

基于深度学习的循环水养殖鳗鲡(*Anguilla*) 行为状态识别的研究*

许志扬^{1,2} 江兴龙^{1,2} 林茜^{1,2} 李凯^{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 对鱼类的行为进行智能监测, 精准地量化与识别其健康状态, 已成为研究热点。为实现养殖鳗鲡行为状态精准识别, 提出一种基于 DenseNet 双流卷积神经网络的鳗鲡行为状态检测方法。利用混合高斯背景模型进行前景提取构建数据集, 针对传统卷积神经网络对于时间动态信息提取能力有限的问题, 搭建关联空间特征与时间特征的双流网络结构(Two-stream), 并使用 DenseNet-121 网络替换原网络, 对比 VGGNet、ResNet 等网络, 通过密集连接实现特征重用, 在搭建更深的网络结构基础上加强了运动特征传递并减少了参数量, 更好地提取具有代表性的行为特征。传统双流网络在两端的 softmax 层仅作简单的决策层平均融合, 无法更深程度关联时空高级特征, 提出在网络卷积层提取空间特征与时间特征后, 加上一层卷积层将时空特征进行卷积融合以提升模型识别精度。实验结果表明: 文中提出的基于 DenseNet 双流卷积神经网络对 6 种鳗鲡行为状态检测方法准确率达到 96.8%, 相较于单通道的空间流与时间流网络, 准确率分别提升了 10.1%和 9.5%; 相较于以 VGGNet、ResNet 搭建的双流网络, 准确率分别提升了 12.4%和 4.2%; 与决策层平均融合、特征层拼接融合的方式相比, 时空特征卷积融合方式准确率分别提升了 2.5%和 1.7%。

关键词 鳗鲡; 行为状态; 混合高斯背景模型; DenseNet-121; 双流网络结构; 卷积融合

中图分类号 Q959.9; S965 doi: 10.11693/hyhz20230300048

我国是世界上最大的鳗鲡生产国, 养殖产量稳居世界首位, 鳗鲡因其具有较高的营养价值, 是我国优良水产养殖种类之一(李岑鹏等, 2009)。目前, 在鳗鲡养殖过程中, 对鱼类养殖状态的监测方法多为人工观察法, 但是人工观察存在效率低、工作量大, 以及判断易出错等缺点。近年来, 随着科学技术的发展和人工智能技术的不断革新, 计算机视觉技术已经应用在许多领域里, 为了适应水产养殖业向智能化方向发展的需求, 在水产养殖行业里利用计算机视觉技术进行鱼类行为智能监测也逐渐成为一个研究热点(Niu *et al.*, 2018; 郑书康, 2020)。通过计算机视觉技术对其进行非入侵式的行为状态监测, 在不影响正常养殖的前提下, 高效精准地量化与识别鳗鲡行

为, 监测养殖鱼类健康状态, 利用鱼类行为表现判断水体环境是否优良, 对于鳗鲡高效养殖具有重要的现实意义。

国内外学者利用计算机视觉技术, 对水产养殖对象的行为进行特征提取、量化、分析的研究日渐加深。Kato 等(1996)开发了以金鱼为研究对象的计算机图像处理系统并进行了模拟分析, 首次对鱼群目标进行群体行为检测, 但该研究只统计金鱼的直行、左转、右转等简单特征作为行为分析依据, 无法充分地反映鱼类行为状态。Sadoul 等(2014)采集虹鳟鱼的图像数据, 并对其进行灰度化和滤波操作, 通过背景模型获取虹鳟鱼群的周长, 以此来计算分散程度并描述其游泳行为, 但该研究的行为分类特征选择较为

* 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。许志扬, 硕士研究生, E-mail: 1227550569@qq.com

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

收稿日期: 2023-03-02, 收修改稿日期: 2023-05-08

简单。范良忠等(2011)在分析比较计算机视觉领域常用的目标检测算法上,将帧间差分、背景建模和光流估计等方法相互结合提出了3种改进算法,提升了鱼体运动的检测精度,但该方法的鲁棒性与泛化能力较差。近年来,深度学习发展迅速,相较于传统图像处理算法,深度学习网络不再使用人为定义特征的方法,而是直接将图像信息送进神经网络自动提取其高维抽象语义特征,进一步挖掘图像数据的时空信息。沈军宇等(2018)结合计算机视觉技术与深度学习方法,提出了一种基于YOLO算法的端到端鱼群检测方法,对于实时视频的鱼群检测精度能够达到90%以上,但该方法对鱼群行为识别能力较弱。李星辉(2021)通过训练目标检测模型Faster RCNN对运动鱼类进行检测,提出轨迹关联方法提取轨迹数据并对其异常轨迹判别,实现水下鱼类异常行为在线监测,但该方法在鱼类密集区域的监测效果较差。

综上,将深度学习技术应用在水产养殖业已成为今后智能化水产养殖的重要研究趋势。本文针对养殖鳗鲡行为监测,提出一种以DenseNet双流卷积神经网络搭建的深度学习改良模型,实现对养殖鳗鲡行为状态的精准识别。由于鳗鲡习性喜暗怕光(龚盼,2015),一经普通光线照射会产生躲避、逃窜等应激行为,但红外光照射对鳗鲡行为无明显干扰和影响(李凯等,2022),因此在拍摄鳗鲡行为图像时选择使用红外相机与红外补光灯进行图像采集。本研究为鳗鲡精准化高效养殖提供可量化的数据参考,有助于及时发现和处理养殖过程中存在的问题,研究成果可为鳗鲡乃至鱼类的精准高效绿色养殖提供有效技术支撑和应用参考。

