doi:10.6041/j.issn.1000-1298.2021.08.025

基于 EfficientDet 网络的湖羊短时咀嚼行为识别方法

陆明洲[1]　梁钊董[1]　NORTON Tomas[2]　张生福[3]　沈明霞[1]

(1.南京农业大学人工智能学院,南京 210031;2.鲁汶大学 M3-BIORES 实验室,海弗莱 B3001;
3.青海民族大学计算机学院,西宁 810007)

摘要:为分析羊进食行为、自动估算其进食量,提出一种从舍饲湖羊采食视频中自动识别其短时咀嚼行为的方法。首先,针对舍饲湖羊采食区域特点,在 EfficientDet 网络架构中增加目标框筛选模块,检测视频帧中羊嘴张开、上下颌错开及闭合 3 种状态,根据羊脸与相机拍摄角度的方位关系检测羊嘴状态,并为各状态赋编码值;然后,利用正则表达式提取连续视频帧中的一次上下颌张合对应的羊嘴状态编码值序列片段;最后,针对羊侧脸面对相机咀嚼、抬头正脸面对相机咀嚼、低头正脸面对相机咀嚼以及鸣叫等一次上下颌张合动作对应的羊嘴状态编码值序列片段构建分类规则,实现短时咀嚼行为的自动识别。对比了基于 EfficientDet-D0 ~ D4、YOLO v5 和 SSD 网络的羊嘴状态检测性能,结果表明,改进的 EfficientDet-D1 网络能以 28.18 f/s 的传输速率,获得 95.64% 和 98.84% 的羊嘴状态检测精确率和均值平均精确率,优于 YOLO v5 和 SSD 网络。利用湖羊采食视频测试 EfficientDet-D1 网络结合正则表达式的湖羊短时咀嚼行为识别分类规则性能,结果表明,分类规则能以 91.42% 的自动识别正确率和 90.85% 的平均正确率直接从视频中提取湖羊短时咀嚼行为发生次数和持续时长。本研究将基于视频的湖羊短时咀嚼行为识别问题转换为羊嘴状态编码值序列分类问题,降低了分类模型的复杂度,为湖羊短时咀嚼行为的自动识别提供了一种新的研究思路。

关键词:湖羊;短时咀嚼行为;羊嘴状态检测;咀嚼行为分析;正则表达式;EfficientDet

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2021)08-0248-07　　**OSID**:

Automatic Identification Method of Short-term Chewing Behaviour for Sheep Based on EfficientDet Network

LU Mingzhou[1]　LIANG Zhaodong[1]　NORTON Tomas[2]　ZHANG Shengfu[3]　SHEN Mingxia[1]

(1. *College of Artificial Intelligence*, *Nanjing Agricultural University*, *Nanjing* 210031, *China*
2. *M3-BIORES-Measure*, *Model and Manage Bioresponses*, *KU Leuven*, *Heverlee B*3001, *Belgium*
3. *Computer College*, *Qinghai Nationalities University*, *Xining* 810007, *China*)

Abstract: Animal's short-term chewing behaviour is accomplished with discrete jaw movements which occurs through a repeating jaw opening-closing cycle. Recognition of short-term chewing behaviour is the foundation of feeding behaviour analysis and feed/pasture intake estimation for sheep. Several attempts have been made to establish models for short-term chewing behaviour recognition based on the jaw pressure or acoustic signal obtained using wearable sensors. However, such data collection methods have shortages such as difficulty in battery replacement, low stability of the data, and the sensors are vulnerable to damage. A short-term chewing behaviour identification method using computer vision technology was presented, which can extract the frequency and duration of each individual short-term chewing from the feeding video of sheep. Firstly, based on the characteristics of sheep feeding area, a module for target box selection was added in the EfficientDet network architecture. This modified EfficientDet network was employed to detect the three status of sheep mouth, that was opening, stagger of the upper/lower jaw, and closing, in each video frame. Once the sheep mouth status in a video frame was determined, a numerical label was assigned. Then, the regular expression was employed to extract the numerical label sequence segment corresponding to each individual jaw opening-closing cycle. Finally, classification rules were constructed for short-term chewing behaviour identification, where

