

基于光流注意力网络的梅花鹿攻击行为自动识别方法

高云^{1,2} 侯鹏飞¹ 熊家军³ 许学林³ 陈斌¹ 李康¹

(1. 华中农业大学工学院, 武汉 430070; 2. 生猪健康养殖协同创新中心, 武汉 430070;

3. 华中农业大学动物科技学院, 武汉 430070)

摘要: 人工养殖的雄性梅花鹿在发情期间攻击行为剧增,易造成鹿茸损伤,自动监测其攻击行为能为研究减少攻击行为提供重要依据。本文基于注意力机制和长短记忆序列研究了一种光流注意力网络(Optical flow attention attacking recognition network, OAAR),对梅花鹿的攻击、采食、躺卧、站立行为进行识别。OAAR网络包括前置网络、基础网络和时序网络,前置网络由LK光流算法(Lucas kanade optical flow algorithm)组成,用于提取RGB数据光流信息;基础网络中采用自注意力模块,将ResNet-152网络改造为ARNet152(Attention ResNet-152),用于将RGB、光流数据集经ARNet152提取特征后输入时序网络;时序网络采用添加注意力模块的长短记忆序列(Attention long short term network, ALST),并通过分类器输出行为得分和分类结果。视频数据集包括10 942段,共310 574帧,划分为攻击、采食、站立和躺卧4个大类,攻击行为又划分为撞击、脚踢和追逐3个小类;训练集、验证集和测试集比例为3:1:1。研究结果显示,OAAR模型在测试集上正确率为97.45%、召回率为97.46%、F1值为97.45%,ROC曲线中各类识别效果良好,特征嵌入图中各类行为特征区分度较高,各项结果均优于LSTM、双流I3D和双流ITSN网络,具有较好的泛化能力和抗干扰性。在本研究算法基础上集成的鹿只行为自动识别采集系统,为提高梅花鹿养殖生产管理水平和生产效率提供了技术基础。

关键词: 梅花鹿; 行为识别; 攻击行为; 深度学习; 光流注意力网络

中图分类号: TP391 文献标识码: A 文章编号: 1000-1298(2022)10-0261-10

OSID:



Automatic Recognition Algorithm for Sika Deer Attacking Behaviors
Based on Optical Current Attention Network

GAO Yun^{1,2} HOU Pengfei¹ XIONG Jiajun³ XU Xuelin³ CHEN Bin¹ LI Kang¹

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China

3. College of Animal Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China)

Abstract: Aggressive attacking behaviors of artificial rearing male sika deer on heat period are increased dramatically, which causes damages to deer's antlers and even deer themselves. Automatic monitoring of their aggressive attacking behaviors can provide an important basis for the research to reduce them. A dual-stream neural network (optical flow attention attacking recognition network, OAAR) was proposed, which was based on the attention mechanism and long-short memory sequences. It was used to achieve automatic recognition and detection of sika deer behaviors, including attacking, feeding, lying down, and standing. The OAAR network consisted of a pre-network, a base network, and a time-sequential network. The pre-network consisted of the LK optical flow algorithm(lucas kanade optical flow algorithm), which was used to extract the information from the RGB data. In the base network, a self-attentive module was added to the ResNet-152 to build a new design ARNet152 (Attention ResNet-152), which was used to combine the RGB and optical flow information, extract the features, and input them into the time-sequential network. The time-sequential network was based on an attention long short term network (ALST), which was composed of an attention long-short memory sequence that can classify the behavior and give scores. The experimental dataset was composed of 10 942 video segments, with a total of 310 574 frames, which were divided into four major categories of behaviors, including aggression,

收稿日期: 2021-11-30 修回日期: 2022-02-17

基金项目: 国家自然科学基金项目(31972533)

作者简介: 高云(1974—),女,副教授,博士,主要从事农业智能检测与控制研究,E-mail: angelclouder@mail.hzau.edu.cn

foraging, standing, and lying. From the aggressive behaviors, three sub-categories were further divided, including hitting, kicking, and chasing. The training, validation, and test sets were divided at a ratio of 3:1:1. The results of the study showed that the OAAR model reached an accuracy of 97.45%, a recall rate of 97.46%, and an F1 value of 97.45% on the test set, and good classification results in ROC curves and improved discrimination in feature embedding maps. All the results of OAAR were better than the results of LSTM, I3D, and ITSN networks. Meanwhile, the online deer behavior identification and recording system based on the OAAR network was developed to improve the management level and production efficiency of the sika deer farming industry.

Key words: sika deer; behavior recognition; attacking behavior; deep learning; optical current attention network

0 引言

人工集约化养殖梅花鹿作为我国一种半驯化的特种经济动物,其鹿茸有着极高的药用价值^[1]。雄性梅花鹿在发情期间,攻击行为发生次数剧增,极易造成鹿茸损伤,打斗激烈时,还会造成鹿只伤残,甚至死亡,给鹿业养殖户造成巨大的经济损失^[2]。目前实际生产中,主要依靠人工观察鹿只的攻击行为并进行人为干涉。人工观察攻击行为费时费力,难以实现长时间监控,且容易漏判,不适合大规模采用。

近年来,机器视觉技术在动物动作监测和识别领域得到长足的发展。在传统攻击行为识别算法研究中,支持向量机(SVM)和聚类算法等模式识别方法被用于区分畜禽攻击行为。改进SVM算法区分生猪攻击行为正确率^[3]达到95.70%^[4]、97.50%^[4]、97.60%^[5]。SVM算法也被用于区分鹿的采食行为,正确率为90.00%^[6]。基于分层聚类改进生猪攻击行为识别算法正确率达85.43%^[7],高度攻击行为识别正确率达95.80%^[8]、97.04%^[9],中度正确率能够达92.30%^[8]、95.82%^[9]。聚类算法中的K-means算法可实现种鸡打斗、采食、饮水、交配、振翅等行为的识别,正确率达84.03%^[10]。此外,也有学者选用线性分类器,即采用线性混合模型实现生猪攻击行为识别,正确率为73.90%和89.00%^[11]。随着神经网络和深度学习的高速发展,其在动物行为识别研究中的应用也成为近年来的热点^[12-16]。

注意力机制于20世纪90年代初次被引入传统机器视觉。2014年后,注意力机制与深度学习相结合在工业和农业上都得到了较好应用^[17]。注意力机制中自注意力和软注意力应用广泛。自注意力也称为内部注意力,目的是计算序列的表达形式。软注意力更关注区域或通道,包含空间域、通道域和混合域(融合空间和通道)等类别。在相关研究中^[18-23],注意力机制被加入到深度网络,分别应用