1 图像采集系统与构建鳗鲡行为状态数据集

1.1 图像采集系统

鳗鲡行为图像数据采集自集美大学水产试验场的鳗鲡循环水养殖车间。如图1所示,将红外监控摄

像机(海康威视 Q/BFW 066)安装于养殖桶正上方1.5 m,并在养殖桶上方周围使用红外灯进行补光。以倾斜的角度设置红外灯避免灯光直射桶底,降低水面反射红外灯光造成的图像噪声。相机通过一根5 m长的千兆网线与计算机相连接,使用H.264编码的方式存储图像数据于计算机硬盘中。采集视频图像信息时让车间处于完全黑暗环境,减少外界光线变化带来的影响。视频采集的帧率均为20 帧/s,位深度8 bits,分辨率2 560×1 440 (16 : 9)。

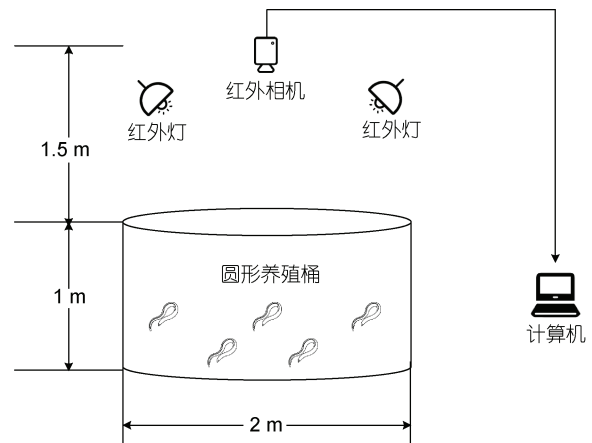


图1 图像采集系统

Fig.1 The image acquisition system

1.2 构建鳗鲡行为状态数据集

实验用鱼为取自集美大学水产试验场自行养殖的澳洲鳗鲡,每组实验选取大小均匀、无病无伤的鳗鲡各10尾。在实验开始前,提前一周将鳗鲡移入养殖桶内适应环境,避免产生应激反应。

通过长时间的实验观察及筛选,参考谢富华(2020)对于鱼群行为状态样本数据库的建立,本文确定了6种具有明显分辨特征的行为状态作为本文的研究对象,分别为正常状态、饥饿状态、缺氧状态、高水温状态、摄食状态、受惊扰状态,见表1。其中,摄食状态及受惊扰状态可直接进行图像采集,取10

表1 鳗鲡行为状态分类表
Tab.1 Classification of eel behavior states

行为状态	行为描述
正常状态	个体在池底游动或静卧,游速正常,群体分布集中于池底边缘
饥饿状态	个体游速较快,鱼群多数在水体中上层游动,偶有少数鱼将头探出水面
缺氧状态	个体游速较慢,出现浮头现象,鱼群多分布于水体上层
高水温状态	个体游速快,鱼鳃呼吸急促,反应过激,体色变淡,鱼群分布无规律
摄食状态	个体游速较快,撕咬饲料团,鱼体旋转翻滚,鱼群集中于饲料附近
受惊扰状态	个体游速多变,先迅速躲避到距惊扰源较远的地方,并静卧观察

条体质量(274.4±8.5) g 的鳊鱼, 用摄像头记录正常状态下摄食饵料的行为作为摄食状态数据集, 并记录通过环境刺激使鳊鱼受惊扰的行为作为受惊扰状态数据集。但采集饥饿状态、缺氧状态、高水温状态的行为图像时, 需确保鳊鱼已处于异常状态中, 因此先设置不同饥饿天数、不同溶解氧浓度、不同水温的梯度实验探究 3 种异常状态与正常状态可区分的临界点, 并与正常状态可区分的异常状态图像作为本实验的训练数据。

将采集视频进行逐帧分解。相机采集的原始分辨率较高, 为加快运算需将其标准化处理, 固定每张图像尺寸为 576×360。相机直接采集的空间位置图(见图 2a)描述鳊鱼运动时序信息能力较弱, 为加强图像行

为特征, 在图 2a 基础上利用混合高斯背景模型提取运动目标, 去除相对不变的背景, 所得前景提取图(图 2b)作为网络模型的空间流图像数据输入。在图 2a 的基础上利用 Farneback 稠密光流算法(Farneback, 2003)提取鳊鱼行为的光流图像, 用光流能量图(图 2c)描述鳊鱼的运动时序信息, 并作为网络模型的时间流图像数据输入。

本实验共选择 1 800 张图像作为 6 种鳊鱼行为状态数据集, 每种状态选取前景提取图和光流能量图各 300 张。将数据集图片的 70%划分为训练集, 30%划分为测试集, 并使用 DataLoader 程序里的 shuffle 将数据集里的图片打乱使其呈现随机分布, 更好地加强模型的泛化能力。



图 2 图像样本示例

Fig.2 Example of the image sample
注: a. 空间位置图; b. 前景提取图; c. 光流能量图

2 基于 DenseNet 双流卷积神经网络的鳊鱼行为状态检测

2.1 DenseNet 网络模型

Huang 等(2017)提出了 DenseNet 网络。DenseNet 网络建立在 ResNet 的基础上(He *et al*, 2016), 基本网络模型由密集连接块(Dense Block)和过渡层(Transition Layer)组成, 采用所有层互相连接的密集连接机制, 见图 3, 实现了特征重用和更强的梯度流动性, 最大

程度保留低维度特征和提升梯度的反向传播, 在参数量和计算量更少的情况下获得比 VGGNet、ResNet 更优的性能。其结构如式(1)所示:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \tag{1}$$

