

基于双流残差卷积神经网络的养殖鳗鲡(*Anguilla*) 摄食强度评估研究*

李 凯^{1,2} 江兴龙^{1,2} 许志扬^{1,2} 林 茜^{1,2}

(1. 集美大学水产学院 福建厦门 361021; 2. 鳗鲡现代产业技术教育部工程研究中心 福建厦门 361021)

摘要 为实现对养殖鳗鲡(*Anguilla*)摄食强度的准确评估,提出了一种基于双流残差卷积神经网络的鳗鲡摄食强度评估方法,该方法针对传统双流网络(Two-stream)中存在的问题做出了相应的改进。首先针对传统双流网络存在网络结构较浅,无法提取到充分的鳗鲡摄食行为特征的问题,选择使用 ResNet50 网络进行替换,以提取到更具代表性的特征。其次针对传统双流网络最后的分类结果是把空间流和时间流的得分取平均值融合而获得,这种方式较为简单,且其空间流和时间流网络为独立进行训练,容易导致网络出现学习不到鳗鲡摄食行为的时空关联特征的问题,选择使用特征层融合方式对空间流和时间流网络提取获得的特征进行融合,让网络能够并行进行训练,以提取到时空信息间的关联特征。试验结果表明:文内提出的基于双流残差卷积神经网络的鳗鲡摄食强度评估方法准确率达到 98.6%,与单通道的空间流和时间流网络相比,准确率分别提升了 5.8%和 8.5%,与传统的双流网络相比准确率也提升了 3.2%。

关键词 鳗鲡;摄食强度;双流残差卷积神经网络;ResNet50;并行训练;特征层融合
中图分类号 Q959.9; S965 doi: 10.11693/hyh20221100291

目前,我国的鱼类养殖过程中饵料的投喂方式主要有两种,一种是人工投喂方式,主要依靠养殖人员自身的养殖经验来控制投喂量,对养殖人员的养殖经验有很高的要求,如果是大规模的养殖基地就会耗费大量的人力成本,大大降低了养殖效益;另外一种为机械投喂方式,它的缺点是忽略了鱼类摄食状态的变化情况,不能根据鱼类的食欲自动调整投喂量,容易导致投喂不当的情况出现,从而影响养殖鱼类的健康生长,降低养殖效益(周应祺等, 2013; 穆春华等, 2015)。而鳗鲡(*Anguilla*)作为我国经济价值较高的优质水产养殖种类,在其养殖过程中,准确检测出鳗鲡的实际摄食状况有助于科学地投喂饲料,对实现鳗鲡精准高效养殖有重要意义。

机器视觉技术作为一种非入侵式、经济且高效的方法,已经成为检测和评估鱼群摄食状态的重要研

究手段(张志强等, 2011; 范良忠等, 2011; 李贤等, 2012; Wishkerman *et al*, 2016)。乔峰等(2015)通过对采集的鱼群摄食图像进行滤波、增强、二值化等图像处理操作,提取出鱼群位置、大小以及聚集程度等摄食行为特征,然后将这些特征输入实时决策的饵料投喂系统,结合鱼群实时的摄食状态进行投喂,但易受室外光线和水面反光等因素的影响,导致系统决策失误。赵建等(2016)提出了一种改进的动能模型来分析鱼群的摄食活动,通过色彩空间转换的方法分割出水面的反光区域,结合光流法和信息熵量化水面反光区域的动能信息,获得了较好的试验结果,但是该方法对光照条件要求较高,在实际养殖环境中难以应用。Sadoul 等(2014)提出了一种通过拍摄的视频来量化循环水养殖条件下鱼群行为的方法,即通过鱼群分散度和鱼群游动活跃度两个指标来量化鱼群

* 国家重点研发计划“特色鱼类精准高效养殖关键技术集成与示范”, 2020YFD0900102 号; 福建省科技厅高校产学研合作项目, 2020N5009 号。李 凯, 硕士研究生, E-mail: 924140205@qq.com

通信作者: 江兴龙, 博士, 教授, E-mail: xinlongjiang@jmu.edu.cn

收稿日期: 2022-11-07, 收修改稿日期: 2023-01-14

行为, 试验发现鱼群在摄食时, 其分散度指标明显减小, 游动活跃度指标明显增大, 这为检测鱼群的摄食状态提供了一种新方法, 但是该试验中鱼群的数量较少。陈彩文等(2017)首先通过背景减法分割出目标鱼群, 接着使用灰度共生矩阵算法对鱼群摄食图像进行处理, 以提取出图像的纹理特征, 最后利用这一特征实现了对鱼群摄食状态的实时检测, 但是该方法仅仅只利用了图像的纹理特征, 较为简单, 难以在环境条件复杂的生产环境中使用。Liu 等(2014)使用改进后的帧间差分法来分析鱼群摄食视频, 将得到的鱼群分散度、相互作用力以及水流场变化幅度这三个鱼群行为特征综合起来, 实现了对鱼群摄食状态的实时评估, 但是该方法要求鱼群和养殖环境之间存在明显色差, 并且检测精度易受水面反光和水花的影响。胡利永等(2015)通过图像滤波、边缘检测以及阈值分割等一系列图像处理技术提取出鱼群摄食图像的面积特征和水花特征, 量化了鱼群的摄食行为, 以此来控制饵料投喂, 但是该方法需要较好的光照条件, 还易受水面波动和反光的影响, 有较大的局限性。

目前, 深度学习技术在鱼类检测方面也得到了应用(Rauf *et al*, 2019; Li *et al*, 2020; Zhang *et al*, 2020), 其中就有研究学者将其用于检测和识别鱼类的摄食状态。刘杨(2021)利用改进后的 YOLOV4 (Bochkovskiy *et al*, 2020)模型对水下残饵进行检测识别, 间接地研究了鱼类的摄食行为, 结果表明模型的检测准确率远高于传统的机器学习方法, 证明了深度学习方法的实用性和优越性。Zhou 等(2019)通过分类模型 LeNet 对鱼类摄食强度进行了分类, 分为无、