于图像中的目标识别、图像分割和视频中的表情识别,与未加入注意力机制的网络对比,注意力机制能够提取重点特征,明显提升网络识别的正确率。如果将注意力机制加入到动作和行为识别网络,将会有效提升重点特征的提取,使识别效果更好。

梅花鹿的攻击行为特征相比其它动物有其特殊性,如常采用鹿角或鹿蹄等部位进行攻击。在行为识别算法的研究中需要专门针对其特点研究可行的识别网络。本文结合光流法和注意力机制,在长期循环卷积网络(Long-term recurrent convolutional networks, LRCN)^[24]的基础上进行改进,研究梅花鹿行为的分类算法,行为包括攻击、采食、躺卧和站立等,其中攻击行为包括角斗、脚踢等特有攻击性行为动作。本文基于该算法设计鹿只行为自动识别系统,以实现各项行为在线自动监测和记录。

1 数据采集与数据集构建

1.1 数据采集

试验数据采集于2020年11月20日至2021年3月10日,采集地点为湖北省某鹿业有限公司。试验圈舍尺寸为30 m×12 m(长×宽),其中休息区96 m²、活动区264 m²,外墙高2.2 m。圈舍中设料槽,圈内饲养5~6岁龄、体况相近的42头雄性梅花鹿。试验对接种了KISS1-GnRH双表达DNA去势疫苗的雄性梅花鹿的攻击行为进行了测试,42头雄性梅花鹿随机均分为2组,一组进行疫苗肌肉注射(试验组),另一组不做处理(对照组),关于该疫苗试验内容另撰文描述^[25]。

视频采集选用海康威视网络红外摄像头(3T56WD-I3型),镜头采用8 mm焦距,F1.6定焦光圈镜头。摄像头通过网线连接到服务器,安装MySQL数据库服务器(MySQL 8.0.24, MySQL AB公司,瑞典),同时连接移动硬盘(WDBU6Y0050BBK-WESN 4TB,西部数据,美国),服务器可通过浏览器对摄像头和数据访问。采集图像尺寸为2 560像素×1 920像素,帧率为4 f/s,以avi格式存

储于移动硬盘。

由于梅花鹿圈舍分为休息区和活动区两部分，为了获取梅花鹿白天主要的活动，试验期间摄像头选择安装在鹿只主要活动区域，即活动区一角，可以俯视拍摄到院内鹿只所有活动。摄像头安装位置距圈舍围墙水平距离 6 m，距地面高度约 8 m，以俯视 15° 拍摄，水平视场角为 37.5°，垂直视场角为 27.8°，对角线视场角为 47.5°。为了方便进行每日对比，试验拍摄时间为 07:00—18:00，以采集鹿只白天院内所有活动。

1.2 雄鹿攻击行为定义

雄鹿在发情期间，攻击行为发生次数剧增。鹿只在进食时，会频繁用角顶撞其他鹿只的身体；在站立时，会发生角斗，即鹿只用鹿角撞击另一只鹿的角。攻击期间，鹿只间会前脚互踢，或一只鹿用前脚踢另一只鹿的身体。攻击末期，还会发生追逐行为。根据攻击发生的部位和发生的阶段将鹿只的攻击性行为细分为撞击、脚踢、追逐。同时，本文为区分鹿只的攻击行为和非攻击行为，将非攻击行为细分为采食、躺卧、站立，各行为的细分说明如表 1 所示。

表 1 鹿行为定义
Tab.1 Definition of deer's behavior

行为类别	行为名称	行为定义
攻击	角斗	鹿只用角撞击另一只鹿的角
	撞击	
	角对身体撞击	鹿只用角撞击另一只鹿的身体
	前脚互踢	鹿只互相用前脚对踢
	脚踢	
非攻击	踢身体	鹿只用前脚踢另一只鹿的身体
	追逐	攻击行为末期发生的行为
	采食	鹿只站立于料槽边，并将饲料卷入口腔中
	躺卧	身体腹部与地面接触，没有使用蹄部和腿部支撑身体
	站立	至少 3 条腿同时立于地面

1.3 数据集制作

梅花鹿 RGB 行为数据集采用行为优先级 (Attacking precedence, AP) 的方法制作，优先标注攻击行为，其次标注采食、站立和躺卧等行为。由于梅花鹿是群居动物，一般表现出群体性行为，而攻击行为发生在不少于两只鹿只之间的小群体性特异行为，持续时间较短，并且通常发生于采食、站立和躺卧行为中，采用 AP 法标注有助于训练模型提高对攻击行为的识别正确度。本研究中采用该方法分别对试验组和对照组截取图像的行为进行标注，网络模型优先对视频段中的攻击行为进行学习。标注鹿群其他日常行为时，多数鹿只同时发生的行为标注为该群体性行为。由于视频邻帧间，鹿只运动幅度较小，所以以 1 f/s 截取视频帧。标注结果如图 1 所

示。上一行为视频中直接截取的图像帧，下一行为图像帧放大后的行为图像。本次试验共标注视频 10 942 段，共 310 574 帧，标注时长 5 175.95 min。



图 1 鹿的行为帧
Fig.1 Frames of deer's behavior

针对鹿只运动时身体的变化，采用 LK 光流算法 (Lucas kanade optical flow algorithm) 进行识别，这是一种计算相邻图像帧鹿只运动所产生像素点变化的计算方法。用此方法制作光流数据集，并按照 3:1:1 随机划分为训练集、验证集和测试集，如表 2 所示。

表 2 光流数据集划分
Tab.2 Division of optical flow dataset

数据集	视频段数/段	视频帧数/帧	总时长/min
训练集	6 672	188 957	3 149.00
验证集	2 270	64 260	1 071.00
测试集	2 000	57 357	955.95
合计	10 942	310 574	5 175.95

2 行为攻击识别算法

2.1 算法设计

梅花鹿攻击行为相比其它动物有其特殊性，如常采用鹿角或鹿蹄等部位进行攻击。为了区分鹿只攻击行为与非攻击行为，解决鹿只之间的遮挡问题，提升视频识别效率和正确率，在算法中加入光流算法和注意力机制。设计光流注意力网络 (Optical flow attention attacking recognition network, OAAR)，改进攻击行为监测网络来对攻击行为和其它行为进行识别，该网络包括前置网络、基础网络和时序网络。前置网络由 LK 光流算法组成，提取 RGB 数据光流信息；基础网络采用自注意力模块将 ResNet-152 网络改造为 ARNet152 (Attention ResNet-152)，用于将 RGB、光流数据集提取特征后输入时序网络；时序网络采用添加注意力模块的长短记忆序列