式中, x_l 为 DenseNet 网络在 l 层的输出, H_l 为非线性转化函数(non-linear transformation), 它包括了 BN (Batch Normalization)、ReLU、Conv 等特征提取操作, 相较于 ResNet 采用的元素级相加操作(addition), DenseNet 采用的是特征通道级连接操作(concatenation)。

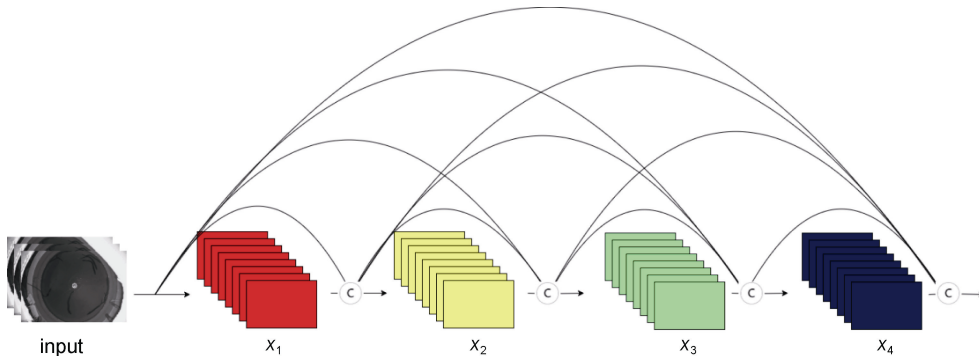


图 3 DenseNet 网络的密集连接机制

Fig.3 The dense connection mechanism of the DenseNet network

在本文实验中采用的是 DenseNet-121 网络, 网络结构由 3 个 Transition Layer 模块连接 4 个 Dense

Block 模块, 增长率 k (Growth rate) 设置为 32。DenseNet-121 网络结构如表 2 所示。

表 2 DenseNet-121 网络结构
Tab.2 The structure of the DenseNet-121 network

层名	121 层网络	特征图尺寸
卷积层	Conv, 7×7, 64, stride2	112×112
池化层	Max pool, 3×3, stride2	56×56
密集连接块 1	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 3 \times 3, \text{Conv} \end{bmatrix} \times 6$	56×56
过渡层 1	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 2 \times 2, \text{Average pool, stride2} \end{bmatrix}$	28×28
密集连接块 2	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 3 \times 3, \text{Conv} \end{bmatrix} \times 12$	28×28
过渡层 2	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 2 \times 2, \text{Average pool, stride2} \end{bmatrix}$	14×14
密集连接块 3	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 3 \times 3, \text{Conv} \end{bmatrix} \times 24$	14×14
过渡层 3	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 2 \times 2, \text{Average pool, stride2} \end{bmatrix}$	7×7
密集连接块 4	$\begin{bmatrix} 1 \times 1, \text{Conv} \\ 3 \times 3, \text{Conv} \end{bmatrix} \times 16$	7×7
分类层	1000D Fully-connected, softmax	1×1

DenseNet 网络结构中每个 Dense Block 模块内部使用 Bottleneck 来降低网络参数量, Bottleneck 具体结构为 BN + ReLU + 1×1 Conv + BN + ReLU + 3×3 Conv。整个网络一共为 121 层, 4 个密集连接块中的 Bottleneck 层分别有 6 层、12 层、24 层以及 16 层, 输出的特征图维度分别为 256 维、512 维、1 024 维、1 024 维。

2.2 DenseNet 双流卷积神经网络

2.2.1 双流卷积网络

传统单帧卷积神经网络 (Single Frame CNN) 对图像信息的空间信息有较好的获取能力, 但无法获取时间信息, 即使利用多帧图像并使用多种不同融合方式如 Late Fusion、Early Fusion、Slow Fusion 融合视频帧进行改进以获取图片上的时间信息, 其获取的能力也有限, 最佳融合方式 Slow Fusion 在 UCF101 数据集上预测准确率达到 60.9% (杨戈等, 2022)。双流卷积网络是 Simonyan 等 (2014) 提出的关联时空特征信息的网络模型, 其网络结构由空间流网络与时间流网络构成, 见图 4。对于给定的行为视频来说, 图像信息被双流卷积网络分成空间信息与时间信息两部分。空间信息以单帧 RGB 图像作为输入, 利用卷积神经网络提取空间位置特征信息, 最后通过全连接层给每个视频帧标注上行为标签。时间信息则以多张相邻帧计算得到的光流图

像作为输入, 同样经过卷积层与全连接层, 对每个光流图像的动作特征进行标注。最后两个网络都要各自经过 Softmax 层, 输出向量进行平均融合, 确定最终分类的结果。双流卷积神经网络在 UCF101 数据集上预测准确率达到了 88.0%。

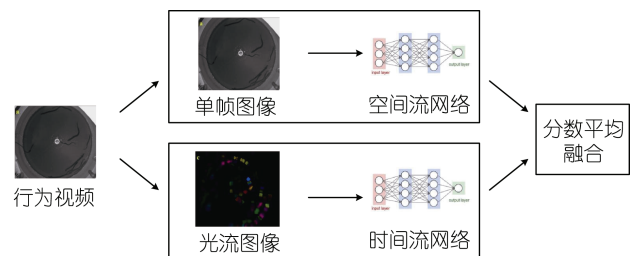


图 4 双流卷积神经网络结构

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

2.2.2 DenseNet 双流卷积神经网络模型设计

上文提到双流卷积神经网络是将行为视频分解成空间图像和光流图像进行输入。对于本文所研究的鳗鲡行为分类任务而言, 关注的重点是鳗鲡本身, 在双流网络中直接输入视频单帧图像和对应的光流图像可能会存在背景噪声干扰, 因此选择利用混合高斯背景模型去除背景的干扰, 仅保留有意义的鳗鲡运动目标, 提取鳗鲡的运动信息。下面对混合高斯背景模型

