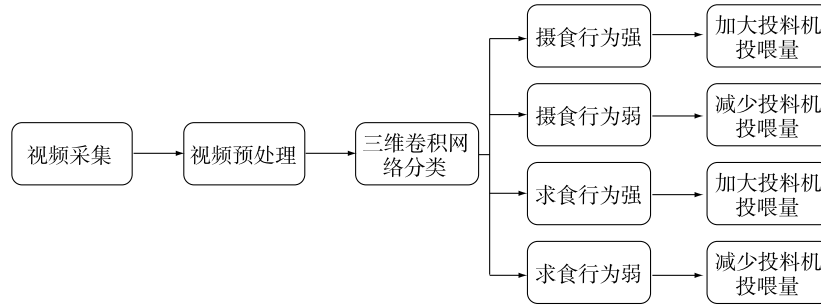


图3 视频分类算法流程图

Fig.3 Flowchart of video classification algorithm

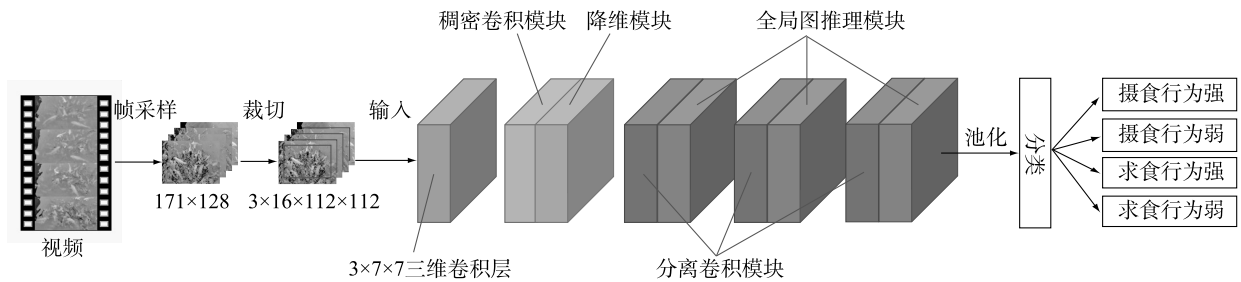


图4 全局图推理与改进三维动态卷积网络结构图

Fig.4 Framework diagram of global graph inference and improved three-dimensional dynamic convolution

1.3.1 DenseBlock 和 Transition DenseBlock 和 Transition 是由 Huang 等^[25]提出的用来解决深度卷积网络梯度消失问题的模块。使用由 6 层 DenseLayer 结构稠密连接组成的 DenseBlock 来对鱼类摄食数据特征进行增强,具体结构见图 5。DenseLayer 结构由 2 层卷积构成,第一层 $1\times1\times1$ 卷积用来对输入特征进行升维,通道数,并输出处理后的特征;第二层 $3\times3\times3$ 卷积用来对处理后的特征降维,缩减通道数,并对卷积的输入输出进行拼接。

DenseBlock 中 DenseLayer 结构的输入由前一层的输出和前几层的输入组合而成。1 个具有 6 层 DenseLayer 结构的 DenseBlock 输入为 x_{int} 时,每一层输出和 DenseBlock 总输出(x_{out})对应公式如下所示:

$$x_{\text{out}(1)} = F(x_{\text{int}}) \quad (1)$$

$$x_{\text{out}(2)} = F[x_{\text{out}(1)} + x_{\text{int}}] \quad (2)$$

$$x_{\text{out}(3)} = F[x_{\text{out}(1)} + x_{\text{out}(2)} + x_{\text{int}}] \quad (3)$$

$$x_{\text{out}(4)} = F[x_{\text{out}(1)} + x_{\text{out}(2)} + x_{\text{out}(3)} + x_{\text{int}}] \quad (4)$$

$$x_{\text{out}(5)} = F[x_{\text{out}(1)} + x_{\text{out}(2)} + x_{\text{out}(3)} + x_{\text{out}(4)} + x_{\text{int}}] \quad (5)$$

$$x_{\text{out}(6)} = F[x_{\text{out}(1)} + x_{\text{out}(2)} + x_{\text{out}(3)} + x_{\text{out}(4)} + x_{\text{out}(5)} + x_{\text{int}}] \quad (6)$$

$$x_{\text{out}} = x_{\text{int}} + x_{\text{out}(1)} + x_{\text{out}(2)} + x_{\text{out}(3)} + x_{\text{out}(4)} + x_{\text{out}(5)} + x_{\text{out}(6)} \quad (7)$$

DenseBlock 中 DenseLayer 结构输出通道数相同,若每层 DenseLayer 输入通道数为 k_0 ,输出通道数为 k ,则 1 个具有 6 层 DenseLayer 的 DenseBlock 输出通道数为 $k_0 + 6k$ 。由此可见,使用稠密卷积模块虽然能通过输入特征的复用提高网络性能,但也会因此带来网络特征通道数的增加,导致模型计算量的增大,因此引入 Transition 降维模块,通过卷积核为 1 的三维卷积网络来对稠密卷积模块处理后的通道数进行降维,经过 Transition 结构下采样后输出通道数为 $\frac{k_0 + 6k}{2}$,有效降低通道数,减少模型计算量。

