

基于运动特征提取和2D卷积的鱼类摄食行为识别研究

张铮¹ 沈彦兵¹ 张泽扬²

(1. 上海海洋大学工程学院, 上海 201306; 2. 上海城市电力发展有限公司, 上海 200123)

摘要: 为了促进渔业装备智能化,近年来基于视频流的鱼类摄食行为识别研究受到了广泛关注。针对基于视频流的传统识别方法模型过于复杂,难以在边缘计算设备部署的问题,提出了一种轻量级的2D卷积运动特征提取网络 Motion-EfficientNetV2,该网络以视频流为输入,能够有效识别鱼类摄食行为。提出的模型以 EfficientNetV2 为主干网络,基于 TEA 和 ECANet 构建了运动特征提取模块 Motion,并将该模块嵌入到 EfficientNetV2 的每个 Fused-MBConv 模块中,使改进后的 EfficientNetV2 具有运动特征提取能力。同时使用 ECANet 对 EfficientNetV2 网络中的 MBConv 进行改进,增强其通道特征提取能力。在此基础上利用空洞卷积扩大感受野,提高大范围特征提取能力。试验结果表明, Motion-EfficientNetV2 的参数量和浮点运算量分别为 9.3×10^6 和 1.31×10^{10} , 优于 EfficientNetV2。在 TSN-ResNet50、TSN-EfficientNetV2、C3D 以及 R3D 模型上进行对比试验,本文模型在降低参数量和浮点运算量的同时,使识别准确率提高到 93.97%。该研究对于渔业装备智能化升级和科学养殖具有推动作用。

关键词: 鱼类摄食行为; 运动特征; 深度学习; 卷积神经网络; 轻量化; EfficientNetV2

中图分类号: S951.2

文献标识码: A

文章编号: 1000-1298(2024)06-0246-08

OSID:



Recognition of Feeding Behavior of Fish Based on Motion Feature Extraction and 2D Convolution

ZHANG Zheng¹ SHEN Yanbing¹ ZHANG Zeyang²

(1. School of Engineering, Shanghai Ocean University, Shanghai 201306, China

2. Shanghai City Electric Power Development Co., Ltd., Shanghai 200123, China)

Abstract: In order to promote the intelligence of fishery equipment, video streaming-based fish feeding behaviour recognition has received extensive attention in recent years. The model of traditional recognition methods based on video streaming is too complex to be realized on edge computing devices. To address this problem, a lightweight 2D convolutional motion feature extraction network, Motion-EfficientNetV2, was proposed which can effectively recognize fish feeding behaviour by using video streams as input. The proposed model used EfficientNetV2 as the backbone network, constructed the motion feature extraction module Motion based on TEA and ECANet, and embedded the Motion module into each Fused-MBConv module of EfficientNetV2, in order to give EfficientNetV2 the ability to extract motion features. The MBConv in the EfficientNetV2 network was also improved by using ECANet to enhance its channel feature extraction capability. Null convolution was used in Motion-EfficientNetV2 to expand the receptive field and improve the wide-range feature extraction capability. The experimental results showed that after introducing the designed Motion module and a series of improvements, the number of parameters and FLOPs of Motion-EfficientNetV2 was 9×10^6 and 1.31×10^{10} , respectively, which were reduced compared with EfficientNetV2. Comparison experiments using the same dataset in the algorithmic models of TSN-ResNet50, TSN-EfficientNetV2, C3D, and R3D, respectively, showed that the present algorithm achieved an accuracy of 93.97% while the number of parameters and FLOPs were lower than the rest of the models. Therefore, the model proposed can effectively identify fish feeding behavior and guide aquaculturists to develop fish feeding strategies.

Key words: feeding behavior of fish; motion feature; deep learning; convolutional neural network; lightweight; EfficientNetV2

收稿日期: 2023-10-13 修回日期: 2024-02-29

基金项目: 上海市崇明区农业科创项目(2021CNKC-05-06)、国家重点研发计划项目(2023YFD2401304)和上海市水产动物良种创制与绿色养殖协同创新中心项目(2021科技02-12)

作者简介: 张铮(1981—),男,副教授,博士,主要从事物联网工程和智能仪器设计研究, E-mail: z-zhang@shou.edu.cn

0 引言

在工厂化循环水养殖系统中,饲料利用率是影响养殖效益的重要因素,过度投喂不仅会造成饲料浪费更会引起水体污染^[1]。因此,合适的投喂策略对于提高大规模水产养殖的效益十分重要。目前对于投喂量的控制极度依赖于养殖工人的经验,很难保证其准确性和实时性,不能满足大规模工厂化养殖的需求^[2]。创建科学合理的数据集,并利用计算机视觉方法对鱼类摄食行为进行识别和分级,可以摆脱个人养殖经验的限制,有利于水产养殖人员及时调整投喂策略,提升水产养殖的效率^[3-4]。

机器视觉在农业领域有广泛的应用场景,包括水产养殖研究中的鱼类计数、物种识别和行为分析等。鱼类摄食行为识别是鱼类行为识别的重要方面^[5],其通过对鱼类摄食行为进行观察和分类来对摄食强度分级。其中群体行为相比于个体行为更能反映鱼类整体的摄食强度,整体摄食强度也更具有现实意义^[6]。文献[7-9]基于图像处理算法或深度学习的方法都只提取了单个图像特征,忽略了鱼类摄食行为是一个动态变化的过程。