(Attention long short term network, ALST), 并通过分类器输出行为得分和分类结果。

2.1.1 前置网络

观察鹿只日常情况可知, 相比于背景鹿只的运动, 发生攻击行为的鹿只前后帧图像变化较激烈。为了从背景中提取鹿只, 并区分攻击行为与非攻击行为, 在前置网络中设计 LK 光流算法, 计算鹿只相邻图像帧(图 2 中相邻的 3 帧图像)之间光流数据 V 。提取光流信息前, 先排除图像不清晰、场景光线暗, 或有拖尾的行为帧, 共排除 435 个。将剩余文件接入前置网络接口后, 根据鹿只行为发生时身体上重点发生变化的位置, 计算两帧之间光流 V 关于 X 、 Y 两个方向的信息。

$$I_x V_x + I_y V_y + I_t = 0 \tag{1}$$

式中 I_x 、 I_y 、 I_t ——图像中像素点灰度关于 X 、 Y 、 T 方向偏导数

V_x 、 V_y —— X 、 Y 方向光流向量

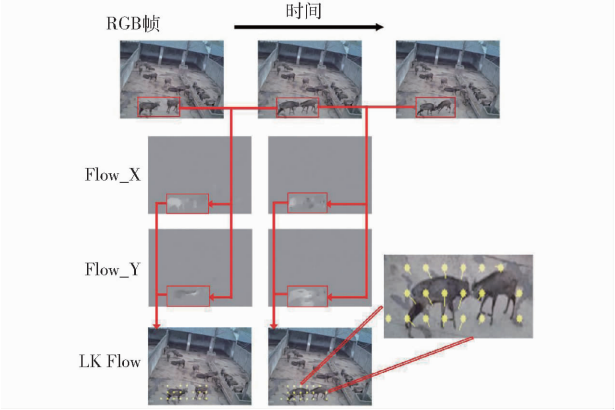


图 2 光流提取过程示意图
Fig.2 Extracted optical flow

结合相邻 $m \times m$ 范围内的像素点, 采用最小二乘法计算出最终光流信息, 并保存光流帧, 如图 2 中 Flow_X 帧和 Flow_Y 帧。将 X 、 Y 方向光流信息结合生成输出矢量 V 并在前一帧图像上显示, 即图 2 中 LK Flow 帧。将 LK Flow 帧放大后, 黄色点(图 2 右图中小圆点)表示像素起始位置, 引出线表示像素点的光流方向。可以直观看到, RGB 帧经前置网络处理后, 输出的光流帧能够捕获到运动的鹿只。由此可知, LK 光流法能够关注行为的位置变化, 忽略计算背景光流信息, 有效节约制作光流数据集的时间成本。

2.1.2 基础网络和时序网络

为了快速提取攻击行为特征信息, 如攻击行为发生时鹿角、鹿蹄等位置变化, 减少背景等非行为特征计算, 节省网络计算成本, 从而提升视频识别的效率和正确率, 在基础网络和时序网络中引入了注意力机制。基础网络和时序网络在长期递归卷积网络

(Long-term recurrent convolutional networks, LRCN)^[24]的基础上针对鹿的攻击行为进行改进和搭建(图 3a 中基础网络和 ALST)。长期递归卷积网络是一种将卷积神经网络和长短时序网络相结合的网络, 能够提取序列视频或图像特征, 并预测目标动作信息。基础网络负责计算 RGB、光流数据集中的行为特征值; 时序网络基于长短记忆序列建模时序信息, 并输出得分; 最终通过分类器按特征权重的差异分类, 输出攻击、采食、躺卧和站立 4 种行为得分。

由于网络训练长视频时, 随着时间维变化会遗忘最初的行为信息, 所以本文通过 ALST 提取行为时序信息。输入数据集视频帧尺寸为 320 像素 $\times$ 240 像素, 每段视频选取首尾 2 帧以及中间随机的 3 帧进行训练。对数据集中每一段 RGB 帧、光流帧进行 transformer 处理后, 尺寸裁剪为 $3 \times 224 \times 224$ 和 $2 \times 224 \times 224$ (深度 $\times$ 帧长 $\times$ 帧宽), 并通过 pytorch 框架的 dataloader 接口分别传入基础网络中, 调用 forward 函数进行训练。

基础网络由空间网络(Spatial ARNet152)和时间网络(Temporal ARNet152)构成。RGB 帧和光流帧分别经过由 ARNet152 搭建的空间网络和时间网络提取特征权重, 并将权重传入 ALST 网络。ARNet152 在 ResNet152 网络基础上插入自注意力卷积模块^[26](Attention key), 即自注意力卷积模块替换 ResNet152 中的 3×3 卷积键, 自注意力模块可建模动、静态上下文信息, 提取鹿只攻击行为特征, 提高鹿只行为识别的精度, 如图 3b 所示。ARNet152 网络中, 数据先经过 3 个卷积核为 3×3 的卷积键卷积计算, 第 1 个 2Dconv 步长为 2, 其余卷积层步长为 1, 填充像素(padding)均为 (1, 1)。卷积键(Convolution key)包含 1 层 2Dconv 卷积层、1 层 GN(Group normalization)层、1 层 ReLU 层, 卷积层左边绿色框为通道数, 右边橙色框为特征图输入、输出尺寸。数据依次经过包含 3、8、36、3 个 Bottleneck 的残差块(Blocks)提取特征权重。经过 4 个循环提取特征后, 通过自适应池化层(AvgPool)输出 $3 \times 1 \times 1$ 特征图, 并输入 ALST 网络中。第 1 至第 4 循环的第 1 个 Bottleneck 的自注意力卷积模块和下采样步长均为 2, 填充像素均为 (1, 1), 以减少输入参数量, 防止参数太多数据爆炸。Bottleneck 由主路径(main path)和捷径(shortcut)组成。主路径包括 2 个 1×1 conv 卷积键和 1 个自注意力卷积模块。自注意力卷积模块输入的特征值 X 重新定义为 Key Map、Query 和 Value Map 3 个值。先将 Key 值进行 3×3 卷积, 输出静态上下文建模矩阵 K_1 。