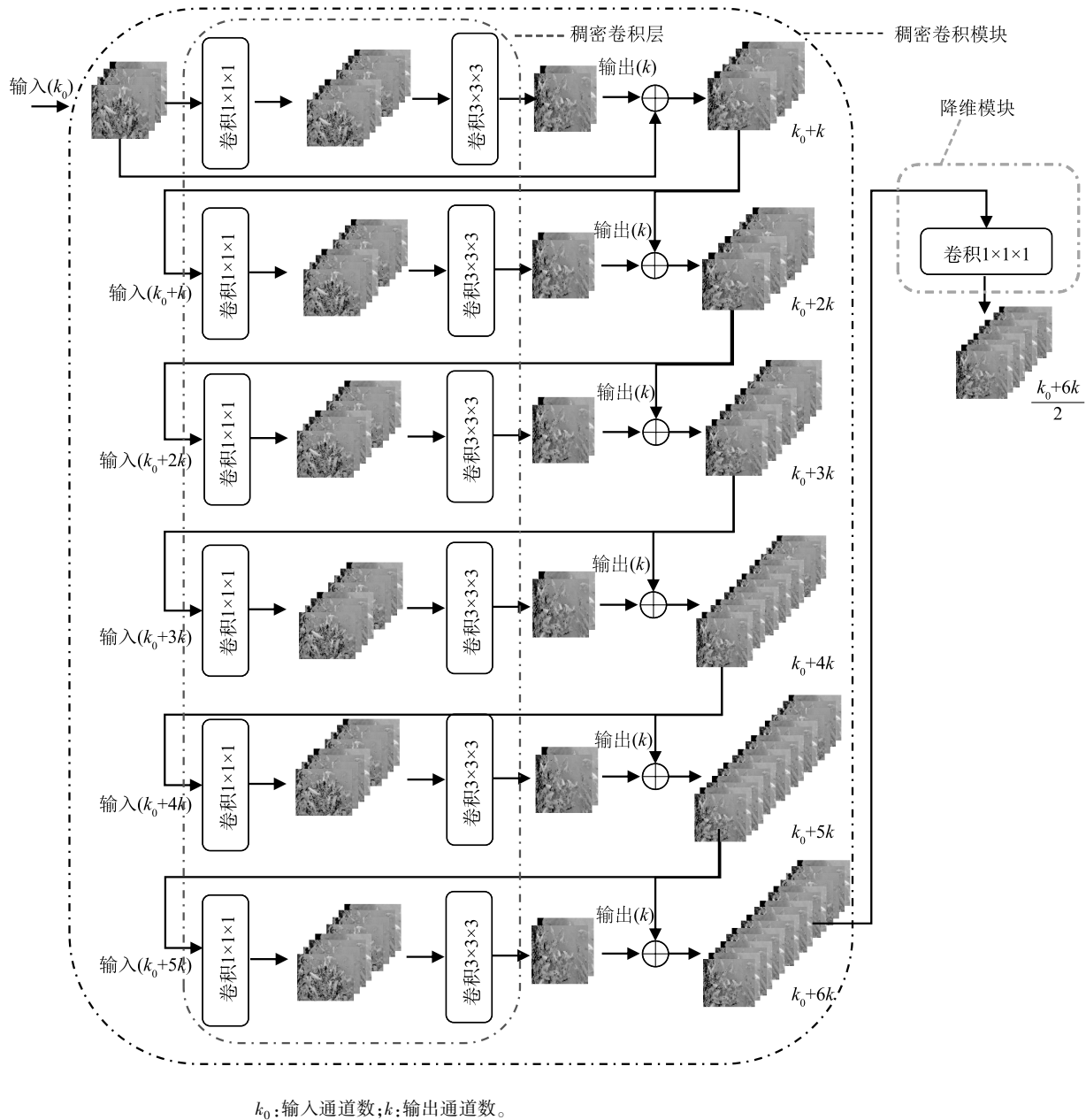


图5 DenseBlock 模块和 Transition 模块结构示意图

Fig.5 Framework diagram of DenseBlock module and Transition module

1.3.2 深度可分离卷积模块 深度可分离卷积^[26]通过区域和通道分离,能够在降低网络参数量和计算量的同时保证网络原有性能损失较少,被广泛运用在轻量化网络模型结构中。深度可分离卷积分为2个步骤,分别为逐通道卷积和逐点卷积,模块结构见图6。

对于输入通道数为 n ,输出通道数为 m 的卷积过程,深度可分离模块先对输入特征进行逐通道卷积,使用与通道数对应的 n 个 $k \times k$ 大小卷积核来分

离通道;其次再对特征进行逐点卷积,使用 m 个 $1 \times 1 \times n$ 卷积输出通道数为 m 的特征,此时深度可分离模块卷积参数量为 $k \times k \times n + n \times m$,远小于普通卷积的 $k \times k \times n \times m$ 。

1.3.3 全局图推理模块(GloRe) 全局图推理模块(GloRe)是由Chen等^[27]提出的,用来解决传统卷积因叠加运算带来的较远关联区域运算效率低下的问题。该模块通过加权平均池化和加权广播实现特征坐标空间和交互空间的相互映射,并在交互空间