收稿日期:2021-03-30　修回日期:2021-05-21

基金项目:国家自然科学基金项目(31972615)、江苏省自然科学基金项目(BK20191315)和青海省科技厅基础研究计划项目(2020-ZJ-716)

作者简介:陆明洲(1981—),男,教授,博士,主要从事畜牧数据智能计算方法与技术研究,E-mail:lmz@njau.edu.cn

chewing with the side face facing the camera, chewing while facing the camera with the head down, chewing while facing the camera with the head up, and sheep chirping were distinguished. The performance of the sheep mouth status detection obtained by the modified EfficientDet－D0～D4 networks were compared with those obtained by the YOLO v5 and SSD networks. The comparison results indicated that the precision rate, mean average precision, and frame rate of the modified EfficientDet－D1 network was 95.64%, 98.84%, and 28.18 f/s, respectively, which were better than those of YOLO v5 and SSD. Short-term chewing behaviour classification rules, which consisted of EfficientDet－D1 network and regular expression, were applied to the testing videos. The testing results indicated that the frequency and duration of short-term chewing can be extracted from the videos with the accuracy of 91.42% and 90.85%, respectively. The developed method transformed the video-based sheep short-term behaviour identification problem into the problem of the status label sequences classification, which reduced the complexity of the short-term chewing behaviour classification task. The presented method provided a solution for the automatic short-term chewing behaviour recognition for sheep.

Key words: sheep; short-term chewing behaviour; sheep mouth status detection; chewing behaviour analysis; regular expression; EfficientDet

0　引言

湖羊是我国太湖平原地区特有的羔皮用绵羊品种,属一级保护地方畜禽品种,其主要饲养方式为规模化舍饲[1]。食欲减退甚至废绝是湖羊常见疾病(如肝片吸虫病、羊黑快疫等)的主要症状之一,监测湖羊采食行为以及时发现食欲异常的湖羊个体在养殖生产实践中至关重要[2]。传统的采食行为监测方法为人工观测,需要具有丰富养殖经验的饲养员连续或定时观测湖羊的采食过程,耗费大量人力与时间,且易对湖羊造成惊扰,无法长期持续监测。随着各类传感器技术、机器学习技术、自动控制技术等被越来越多地应用于养殖业[3],精准畜牧养殖(Precision livestock farming, PLF)成为养殖业的重要发展方向之一。采食行为的自动监测和分析是 PLF 的重要组成部分[4],研究人员针对动物采食行为自动监测开展了相应研究工作。

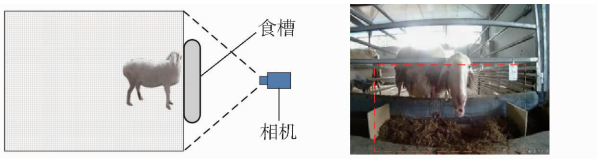
放牧养殖模式下的动物活动范围广,动物牧食行为的自动监测主要通过穿戴式传感装置采集牧食行为相关的动物头部、颚部运动[5-7]及音频信号等数据[8-9]。穿戴式传感装置满足连续、自动监测动物牧食行为的技术需求,但这种接触式数据采集方式也给动物带来应激,同时,动物对穿戴传感装置的破坏及复杂养殖环境下传感器元器件的腐蚀等也会导致传感数据的丢失。因此,对于活动空间固定的舍饲动物,无接触式的视频数据采集与分析方式更具优势。近年来,研究人员尝试利用计算机视觉技术自动监测分析圈养反刍动物饮食的相关行为。在奶牛饮水行为[10]、反刍行为[10-12]的自动监测方面,CHEN 等[11]提出了一种基于时空上下文信息学习的反刍奶牛嘴部区域动作跟踪算法。MAO 等[13]利用光流法提出一种奶牛嘴部的识别与定位方法。WU 等[10]和 SAFA 等[12]分别利用深度学习框架 CNN－LSTM 及 CNN 训练奶牛反刍行为自动识别分类器,这 2 种分类器能分别以 95% 和 97.1% 的精度识别奶牛的反刍行为。JIANG 等[14]在圈舍群养羊饮水采食行为的自动监测方面进行了探索,利用 YOLO v4 深度学习框架定位羊只,通过羊只与食槽、饮水区域的位置关系识别羊只的采食与饮水行为,对两种行为的识别精度分别达到 97.82% 和 98.27%。