大口黑鲈摄食活动激烈,摄食时通常会冲出水面,造成大面积水面波动。但工厂化高密度养殖场景下,识别单个鱼体的运动特征是极其困难的,因此摄食行为的强弱更加难以区分。针对这一问题,本文通过提取整个养殖桶水面的运动特征来反映鱼群摄食活跃程度。运动特征可以捕捉一段时间内鱼类活动的变化情况,当摄食行为较强时,由于循环水的作用,水面会在局部持续产生移动的剧烈水花,并且鱼群的聚集也在随水流移动。此时运动特征的变化范围较大,但整体强度较稳定。当摄食强度为弱时,鱼类摄食行为不连续,水面由于外部或内部环境刺激会间断产生鱼群聚集和波动,这时单帧图像已经不能反映当前时刻鱼类摄食强度,视频流可以更准确地反映此时的摄食强度。

刘世晶等^[10]使用光流法结合循环神经网络(Recurrent neural network, RNN)对摄食行为视频进行分类,黄志涛等^[11]使用背景差分法和光流法提取图像运动特征,这些方法都可以获取到丰富的运动特征,但光流法计算量大,不能保证实时性。WEI等^[12]使用二维卷积神经网络(Two-dimensional convolution neural network, 2DCNN)结合长短时记忆网络(Long short-term memory, LSTM)的方式对鱼类摄食行为进行分级,通过提取时间特征提高了分类准确率。MÅLØY等^[13]使用时空递归神经网络进行水下鲑鱼行为识别,采用卷积和LSTM来提取运动

特征,对鲑鱼行为进行分类。这些方法通常需要引入3D卷积或者LSTM,前者会导致模型复杂度增加,后者会导致模型计算量增加,都不利于鱼类摄食行为识别的工程化^[14-17]。

时间分段网络^[18](Temporal segment networks, TSN)通过光流图像和RGB图像提取视频中时空特征,并提出了时间分段策略,将一个样本按时序进行均匀分段,采集每段随机帧以减少运算量,实现了对人类行为动态过程的识别。但光流图像的复杂性导致模型不能进行实时性检测。时间激励与聚合网络^[19](Temporal excitation and aggregation for action recognition, TEA)模型通过对卷积网络提取的特征做差分运算来获取相邻帧的运动特征,避免了光流图像这一复杂操作,降低了模型复杂度。EfficientNetV2网络是TAN等^[20]提出的轻量化图像分类网络,该网络合理组合了深度可分离卷积和普通卷积,并设计了Fused-MBConv模块,模型训练速度快,参数量少。

鱼类摄食行为是一个动态过程,其识别需要保证运动特征的获取。但过于复杂的预处理和模型结构会导致成本增加,不适用于工厂化水产养殖。针对上述问题,本文提出面向视频流的鱼类摄食行为轻量化识别模型Motion-EfficientNetV2。

1 数据采集与处理

1.1 试验环境与数据采集

为了评估所提出模型的性能,试验以摄食活动强烈的大口黑鲈为目标,进行摄食行为识别试验,模型同样适用于其他有明显摄食活动差异的鱼类。试验地点为上海城市电力发展有限公司水产养殖基地的工厂化养殖车间,试验装置架构如图1所示。单个养殖桶直径为4 m,高1.3 m,水深约90 cm。养殖桶正上方有自动投料机,投料机通过可编程控制器设置投喂量。录像设备为明创达无畸变摄像头,以分辨率1080 p、每秒30帧进行视频录制。大口黑鲈养殖密度约800尾每桶,体长10~15 cm,水中放置曝气盘并输送液氧,以保持合适的溶解氧浓度。

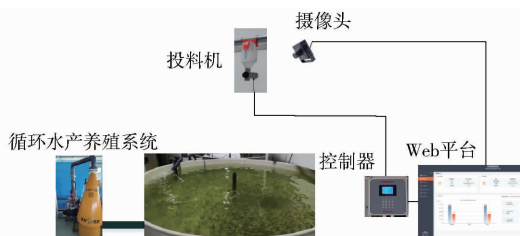


图1 水质监测整体架构图

Fig. 1 Overall framework diagram of water quality monitoring

投喂饲料采用通威饲料有限公司的膨化饲料,饲料的粗蛋白质含量大于48%,粗脂肪含量大于6%,颗粒直径为3 mm左右。饲料投喂时大口黑鲈鱼会浮出水面,便于监控鱼类摄食情况。每天早晚各投料一次,单次投料时长约30 min,视频采集开始于投料前5 min,结束于投料后5 min,以充分记录完整摄食过程。