进行介绍。

混合高斯背景模型是基于像素样本统计信息的背景表示方法,对于各像素点的处理都是独立的。图像中的每个像素值短时间内在一定的距离里满足围绕某一中心值分布,如果数据点足够多的话认为其满足高斯分布,如果像素值偏离中心值较远,则认为其属于前景;如果像素值偏离中心值较近,则认为其属于背景。利用像素在较长时间内大量样本值的概率密度等统计信息,使用统计差分进行目标像素判断,对复杂动态背景进行建模。

将图片的每个像素值由 k 个高斯模型叠加的混合高斯背景模型来建模,通常 k 的取值为 3~5。即:

$$P(x_j) = \sum_{i=1}^k \omega_{j,t}^i \times \eta(x_j, \mu_{j,t}^i, \tau_{j,t}^i) \quad (2)$$

式中, $\omega_{j,t}^i$ 为 t 时刻的第 i 个高斯分布的权系数的估计值; x_j 为像素 j 在 t 时刻的值; $\mu_{j,t}^i$ 和 $\tau_{j,t}^i$ 分别表示在 t 时刻中第 i 个高斯分布的均值向量和协方差矩阵; η 为高斯分布概率密度函数,即:

$$\eta(x_j, \mu_{j,t}^i, \tau_{j,t}^i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\tau_{j,t}^i|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2}(x_j - \mu_{j,t}^i)^T \tau_{j,t}^i^{-1} (x_j - \mu_{j,t}^i)\right] \quad (3)$$

每个新像素值 x_t 和 k 个高斯模型比较,找到匹配 x_t 的高斯分布模型,即和该模型的偏差在 2.5σ 内。如果匹配的高斯模型符合要求,则认为该像素是背景,否则认为该像素属于前景。本文实验 k 的取值为 3。

将经过混合高斯背景模型处理的鳗鲡行为单帧空间图像与光流图像作为双流卷积神经网络的输入,并对双流卷积神经网络进行改进。原双流卷积神经网络所使用的神经网络模型为 VGG-16,网络深度为 16 层,为加深网络层数以获取更深层的图像目标运动特征,使用 DenseNet-121 替换神经网络,将网络层数加深至 121 层。原模型使用的是决策层融合,时空网络分别在 Softmax 层得出分数后再进行平均分数融合。为更大程度地关联时空特征,采用卷积融合的特征融合方式,在网络卷积层提取空间特征与时间特征后,再加上一层卷积层将时空特征进行卷积融合以提升模型识别精度。本文所使用的 DenseNet 双流卷积神经网络模型结构如图 5 所示。

3 实验与结果分析

3.1 环境框架与超参数设置

本实验使用的是 Pytorch 框架,以 Python 为开发

语言。深度学习运算平台的硬件配置信息为:处理器为 Inter i9-12900KF,显卡为 NVIDIA GeForce RTX 3090Ti (24GB);使用的软件环境配置为:系统为 Window 11,运行内存为 32GB,python 版本为 3.8,CUDA 版本为 11.3。

在 DenseNet 双流卷积网络模型中的超参数设置如下:批尺寸(Batch size)为 16;迭代轮次(Epoch)为 100;模型优化器选择带动量的随机梯度下降算法 SGD,动量因子(momentum)设置为 0.9;学习率(learning rate)为 0.001;权重衰减参数(weight_decay)设置为 0.000 4。

为加快并优化 DenseNet 双流卷积网络模型的学习效率,采用迁移学习(transfer learning)的方式,选择使用 ImageNet 数据库进行预训练处理,得到预训练网络模型之后再使用自制的鳗鲡行为状态数据集进行训练和优化。

3.2 三种状态下的鳗鲡异常行为区分点

上文提到针对鳗鲡状态分类确立 6 种具有较为明显的分辨特征的行为状态作为研究对象,但在饥饿、缺氧、高水温这 3 种状态中,如何选定合适的饥饿天数、溶氧量、水温,使鳗鲡的行为状态处于异常,是一项需要预先进行的实验。实验环境设置参考杨小东等(2019),将不同饥饿天数、不同溶解氧浓度、不同水温各自设置 7 个梯度,每个梯度的鳗鲡行为状态都与正常状态的图像数据一起送入 DenseNet 双流卷积神经网络进行二分类识别,以二分类网络识别准确率达 90%以上的状态为行为异常区分点,并以此设定为本文研究所需的异常状态训练数据。

每个梯度使用 200 张空间图像与光流图像用作训练与测试;使用 DenseNet-121 的权重参数初始化 DenseNet 双流卷积神经网络的权重参数;网络迭代轮次设置为 50 轮,每迭代 1 轮记录训练过程中的网络损失函数值与测试集的准确率,选取网络模型最优状态。

3.2.1 饥饿状态异常行为区分 将体质量(275.6±6.2) g 的 10 条成鳗放入养殖桶内暂养一周消除鳗鲡的应激反应对实验结果的影响,保持溶氧 6.5~7.5 mg/L, pH 7.0±0.2,控制水温 25 °C,正常喂食。一周后停止喂食并使用摄像头记录鳗鲡活动情况,每隔 7 d 记录一次,记录总次数为 7 次,将图像数据送入 DenseNet 双流卷积神经网络模型中与正常喂食状态下的鳗鲡活动图像数据作二分类识别,识别结果如表 3 所示。在第 7 天时记录的图像数据与正常状态