弱、中等和强四类, 该模型的分类准确率达到 90%。Måløy 等(2019)使用双流递归卷积神经网络对水下鱼类摄食行为进行了识别, 该网络将鱼类游动过程的空间图像信息和时间序列的运动信息结合起来, 使得网络有更强的鲁棒性(Robust), 从而能更准确地区分水下鱼类的摄食状态和非摄食状态。深度学习模型的网络结构更深, 并且它所提取的特征是神经网络经过大量的学习得到, 能够挖掘出隐藏在图像中的高层语义信息, 所以不易受到养殖鱼类的大小、种类以及养殖环境中的光照不均匀等因素的影响, 相较于传统的机器视觉方法有更强的鲁棒性。综上, 针对传统双流网络(Two-stream)中存在的问题, 本研究提出一种基于残差结构的双流卷积神经网络, 实现了对鳗鲡摄食强度的准确评估。

1 图像采集系统与数据集构建

1.1 图像采集系统

试验数据采集于福建省三明市清流县和南平市延平区的工厂化鳗鲡养殖基地及广东省台山市的土池鳗鲡养殖基地。鳗鲡养殖过程中饵料投喂都是定时定点进行的, 即到了喂食时间养殖池中的鳗鲡会基于条件反射习惯性地聚集于饵料台进行摄食。因此本研究的图像采集系统如图 1 所示, 选择将相机安装于养殖池中饵料台的正上方, 由于工厂化养殖基地的养殖池位于室内, 光线条件较差, 且鳗鲡对于红外光线不敏感, 所以试验相机采用近红外工业相机(MER-232-48NIR), 并同时使用两个红外灯进行补光。此外, 由于土池养殖基地的养殖池位于室外, 光线条件较

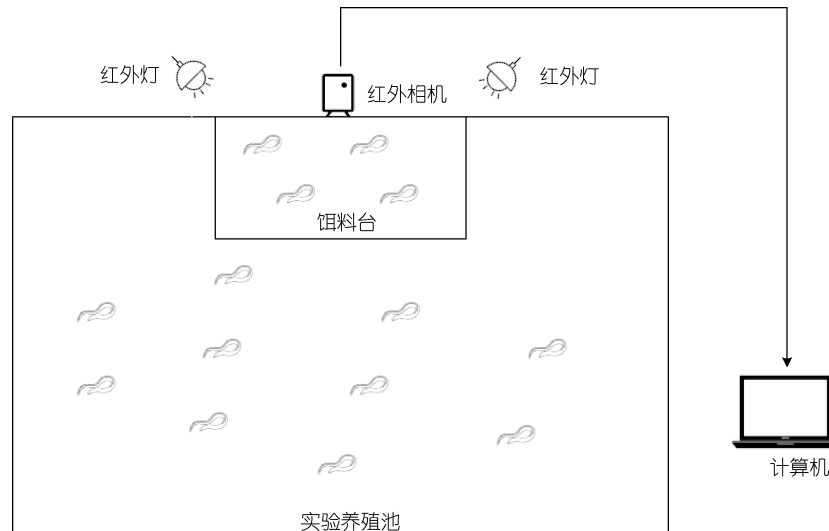


图 1 图像采集系统

Fig.1 The image acquisition system

好, 因此只需使用普通监控相机(海康威视)即可。试验相机均通过一根长 4 m 的千兆双绞线与计算机相连, 其中近红外工业相机采集的是黑白视频, 普通相机采集的为彩色视频, 视频采集帧率均为 24 帧/s。

1.2 数据集构建

试验首先从 13 口鳗鲡养殖池中采集了多个鳗鲡摄食视频, 然后将这些视频进行分解, 分解方式为每秒截取一帧图像, 并且使用 Farneback 稠密光流算法 (Farneback, 2003) 提取对应的鳗鲡摄食行为的光流图像。参考 Overli 等(2006)制定的鱼类摄食强度分类规则, 并通过作者长期的观察后, 发现不同摄食状态下的鳗鲡摄食行为有较大的变化, 综上本研究将鳗鲡的摄食强度分为强、较强、正常、较弱和弱这 5 个等级, 其具体的区分规则如表 1 所示。

试验共选择了 2 000 张图像作为本研究的鳗鲡摄食强度数据集, 其中 1 000 张图像为工厂化养殖池中采集的, 每种摄食强度包含空间位置图及其对应的光流能量图各 200 张, 示例如图 2 所示, 需要说明的是为了避免无关区域对图像检测的影响, 选择将图中除饵料台之外的区域做置黑处理; 另 1 000 张图像为池塘养殖池中采集的, 同样每种摄食强度包含空间位置图及其对应的光流能量图各 200 张, 示例如

图 3 所示。在制作训练集和测试集之前, 使用自编写的 shuffle_data 程序将原数据集中的各类图像打乱, 然后划分训练集和测试集, 划分比例为 7 : 3。

表 1 鳗鲡的摄食强度区分标准

Tab.1 Criteria for distinguishing the eel feeding intensity

级别	行为描述
强	鱼群迅速涌向饲料并出现很激烈的抢食行为
较强	鱼群对饲料有较强反应, 且出现明显的抢食行为
正常	鱼群主动游向饲料, 且出现抢食行为
较弱	鱼群会摄食面前的饲料, 但无抢食行为
弱	鱼群对饲料无明显反应

2 基于双流残差卷积神经网络的鳗鲡摄食强度评估

2.1 双流卷积神经网络

视频信息相较于二维的静态图像而言多了时间维度这一概念, 因此要对视频中物体的行为进行分析就需要融合物体的空间流和时间流信息。随着卷积神经网络的不断发展, 研究学者们通过模仿人脑视觉系统的双通路结构设计了双流卷积神经网络 (Simonyan *et al*, 2014), 其具体的网络结构如图 4 所示。该网络由空