1.2 数据集制作

首先对采集到的长视频进行处理,按照摄食

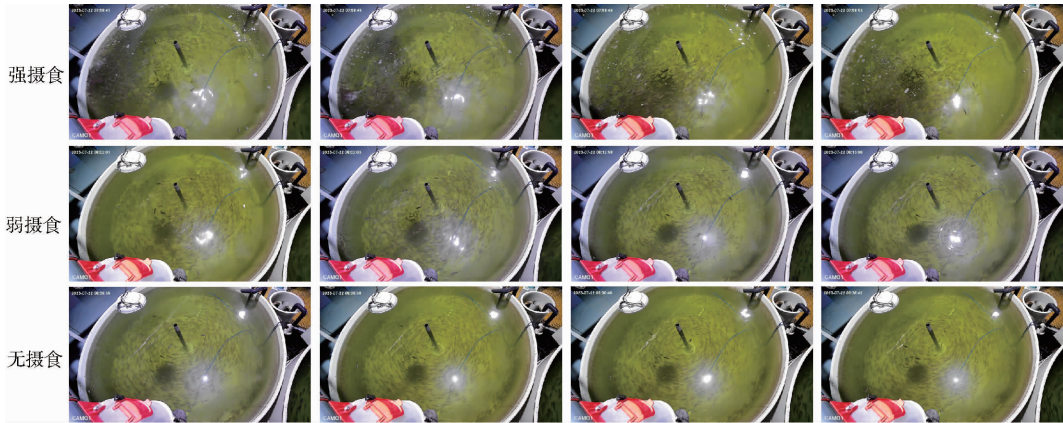


图2 摄食行为视频帧

Fig.2 Frames of feeding behavior videos

通过观察,每个摄食行为类别特征如表1所示,作为划分数据集的依据。对处理后得到的视频片段按比例6:2:2划分为训练集、验证集和测试集,以确保模型泛化能力。鱼类摄食行为视频数据集包括743段视频片段,每个片段时长为5~10 s,数据集具体划分如表2所示。

表1 鱼类摄食行为特点

Tab.1 Characteristics of fish feeding behavior

摄食行为类别	行为特点
无摄食	鱼群均匀游动,水面平静
弱摄食	少部分鱼类快速游动进食,水面激起涟漪,但摄食活动并不连续
强摄食	鱼群持续激烈地进食,水面剧烈波动

表2 摄食行为数据集划分

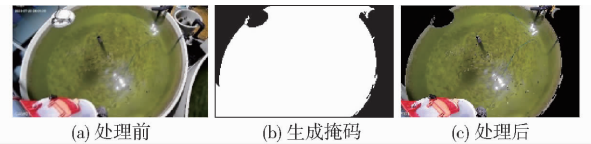
Tab.2 Feeding behavior dataset segmentation

摄食行为	训练集	验证集	测试集	总计
强摄食	144	50	50	244
弱摄食	144	47	47	238
无摄食	144	51	51	246
总计	432	148	148	728

1.3 数据预处理

从图3可以看出,由于养殖桶较大,摄像机拍摄的图像边缘会有外部干扰。本研究使用高斯滤波、开运算对图像进行预处理,利用Canny边缘检测和霍夫变换检测识别养殖桶的曲线,将曲线内的区域

强度进行分类剪辑,再参考TSN中使用的UCF101数据集,将长视频片段剪辑为时长5~10 s的短视频片段。参考徐立鸿等^[21]的方法将鱼类摄食行为分为3个阶段,分别是强摄食、弱摄食和无摄食。每个摄食强度视频帧如图2所示,从连续的多幅图中可以看出,强摄食水面出现大范围的水花,鱼体聚集引起整体色彩变化剧烈,弱摄食水面波动面积较小,且不连续,无摄食水面平静且多帧间无明显变化。



(a) 处理前

(b) 生成掩码

(c) 处理后

图3 图像预处理

Fig.3 Image preprocessing

作为生成一个养殖桶范围的掩码A。将掩码A置0并覆盖到原图中,生成养殖桶外部范围的掩码B。图3b中黑色部分为掩码B,白色部分为掩码A。将掩码B通过或运算覆盖到原图像中,养殖桶外的范围被二值化置0,处理后图像如图3c所示。处理后的图像没有外部干扰,可以更精确地识别鱼类摄食行为。在进行网络训练时,对输入图像进行预处理,并调整图像尺寸为224像素×224像素。

2 基于运动特征提取的分类模型

2.1 网络总体结构

Motion-EfficientNetV2网络结构如图4所示,其总体结构是由多个MF-FusedConv和E-MBConv构成,为了轻量化设计,在保证模型性能的同时,减少了模型层数。将设计的Motion模块插入到Fused-MBConv的第1个卷积层,构建了MF-FusedConv,由于Motion模块已经包含ECE(高效通道激励)模块,Fused-MBConv中的SE模块^[22]被移除,以避免重复处理造成模型复杂度增加。MBConv

中的 SE 模块被替换为 ECA 模块^[23] 构建了 E-MBConv, 降低了模型参数量和 FLOPs (浮点运算量)。本研究参照 TEA 的方式对输入输出进行处理, Motion 模块外所有张量都是 4 维, 即 $(N \times T, C, H, W)$ 。将 Motion 的输入重塑为 5 维张量 (N, T, C, H, W) 。

H, W), 以便在 Motion 内的特定维度上进行操作, 最后将 Motion 的输出张量重塑为 4 维, 输入到下一个 2 维卷积块。 N 表示批量大小, T 表示段数, C 表示通道数, H 表示高度, W 表示宽度, r 表示通道缩减率。