中运用图卷积进行关系推理,原理和结构见图 7、图 8。在本研究中,通过提取视频特征数据中各摄食鱼

群特征和饵料特征,构建其对应关系,并将其投影到交互空间,便于卷积时获取区域之间的关联。

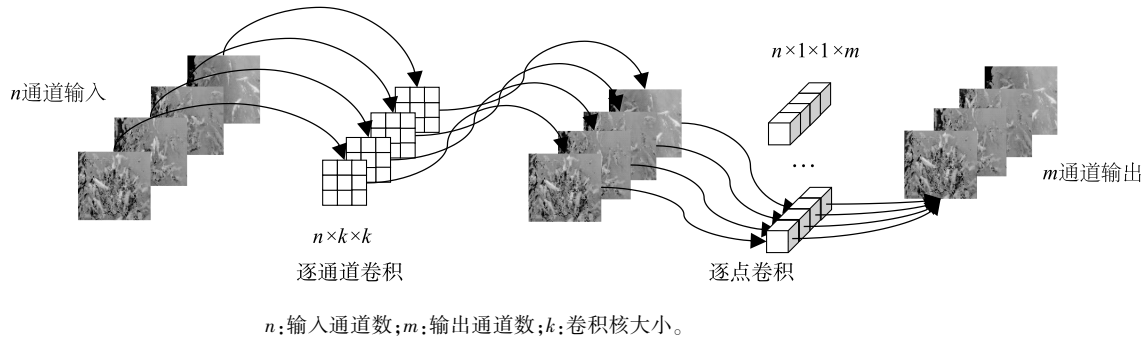


图 6 深度可分离卷积模块运算过程

Fig.6 Operation process of depthwise separable convolution module

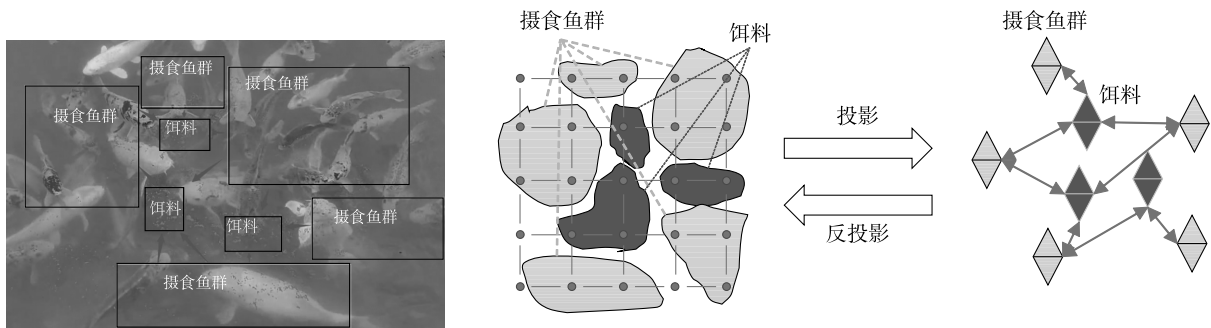
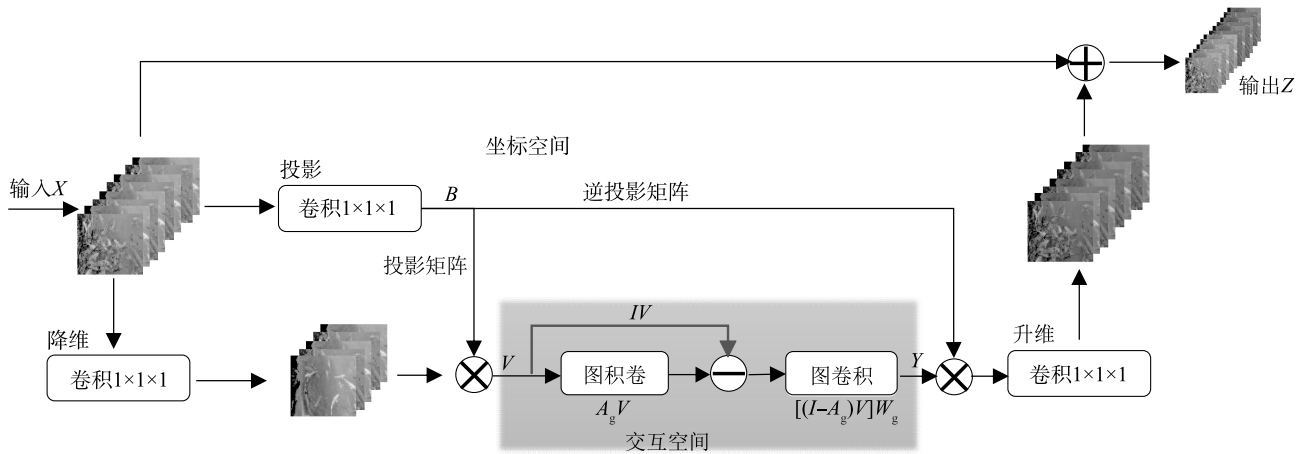