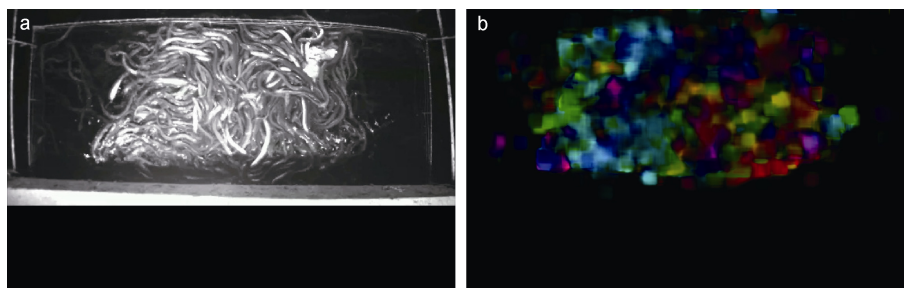


图 2 工厂化养殖池中采集图像
Fig.2 Photos of industrial culture ponds
注: a. 空间位置图; b. 光流能量图

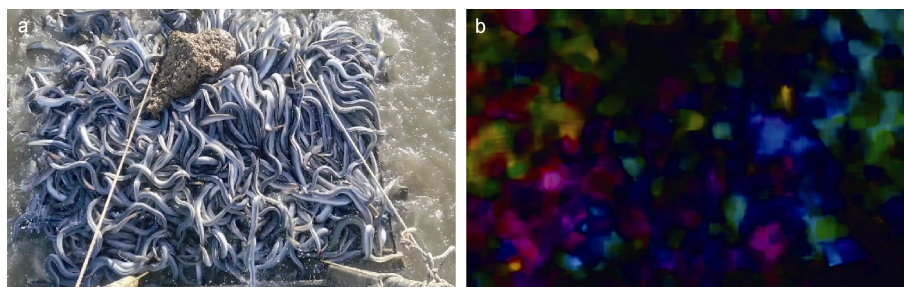


图 3 土池养殖池中采集图像
Fig.3 Photos of soil culture ponds
注: a. 空间位置图; b. 光流能量图

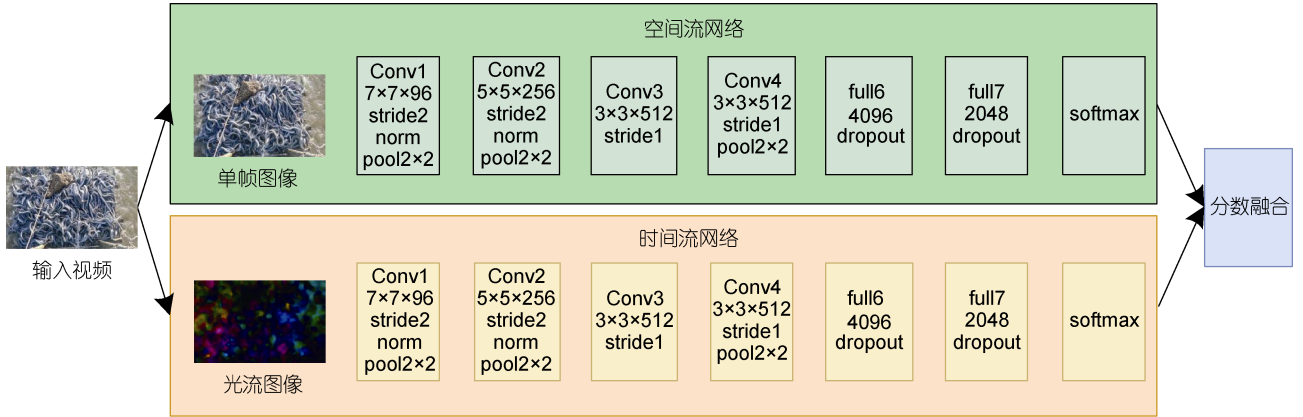


图 4 双流卷积神经网络具体结构

Fig.4 The detail structure of double-flow convolution neural network

间流和时间流这两部分神经网络构成。其中空间流网络将视频中的单帧图像作为输入，用于提取图像中的空间位置特征；而时间流网络则将视频中相邻帧间计算得到的光流图像当作输入，用于提取出视频帧序列中的运动特征。在双流网络的最后利用平均值融合法将两个流的分类结果进行融合，将融合后的分数作为网络的最终分类结果。

2.1.1 空间流网络 空间流网络以视频中的单帧图像作为网络的输入，它通过提取静态帧图像的特征来识别视频中物体的信息。那么对于本研究的鳙鲮摄食强度评估任务而言，空间流网络可以通过提取单帧鳙鲮摄食图像中摄食鱼群的形状、大小以及空间位置等特征来评估视频中鳙鲮的摄食强度。因为鳙鲮摄食行为是一个过程，所以理论上来说摄食视频中的每一帧摄食图像都能作为空间流网络的输入，但是为了使网络能够更好提取出鳙鲮摄食时的特征，在选取摄食图像帧时选择图像较为清晰的那一帧作为网络的输入。空间流网络使用的是 VGG-16 网络 (Simonyan *et al.*, 2015)。

2.1.2 时间流网络 时间流网络的与空间流网络的结构一致，都是使用 VGG-16 作为特征提取网络，不同之处在于时间流网络将视频中相邻帧之间计算得到的光流图像当作输入。光流图像能有效地描述鳙鲮摄食过程中的运动趋势和速度等运动信息，有助于提升鳙鲮摄食强度的评估准确率。

光流法是视频运动目标识别领域中常用的算法，包括稀疏光流法和稠密光流法这两类。其中稠密光流法能对视频帧图像中的所有像素点进行计算，得到的光流位移场也包含了所有运动位移向量，所以这类方法不但计算精度高，而且所呈现出的光流图像

效果也好。因此本研究使用 Farneback 稠密光流算法提取鳙鲮摄食行为的光流图像，下面将对 Farneback 光流算法做详细介绍。

Farneback 光流算法通过估计连续两帧图像间的全局位移的方式解决了图像中存在灰度值快速变化的问题，所以该方法不要求视频中场景空间一定静止的，十分适用于提取物体大尺度运动产生的光流信息。其主要原理在于利用一个扩展多项式近似表示每个像素点的邻域值，如式(1)所示：

$$f(x) = x^T A x + b^T x + c, \quad (1)$$

式中， $f(x)$ 为像素点的邻域值， T 表示矩阵的转置， A 为对称矩阵， b 为向量， c 为标量。假如前一帧图像表示为

$$f_1(x) = x^T A_1 x + b_1^T x + c_1, \quad (2)$$

那么下一帧图像在全局位移 d 后变为

$$\begin{aligned} f_2(x) &= f_1(x-d) = (x-d)^T A_1 (x-d) + b_1^T (x-d) + c_1 \\ &= x^T A_1 x + (b_1 - 2A_1 d)^T x + d^T A_1 d - b_1^T d + c_1 \\ &= x^T A_2 x + b_2^T x + c_2 \end{aligned} \quad (3)$$