目前,基于计算机视觉技术进行圈养反刍动物饮食行为的研究重点多集中于动物的长时行为(如采食、饮水、反刍等)。已有研究表明,动物的短时采食行为特征(如短时咀嚼次数、短时咀嚼时长、短时咀嚼声音能量通量密度等)与其采食量具有较高的相关性[8-9,15-17]。本文根据鸣叫与短时咀嚼行为在上下颌张合状态持续时间的不同,利用深度学习方法检测舍饲湖羊采食视频帧中上下颌张合状态,提出一种基于上下颌张合状态时变特点的湖羊短时咀嚼行为分类规则,以自动识别湖羊的短时咀嚼行为。

1　数据与方法

1.1　湖羊饲养及采食行为视频获取

在江苏省农业科学院六合动物实验基地种羊场随机选择 2 只 300 日龄、体质量约 45 kg 的公湖羊用于视频数据采集。采用 2 个羊圈饲养试验湖羊,羊圈面积为 2 m×1.5 m。选用 SARGO 运动相机(T1 型)记录湖羊采食行为视频,视频分辨率为 2 880 像素(水平)×2 160 像素(垂直),帧率为 24 f/s。利用高度设定为 0.3 m 的三脚架将相机固定在羊圈正前方,相机距离羊圈食槽 0.5 m,相机与羊圈的位置关系示意图如图 1a 所示。



(a) 相机与羊圈位置关系示意图 (b) 羊采食短时咀嚼行为视频帧

图 1 舍饲湖羊采食行为视频数据采集方式示意图

Fig. 1 Schematics of video data collection of feeding behaviour of sheep under stall-feeding condition

视频采集时间为 2020 年 8 月 21—25 日07:30—10:00 及 16:00—18:30。湖羊在正常采食过程中鸣叫次数较少,但在等待饲养员投喂饲料的时间段内鸣叫行为高发。在 07:30—7:50 以及 16:00—16:20 两个时间段,饲养员接近羊圈但不投饲料,开启相机采集湖羊在食槽位置的鸣叫视频。07:50—10:00 及 16:20—18:30 两个时间段饲养员向食槽投放青贮饲料,采集湖羊采食行为视频。视频以 MP4 格式存储在相机 SD 卡中,视频数据采集结束后,从相机 SD 卡读取所有视频并转存到计算机硬盘构建视频库。视频库共有湖羊鸣叫行为视频 32 段,采食行为视频 104 段,共 183 GB,其中最长、最短行为视频段时长分别为 25 min 和 53 s。分别从鸣叫、采食行为视频段中随机选择 6 段和 20 段视频数据用于构建羊嘴状态检测模型。从 26 段视频中截取图像帧,经过数据清洗后,获得属于鸣叫和采食行为的图像 1 271 幅和 7 123 幅。采用精灵标注助手软件完成羊嘴状态标注,构建一个包含 8 394 幅图像的羊嘴状态数据集,共 5.5 GB。

1.2 短时咀嚼行为识别

湖羊从食槽叼起草料后通过多次上下颌张合(即短时咀嚼)磨碎草料,然后通过吞咽将磨碎后的草料吞入瘤胃、网胃。从正对羊脸角度观察,羊的一次短时咀嚼行为共有 3 类羊嘴状态:张嘴、上下颌错开和闭嘴。但羊在采食过程中其脸部与相机视角方位关系不固定,张嘴、上下颌错开和闭嘴状态在不同的相机视角下呈现不同的空间特征。引入羊脸与相机的视角方位关系,将湖羊短时咀嚼过程中的羊嘴状态分为如表 1 所示的 7 种类别并分别打上相应标签。