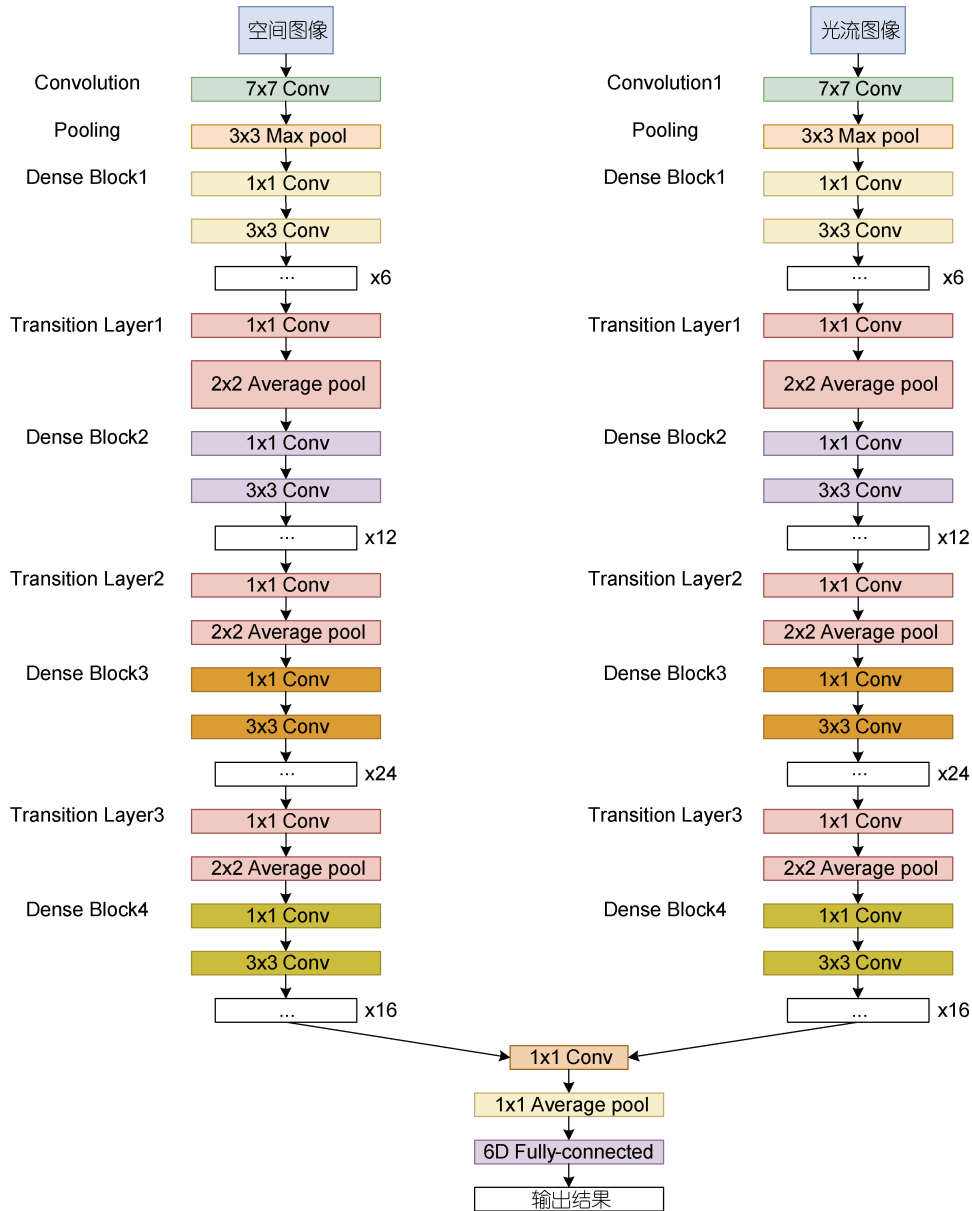


图 5 DenseNet 双流卷积神经网络结构

Fig.5 The structure of the DenseNet double-flow convolution neural network

区别不大，识别准确率低，训练集预测准确率剧烈震荡，网络无法收敛；第 14 天记录的图像数据送进网络模型中发现虽然开始收敛，但是识别准确率仅为 68.1%，说明与正常状态下的行为区分度依然较低；第 35 天的记录数据异常检测准确率首次突破 90%，达到 93.5%，说明将鳗鲡停食 35 d 后其活动行为与正常状态下可通过 DenseNet 双流卷积神经网络模型作出区分，因此选择停食 35 d 后的鳗鲡活动状态作为本文研究的饥饿异常状态。

3.2.2 缺氧状态异常行为区分 将体质量为 (291.2±9.2) g 的 10 条成鳗放进养殖桶内暂养一周消

表 3 不同饥饿天数对鳗鲡饥饿状态识别准确率的影响结果

Tab.3 Effect of the length of starvation (number of days) on the accuracy of recognition of the hungry state of eel