利用相邻帧间亮度值恒定不变原理，使得 $f_1(x)$ 和 $f_2(x)$ 对应项系数相等，得到以下等式：

$$\begin{aligned} A_2 &= A_1, \\ b_2 &= b_1 - 2A_1 d, \\ c_2 &= d^T A_1 d - b_1^T d + c_1 \end{aligned} \quad (4)$$

若式中 A_1 为非奇异矩阵，则可解出全局位移量 d ：

$$d = -\frac{1}{2} A_1^{-1} (b_2 - b_1) \quad (5)$$

在计算得到稠密光流场之后，通过使用孟赛尔颜色系统对其进行上色，可将光流场可视化光流图像。

2.2 双流残差网络模型

2.2.1 残差网络 通常情况下, 增加网络的深度可以使网络提取到更高级的图像特征, 进而有效提升网络的性能。然而, 如果只是通过简单地堆叠网络层的方式来加深网络的深度, 不但起不到作用, 反而会使网络的性能退化。研究发现, 增加卷积神经网络的深度之后, 容易在网络反向传播过程中引起梯度消失和爆炸等问题, 导致训练网络无法收敛, 网络中的权重参数得不到优化, 并最终影响网络的性能。

为了消除加深网络深度对网络性能造成的影响, He 等(2016)提出了残差神经网络, 主要原理是将深层网络中的多个网络层拟合为一个非线性残差映射, 通过引入多个这样的结构以进行整体映射关系的拟合。假设非线性网络层的输入为 x , 期望要获得的实际映射函数为 $H(x)$, 而需要逼近的残差函数为 $F(x)$, 则整体映射关系表达式如式(6)所示:

$$H(x) = F(x) + x \quad (6)$$

在实际的残差网络中引入了恒等映射的概念, 具体来说就是通过一种跳跃连接的方式, 直接跳过两层或多层, 将每个网络层输入和输出连接在一起, 这意味着每层网络的输出不再是像以前的神经网络一样是输入的映射, 而是代表输入与映射的相加。残差结构如图 5 所示。

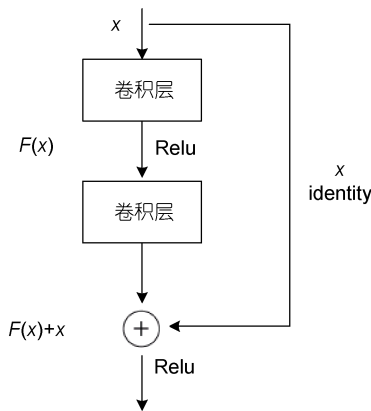


图 5 残差结构示意图

Fig.5 Schematic diagram of residual structure

实际映射关系如式(7)所示, $f(x, \{W_i\})$ 表示残差映射函数, 如式(8)所示。

$$H(x) = f(x, \{W_i\}) + x, \quad (7)$$

$$f(x, \{W_i\}) = W_2 \cdot \delta(W_1 x), \quad (8)$$

式中, δ 为激活函数 Relu, W_1 和 W_2 分别表示两个卷积层(如图 5 所示的一个两层的残差结构, 即包含两个卷积层)的权重参数。此外, 式(6)中的 F 与输入 x 的

维度必须相同, $F+x$ 表示通过跳跃连接将 F 与 x 中的所有像素点进行逐行元素相加。如果出现 F 与 x 维度不相同的情况, 则必须对跳跃连接处进行线性投影 W_s , 其操作如式(9)所示。

$$H(x) = f(x, \{W_i\}) + W_s x \quad (9)$$

在本研究的试验中, 将使用网络结构较深的残差神经网络 ResNet50 来替换原 Two-stream 网络中的 VGG-16 网络。ResNet50 的网络结构如表 2 所示。

表 2 ResNet50 网络结构
Tab.2 The ResNet50 network structure

层名	50 层网络	特征图尺寸
卷积层 1	Conv, 7×7, 64, stride2	112×112
池化层	Max_pool, 3×3, stride2	56×56
卷积层 2	Conv, $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	56×56
卷积层 3	Conv, $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	28×28
卷积层 4	Conv, $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	14×14
卷积层 5	Conv, $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	7×7
分类层	Average_pool, fc-1000, softmax	1×1

在 ResNet50 网络中, 包含了 49 层卷积层, 以及最后分类的 1 层全连接层。从表 2 中可以看出, ResNet50 的卷积层可分为 5 个卷积模块, 即卷积层 1 至卷积层 5。其中卷积层 1 只包含一个 7×7 的卷积层, 在其之后又连接了一个最大池化层; 而后四个卷积层由多个残差结构堆叠而成, 每个残差结构由三层卷积层组成, 卷积核大小分别为 1×1、3×3 以及 1×1, 卷积层 2 至卷积层 5 包含的残差结构数量分别 3 个、4 个、6 个以及 3 个, 各卷积层输出的特征图维度分别为 256 维、512 维、1 024 维以及 2 048 维。在卷积层 5 后面连接了一个平均池化层, 能够加强特征映射和分类类别的一致性, 最后就是分类的全连接层, 其包含的神经元数量与所需分类的类别数一致, 一般为 1 000 个, 而在本研究中为 5 个。

2.2.2 双流残差网络模型设计 在 2.1 节的内容中介绍了双流卷积神经网络, 该网络能够融合鳗鲡摄食行为的空间流信息和时间流信息。其中空间流网络以视频中的单帧鳗鲡摄食图像作为输入, 用于提取图像中摄食鳗鲡的空间分布特征; 时间流网络以