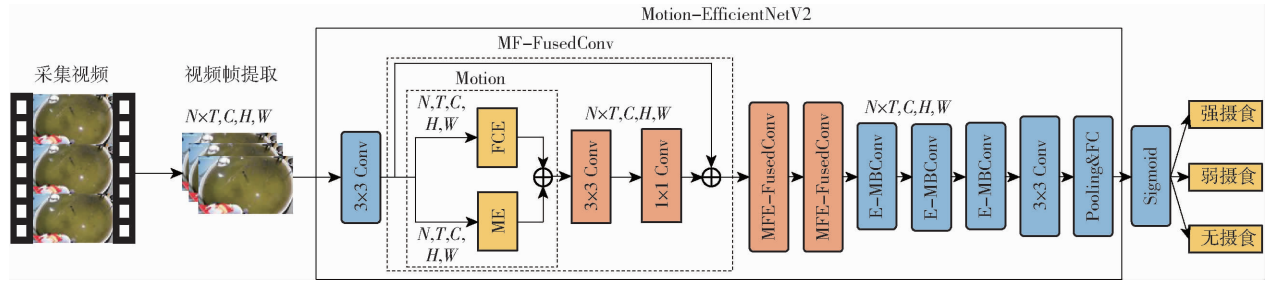


图 4 Motion-EfficientNetV2 网络结构图

Fig. 4 Structure of Motion-EfficientNetV2

从图 3 中可以看出, 鱼群摄食活动范围较大, 摄食行为分类依赖大范围特征, 因此扩大感受野, 提取更丰富的多尺度特征更利于摄食行为的准确识别。如图 5 所示, 本文采用膨胀率为 3 的空洞卷积对 EfficientNetV2 中的 3×3 卷积进行改进, 以在不增加运算量和参数量的情况下扩大感受野。

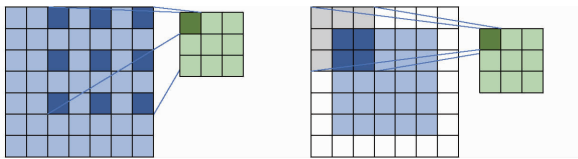


图 5 空洞卷积

Fig. 5 Dilated convolution

与传统 2 维卷积网络不同的是, 本研究构建的网络可以同时处理多幅图像, 并将多幅图像间的特征做差分运算以提取运动特征。在分类结果输出阶段, 每帧图像会有一个包含与相邻帧特征差分的置信度分数, 最终短视频片段的识别结果再由多帧置信度分数加权得到, 视频的识别结果与视频中的每帧图像特征、视频帧间的动作特征紧密相关。

2.2 动作特征提取模块

如图 6 所示, Motion 模块由运动激励 (Motion excitation, ME)^[19] 模块和高效通道激励 (Efficient channel excitation, ECE)^[21] 模块组成。ME 模块能够提取运动特征, ECE 模块能够提取通道特征, ME 和 ECE 模块通过并行连接分别提取运动特征和通道特征, 并将输出拼接后输出给下一个卷积层。

2.2.1 动作激励模块

ME 模块结构如图 7 所示, 通过卷积提取每帧图像的特征, 对图像的特征做差分运算以提取帧与帧之间的变化, 即运动特征。在进行差分运算之前, 对每帧图像的特征都进行变换以减少运算量。使用 2 个 1×1 卷积层, 第 1 个 1×1 卷积层将输入通道

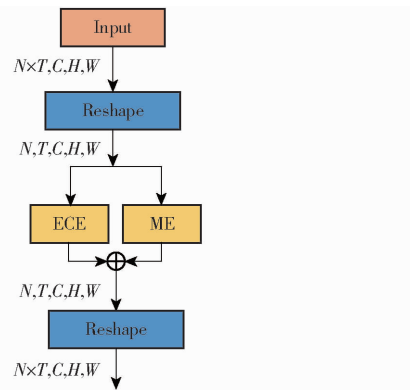


图 6 Motion 模块结构图

Fig. 6 Motion module structure

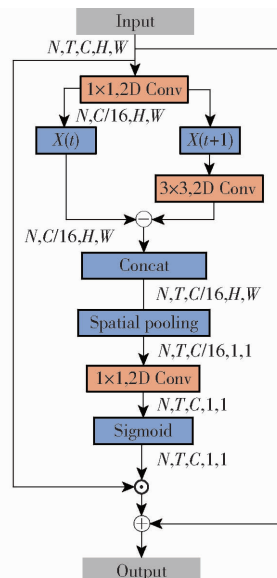


图 7 ME 模块结构图

Fig. 7 ME module structure

数降为原来的 $1/16$, 以减少运算量, 第 2 个 1×1 卷积层将提取的运动特征恢复成与输入相同维度的张量, 便于下一个模块处理。

对两帧进行的特征级差分运算表示为

$$F_m = K * F_r[:, t+1, :, :, :] - F_r[:, t, :, :, :] \quad (1)$$

式中 K —— 3×3 卷积

F_r ——提取的特征图

t ——输入序列

F_m ——两帧图像运动特征

F_m 为得到的能够替代光流的运动特征,此时该张量结构为 $N \times 1 \times \frac{C}{r} \times H \times W$ 。运动特征根据时间维度相互连接,并将 0 填充到最后一个元素,得到多帧运动特征构成张量 F_M , F_M 表达式为

$$F_M = (F_m(1), F_m(2), \dots, F_m(t-1), 0) \quad (2)$$

式中 F_M ——多帧图像运动特征

此时 F_M 为 $N \times T \times \frac{C}{r} \times H \times W$ 张量。通过空间平均池化处理 F_M , 可得

$$F_{temp} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W F_M[:, :, :, i, j] \quad (3)$$