时间/d	检测准确率/%
0	无法收敛
7	无法收敛
14	68.1
21	83.6
28	87.7
35	93.5
42	92.8

除鳗鲡的应激反应,保持溶氧 6.3~7.3 mg/L, pH 7.0±0.2, 控制水温 25 °C, 正常喂食。一周后停止曝气, 使用聚酯薄膜(PET)将养殖桶与摄像头一起密封, 确保桶内环境与空气隔绝, 使用溶氧仪(哈纳 HI98193)和摄像头记录水的溶解氧浓度与鳗鲡活动情况。起始溶解氧含量为 6.9 mg/L, 后待溶氧仪显示溶解氧浓度分别为 6.0、5.0、4.0、3.0、2.0、1.0 mg/L 时用摄像头记录此时鳗鲡活动状态。所得图像数据送入 DenseNet 双流卷积神经网络模型中与正常状态下的鳗鲡活动图像数据作二分类识别, 识别结果如表 4 所示。当溶解氧浓度降至 5.0 mg/L 时, 其行为状态与正常状态的行为基本无区分, 网络不收敛; 当溶解氧浓度下降至 4.0 mg/L 时, 网络开始收敛, 模型检测准确率为 69.4%; 当溶解氧浓度下降至 2.0 mg/L 时, 鳗鲡的行为状态与正常状态具有明显区分, 识别准确率达到 99.6%; 而当溶解氧浓度下降至 1.0 mg/L 时, 识别准确率反而下降至 88.7%, 分析其原因是当溶解氧浓度为 1.0 mg/L 时, 已经接近澳洲鳗鲡成鳗窒息点 0.90 mg/L (魏金生等, 2021), 此时已有 3 条鳗鲡出现死亡而无行为活动, 影响了网络模型的识别效果。因此选择 2.0 mg/L 的溶解氧浓度下的鳗鲡行为状态作为本文研究的缺氧异常状态。

表 4 不同溶解氧浓度对鳗鲡缺氧状态识别准确率的影响结果

Tab.4 Effect of dissolved oxygen concentration on the accuracy of recognition of anoxic state of eel

溶解氧浓度/(mg/L)	检测准确率/%
6.9	无法收敛
6.0	无法收敛
5.0	无法收敛
4.0	69.4
3.0	86.6
2.0	99.6
1.0	88.7

3.2.3 高水温状态异常行为区分 将体质量 (246.9±5.3) g 的 10 条成鳗放入养殖桶内暂养一周, 保持溶氧 6.4~7.4 mg/L, pH 7.0±0.2, 正常喂食。一周后测量初始水温为 27 °C, 用摄像头记录鳗鲡活动状态。此后利用加热棒将水温每上升 1 °C 后保持目标水温 4 h, 并用摄像头记录鳗鲡行为活动。后将所有图像数据送入 DenseNet 双流卷积神经网络模型中与正常状态下的鳗鲡活动图像数据作二分类识别, 识别结果如表 5 所示。水温对鳗鲡活动行为的影响较大,

当水温从初始温度 27 °C 提高 1~28 °C 时, 网络模型识别准确率为 75.4%; 当温度达到 31 °C 时网络模型识别准确率达到 98.7%, 表明鳗鲡在这个温度下行为状态与正常温度下具有明显区分, 因此选择水温 31 °C 下的鳗鲡行为状态作为本文研究的高水温异常状态。

表 5 不同水温对鳗鲡高水温状态识别准确率的影响结果
Tab.5 Effects of water temperature on the accuracy of recognition of eel under high water temperature

水温/°C	检测准确率/%
27	无法收敛
28	75.4
29	84.6
30	89.1
31	98.7
32	98.9
33	98.2

3.3 DenseNet 双流卷积神经网络状态识别训练结果

本文使用的鳗鲡行为状态数据集一共有 1 800 张图像, 将其 70% 作为训练集, 30% 作为测试集; 设置网络迭代轮次(epoch) 100 轮; 使用 DenseNet-121 的权重参数初始化 DenseNet 双流卷积神经网络的权重参数; 在空间流网络与时间流网络进行并行训练, 每迭代 1 轮记录训练过程中的网络损失函数值(Loss)与测试集的准确率(Accuracy); 双流融合方式为卷积融合, 即将空间流网络与时间流网络各自通道数为 2 048 的特征层, 使用 1×1 卷积核的卷积层将其融合为通道数 1 024 的特征层, 再通过全连接层与需要进行分类的 6 种状态进行分类。

模型训练过程中网络损失函数值与测试集准确率的变化趋势如图 6 和图 7 所示。网络损失函数值在前 20 个 epoch 下降较快, 此后放缓下降速度, 在第 45 个 epoch 趋于稳定, 说明此时模型达到收敛。模型的测试准确率在前 30 个 epoch 上升较快, 在第 55 个 epoch 时趋于稳定, 准确率达到 95% 以上, 此后模型经过多轮训练, 测试准确率缓慢上升, 最终稳定于 96.8%。

3.4 不同网络模型检测结果比较和分析

将 DenseNet 双流卷积神经网络的检测准确率与单空间流网络与单时间流网络的检测准确率作对比, 验证双流网络结构相比单流网络结构在鳗鲡行为状态识别检测的优势, 其结果如表 6 所示。

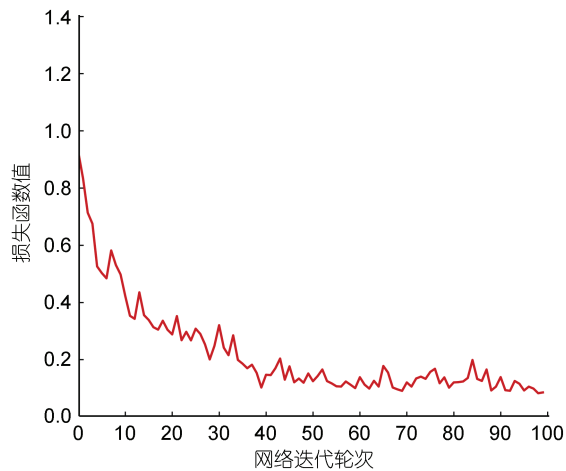


图 6 DenseNet 双流卷积网络训练损失变化图
Fig.6 Variation in the training loss of the DenseNet double-flow convolutional network