Farneback 稠密光流算法提取鳗鲡摄食行为的光流图像作为输入，用于提取鳗鲡摄食时的运动趋势和速度信息。在本研究中，首先对空间流和时间流使用的卷积神经网络进行改进，即使用 ResNet50 网络代替了网络结构较浅的 VGG-16 网络；然后由于原 Two-stream 的空间流和时间流网络为各自分开训练的，会导致网络出现无法学习到时空信息之间的关联特征的问题，且双流网络最后的分类分数是把空间流和时间流的分数按一定比例融合得到，最常见

的为平均融合法，这种融合方式较为简单，容易导致网络分类结果不准确。因此本研究对 Two-stream 网络的融合方式进行了改进，即使用特征层融合的方式来代替 Two-stream 网络的决策层融合方式。综上，本研究设计了一种基于残差结构的双流卷积神经网络，它的网络结构更深，而且能够让空间流网络和时间流网络并行进行训练，使网络能学习到时空信息的关联特征，提高网络的评估准确率。其具体结构如图 6 所示。

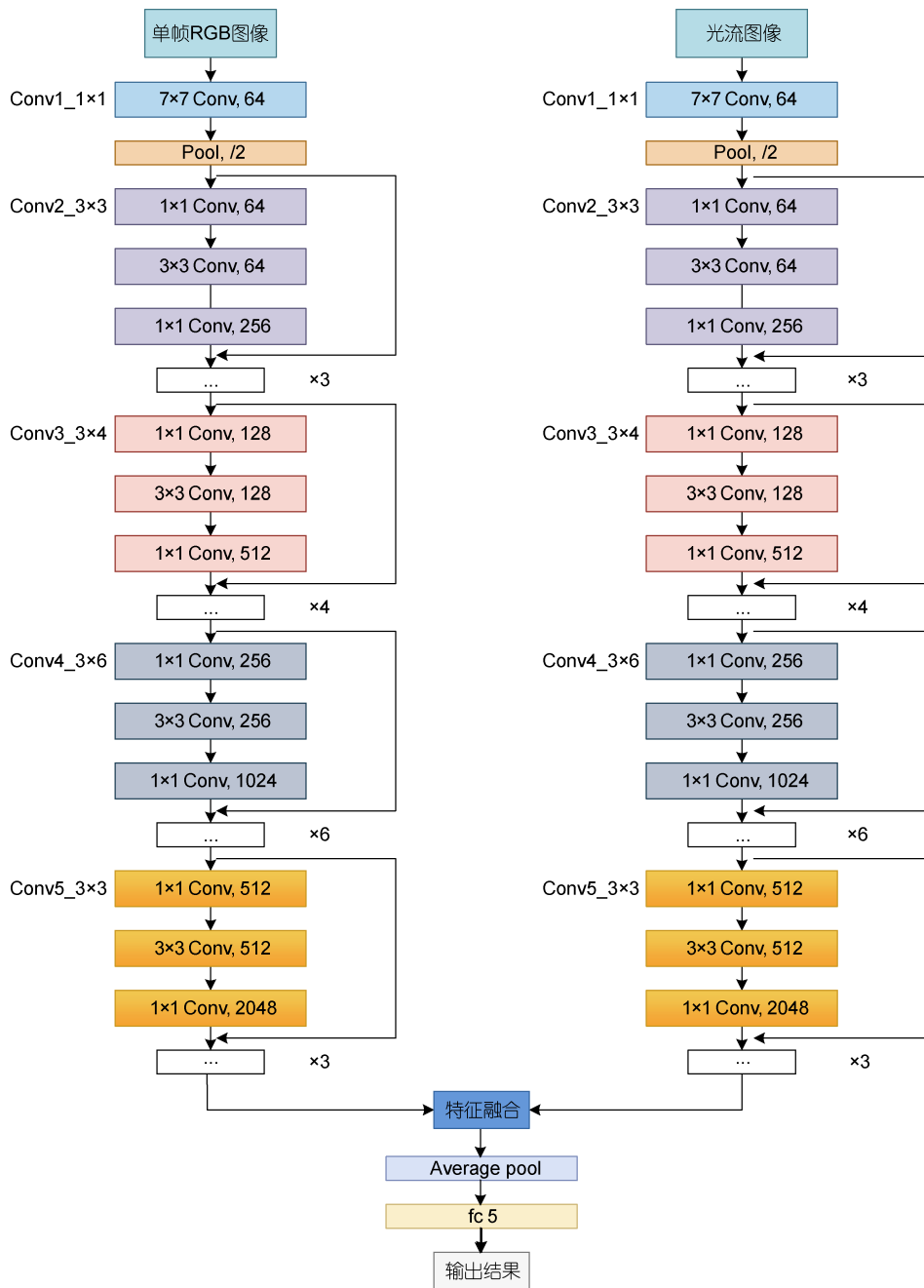


图 6 本文双流残差网络的具体结构

Fig.6 The flowchart of double-flow residual network

3 试验与结果分析

3.1 试验环境与参数设置

本研究试验基于 PyTorch 框架, 使用硬件环境配置为: 处理器为 Intel 酷睿 i9-9900X, 显卡为 NVIDIA RTX2080Ti, 显存为 11 G; 使用软件环境配置为: 系统为 Ubuntu18.04.1, 运行内存为 128 G, Python 版本为 3.8, CUDA 版本为 11.2。

本研究试验相关网络模型参数设置: 模型迭代次数设置为 100 个 epoch; batch_size 设置为 4; 模型优化器选择带动量的 SGD 算法, 其中动量因子参数设置为 0.9; 权重衰减参数设置为 0.000 5; 模型学习率参数初始化为 0.001, 学习率每迭代完一个 epoch 就衰减为原来的 0.33 倍。

3.2 双流残差卷积神经网络训练结果

本研究的双流残差卷积神经网络的融合位置选择在空间流和时间流 ResNet50 的卷积层 5 之后。使用特征层融合的方式进行融合, 具体操作是将两个通道数为 2 048 的特征层在相同的空间位置上进行拼接, 融合得到的特征层通道数为 4 096。

鳊鱼摄食强度数据集共有 2 000 张图像, 其中 70%用作训练, 30%用作测试。在数据集输入网络进行训练之前, 使用经过预训练的空间流和时间流 ResNet50 网络的权重来初始化双流残差网络的权重。在双流残差网络中时间流和空间流网络并行进行训练, 网络共迭代 100 轮。