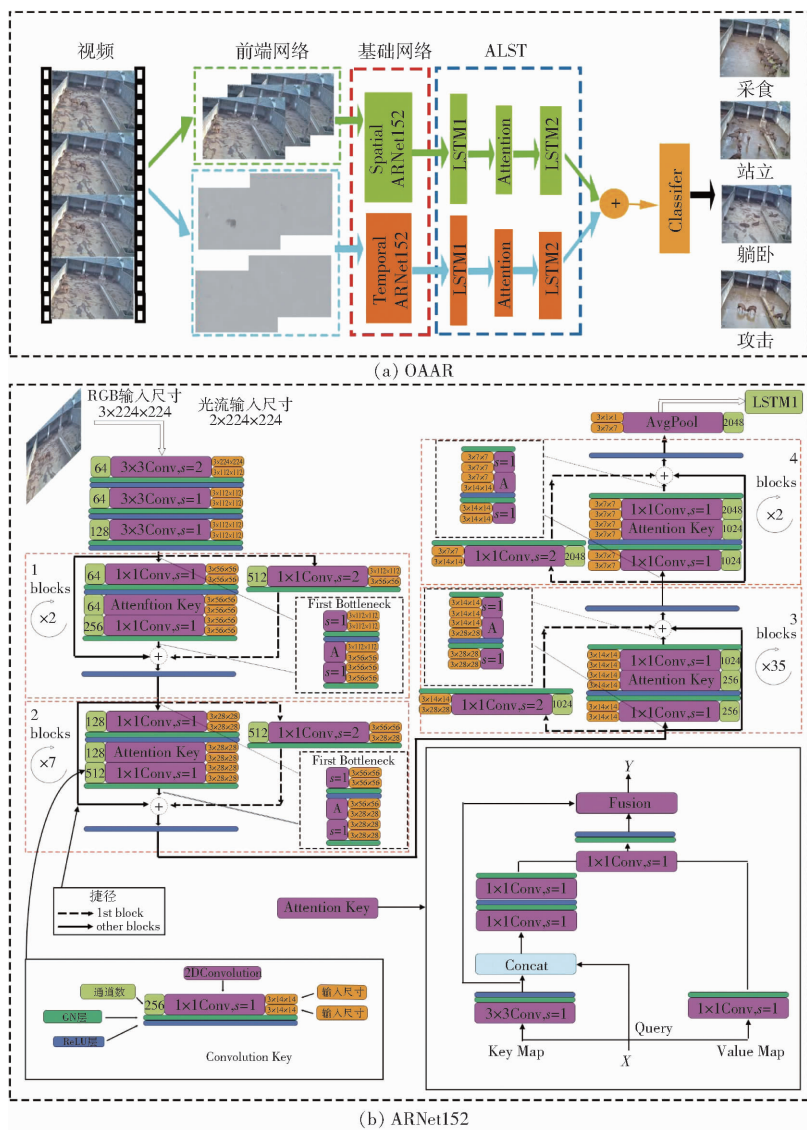


图 3 OAAR 结构示意图

Fig. 3 Structure diagram of OAAR

融合计算 K_i 和 Query 值后,进行 2 次 1×1 卷积计算,输出动态上下文 Attention 矩阵。经 Attention 矩阵与 Value 值卷积计算后,由 Softmax 输出自注意力特征值 Y 。每个循环中第 1 个 Bottleneck 的捷径由 1 个 1×1 conv 卷积键构成(图 3b 虚线框),实现对数据降维,输出维度即卷积键通道数,其余捷径(图 3b 实线框)不经过降维,直接将主路径输入数据与输出结果相加,blocks 下的数字表示循环次数。

ALST 网络由 2 层 512 个 LSTM 单元组成的时序结构和 1 层注意力机制层组成,注意力机制层计算公式为

$$a_i(s) = \text{align}(h_i, \bar{h}_s) \quad (2)$$

式中 a_i ——源端对齐位置(aligned position)向量

h_i ——decoder 隐状态

$\bar{h}_s$ ——源端隐状态

$\text{align}()$ ——对齐函数

每个 LSTM 单元由输入门、遗忘门和输出门构

成,实现对数据特征时序信息的提取计算,注意力机制层的作用是为了减少遗忘门遗忘量。时序信息建模完成后输入分类器。分类器(Classifier)由 1 层全连接层和 1 层 softmax 层构成,将双流信息融合计算后,对不同类别的特征权重分类,最终输出 4 类行为得分。

OAAR 网络的主要改进包括:①加入前置网络。前置网络由 LK 光流算法提取前后帧中的动作形成光流数据帧,增强后续网络输入的运动信息,减少背景的影响。②用自注意力模块替换 ResNet152 网络中残差结构的 3×3 卷积键,同时用 GN 层替换 BN 层,加上 ReLU 层和 dropout 层组成新的残差结构。自注意力模块可建模动、静态上下文信息,提取鹿只攻击行为特征,提高鹿只行为识别的精度。③LSTM 层中间插入注意力机制层,减少 LSTM 单元中遗忘门的遗忘量,改进时序网络,加强对时序信息的提取。除了对网络结构进行改进外,本文

在网络训练方式中也进行了改进,每段训练样本视频选取首尾 2 帧以及中间随机 3 帧进行训练,突出行为视频中关键动作特征的提取,同时节省网络训练成本。

2.2 网络参数设置

2.2.1 损失函数设置

OAAR 结合双流信息得分通过标准分类交叉熵损失 (Cross-entropy loss) 得到损失函数。采用交叉熵损失函数的优势在于,输出值和真实值的差值仅与最后一层权重梯度成正比,能够加快网络收敛速度。同时,网络反向传播的连乘计算,会加快更新整个权重矩阵。另外,多分类交叉熵损失函数求导更易求解,其损失率仅与正确类别的概率相关,降低了网络的计算量。损失函数公式为

$$L(y_i, \hat{y}_i) = - \sum_{i=1}^C y_i \lg \hat{y}_i \tag{3}$$

式中 y_i ——类别 i 的真实值
 $\hat{y}_i$ ——类别 i 的预测值
 C ——总类别数 L ——损失函数

2.2.2 网络参数初始化

网络训练平台使用 64 位 ubuntu18.04 系统,搭建 GPU 版 pytorch 框架,采用单 GPU (GeForce GTX 1080 12GB, NVIDIA, 美国) 进行训练。网络采用随机梯度下降法 (SGD) 作为优化器,动量设为 0.9,能够有效抑制振荡,加快收敛速度。批量大小 (Batch size) 设为 8,批量大小过小会导致训练速度太慢,过大会导致显卡占用内存溢出。迭代周期 (Epoch) 设为 100,初始学习率设为 1×10^{-3} ,Dropout 设为 0.8,增加迭代次数,降低学习率,能够使网络输出模型识别效果更精确,提升训练速度。但网络训练到一定程度时,模型趋于饱和且学习停滞,所以在第 30 和第 60 个迭代周期,学习率为原来的 1/10。

3 结果与分析

3.1 模型训练过程分析

为解决网络深度增大时鲁棒性差的问题,在正式训练前,先预训练 OAAR 网络。将空间网络加载 ImageNet 预训练权重初始化网络,输出空间网络权重后,时间网络加载该权重完成初始化训练。采用 pytorch 自带 tensorboardX 记录模型训练过程,模型正确率和损失值如图 4 所示。由于设置训练迭代次数到达第 30 和第 60 个迭代周期时,学习率下降至原来的 1/10,所以在第 30 个迭代周期时,正确率出现大幅增加,损失值大幅减小;在到达第 60 个迭代周期前训练集正确率和损失值逐渐趋于稳态;第 60