式中 F_{temp} ——运动特征

使用一维卷积对池化处理后的运动特征张量 F_{temp} 进行处理,以恢复张量通道数,可得

$$F_o = K_3 F_{temp} \quad (4)$$

式中 K_3 —— 1×1 卷积 F_o ——输出张量

此时 F_o 为 $N \times T \times C \times 1 \times 1$ 张量。使用 Sigmoid 激活函数对 F_o 处理后得到 M , 其计算公式为

$$M = \delta(F_o) \quad (5)$$

其中 M 为 $N \times T \times C \times 1 \times 1$ 的张量,将输入与 M 作通道级乘法并与输入相加,得到模块最终输出。最终输出为

$$Y = X + X \odot M \quad (6)$$

式中 Y ——输出 X ——输入

2.2.2 高效通道激励模块

ECE 模块是通过 ECANet 得到的通道特征提取模块,其结构如图 8 所示。SENet 在提取通道特征时,通过 2 个全连接层对每个通道进行先升维后降维操作,这会造成部分信息的丢失,增加计算量,影响通道学习效果。ECANet 使用一维卷积实现跨通道交互,避免了 SENet 中的降维操作,并设计了自适应 k 值以减少运算量。图 8 中的 k 表示参与跨通道交互的通道个数。

上述推理通过在网络中加入一个卷积核为 k 的一维卷积实现,同时,为了避免通过交叉验证手动调整 k , k 计算公式为

$$k = \left\lfloor \frac{\text{lb}C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (7)$$

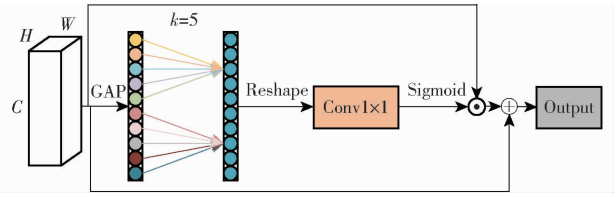


图 8 ECE 模块结构图

Fig. 8 ECE module structure

式中 $\lfloor \cdot \rfloor_{\text{odd}}$ ——绝对值最接近的奇数

b, γ ——比例系数

2.3 多帧同步处理

以 2 维卷积为主体的 Motion - EfficientNetV2 不能够直接读取视频文件进行训练,本研究采用 TSN^[18] 中的方式进行数据读取和 STM^[24] 中的方式进行识别。在进行数据读取时,将每个鱼类摄食行为视频样本分为 T 个小片段,对每个片段随机采样一帧,将采样得到的 T 帧输入到网络中。对于 T 的取值,参考 TSN 模型并进行试验, $T=8$ 时,模型准确率可以达到较高的水平,且计算量最小,性能最优。Motion - EfficientNetV2 的输入为 4 维张量 ($N \times T, C, H, W$),在输入到 Motion 模块时,将输入调整为 N, T, C, H, W 的 5 维张量,每个视频样本采样帧均为一个 4 维张量,每个 4 维张量均被单独输入到 Motion 模块。Motion 模块可以通过每个视频样本采样帧构成的 4 维张量提取视频样本的动作和通道特征,并将输出重塑为 $N \times T, C, H, W$ 的 4 维张量输入到下一卷积层。在视频样本分类时,直接对单帧识别结果进行加权平均得到的结果会存在误差,因此在全连接后再增加一个节点为 T 的全连接层,以获得更准确的权重值。

2.4 评价指标

在深度学习分类任务模型中,通常使用精确率、召回率、F1 值、准确率来评估模型性能。精确率是指预测为正的单次喂食强度的样本中的正确预测比例。召回率是指实际标记为阳性的单次喂食强度样本中的正确预测比例。F1 值为精确率和召回率调和平均值,可以反映模型总体性能^[25]。每秒识别的平均帧数可以评估在同一硬件平台下模型的实时性能^[26]。

3 试验与结果分析

试验硬件平台为英特尔 i7 - 8700 CPU, 32GB 内存,英伟达 RTX 3060Ti 8G 显卡。试验的软件平台为 Windows 10 系统、Python 3.6 解释器和 PyTorch 1.8.0 深度学习框架。

3.1 模型训练

EfficientNetV2 模型含有在公共数据集上训练

的预训练权重,为了提高模型训练速度,在模型训练时,先将 EfficientNetV2 模型在公共数据集上的部分预训练权重导入,再在所采集的数据集上进行训练和测试,模型训练曲线如图 9 所示。由于导入了预训练权重,模型训练曲线在迭代 10 次左右基本趋于稳定,训练准确率在 90% 上下振荡。模型测试准确