双流残差网络在训练过程中每迭代完一个 epoch 就记录一下网络损失函数值, 其变化趋势如图 7 所示。从图 7 中可以看出, 训练过程中网络损失在大概第 18 个 epoch 之前下降很快, 而在第 50 个 epoch 之后逐渐趋于稳定, 证明网络模型已经达到收敛。

双流残差网络每训练完一个 epoch 就通过测试集进行评估, 并记录下此时网络的准确率。网络在训练阶段的准确率变化如图 8 所示, 从图 8 中可知, 训练时网络的准确率上升很快, 在第 10 个 epoch 准确率就达到 95%以上, 随着训练的进行准确率缓慢上升, 在第 20 个 epoch 之后逐渐平稳, 最后网络的准确率达到 98.6%。

3.3 模型检测结果比较与分析

为了验证本研究的双流残差网络在鳊鱼摄食强度评估上的有效性, 首先与空间流和时间流网络的评估准确率进行了比较, 如表 3 所示。其次还比较了双流残差网络与 Two-stream 以及双流网络的评估准确率, 如表 4 所示。

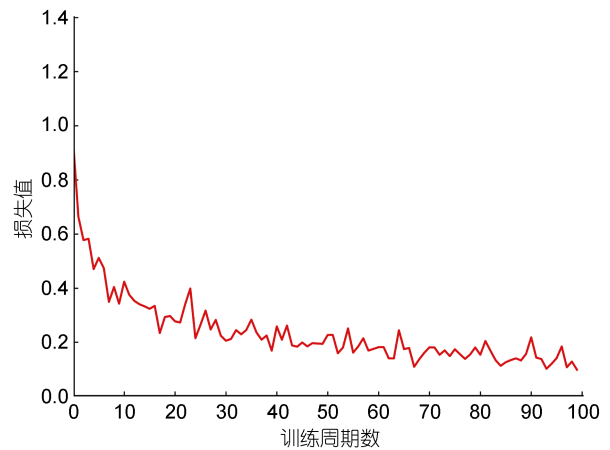


图 7 双流残差网络的训练损失变化

Fig.7 The variation in training loss of the dual-flow residual network

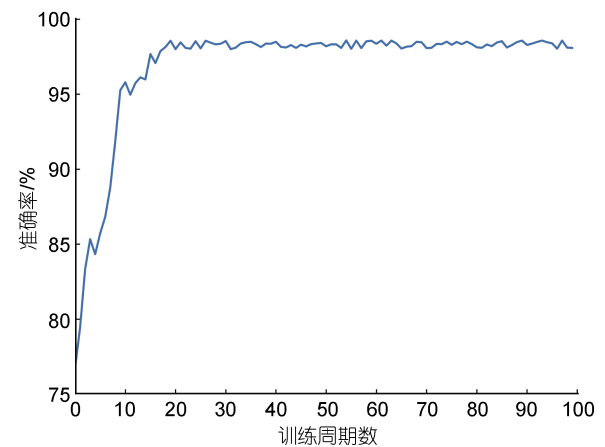


图 8 双流残差网络在测试集上的准确率变化

Fig.8 The variation in the classification accuracy of the two-stream residual network on the test set

表 3 双流残差网络与单流网络准确率比较
Tab.3 Comparison in the classification accuracy between dual-stream residual network and single-stream network

方法	评估准确率/%
空间流网络	92.8
时间流网络	90.1
本文双流残差网络	98.6

注: 表中各方法的分类算法均为 ResNet50

表 4 双流残差网络与其他双流网络准确率比较
Tab.4 Comparison accuracy between dual-stream residual network and other dual-stream networks

方法	分类算法	融合方式	评估准确率/%
Two-stream	VGG-16	平均融合	95.4
双流网络	ResNet50	平均融合	96.2
本文双流残差网络	ResNet50	特征层融合	98.6

注: 平均融合方式是指将空间流和时间流网络的检测结果之和取平均值的方式; 特征层融合方式在 2.2.2 双流残差网络模型设计中有详细介绍

从表 3 中可以看出,空间流网络的评估准确率为 92.8%,时间流网络的评估准确率为 90.1%,前者比后者准确率高了 2.7%,说明在网络结构一致的情况下,空间流网络的表现要优于时间流网络。分析原因可能是由于时间流网络的输入即光流图像容易受到水面反光以及水面波动等因素的影响,导致光流图像对摄食鱼群的运动信息描述不准确,从而影响网络的评估结果。而二者使用特征层融合方式得到的双流残差网络的评估准确率为 98.6%,相比空间流网络和时间流网络分别提升了 5.8%和 8.5%,说明本研究的双流残差网络能有效区分鳗鲡的各类摄食强度。

从表 4 中可以看出,Two-stream 网络的评估准确率为 95.4%,在将 Two-stream 的分类算法 VGG-16 替换为 ResNet50 之后得到双流网络,其评估准确率为 96.2%,相较于 Two-stream 网络准确率有些许提升,提升了 0.8%,说明网络结构较深的 ResNet50 能比 VGG-16 提取到更有效的特征。而本研究的双流残差网络准确率为 98.6%,比 Two-stream 网络和双流网络分别提升了 3.2%和 2.4%,说明在特征层融合的方式能使网络充分学习到鳗鲡摄食行为的时空关联特征,从而使得本研究双流残差网络对鳗鲡摄食强度有较高的评估准确率。

由于本研究的双流残差网络是在原 Two-stream 网络基础之上改进的,所以为了进一步分析本研究改进的双流残差网络对鳗鲡摄食强度评估准确率的影响,本研究还计算了 Two-stream 和双流残差网络对于鳗鲡摄食强度评估的混淆矩阵,如图 9 和图 10 所示。混淆矩阵能够直观地展示出网络对于每类摄食强度的评估准确率,其中横轴表示摄食强度的真实标签,纵轴表示网络预测标签,对角线上的数值为各类摄食强度的评估准确率。由图 9 和图 10 可知,本研究双流残差网络在鳗鲡摄食强度数据集上各类别的评估准确率均有不同程度的提升,其中摄食强度为较强、正常、较弱和弱的准确率分别提升了 5%、4%、2%和 5%,并且本研究双流残差网络在鳗鲡摄食强度为强、正常和较弱的评估准确率均达到 100%,在摄食强度为较强和弱的评估准确率也都在 95%以上。综上所述,与原 Two-stream 网络相比,本研究的双流残差网络能充分利用鳗鲡摄食行为的时空关联信息来实现对鳗鲡摄食强度的准确评估,反映出本研究双流残差网络的改进策略显著有效。