图 7 GloRe 原理图

Fig.7 Schematic diagram of GloRe



X : 输入特征; B : 投影后特征; V : 降维投影后特征; I : 图推理转换邻接矩阵; A_g : 第一层图卷积; W_g : 第二层图卷积; Y : 处理后的输出特征; Z : 模块输出特征。

图 8 GloRe 单元结构图

Fig.8 GloRe unit framework diagram

GloRe 模块对数据的处理分为 3 个阶段: 第一, 将坐标空间的输入特征通过降维投影方式映射到交互空间; 第二, 在交互空间中运用图卷积进行全局关

系推理; 第三, 将交互空间中处理后的特征通过升维和逆投影映射回原坐标空间。

第一阶段使用 2 个 $1 \times 1 \times 1$ 卷积, 分别用于提取

输入数据的局部特征和对输入数据进行降维。输入特征 X 经投影函数 $\theta(X)$ 得到双投影矩阵 B , 矩阵 B 再与降维后的特征 $\varphi(X)$ 相乘得到映射到交互空间的鱼类摄食特征 V , 对应公式如下:

$$B = \theta(X) \quad (8)$$

$$V = \varphi(X)B \quad (9)$$

第二阶段对进入交互空间的特征进行图卷积推理, 构建全体和局部关系。输入特征 V 通过邻接矩阵 A_g 和 I 转换为全连通图的节点, 再通过状态更新函数 W_g 将节点间关系转化为全连通图边权值, 此时得到处理后的输出特征 Y , 对应公式如下:

$$Y = (I - A_g) V W_g \quad (10)$$

第三阶段为将交互空间处理后的特征经过关系推理后映射回原坐标空间。特征 Y 经双投影矩阵 B

映射回原坐标空间, 再使用升维函数 $f(x)$ 升维, 并将输出与原输入结合得到总输出 (Z), 公式如下:

$$Z = f[(I - A_g) V W_g B] + X \quad (11)$$

经过全局图推理模块的鱼类摄食特征构建摄食鱼群与饵料的关系, 在训练过程中能够加快网络训练速度, 提高网络训练的准确率。

1.3.4 三维动态卷积模块 动态卷积模块是由 Chen 等^[28]提出, 通过注意力机制根据输入特征自适应调整卷积核的模块, 静态卷积和动态卷积对比如图 9 所示。采用三维动态卷积来减少网络的计算量和参数量, 相较于传统的静态卷积, 使用注意力机制聚合卷积核的三维动态卷积能够在较小的计算量和参数量下提升模型性能, 并避免了过大的卷积核导致的过拟合等问题。