个迭代周期后,训练集正确率继续小幅提高,损失值小幅下降。验证集验证过程中的正确率、损失值和训练集对应的值贴合较好,较为近似。在模型中适当调低学习率有助于网络加快学习速度和对特征点细节的继续学习。最终得出 OAAR 网络模型在训练集上正确率为 99.16%,损失值为 0.026 4,在验证集上正确率为 97.91%,损失值为 0.061 2。

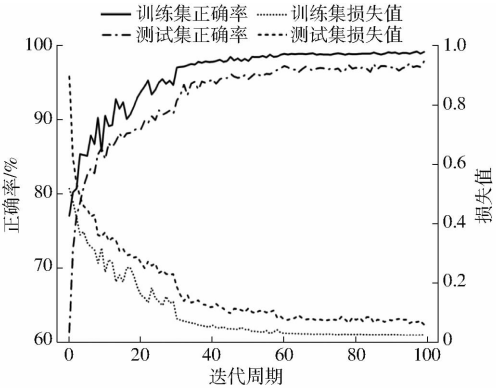


图 4 OAAR 网络识别正确率和损失值
Fig. 4 Recognition accuracy and loss value of OAAR

3.2 行为识别评价指标与结果

采用正确率 (Accuracy)、召回率 (Recall)、精确率 (Precision) 和 F1 值 (F1 score)^[15] 作为评价模型性能的指标。

正确率常被用于评价鹿只所有行为识别正确的比重,但当样本不平衡时,仅依靠正确率不能正确评价模型。因此采用召回率、精确率和 F1 值对模型进行综合评价。PR 曲线中的平衡点 (BEP) 对应 F1 值的计算结果,表示模型的综合性能。

测试集中每类行为包含 500 段视频,合计 2 000 段视频,其中共 1 949 段视频被正确识别,51 段视频未能被正确识别,OAAR 网络混淆矩阵如图 5 所示。模型对攻击、采食、躺卧、站立行为识别精确率分别为 96.00%、96.60%、100%、97.20%。

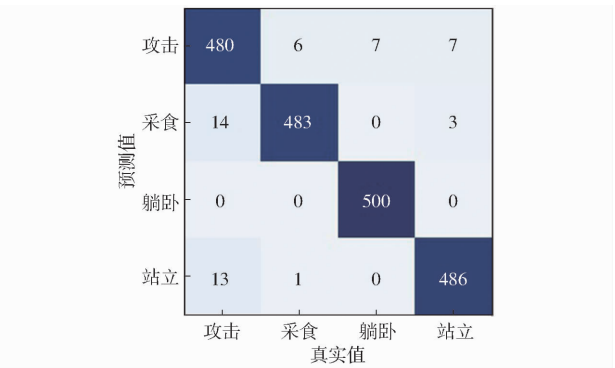


图 5 OAAR 分类结果混淆矩阵
Fig. 5 Confusion matrix of OAAR

由图 5 可知,攻击行为的精确率较其它 3 种行为略低。原因可能是:①鹿只严重重叠或遮挡时,识

别较为困难。部分视频帧中出现鹿只之间严重重叠、遮挡问题,如鹿只的头部、脚部被其他鹿只或树等物体全程遮挡时,模型未能检测到鹿只攻击动作。

②攻击行为发生时间极短时难以识别。当鹿只攻击行为持续时间太短,部分视频只持续 1~2 帧攻击动作,这部分行为为较难被模型识别。

③发生行为时前后帧图像变化幅度较小时较难识别。攻击动作发生的前后两帧变化幅度很小时,前置网络难以获取其动作,模型的注意力在鹿只的其他行为,导致攻击行为的置信度低于其他行为,所以分类结果输出为非

攻击行为。

从特征可视化角度验证 OAAR 网络识别效果,如图 6 所示。图中第 1 行为测试集中随机选取的一段连续的攻击行为帧,第 2 行为通过网络最后一层卷积操作输出的特征可视化结果,第 3 行在第 2 行基础上通过特征注意力机制生成热力图,颜色越偏近暖色表示攻击动作越激烈。从热力图中可以观察到,高亮部分集中在发生不同程度攻击行为的 3 个区域。因此,验证了自注意力模块能够有效识别鹿只攻击行为。

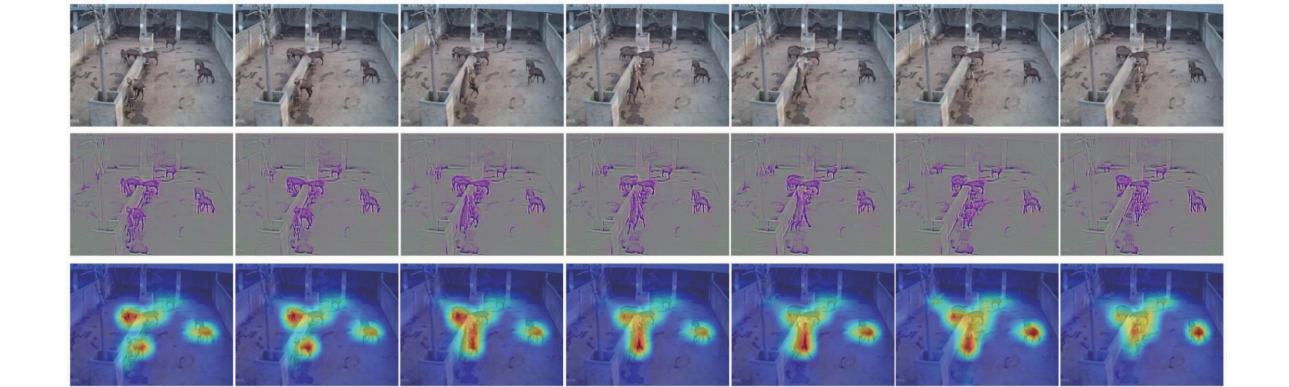


图 6 特征可视化

Fig. 6 Feature visualization

将网络最后 1 层特征映射在二维空间,如图 7 所示。图中每个点代表测试集的 1 段视频,每个行为类别采用一种颜色表示,类别 A (攻击)、F (采食)、L (躺卧) 和 S (站立) 特征映射点分别采用红色、绿色、蓝色、紫色表示,图 7 中不同行为分别聚集于 4 个角,攻击行为与躺卧、站立、采食有很少量混杂,且每种行为自身的聚集度较高。特征嵌入图说明了 OAAR 网络对攻击、采食、躺卧和站立行为类别的区分度较高。