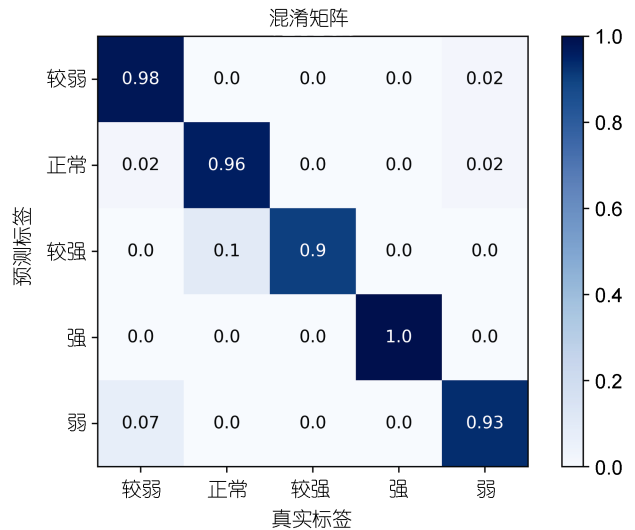


图 9 Two-stream 网络的混淆矩阵

Fig.9 The confusion matrix of the two-stream network

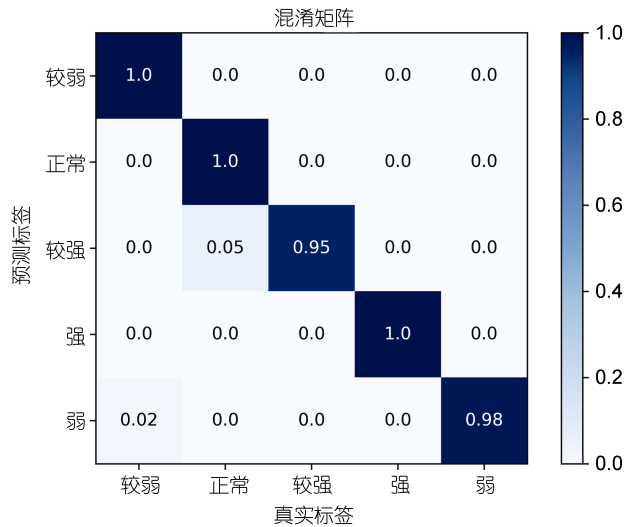


图 10 双流残差网络的混淆矩阵

Fig.10 The confusion matrix of the two-stream residual network

此外,与以往的研究相比本研究方法也有更好的效果。周超等(2019)提出了一种基于近红外机器视觉的方法,将鱼群摄食强度分为弱、一般、中和强 4 类,该方法通过灰度共生矩阵来提取鱼群摄食图像的纹理特征,再利用支撑向量机进行分类,实现了对镜鲤摄食强度的评估,其评估准确率为 87.78%。张佳林等(2020)首先利用变分自动编码器提取水下鱼群图像的特征,然后将得到的特征矩阵输入卷积神经网络进行分类,以区分水下大西洋鲑的摄食行为和非摄食行为,其分类准确率达到 89%。Ubina 等(2021)提出了一种基于三维卷积神经网络的鱼类摄食强度

分类方法, 将鱼类的摄食强度分为无、弱、中和强四类, 该方法的分类准确率为 95%。与以上研究相比, 本方法的效果更好, 评估准确率达到 98.6%, 并且在本研究中将鱼群的摄食强度等级划分更为详细, 分为了强、较强、正常、较弱和弱这 5 个等级。本研究提出的方法对于鱼类摄食强度有着更好的评估效果。

4 结论

本研究将双流残差卷积神经网络应用于养殖鳗鲡的摄食强度评估。首先对传统的 Two-stream 网络进行了介绍, 然后针对该网络存在的弊端提出了改进方法。针对原 Two-stream 网络存在网络结构较浅, 无法提取到充分的鳗鲡摄食行为特征的问题, 本研究使用网络结构较深 ResNet50 网络代替了 VGG-16 网络以提取到更充分、更具代表性的特征; 针对原 Two-stream 的空间流和时间流网络为独立进行训练, 且网络最后的分类结果是把空间流和时间流的得分求和然后取平均值得到, 容易导致网络出现无法学习到时空信息之间的关联特征的问题, 本研究使用特征层融合的方式替代平均值融合的方式, 使网络能够学习到时空信息之间的关联特征, 并能进行并行训练。

试验结果表明, 使用网络结构更深的 ResNet50 之后, 鳗鲡摄食强度评估准确率达到 96.2%, 准确率有小幅提升, 提升了 0.8%; 再使用特征层融合方式之后得到本研究的双流残差网络, 其评估准确率达到 98.6%, 准确率又提升了 2.4%, 并且与原 Two-stream 网络相比, 准确率提升了 3.2%。本研究的成果可以用于指导鳗鲡养殖全过程的高效精准投饲, 提高饲料的利用与转化效率。例如, 可以根据本研究双流残差卷积神经网络的评估结果对鳗鲡饲料的投喂量进行科学精准的调整, 当鳗鲡摄食强度为弱和较弱时, 采取降低投饵量的措施; 当鳗鲡摄食强度为正常时, 采取维持投饵量的措施; 当摄食强度为较强和强时, 采取增加投饵量的措施。综上, 本研究的双流残差网络在鳗鲡摄食强度评估研究上有重要的参考价值。

参 考 文 献

乔峰, 郑堤, 胡利永, 等, 2015. 基于机器视觉实时决策的智能投饵系统研究[J]. 工程设计学报, 22(6): 528-533.
刘杨, 2021. 基于深度学习的水下残饵检测方法研究与实现[D]. 扬州: 扬州大学: 34-37.
李贤, 范良忠, 刘子毅, 等, 2012. 基于计算机视觉的大菱鲆