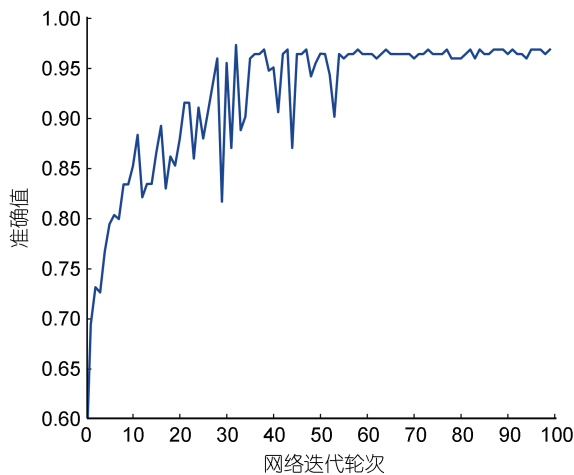


图 7 DenseNet 双流卷积网络在测试集上准确率变化图
Fig.7 Variation in the accuracy of the DenseNet double-flow convolutional network

表 6 DenseNet 双流卷积网络与单流网络准确率比较
Tab.6 Comparison in recognition accuracy between the DenseNet double-flow convolutional network and single-stream network

网络模型	检测准确率/%
DenseNet 空间流网络	86.7
DenseNet 时间流网络	87.3
本文 DenseNet 双流卷积网络	96.8

本文所使用的双流卷积网络模型为 DenseNet 网络, 如果采用其他网络如 VGGNet、ResNet 的双流卷积网络, 并使用不同策略的双流融合方式, 其模型检测准确率与本文的模型作对比, 其结果如表 7 所示。

从表 6 可以看出, 在模型结构相同的情况下, 使用 DenseNet 网络的空间流网络模型在鳗鲡行为状态

识别检测准确率为 86.7%, 而时间流网络的检测准确率为 87.3%, 两者的检测准确率相近, 后者要比前者稍高, 证明了在行为识别检测上, 即使作为输入的光流图像更容易受到光照变化、水面扰动、水面反光等情况的干扰, 但是依然比空间位置图像更能描述鳗鲡的行为运动信息, 包含了更多的运动特征。使用本文的 DenseNet 双流卷积网络在特征层上通过卷积融合的方式将双流各自的特征信息进行融合, 检测准确率达到 96.8%, 对比空间流网络与时间流网络分别提升了 10.1%和 9.5%, 证明了双流网络相比单流网络对于运动特征信息的提取具有更好的效果, 有效区分了鳗鲡在不同状态下的行为模式。

表 7 DenseNet 双流卷积网络与其他双流网络准确率比较

Tab.7 Comparison in recognition accuracy between the DenseNet double-flow convolutional network and other two-stream networks

双流网络分类算法	融合方式	检测准确率/%
VGG-16	决策层平均融合	84.4
ResNet-50	决策层平均融合	92.6
DenseNet-121	决策层平均融合	94.3
DenseNet-121	特征层拼融合	95.1
DenseNet-121	特征层卷积融合	96.8

从表 7 可以看出, 在传统双流网络中使用 VGG-16 的检测准确率为 84.4%, 将 VGG-16 替换成网络层数更深的 ResNet-50, 检测准确率提升为 92.6%; 之后再将 ResNet-50 替换成 DenseNet-121, 检测准确率提升为 94.3%, 说明网络层数加深有利于提取更深层的图像特征。而后在 DenseNet 双流卷积神经网络中使用不同的双流融合方式, 从决策层平均融合替换成特征层拼融合, 检测准确率提升为 95.1%, 说明特征层融合更能使网络学习图像的时空关联特征; 替换成特征层卷积融合, 检测准确率提升为 96.8%, 在特征层进行卷积融合的方式相比拼融合效果更佳。

而相较于以往对鱼类行为识别的研究方法, 本文方法也具有更好的效果。Zhou 等(2017)用背景差分法提取鱼体, 通过狄罗尼三角剖分计算鱼群的群聚指数来描述鱼群的活跃程度, 识别准确率达到 94.5%。黄志涛等(2022)以大西洋鲑为研究对象, 利用背景差分法、光流法、灰度共生矩阵提出一种基于鱼体运动特征和图片纹理特征鱼群活动强度量化方法, 其检测准确率达到 94.17%。李济泽等(2022)将鱼群活动行为图像通过归一化、PCA 降维和支持向量机

的方法提出一种基于机器视觉的摄食行为识别方法,其平均准确率为 92.3%。与以上研究相比,本文的研究方法利用深度学习这一工具,深入挖掘鳗鲡活动图像的时空特征信息,且将研究范围进一步扩大到 6 种不同的行为状态,检测准确率为 96.8%。

4 结论

本文使用了 DenseNet 双流卷积神经网络对鳗鲡的 6 种行为状态进行识别分类。首先通过摄像机采集不同状态下鳗鲡行为视频,并通过混合高斯背景模型和光流算法对视频数据进行处理得到双流图像数据集。针对传统卷积神经网络对于时间信息获取能力有限的问题提出以 DenseNet 网络进行改进的双流卷积神经网络,采用所有层互相连接的密集连接机制实现了特征重用和更强的梯度流动性,网络层数更深,性能更优;使用并行训练,在特征层进行卷积融合的方式关联时空信息特征。使用 DenseNet 双流卷积神经网络对饥饿状态、缺氧状态、高水温状态进行正常状态与异常状态二分类识别,确定 3 种状态的异常行为区分点,并将其作为本文识别任务中的所需异常状态。最后对于 6 种行为状态使用 DenseNet 双流卷积神经网络进行识别分类并与其他网络模型的检测结果进行比较和分析。