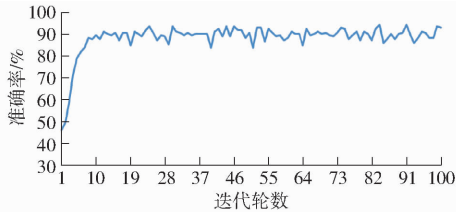


图 9 模型训练曲线

Fig. 9 Model training curve

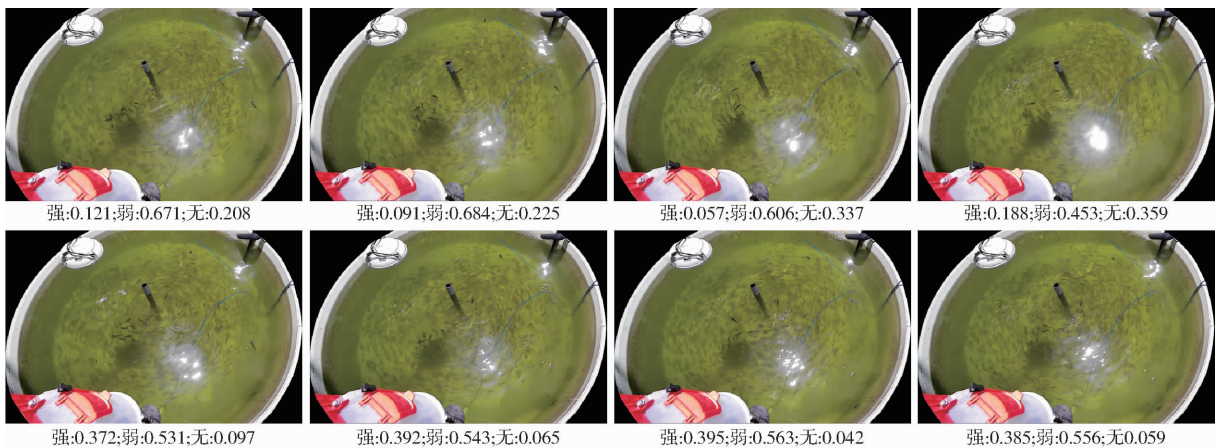


图 10 测试样本置信度分数

Fig. 10 Test sample confidence score

3.2 对比试验结果

使用制作的数据集对 C3D^[27]、R3D^[28]、TSN - ResNet50 和 TSN - EfficientNetV2 网络进行了对比试验,模型内存占用量和测试准确率如表 3 所示。本文提出的 Motion - EfficientNetV2 测试准确率相对于 C3D、R3D、TSN - ResNet50、TSN - EfficientNetV2 模型分别提高 4.19、2.41、15.50、9.69 个百分点,同时模型内存占用量仅相当于 C3D、R3D、TSN - ResNet50、TSN - EfficientNetV2 的 4.6%、10.9%、15.4%、33.5%。与 TSN - ResNet50 和 TSN - EfficientNetV2 相比, Motion - EfficientNetV2 由于其运动特征和更高

表 3 模型参数

Tab. 3 Parameters of models

模型名称	内存占用量/MB	浮点运算量	参数量	准确率/%
Motion - EfficientNetV2	27.5	1.31×10^{10}	9.3×10^6	93.97
C3D	595.0	2.246×10^{10}	7.8×10^7	89.78
R3D	253.0	2.534×10^{10}	3.32×10^7	91.56
TSN - ResNet50	179.0	3.3×10^{10}	2.368×10^7	78.47
TSN - EfficientNetV2	82.2	1.8×10^{10}	2×10^7	84.28

率达到 93.97%,在参数量和浮点运算量较低的同时,保证了较高的识别准确率。本研究同样检测了模型实时性能,在包含图像预处理的同时,模型平均每秒可以处理 25 帧,在分段值取 8 时,每秒可以处理 3 个样本,能够满足应用需求。

为了进一步观察和分析模型的效果,在测试集中随机抽取一个样本进行识别,检测结果如图 10 所示。通过将每帧的置信度分数打印并显示,置信度分数表示识别为该类别的可能性。从该样本的第 4 帧可以看出,即使部分图像帧没有水面波动,弱摄食置信度分数仍大于无摄食类别置信度分数。这说明模型能够获取到视频样本的前后帧变化特征,并在分类过程中作为权重提高了分类准确率。

效的通道特征,模型准确率大大提高。在替换了 SENet 后,模型内存占用量也显著减小。相对 3 维卷积网络而言, Motion - EfficientNetV2 准确率稍高,而其轻量化的特点更有利于实际工程应用。