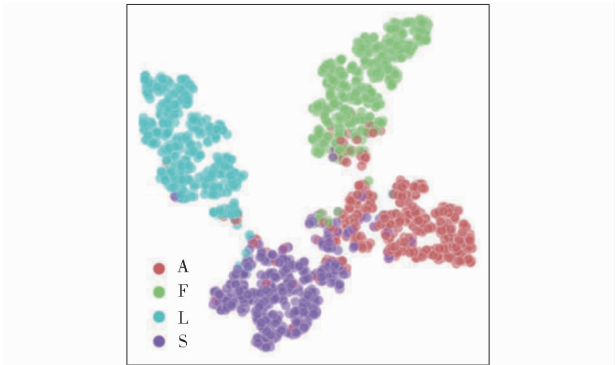


图 7 特征嵌入

Fig. 7 Feature embedding

4 种行为的受试者特性 (ROC) 曲线如图 8 所示。ROC 曲线均位于随机猜测线左上方,曲线距左上角越近,显示分类结果越准确。采用宏平均和微平均综合评价网络分类性能。宏平均先计算每类行

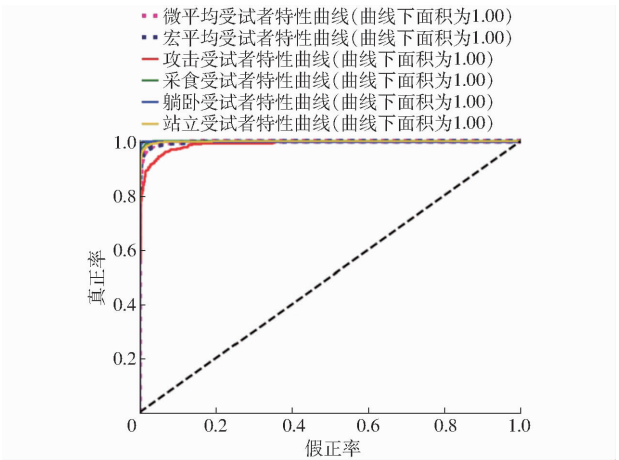


图 8 ROC 曲线

Fig. 8 ROC curve

为 ROC 值,再计算算术平均值;微平均先计算所有类真正率和假正率的平均值,再作 ROC 插值。从图 8 可知,采食、躺卧和站立行为接近理想指标,即对某类识别效果极好,攻击行为趋于理想指标,各分类的曲线下方面积 (AUC) 均趋于 1。综上所述,OAAR 能够对各类行为进行较好的识别和分类。

3.3 不同网络对比分析

为了验证 OAAR 网络对梅花鹿行为识别的有效性,选取长短时记忆网络 (Long short term memory network, LSTM)^[27]、双流 I3D 网络 (Two-stream inflated 3d convNets)^[28] 和双流 ITSN 网络

(Improvement temporal segment network)^[29]与 OAAR 网络进行对比验证。LSTM 采用 decoder 和 encoder 两个环节配合完成行为识别。双流 I3D 网络采用膨胀 3D 卷积计算 RGB、光流数据特征,利用空间感受野融合时间感受野,实现特征信息识别和分类。双流 ITSN 基于长范围时间建模,在原本 TSN 网络的基础上替换 2D 卷积为 3D 卷积。用视频数据集分别训练各网络,并将各网络在测试集上的识别结果与本文中的 OAAR 网络的分类结果进行对比分析。结果如表 3 所示,各网络模型正确率分别为 82.65% (LSTM)、84.33% (双流 I3D)、96.80% (双流 ITSN)和 97.45% (OAAR)。

表 3 当前常用网络模型比较

Tab.3 Comparisons with state-of-the-art results

模型	正确率/ %	平均识别 时间/s	参数量/ MB	浮点运算 次数/次
LSTM ^[27]	82.65	0.21	27.48	4.37×10^9
双流 I3D ^[28]	84.33	2.81	46.19	8.88×10^9
双流 ITSN ^[29]	96.80	1.44	42.40	7.85×10^9
OAAR	97.45	0.17	38.27	6.10×10^9

由表 3 可知,OAAR 识别单个视频的平均时间明显短于 LSTM、双流 I3D 和双流 ITSN。由于双流 I3D 网络基于 3D 卷积的 InceptionV1 网络提取特征^[28],网络的参数量和浮点运算次数剧增,同时识别单个视频的时间也会增加。双流 ITSN 网络中,采用视频分段的形式训练模型网络^[29],并使用 Resnet3d-50 作为提取特征的基础网络,基础网络中的残差结构能够有效降低实际训练的参数量和浮点运算次数,同时提升网络的识别正确率。LSTM 的基础网络采用传统 2D 卷积的 CNN 网络^[27],提取单帧图像特征后,传入 LSTM 单元中关联前后行为

特征信息,参数量和浮点运算次数在 4 种网络中最小,但由于基础网络没有提取到有效的攻击行为特征,该网络的识别正确率较其他网络略低。在 OAAR 网络中,基础网络基于注意力机制模块和残差结构搭建的 ARNet152,侧重提取鹿角、鹿蹄等攻击行为特征,同时时序网络中的注意力机制能够有效减少遗忘门的遗忘量,从而提高了视频行为识别的效率和正确率,验证了注意力机制的有效性。

表 4 分别列出了 4 种网络对测试集识别的结果,OAAR 精确率为 97.45%,召回率为 97.46%,远高于 LSTM 精确率(82.65%)、召回率(83.65%)和双流 I3D 精确率(86.57%)、召回率(86.58%),略高于双流 ITSN 精确率(96.81%)、召回率(96.74%)。相比于 LSTM 网络、双流 I3D 网络和双流 ITSN 网络,OAAR 网络主要从 4 方面优化了其性能:①前置网络提取了行为随时间变化的动作特征。前置网络提取鹿只身体的位置变化,抑制背景噪声的影响。通过训练双流网络后,取空间和时间网络输出得分的均值,有效提升网络行为识别的准确度。②自注意力模块使发生攻击行为的鹿只更加突显。由于在基础网络中使用自注意力卷积模块加工动、静态上下文信息,提取鹿只攻击行为特征,较大提升了鹿只攻击行为识别正确率。③残差结构能够获取更多行为特征信息。残差结构保留了更多的低层特征信息,在网络加深过程中,有效避免了过拟合、梯度爆炸和消失等问题,加快网络训练进程。④每段视频选取首尾 2 帧以及中间随机的 3 帧进行稀疏训练。为避免行为视频过长时导致重要行为特征削弱的现象,每段视频选取首尾 2 帧以及中间随机的 3 帧进行训练,以便在训练过程中突出视频中的关键动作特征,节省训练成本。