实验结果表明,使用本文的 DenseNet 双流卷积神经网络对鳗鲡 6 种行为状态进行识别,准确率达到了 96.8%;相较于使用 VGG-16 的传统双流网络准确率提升了 12.4%;相较于使用 ResNet-50 的双流卷积神经网络准确率提升了 4.2%;相较于决策层平均融合、特征层拼接融合方式的 DenseNet 双流卷积神经网络,特征层卷积融合方式准确率分别提升了 2.5%和 1.7%。综上,本文的 DenseNet 双流卷积神经网络在鳗鲡行为状态识别研究具有重要的参考价值。

参 考 文 献

- 李岑鹏, 关瑞章, 江兴龙, 等, 2009. 循环水处理系统处理鳗鲡养殖污水的应用实验[J]. 集美大学学报(自然科学版), 14(2): 126-130.
- 李凯, 江兴龙, 陈尔康, 等, 2022. 基于深度学习的循环水养殖鳗鲡(*Anguilla*)计数研究[J]. 海洋与湖沼, 53(3): 664-674.
- 李星辉, 2021. 基于轨迹提取的鱼类异常行为监测[D]. 杭州: 浙江大学: 11-35.
- 李济泽, 位威, 张凯凯, 2022. 基于机器视觉的养殖鱼摄食行为识别方法[J]. 福建工程学院学报, 20(4): 378-382.
- 杨小东, 江兴龙, 乐普敏, 等, 2019. 南美洲鳗鲡(*Anguilla rostrata*)的耗氧率(ROC)、窒息点(AP)和适温范围(RT)及对非离子氨(NIA)、NO₂⁻的 LC₅₀和 SC 的研究[J]. 海洋与湖沼, 50(2): 455-464.
- 杨戈, 邹武星, 2022. 基于深度学习的视频行为分类方法综述[J]. 电子技术应用, 48(7): 1-7, 12.
- 沈军宇, 李林燕, 夏振平, 等, 2018. 一种基于 YOLO 算法的鱼群检测方法[J]. 中国体视学与图像分析, 23(2): 174-180.
- 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.
- 郑书康, 2020. 基于计算机视觉的鱼类行为学规律与精准投喂的研究[D]. 杭州: 浙江大学: 3-5.
- 黄志涛, 何佳, 宋协法, 2022. 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, 52(1): 32-41.
- 龚盼, 2015. 日本鳗鲡幼体对环境因子的选择行为研究[D]. 上海: 上海海洋大学: 35-36.
- 谢富华, 2020. 基于深度学习的鱼群行为状态分类研究[D]. 天津: 天津理工大学: 22-23.
- 魏金生, 江兴龙, 王泽旭, 等, 2021. 澳洲鳗鲡(*Anguilla australis*)不同生长阶段的生物学耐受性特征及其演变趋势[J]. 海洋与湖沼, 52(4): 983-993.
- FARNEBÄCK G, 2003. Two-frame motion estimation based on polynomial expansion [C] // Proceedings of the 13th Scandinavian Conference on Image Analysis. Berlin, Heidelberg: Springer: 363-370.
- HE K, ZHANG X, REN S, *et al*, 2016. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.
- HUANG G, LIU Z, VAN DER MAATEN L, *et al*, 2017. Densely connected convolutional networks [C] // Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE.
- KATO S, TAMADA K, SHIMADA Y, *et al*, 1996. A quantification of goldfish behavior by an image processing system [J]. Behavioural Brain Research, 80(1/2): 51-55.
- NIU B S, LI G Y, PENG F, *et al*, 2018. Survey of fish behavior analysis by computer vision [J]. Journal of Aquaculture Research & Development, 9(5): 1000534.
- SADOUL B, EVOUNA MENGUES P, 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: MIT Press: 568-576.
- ZHOU C, ZHANG B H, LIN K, *et al*, 2017. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture [J]. Computers and Electronics in Agriculture, 135: 233-241.

RECOGNITION OF BEHAVIORAL STATE OF EEL *ANGUILLA* IN RECIRCULATING AQUACULTURE SYSTEM VIA DEEP LEARNING

XU Zhi-Yang^{1,2}, JIANG Xing-Long^{1,2}, LIN Qian^{1,2}, LI Kai^{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 In order to achieve accurate identification of behavioral states of cultured eel *Anguilla*, a recognition of behavioral state of eel in recirculating aquaculture system was proposed based on the DenseNet double-flow convolutional neural network. The Gaussian Mixture Model was used for foreground extraction to construct the dataset. Aiming at the problem in traditional convolutional neural network whose ability to extract time dynamic information is limited, a double-flow network structure associated with spatial and temporal features was built, in which the DenseNet-121 network was used to replace the original network. Compared to VGGNet, ResNet, and other networks, the proposed system can reuse the characteristic attributes via dense connection by building a deeper network structure, by which the motion feature transmission can be strengthened and the number of parameters reduced, thus achieving better extraction of representative behavioral features. The softmax layer at both ends of a traditional two-stream network makes simple average fusion of decision layers only, which cannot correlate spatio-temporal high-level features to a deeper extent. Therefore, after the network convolution layer extracts the spatial and temporal features, a convolution layer is added to the convolutional fusion of spatio-temporal features to improve the recognition accuracy. Results show that the accuracy of the proposed system for the behavior state detection of six eels reached 96.8%, which is 10.1% and 9.5% greater than that of the single-channel spatial stream and temporal stream networks. Compared with the two-stream network built with VGGNet and ResNet, the accuracy was increased by 12.4% and 4.2%, respectively. Compared with the average fusion method of decision layer and the concatenation fusion method of feature layer, the accuracy of the spatio-temporal feature convolution fusion was increased by 2.5% and 1.7%, respectively.

Key words eel *Anguilla*; behavioral state; Gaussian Mixture Model; DenseNet-121; double-flow network structure; convolutional fusion