从表 3 可以看出,模型参数量和浮点运算量的增加并不一定能提高模型性能。合理的网络结构可以提高准确率并降低模型复杂度。基于 2 维卷积的 ResNet50 和 EfficientNetV2 - S 不具有从视频流中提取运动特征的能力,不能准确识别摄食行为的强弱,因此其准确率明显低于 R3D、C3D 和 Motion - EfficientNetV2。TSN - EfficientNetV2 由于结构设计合理,并通过 SE 注意力增强了信道特征,因此其准确率高于 ResNet50。R3D 和 C3D 通过 3 维卷积提取时空特征,使其浮点运算量和参数量高于 MFE - MobileViT 和 TSN - EfficientNetV2,准确率也高于 ResNet50 和 EfficientNetV2 - S。但 R3D 和 C3D 由于通道特征提取的限制影响了识别准确率,导致其低于 Motion - EfficientNetV2。从表 3 也可以看出, EfficientNetV2 - S 中的 SE 模块分别被 Motion 和 ECA 模块替换后, Motion - EfficientNetV2 模型

参数量和浮点运算量大大降低,同时准确率显著提高。

3.3 消融试验

图 11 为所建模型在强、弱和无 3 种摄食强度下的召回率、精确率和 F1 值。从图 11a 中可以看出, Motion - EfficientNetV2 能准确区分鱼类是否摄食,无摄食强度的所有指标都超过 94%。除了弱摄食强度召回率较低外,强摄食和无摄食所有指标都在 85% 以上。Motion 模块通过运动特征,可以显著提高模型对于强弱摄食的识别准确率,结合 ECE 高效通道激励模块,也提高了鱼类是否有摄食行为的识别准确率。总体而言,该模型在检测摄食行为方面表现良好,尤其是在区分鱼类摄食行为强弱方面。如图 11b 所示,针对空洞卷积的有效性进行了消融试验,空洞卷积通过扩大感受野,提升了模型整体性能,尤其是针对强摄食和弱摄食。部分强摄食样

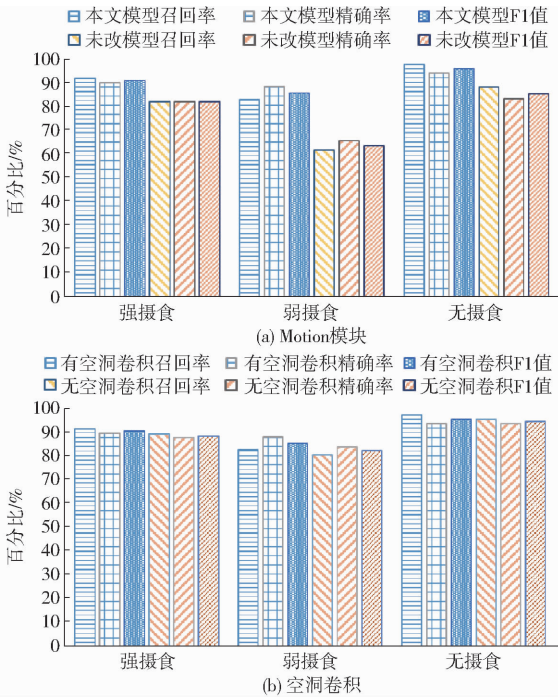


图 11 消融试验评价指标

Fig. 11 Evaluation indicators for ablation experiment

本由于鱼群大面积聚集,本身就具有较为明显的颜色特征,而弱摄食的鱼群聚集范围相对较小,在扩大感受野后,可以增强这种大范围的颜色特征,进而提高模型对鱼类摄食行为识别的性能。

3.4 分段数

为了确定 Motion - EfficientNetV2 模型中样本切割小段的段数 T ,本研究使用不同的 T 和随机帧数比较训练曲线。如图 12 所示,当 T 大于 8 时,准确率基本保持不变,训练曲线的波动减小。主要原因是鱼类摄食行为视频数据的背景相对单一,相邻帧样本之间的相似度相对较高。增加 T 会使训练曲线更加平滑,但模型训练时间也会增加,同时降低了模型实时性能,训练准确率提升不显著。因此,当 T 大于等于 8 时,只提取 1 帧样本,训练曲线的波动范围小,模型识别准确率趋于稳定。

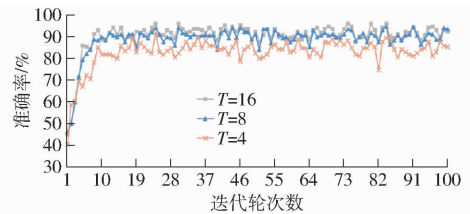


图 12 分段数对模型训练的影响

Fig. 12 Effect of number of segments on model training

4 结论

(1) Motion - EfficientNetV2 模型参数量和浮点运算量仅为 9.3×10^6 和 1.31×10^{10} , 优于 EfficientNetV2 - S, 有利于在边缘设备上部署。

(2) Motion 运动特征提取模块基于 ME 和 ECA 注意力模块设计,能够从视频流中充分提取运动和通道特征,有效增强了对动态鱼类摄食行为的准确率,通过消融试验验证了 Motion 运动特征提取模块的有效性。