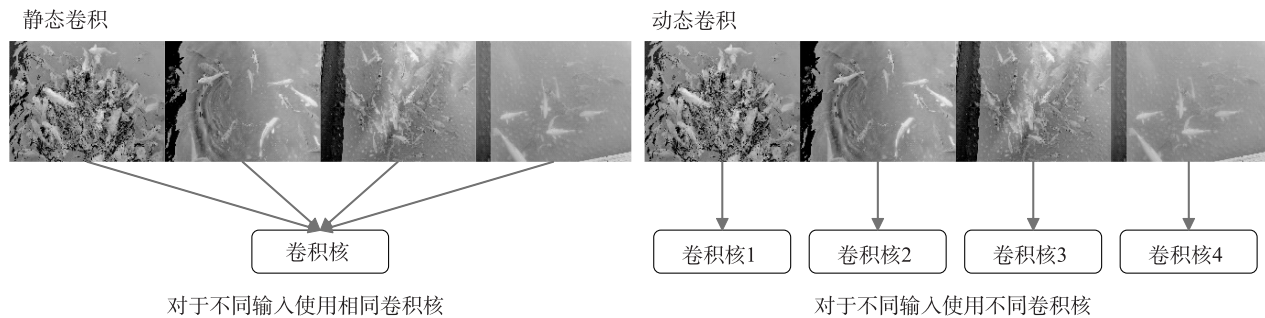


图9 静态卷积和动态卷积对比图

Fig.9 Comparison of static convolution and dynamic convolution

动态卷积模块结构图(图 10)显示, 采用压缩激励操作计算卷积核的注意力权重, 调整所需卷积核的大小和形状。对于输入特征 X , 首先采用全局平

均池化来压缩全局空间信息, 其次使用全连接层对压缩后的特征进行降维和权重计算, 最后根据权重聚合并行卷积核得出输入特征 X 所对应的卷积核。

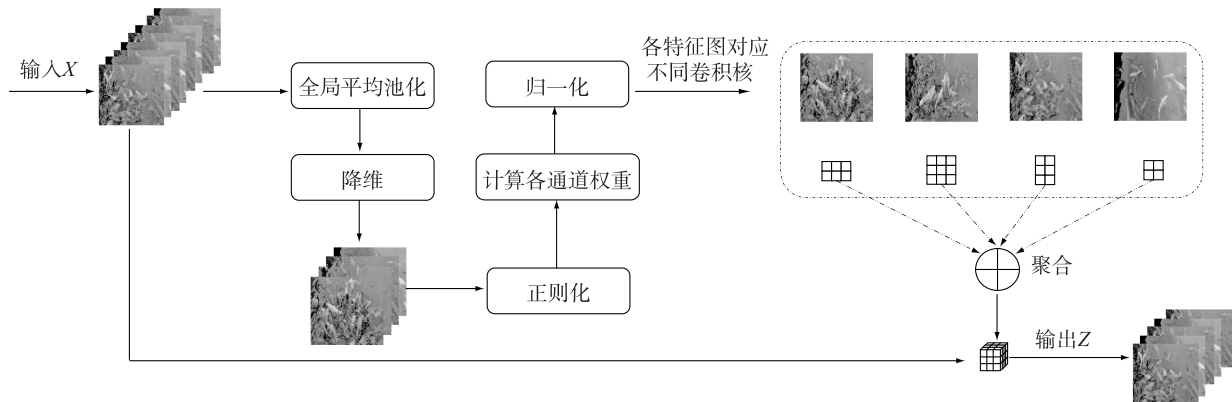


图10 动态卷积模块结构图

Fig.10 The structure diagram of dynamic convolution module

1.3.5 算法性能评估 本研究采用精确率、召回率、F1 值和准确率作为模型评价指标^[30]。精确率是分

类正确的正样本数与分类器判定为正样本的样本数之比;召回率是分类正确的正样本数与实际正样本数之比; $F1$ 值是精确率和召回率的调和平均值;准确率是分类正确的样本数与总样本数之比。各评价指标对应公式如下:

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (12)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (13)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (14)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (15)$$

式中, $Precision$ 为精确率; $Recall$ 为召回率; $F1$ 为 $F1$ 值; $Accuracy$ 为准确率; TP 为模型检测类别和实际类别均为摄食行为强的样本数; FP 为模型检测类别为摄食行为强,实际类别为其他类别的样本数; TN 为模型检测类别和实际类别相似且为非摄食行为强的样本数; FN 为模型检测为非摄食行为强类别,但实际类别为摄食行为强类别的样本数。

1.4 网络环境和参数设置

为减少数据量,保证提取信息丰富,提高网络计算效率,采用抽帧和调整图片像素操作对视频数据进行处理。视频每隔 6 帧抽取 1 次并调整像素大小为 171×128 ,使用窗口滑动选取 16 张连续图片并随机裁剪为 3 通道 112×112 像素图片作为神经网络的输入。

网络超参数设置如下:批处理设置为 16;训练周期设置为 500;初始学习率设置为 0.001,并选择 StepLR 方法动态调整学习率;使用 Adam 优化器优化参数,提高模型收敛性能;选择交叉损失函数计算模型损失,衡量模型预测与真实标签之间的差距。

2 结果与分析

2.1 网络模型训练结果分析

本研究训练集和测试集迭代情况如图 11、图 12 所示,模型在前 15 次训练周期中,损失值迅速下降,准确率快速提升。表明模型在训练初期能较快学习样本的特征。随着迭代周期的增加,损失值缓慢下降,并稳定在 0.80 左右,同时验证集准确率稳定在 95.00% 左右。这说明模型已经逐渐收敛,并且在验证集上也能够达到较高的准确率。

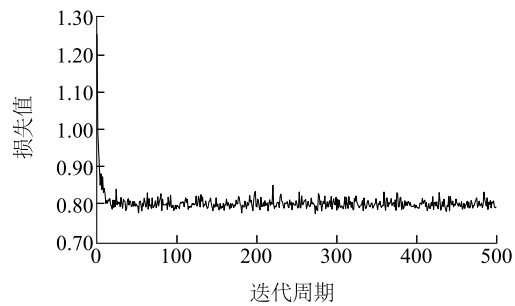


图 11 网络模型训练损失值

Fig.11 The loss values in the process of network model training

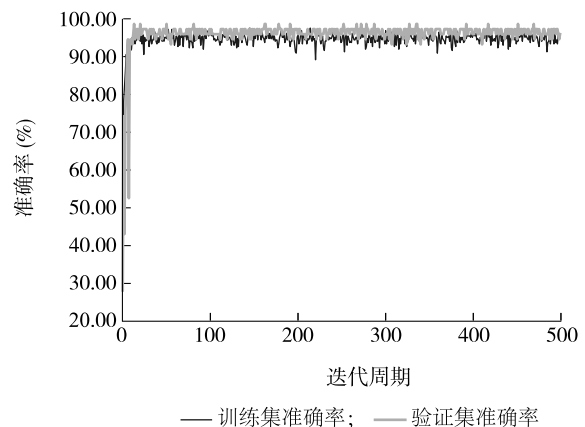


图 12 训练集和验证集的准确率

Fig.12 Accuracy of training set and validation set

用测试集数据进行验证,发现模型总体准确率达到 96.70%。各类别对应的精确率、召回率和 $F1$ 值见表 2,模型对摄食行为强、求食行为强和求食行为弱类别的视频识别效果较好,对摄食行为弱类别的视频识别效果较差。

表 2 各分类精确率、召回率和 $F1$ 值结果

Table 2 Results of precision, recall and $F1$ score

鱼类摄食行为	精确率 (%)	召回率 (%)	$F1$ 值 (%)
摄食行为强	100.00	95.80	97.90
摄食行为弱	92.00	95.80	93.90
求食行为强	95.60	100.00	97.80
求食行为弱	100.00	95.20	97.60