表 4 各模型性能统计

Tab.4 Performance statistics of models

行为 类别	LSTM ^[27]			双流 I3D ^[28]			双流 ITSN ^[29]			OAAR			%
	召回率	精确率	F1 值	召回率	精确率	F1 值	召回率	精确率	F1 值	召回率	精确率	F1 值	
攻击	65.10	77.00	77.71	70.91	78.88	73.92	96.04	93.07	94.55	94.67	96.00	95.33	
采食	90.90	82.07	81.84	90.20	87.30	88.73	97.21	98.22	97.72	98.57	96.60	97.58	
躺卧	96.27	97.82	95.65	97.00	96.66	96.80	98.66	99.67	99.12	98.62	100	99.30	
站立	82.21	73.83	82.39	88.22	85.20	86.67	95.33	96.43	95.84	97.98	97.20	97.59	
合计	83.65	82.65	82.88	86.58	86.57	86.53	96.74	96.81	96.77	97.46	97.45	97.45	

用精确率-召回率曲线直观比较 4 种网络识别效果,如图 9 所示。点 A、B、C、D 为 4 种网络的精确率-召回率曲线的平衡点,对应各网络的 F1 值。图 9 中,OAAR 网络的平衡点 D 高于点 A、B、C,说明 OAAR 网络识别效果优于 LSTM 网络、双流 I3D 网络和双流 ITSN 网络。

4 模型部署及应用

基于 OAAR 网络搭建的鹿只行为自动识别系统如图 10 所示。该系统采用摄像头获取视频上传至服务器,后端框架调用服务器中部署的 OAAR 网络模型,计算视频鹿群中是否发生包括攻击、采食、

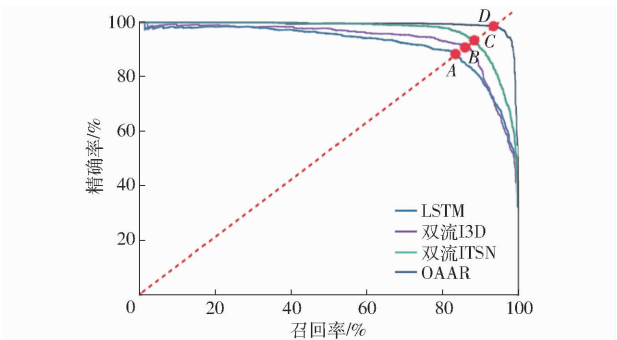


图 9 精确率-召回率曲线

Fig. 9 Precision – recall curves

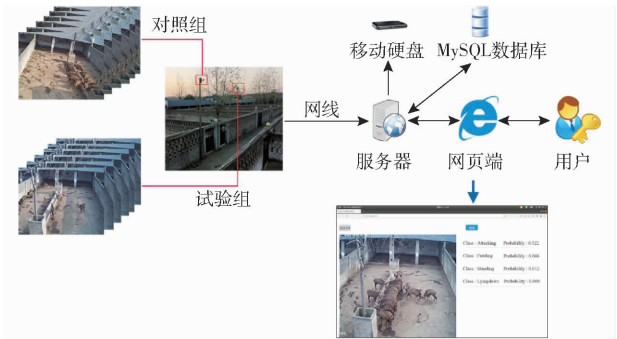


图 10 鹿只行为自动识别系统

Fig. 10 Sika deer behaviors' automatically recognition system

站立、躺卧行为,并进行置信度打分,并将鹿群中发生的行为打上时间戳存储在服务器中的 MySQL 数据库中,相关信息可通过网页发布,并供用户远程访问。

5 结论

(1)新增了前置网络提取光流数据。前置网络采用 LK 光流算法计算光流数据帧,提取前后帧中鹿只身体位置的变化,抑制背景噪声的影响。其输出的光流数据帧与 RGB 数据帧共同作为基础网络的输入数据,进一步加强了行为特征,削弱了背景噪声,试验证明了本文方法能够有效解决行为被遮挡的问题。

(2)将注意力机制分别加入基础网络和时序网络。基础网络中将自注意力卷积模块替换为 ResNet152 中的 3 × 3 卷积层,分别加工动、静态上下文信息。时序网络在 2 层 LSTM 层中加入注意力机制层,减少 LSTM 单元中遗忘门的遗忘量,提升网络处理时序信息的能力。

(3)OAAR 网络在测试集上的识别正确率高达 97.45%,高于 LSTM 正确率(82.65%)、双流 I3D 正确率(84.33%)和双流 ITSN 正确率(96.80%)。OAAR 精确率和召回率也分别优于其它 3 个网络。受试者特性曲线和特征嵌入图均显示了 OAAR 网络具有较好的区分性、泛化和抗干扰能力。

(4)集成 OAAR 网络的鹿只行为自动识别采集系统实现了梅花鹿行为的自动识别功能,为提高梅花鹿养殖生产管理水平和生产效率提供了可行的途径。

参 考 文 献

[1] 杨娇. 公鹿的饲养管理技术[J]. 畜牧兽医科技信息, 2017, 4(4): 114.
YANG Jiao. Breeding management techniques for antler deer[J]. Chinese Journal of Animal Husbandry and Veterinary Medicine, 2017, 4(4): 114. (in Chinese)

[2] 丁修俊. 人工圈养梅花鹿的饲养繁育及日常疾病预防[J]. 上海畜牧兽医通讯, 2018, 4(6): 53 – 55.
DING Xiujun. Human breeding and daily disease prevention[J]. Shanghai Journal of Animal Husbandry and Veterinary Medicine, 2018, 4(6): 53 – 55. (in Chinese)

[3] LEE J, JIN L, PARK D, et al. Automatic recognition of aggressive behavior in pigs using a kinect depth sensor[J]. Sensors, 2016, 16(5): 631.

[4] CHEN Chen, ZHU Weixing, LIU Dong, et al. Detection of aggressive behaviours in pigs using a realsence depth sensor[J]. Computers and Electronics in Agriculture, 2019, 166(C): 105003.

[5] 刘冬, 何东健, 陈晨, 等. 基于 ALR – GMM 的群养猪攻击行为识别算法研究[J]. 农业机械学报, 2021, 52(1): 201 – 208.
LIU Dong, HE Dongjian, CHEN Chen, et al. Recognition of aggressive behaviour in group-housed pigs based on ALR – GMM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1): 201 – 208. (in Chinese)

[6] 杨婷. 基于最小二乘支持向量机的梅花鹿进食行为识别的研究[D]. 长春: 吉林农业大学, 2017.
YANG Ting. Research on facial recognition of sika deer based on least squares support vector machine[D]. Changchun: Jilin Agricultural University, 2017. (in Chinese)

[7] 张圆成. 基于视觉显著性的攻击行为的猪的识别方法研究[D]. 镇江: 江苏大学, 2020.
ZHANG Yuancheng. Research on the recognition method of pigs with aggressive behavior based on visual saliency[D]. Zhenjiang: Jiangsu University, 2020. (in Chinese)

[8] CHEN Chen, ZHU Weixing, GUO Yizheng, et al. A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs[J/OL]. Livestock Science, 2018. DOI:10.1016/j.livsci.2018.10.013.