对背景色选择习性研究[J]. 农业工程学报, 28(10): 189-193.
张志强, 牛智有, 赵思明, 2011. 基于机器视觉技术的淡水鱼品种识别[J]. 农业工程学报, 27(11): 388-392.
张佳林, 徐立鸿, 刘世晶, 2020. 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, 36(13): 158-164.
陈彩文, 杜永贵, 周超, 等, 2017. 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, 33(5): 232-237.
范良忠, 刘鹰, 余心杰, 等, 2011. 基于计算机视觉技术的运动鱼检测算法[J]. 农业工程学报, 27(7): 226-230.
周应祺, 王军, 钱卫国, 等, 2013. 鱼类集群行为的研究进展[J]. 上海海洋大学学报, 22(5): 734-743.
周超, 徐大明, 吝凯, 等, 2019. 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, 1(1): 76-84.
赵建, 朱松明, 叶章颖, 等, 2016. 循环水养殖游泳型鱼类摄食活动强度评估方法研究[J]. 农业机械学报, 47(8): 288-293.
胡利永, 魏玉艳, 郑堤, 等, 2015. 基于机器视觉技术的智能投饵方法研究[J]. 热带海洋学报, 34(4): 90-95.
穆春华, 范良忠, 刘鹰, 2015. 基于计算机视觉的循环水养殖系统残饵识别研究[J]. 渔业现代化, 42(2): 33-37.
BOCHKOVSKIY A, WANG C Y, LIAO H Y M, 2020. YOLOV4: optimal speed and accuracy of object detection [J]. arXiv: 2004.10934.
FARNEBÄCK G, 2003. Two-frame motion estimation based on polynomial expansion [C] // Proceedings of the 13th Scandinavian Conference on Image Analysis. Halmstad, Sweden: Springer: 363-370.
HE K M, ZHANG X Y, REN S Q, *et al*, 2016. Deep residual learning for image recognition [C] // Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 770-778
LI D L, WANG Z H, WU S Y, *et al*, 2020. Automatic recognition methods of fish feeding behavior in aquaculture: a review [J]. Aquaculture, 528: 735508.
LIU Z Y, LI X, FAN L Z, *et al*, 2014. Measuring feeding activity of fish in RAS using computer vision [J]. Aquacultural Engineering, 60: 20-27.
MÅLØY H, AAMODT A, MISIMI E, 2019. A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture [J]. Computers and Electronics in Agriculture, 167: 105087.
ØVERLI Ø, SØRENSEN C, NILSSON G E, 2006. Behavioral indicators of stress-coping style in rainbow trout: do males and females react differently to novelty? [J]. Physiology & Behavior, 87(3): 506-512.
RAUF H T, LALI M I U, ZAHOR S, *et al*, 2019. Visual features based automated identification of fish species using deep convolutional neural networks [J]. Computers and Electronics in Agriculture, 167: 105075.
SADOUL B, MENGUES P E, FRIGGENS N C, *et al*, 2014. A new method for measuring group behaviours of fish shoals from recorded videos taken in near aquaculture conditions [J]. Aquaculture, 430: 179-187.

- SIMONYAN K, ZISSERMAN A, 2014. Two-stream convolutional networks for action recognition in videos [C] // Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Quebec, Canada: MIT Press: 568-576.
- SIMONYAN K, ZISSERMAN A, 2015. Very deep convolutional networks for large-scale image recognition [C] // Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA: ICLR.
- UBINA N, CHENG S C, CHANG C C, *et al*, 2021. Evaluating fish feeding intensity in aquaculture with convolutional neural networks [J]. *Aquacultural Engineering*, 94: 102178.
- WISHKERMAN A, BOGLINO A, DARIAS M J, *et al*, 2016. Image analysis-based classification of pigmentation patterns in fish: a case study of pseudo-albinism in Senegalese sole [J]. *Aquaculture*, 464: 303-308.
- ZHANG S, YANG X T, WANG Y Z, *et al*, 2020. Automatic fish population counting by machine vision and a hybrid deep neural network model [J]. *Animals*, 10(2): 364.
- ZHOU C, XU D M, CHEN L, *et al*, 2019. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision [J]. *Aquaculture*, 507: 457-465.

EVALUATION ON FEEDING INTENSITY OF AQUACULTURE EEL (*ANGUILLA*) BY DOUBLE-FLOW RESIDUAL CONVOLUTION NEURAL NETWORK

LI Kai^{1,2}, JIANG Xing-Long^{1,2}, XU Zhi-Yang^{1,2}, LIN Qian^{1,2}

(1. Fisheries College, Jimei University, Xiamen 361021, China; 2. Engineering Research Center of the Modern Technology for Eel Industry, Ministry of Education, Xiamen, 361021, China)

Abstract To accurately evaluate the feeding intensity in eel (*Anguilla*) culture, the eel intensity evaluation method based on double-flow residual convolution neural network was proposed, by which the problems existing in traditional double-flow network (Two-stream) was solved. The traditional two-flow network is shallow in network structure and not able to extract sufficient eel feeding behavior information. Therefore, ResNet50 network was chosen to extract more representative features. The final classification score of the traditional double-flow network could be obtained by combining the average scores of spatial flow and time flow, and the fusion method was relatively simple, and the spatial flow and time flow network were trained independently, which could lead to an issue that the network cannot learn the spatio-temporal correlation characteristics of eel feeding behavior. We chose to use the feature layer fusion method to fuse the features extracted from the spatial flow and time flow network, by which the network was trained in parallel to extract the correlation features of the spatio-temporal network. Results show that the classification accuracy of the eel feeding intensity evaluation in double-flow residual convolution neural network reached 98.6%, which was 5.8% and 8.5% higher than that of single-channel spatial flow and time flow network, respectively. Compared with the traditional double-flow network, the classification accuracy was improved by 3.2%.

Key words eel; feeding intensity; double-flow residual convolution neural network; ResNet50; parallel training; feature layer fusion