混淆矩阵能直观地显示模型在各类别上的表现,本研究采用模型在测试集验证结果的混淆矩阵如图 13 所示。混淆矩阵灰度的深浅代表了模型预测准确率的高低,主对角线灰度较浅,模型预测准确率较高。

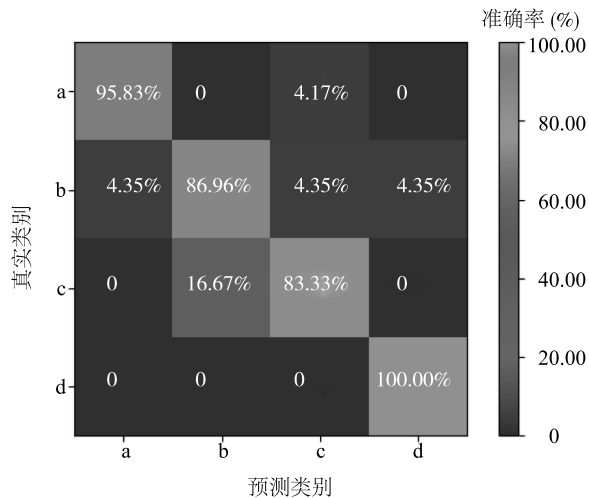


图 13 混淆矩阵

Fig.13 Confusion matrix

表 3 对比试验

Table 3 Comparative experiments

网络	综合准确率 (%)	平均阳性准确率 (%)	平均召回率 (%)	平均 F1 值 (%)	参数量 (M)	计算量 (G)
C3D	83.50	88.05	83.12	82.37	78.02	38.64
ResNet 3D	89.00	89.41	89.05	89.21	33.21	8.33
R2+1D	92.30	92.59	92.32	92.37	33.18	41.85
变分自动编码器卷积网络	89.00	89.00	90.00	87.00		
3D ResNet-GloRe	92.30	92.57	92.37	92.42	15.71	5.91
本研究方法	96.70	96.91	96.72	96.77	1.10	3.87

R2+1D 网络通过将 3D 卷积分解为 2D 空间卷积和 1D 时间卷积,提高了网络的非线性能力,降低了训练损失并提高网络综合准确率。但卷积核的分解需要将 2D 空间卷积和 1D 时间卷积进行参数组合,导致参数量和计算量的增加。

此外,采用变分自动编码器卷积网络对鱼类摄食数据预处理后投入卷积网络进行训练所得综合准确率也较差,仅为 89.00%。变分自动编码器对视频数据进行编解码过程中会损失特征信息,带来网络识别综合准确率的下降。

最后,3D ResNet-GloRe 网络使用全局图推理,虽然提高了网络训练的综合准确率,但并未考虑特征输入的增强,导致综合准确率低于本研究所提网络。

综上所述,本研究所提网络的综合准确率达 96.70%,且能有效降低网络参数量和计算量,适用于鱼类摄食行为分析,为水产智能化养殖中饵料的精准投喂提供可靠依据,具备轻量化部署的优势。

2.2 对比试验

将本研究方法与常见鱼类摄食视频分类网络 C3D^[31]、R2 + 1D^[32]、ResNet 3D^[24]、3D ResNet-GloRe^[22]和变分自动编码器卷积网络^[20]在相同参数条件下进行对比试验,结果(表 3)表明,在对比试验中,C3D 网络分类综合准确率最差,同时模型参数量也最多。其因为 C3D 网络结构中大量使用普通叠加卷积和全连接层,这导致网络计算量会因网络深度的增加而增大,带来梯度消失和过拟合等问题,造成准确率下降。

ResNet 3D 网络与 C3D 网络相比,综合准确率提高至 89.00%。该网络通过使用残差块结构、引入层间跳跃连接,有效缓解了梯度爆炸问题,并减少了所需的卷积层数。因此,网络参数量和计算量也相应减少。

2.3 消融试验

对全局图推理模块(GloRe)、稠密卷积模块(DenseBlock)、深度可分离卷积模块和动态卷积模块进行消融试验(表 4)。ResNet 3D 网络添加 DenseBlock 模块后,模型准确率提高,参数量减少。DenseBlock 通过层间稠密连接实现对输入特征的复用,使得网络能够更好地捕捉和利用特征信息,提升网络特征表达能力。此外,通过复用特征还能减少网络所需参数量,降低模型的复杂度。

添加 GloRe 后,模型准确率进一步提升,参数量和计算量下降。GloRe 通过构建全体特征与局部特征的关系,增强特征关联性,提高网络准确率。此外,整体特征与局部特征关系的构建也减少了非必要的叠加运算,使网络参数量和计算量降低。

添加深度可分离卷积模块和动态卷积模块后,观察到模型在维持原有准确率的同时,参数量和计算量进一步下降。深度可分离卷积模块通过对原有

卷积进行区域和通道分离,而动态卷积模块则根据输入动态调整卷积核,使得模型在减少参数量和计算量的同时能够维持较高的准确率。

表 4 消融试验

Table 4 Ablation experiments

网络	准确率 (%)	参数量 (M)	计算量 (G)
ResNet 3D	89.00	33.21	8.33
ResNet 3D+DenseBlock	94.50	5.71	12.22
ResNet 3D+DenseBlock+GloRe	96.70	4.16	10.78
ResNet 3D+DenseBlock+GloRe+DCN+Dynamic conv	96.70	1.10	3.87