[9] CHEN Chen, ZHU Weixing, MA Changhua, et al. Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs[J/OL]. Computers and Electronics in Agriculture, 2017. DOI:10.1016/j.compag.2017.09.013.

[10] 李丽华, 刘志伟, 赵学谦, 等. 基于加速度传感器的本交笼种鸡个体行为监测与识别[J]. 农业机械学报, 2019, 50(12):

- 247–254.
- LI Lihua, LIU Zhiwei, ZHAO Xueqian, et al. Monitoring and identification of natural mating cage breeding chickens individual behavior based on acceleration sensor[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 247–254. (in Chinese)
- [11] VIAZZI S, ISMAYILOVA G, OCZAK M, et al. Image feature extraction for classification of aggressive interactions among pigs[J]. Computers and Electronics in Agriculture, 2014, 104: 57–62.
- [12] OCZAK M, VIAZZI S, ISMAYILOVA G, et al. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network[J]. Biosystems Engineering, 2014, 119: 89–97.
- [13] 张苏楠, 田建艳, 营莹, 等. 基于帧间差分法-单点多框检测器的圈养生猪打斗行为识别方法[J]. 江苏农业学报, 2021, 37(2): 397–404.
- ZHANG Su'nan, TIAN Jianyan, JIAN Long, et al. Recognition method of aggressive behaviors among pigs in pigpens based on frame difference (FD)-single shot MultiBox detector(SSD)[J]. Jiangsu Journal of Agricultural Sciences, 2021, 37(2): 397–404. (in Chinese)
- [14] 高云, 陈斌, 廖慧敏, 等. 群养猪侵略性行为的深度学习识别方法[J]. 农业工程学报, 2019, 35(23): 192–200.
- GAO Yun, CHEN Bin, LIAO Huimin, et al. Recognition method for aggressive behavior of group pigs based on deep learning[J]. Transactions of the CSAE, 2019, 35(23): 192–200. (in Chinese)
- [15] CHEN Chen, ZHU Weixing, STEIBEL J, et al. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory[J/OL]. Computers and Electronics in Agriculture, 2020. DOI:10.1016/j.compag.2019.105166.
- [16] 甘海明, 薛月菊, 李诗梅, 等. 基于时空信息融合的母猪哺乳行为识别[J]. 农业机械学报, 2020, 51(增刊1): 357–363.
- GAN Haiming, XUE Yueju, LI Shimei, et al. Automatic sow nursing behaviour recognition based on spatio-temporal information fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp.1): 357–363. (in Chinese)
- [17] 刘建伟, 刘俊文, 罗雄麟. 深度学习中注意力机制研究进展[J]. 工程科学学报, 2021, 43(11): 1499–1511.
- LIU Jianwei, LIU Junwen, LUO Xionglin. Research progress in attention mechanism in deep learning[J]. Chinese Journal of Engineering, 2021, 43(11): 1499–1511. (in Chinese)
- [18] 胡旭. 基于注意力机制的金丝猴面部识别研究与实现[D]. 西安: 西安电子科技大学, 2019.
- HU Xu. Research and implementation of facial recognition of golden monkey based on attention mechanism[D]. Xi'an: Xidian University, 2019. (in Chinese)
- [19] 张宏鸣, 汪润, 董佩杰, 等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. 农业机械学报, 2021, 52(4): 248–256.
- ZHANG Hongming, WANG Run, DONG Peijie, et al. Beef cattle multi-target tracking based on DeepSORT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 248–256. (in Chinese)
- [20] 方鹏, 郝宏运, 李腾飞, 等. 基于注意力机制和可变形卷积的鸡只图像实例分割提取[J]. 农业机械学报, 2021, 52(4): 257–265.
- FANG Peng, HAO Hongyun, LI Tengfei, et al. Instance segmentation of broiler image based on attention mechanism and deformable convolution[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 257–265. (in Chinese)
- [21] 胡志伟, 杨华, 娄甜田. 采用双重注意力特征金字塔网络检测群养生猪[J]. 农业工程学报, 2021, 37(5): 166–174.
- HU Zhiwei, YANG Hua, LOU Tiantian. Instance detection of group breeding pigs using a pyramid network with dual attention feature[J]. Transactions of the CSAE, 2021, 37(5): 166–174. (in Chinese)
- [22] 张宏鸣, 周利香, 李永恒, 等. 基于改进 MobileFaceNet 的羊脸识别方法[J]. 农业机械学报, 2022, 53(5): 267–274.
- ZHANG Hongming, ZHOU Lixiang, LI Yongheng, et al. Sheep face recognition method based on improved MobileFaceNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 267–274. (in Chinese)
- [23] 温长吉, 张笑然, 吴建双, 等. 基于多注意力机制级联 LSTM 模型的猪脸表情识别[J]. 农业工程学报, 2021, 37(12): 181–190.
- WEN Changji, ZHANG Xiaoran, WU Jianshuang, et al. Pig facial expression recognition using multi-attention cascaded LSTM model[J]. Transactions of the CSAE, 2021, 37(12): 181–190. (in Chinese)
- [24] DONAHUE J, HENDRICKS A, ROHRBACH M, et al. Long-term recurrent convolutional networks for visual recognition and description[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. DOI:10.1109/TPAMI.2016.2599174.
- [25] 许学林, 杨波, 荆焕松, 等. 双表达去势 DNA 疫苗的构建及其稳定性研究[J]. 中国畜牧兽医, 2021, 48(5): 1745–1754.
- XU Xuelin, YANG Bo, JING Huansong, et al. Constructon and stability study of co-expression castration DNA vaccine[J]. China Animal Husbandry & Veterinary Medicine, 2021, 48(5): 1745–1754. (in Chinese)
- [26] LI Yehao, YAO Ting, PAN Yingwei, et al. Contextual transformer networks for visual recognition[J]. arXiv preprint:2107.12292, 2021.
- [27] WANG Xuanxuan, GAO Lianli, SONG Jingkuan, et al. Beyond france-level CNN: saliency-aware 3-D CNN with LSTM for video action recognition[J]. IEEE Signal Processing Letters, 2017, 24(4): 510–514.
- [28] CARREIRA J, ZISSERMAN A. Quo vadis, action recognition? A new model and the Kinetics dataset[J]. arXiv preprint: 1705.07750v3, 2018.
- [29] WANG Limin, XIONG Yuanjun, WANG Zhe, et al. Temporal segment networks: towards good practices for deep action recognition[J]. arXiv preprint:1608.00859v1, 2016.