(3) 本文模型的大口黑鲈摄食行为系列试验的摄食强度分类准确率达到 93.97%, 优于对比模型。

参 考 文 献

- [1] FØRE M, ALVER M, ALFREDSEN J A, et al. Modelling growth performance and feeding behaviour of Atlantic salmon (*Salmo salar* L.) in commercial-size aquaculture net pens: model details and validation through full-scale experiments [J]. Aquaculture, 2016, 464: 268 - 278.
- [2] HU J, ZHAO D, ZHANG Y, et al. Real-time nondestructive fish behavior detecting in mixed polyculture system using deep-learning and low-cost devices [J]. Expert Systems with Applications, 2021, 178: 115051.
- [3] 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.
- [4] BANERJEE S, ALVEY L, BROWN P, et al. An assistive computer vision tool to automatically detect changes in fish behavior in response to ambient odor [J]. Scientific Reports, 2021, 11(1): 1002.
- [5] LI D, WANG Z, WU S, et al. Automatic recognition methods of fish feeding behavior in aquaculture: a review [J]. Aquaculture, 2020, 528: 735508.
- [6] YANG L, YU H, CHENG Y, et al. A dual attention network based on efficientNet - B2 for short-term fish school feeding

- behavior analysis in aquaculture[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106316.
- [7] 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.
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)
- [8] 陈雨琦,冯德军,桂福坤. 采用机器视觉和傅里叶频谱特征的循环水养殖鱼类摄食状态判别[J]. *农业工程学报*, 2021, 37(14): 155–162.
CHEN Yuqi, FENG Dejun, GUI Fukun. Discrimination of the feeding status of recirculating aquaculture fish via machine vision and reflective corrugated Fourier spectrum[J]. *Transactions of the CSAE*, 2021, 37(14): 155–162. (in Chinese)
- [9] 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: 457–465.
- [10] 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.
LIU Shijing, TU Xueying, QIAN Cheng, et al. Feeding state classification of grass carp based on optical flow and improved RNN[J]. *Acta Hydrobiologica Sinica*, 2022, 46(6): 914–921. (in Chinese)
- [11] 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.
HUANG Zhitao, HE Jia, SONG Xiefa. Recognition and quantification of fish feeding behavior based on motion feature of fishbody and image texture[J]. *Periodical of Ocean University of China*, 2022, 52(1): 32–41. (in Chinese)
- [12] WEI D, BAO E, WEN Y, et al. Behavioral spatial-temporal characteristics-based appetite assessment for fish school in recirculating aquaculture systems[J]. *Aquaculture*, 2021, 545: 737215.
- [13] 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.
- [14] JUN K, LEE D W, LEE K, et al. Feature extraction using an RNN autoencoder for skeleton-based abnormal gait recognition [J]. *IEEE Access*, 2020, 8: 19196–19207.
- [15] SHAHID F, ZAMEER A, MUNEEB M. Predictions for COVID–19 with deep learning models of LSTM, GRU and Bi–LSTM [J]. *Chaos, Solitons & Fractals*, 2020, 140: 110212.
- [16] FENG S, YANG X, LIU Y, et al. Fish feeding intensity quantification using machine vision and a lightweight 3D ResNet–GloRe network[J]. *Aquacultural Engineering*, 2022, 98: 102244.
- [17] 徐访,黄俊,陈权. 基于3D卷积神经网络的动态手势识别模型[J]. *计算机工程*, 2021, 47(11): 283–291.
XU Fang, HUANG Jun, CHEN Quan. Dynamic gesture recognition model based on 3D convolutional neural network [J]. *Computer Engineering*, 2021, 47(11): 283–291. (in Chinese)
- [18] WANG L, XIONG Y, WANG Z, et al. Temporal segment networks: towards good practices for deep action recognition [C] // *European Conference on Computer Vision*. Springer, Cham, 2016: 20–36.
- [19] LI Y, JI B, SHI X, et al. Tea: temporal excitation and aggregation for action recognition [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 909–918.
- [20] TAN M, LE Q. EfficientNetv2: smaller models and faster training [C] // *International Conference on Machine Learning*. PMLR, 2021: 10096–10106.
- [21] 徐立鸿,黄薪,刘世晶. 基于改进LRCN的鱼群摄食强度分类模型[J]. *农业机械学报*, 2022, 53(10): 236–241.
XU Lihong, HUANG Xin, LIU Shijing. Recognition of fish feeding intensity based on improved LRCN [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(10): 236–241. (in Chinese)
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [23] WANG Q, WU B, ZHU P, et al. ECA–Net: efficient channel attention for deep convolutional neural networks [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11534–11542.
- [24] JIANG B, WANG M M, GAN W, et al. STM: spatiotemporal and motion encoding for action recognition [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 2000–2009.
- [25] 孙道宗,丁郑,刘锦源,等. 基于改进SqueezeNet模型的多品种茶叶叶片分类方法[J]. *农业机械学报*, 2023, 54(2): 223–230, 248.
SUN Daozong, DING Zheng, LIU Jinyuan, et al. Classification method of multi-variety tea leaves based on improved SqueezeNet model [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(2): 223–230, 248. (in Chinese)
- [26] 张万枝,曾祥,刘树峰,等. 基于改进YOLO v5s的马铃薯种薯芽眼检测方法[J]. *农业机械学报*, 2023, 54(9): 260–269.
ZHANG Wanzhi, ZENG Xiang, LIU Shufeng, et al. Detection method of potato seed bud eye based on improved YOLO v5s [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(9): 260–269. (in Chinese)
- [27] HARA K, KATAOKA H, SATOH Y. Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and imagenet? [C] // *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018: 6546–6555.
- [28] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 4489–4497.