GloRe:全局图推理模块;DenseBlock:稠密卷积模块;DCN:深度可分离卷积模块;Dynamic conv:动态卷积模块。

2.4 模型应用测试与分析

将训练好的模型投入应用并进行测试,并对结

果(图 14)进行分析。未到达投喂时间段时,鱼群四散于养殖池,投喂区域出现鱼群较少,此时采集鱼群状态,如图 14a 所示,模型识别判断鱼群状态为求食行为弱。到达投喂时间段时,投喂区域鱼群数量增多,并以较快速度游动等待投喂,此时采集鱼群状态,如图 14b 所示,模型识别判断鱼群状态为求食行为强。开始投喂阶段,装置投喂饵料,鱼群对洒下饵料反应激烈,大量上浮水面抢食致使水面溅起水花,此时采集鱼群状态,如图 14d 所示,模型识别判断鱼群状态为摄食行为强。投喂后半段,多数鱼群饱腹,较少鱼类仍在摄食,投喂区域鱼群变少,此时采集鱼群状态,如图 14c 所示,模型识别判断鱼群状态为摄食行为弱。结束投喂后,鱼群结束摄食,四散而去,投喂区域出现鱼群较少,此时采集鱼群状态,如图 14a 所示。

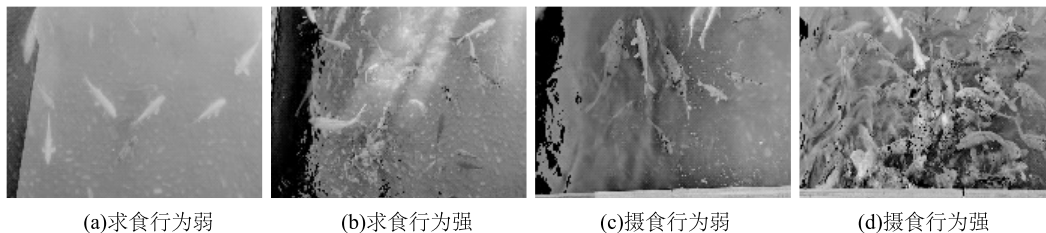


图 14 鱼类摄食行为正确识别效果图

Fig.14 A display of the correct recognition of fish feeding behaviors

本研究训练模型虽能较好识别鱼类摄食行为状态,但仍存在误识别情况。当鱼群处于非摄食阶段时,系统会将鱼群求食行为强状态误识别为摄食行为强状态,误识别案例如图 15a 所示,误识别原因为鱼群聚集过程中白色鱼类颜色和泛起的水花颜色相似,系统会将白色鱼类错误判断为水面泛起的白色水花,致使识别错误。当鱼群处于摄食阶段时,系统会将鱼

群摄食行为弱状态误识别为摄食行为强状态,误识别案例如图 15b 所示,误识别原因为鱼群饱腹程度不一,在投喂阶段后期,多数鱼类摄食完毕,仍有少数鱼类继续摄食,部分鱼类的激烈摄食导致水面溅起水花,系统误识别为鱼群状态为摄食行为强。此外,鱼群摄食完毕未及时离开检测区域也会被系统误识别为求食行为强,而此时应为摄食行为弱。

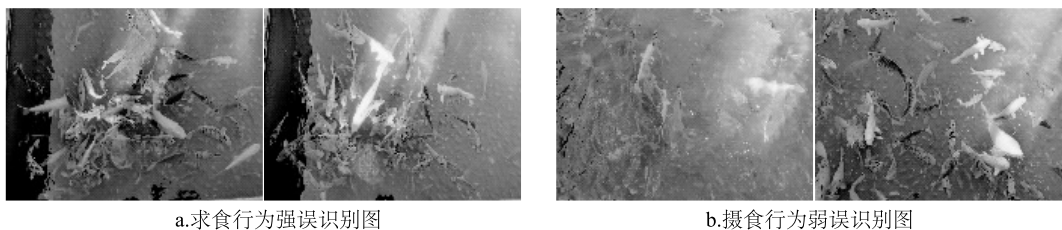


图 15 鱼类摄食行为错误识别效果图

Fig.15 A display of the error recognition of fish feeding behaviors

综上所述,本研究所训练的模型在识别鱼类摄

食行为状态方面虽有一定的准确性,但由于受到部

分鱼类颜色与水花相似、鱼群饱腹程度差异以及未能及时离开检测区域等因素的影响,仍存在明显的误识别问题。在未来的研究中,应致力于进一步优化模型,以提高其准确性和可靠性。

3 结论

本研究提出一种基于时间动作检测定位动作标签的视频分类算法,分析鱼类摄食行为,实现水产养殖中饵料的精准投喂。引入全局图推理和三维动态卷积等模块,在减少模型参数量和计算量的同时提高了准确率。此外,模型使用深度学习方法对输入视频数据分类,自动提取特征,识别精度高;引入时间序列信息能够全面获取鱼类行为信息,加强判断依据,提高模型准确率。经测试研究,该网络对鱼类摄食行为识别的准确率可达 96.70%,与同类型网络 ResNet 3D 网络相比,在准确率提高 7.7 个百分点的同时,参数量和计算量分别下降 96.69% 和 53.54%。参数量和计算量的减少使得模型在轻量化部署方面具有明显优势,为水产养殖中饵料的精准投喂提供可靠参考,具有广泛的应用场景。

参考文献:

- [1] 莽琦,徐钢春,朱健,等. 中国水产养殖发展现状与前景展望[J]. 渔业现代化,2022,49(2):1-9.
- [2] ATOUM Y, SRIVASTAVA S, LIU X M. Automatic feeding control for dense aquaculture fish tanks[J]. IEEE Signal Processing Letters, 2014, 22(8):1089-1093.
- [3] 俞国燕,张宏亮,刘峰春,等. 水产养殖中鱼类投喂策略研究综述[J]. 渔业现代化,2020,47(1):1-6.
- [4] ZHAO S P, DING W M, ZHAO S Q, et al. Adaptive neural fuzzy inference system for feeding decision-making of grass carp (*Ctenopharyngodon idellus*) in outdoor intensive culturing ponds[J]. Aquaculture, 2019, 498:28-36.
- [5] 胡金有,王靖杰,张小栓,等. 水产养殖信息化关键技术研究现状与趋势[J]. 农业机械学报,2015,46(7):251-263.
- [6] LI D L, WANG Z H, WU S Y, et al. Automatic recognition methods of fish feeding behavior in aquaculture: a review[J]. Aquaculture, 2020, 528:735508.
- [7] 朱明,张镇府,黄凰,等. 鱼类养殖智能投喂方法研究进展[J]. 农业工程学报,2022,38(7):38-47.
- [8] 周超,徐大明,吝凯,等. 基于近红外机器视觉的鱼类摄食强度评估方法研究[J]. 智慧农业,2019,1(1):76-84.
- [9] 陈彩文,杜永贵,周超,等. 基于图像纹理特征的养殖鱼群摄食活动强度评估[J]. 农业工程学报,2017,33(5):232-237.
- [10] 陈明,张重阳,冯国富,等. 基于特征加权融合的鱼类摄食活动强度评估方法[J]. 农业机械学报,2020,51(2):245-253.
- [11] 陈志鹏,陈明. 基于光流法与图像纹理特征的鱼群摄食行为检测[J]. 南方农业学报,2019,50(5):1141-1148.
- [12] 刘世晶,涂雪滢,钱程,等. 基于帧间光流特征和改进 RNN 的草鱼摄食状态分类[J]. 水生生物学报,2022,46(6):914-921.
- [13] 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.
- [14] 朱明,张镇府,黄凰,等. 基于轻量级神经网络 MobileNetV3-Small 的鲈鱼摄食状态分类[J]. 农业工程学报,2021,37(19):165-172.
- [15] 钱阳. 基于图像动态获取的水产养殖智能投饵机控制系统研究[D]. 镇江:江苏大学,2017.
- [16] LI D W, XU L H, LIU H Y. Detection of uneaten fish food pellets in underwater images for aquaculture[J]. Aquacultural Engineering, 2017, 78:85-94.
- [17] LIU H Y, XU L H, LI D W. Detection and recognition of uneaten fish food pellets in aquaculture using image processing[C]//SPIE. Sixth International Conference on Graphic and Image Processing (ICGIP 2014). Bellingham: SPIE, 2015:86-92.
- [18] LORENTE S, RIERA I, RANA A. Image classification with classic and deep learning techniques[J]. ArXiv Preprint ArXiv, 2021. DOI:10.48550/arXiv.2105.04895.
- [19] CHEN L Y, LI S B, BAI Q, et al. Review of image classification algorithms based on convolutional neural networks[J]. Remote Sensing, 2021, 13(22):4712.
- [20] 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.
- [21] 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.
- [22] FENG S X, YANG X T, LIU Y, et al. Fish feeding intensity quantification using machine vision and a lightweight 3D ResNet-GloRe network[J]. Aquacultural Engineering, 2022, 98:102244.
- [23] SHOU Z, WANG D, CHANG S F. Temporal action localization in untrimmed videos via multi-stage cnns[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:1049-1058.
- [24] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:770-778.
- [25] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:2261-2269.
- [26] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:1251-

- 1258.
- [27] CHEN Y P, ROHRBACH M, YAN Z C, et al. Graph-based global reasoning networks[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle;IEEE,2020;433-442.
- [28] CHEN Y P, DAI X Y, LIU M C, et al. Dynamic convolution;attention over convolution kernels[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach;IEEE,2020;11021-11030.
- [29] ØVERLI Ø, SØRENSEN C, NILSSON G E. Behavioral indicators of stress-coping style in rainbow trout;do males and females react differently to novelty? [J]. Physiology & Behavior,2006,87(3): 506-512.
- [30] 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.
- [31] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision. Sydney;IEEE,2015;4489-4497.
- [32] TRAN D, WANG H, TORRESANI L, et al. A closer look at spatiotemporal convolutions for action recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas;IEEE,2018;6450-6459.

(责任编辑:王 妮)