表 1 短时咀嚼行为对应的羊嘴状态类别及标签

Tab. 1 Mouth status categories and corresponding label during short-term chewing behaviour

羊头姿态	羊脸与相机视角方位关系	羊嘴状态(标签)
抬头	侧脸面对相机	侧头张嘴(S1)、侧头闭嘴(S3)
	正脸面对相机	张嘴(PU1)、上下颌错开(PU2)、闭嘴(PU3)
低头	侧脸面对相机	侧头张嘴(S1)、侧头闭嘴(S3)
	正脸面对相机	低头上下颌错开(PD2)、低头闭嘴(PD3)

羊在抬头或低头姿态下发生短时咀嚼行为时,如果其侧脸面对相机,则羊嘴张开与上下颌错开在视频中呈现相同的空间特征,因此,表 1 中将羊侧脸面对相机时的张嘴与上下颌错开状态统一分类为侧头张嘴状态。羊低头正脸面对相机时,从视频中观察不到其张嘴状态下的下颌,张嘴和闭嘴状态具有相同的空间特征,因此,在羊低头正脸面对相机的视角关系下,将张嘴与闭嘴状态统一分类为低头闭嘴状态。另外,羊上下颌的一次张合也可能对应一次鸣叫行为,从时间特征上看,鸣叫对应的上下颌张闭的持续时间要明显长于短时咀嚼行为。本文构建的短时咀嚼行为识别方法利用时空特征识别短时咀嚼及鸣叫行为。

短时咀嚼行为识别属于时序动作识别(Temporal action detection)范畴。目前,Two-Stream[18]、C3D[19]等卷积神经网络展现出了优良的时序动作识别精度,但这类算法直接将运动行为短视频作为输入,网络结构较为复杂且算法运行时间偏长。针对湖羊短时咀嚼及鸣叫行为构成元素——羊嘴状态数量少的特点,本文提出一种基于连续视频帧中羊嘴状态标签序列的短时咀嚼行为识别方法。该方法由 2 个模块构成,第 1 个模块是基于 EfficientDet 网络的羊嘴状态检测,第 2 个模块为基于羊嘴状态标签序列分类的短时咀嚼行为识别。

1.2.1 羊嘴状态检测

涵盖嘴部定位及其状态识别的羊嘴状态检测是构建短时咀嚼行为识别分类器的基础,视频空间中的羊嘴状态检测问题属于基于计算机视觉的目标检测问题范畴。目前常用的目标检测方法主要包括基于预设锚框(Bounding box)的单级检测法[20-21]和基于感兴趣区域候选(ROI proposal)的两级检测法[22-23],前者的典型算法包括 YOLO[20]、EfficientDet[21]等,后者典型算法为 R-CNN 系列。其中,TAN 等[21]提出的 EfficientDet 算法以 EfficientNet[24]作为骨架网络,引入加权双向特征金字塔网络(Bi-FPN)及类别/目标框预测网络(Class/Box prediction net)。

EfficientDet 网络的 Bi-FPN 模块接收 EfficientNet 提取的湖羊视频帧图像特征中的第 3~7 层特征(即图 2 中的 P3~P7),完成多次自上而下及自下而上的双向特征融合。融合后的特征分别送入类别/目标框预测网络以完成目标框筛选以及目标类别识别任务。由于相邻圈羊只的嘴部可能会通过栏架空隙进入到相机视角范围,同时,羊只采食过程中可能出现的羊头快速运动会导致少量视频帧中出现嘴部重影,这些都将导致 EfficientDet 网络在同一视频帧图像中可能检测到多个羊嘴区域。为确保

羊嘴状态检测的准确性，将图像中水平、垂直坐标处于[440 像素，2 440 像素]及[1 160 像素，2 160 像素]范围内的矩形区域（如图 1b 中红色矩形框所示）作为采食区域。在此基础上对 EfficientDet 网络做如下改进：在目标框预测网络输出端增加目标框筛选模块，该模块利用检测结果目标框与采食区域的位置关系筛选目标框。增加目标框筛选模块后的 EfficientDet 网络结构如图 2 所示。

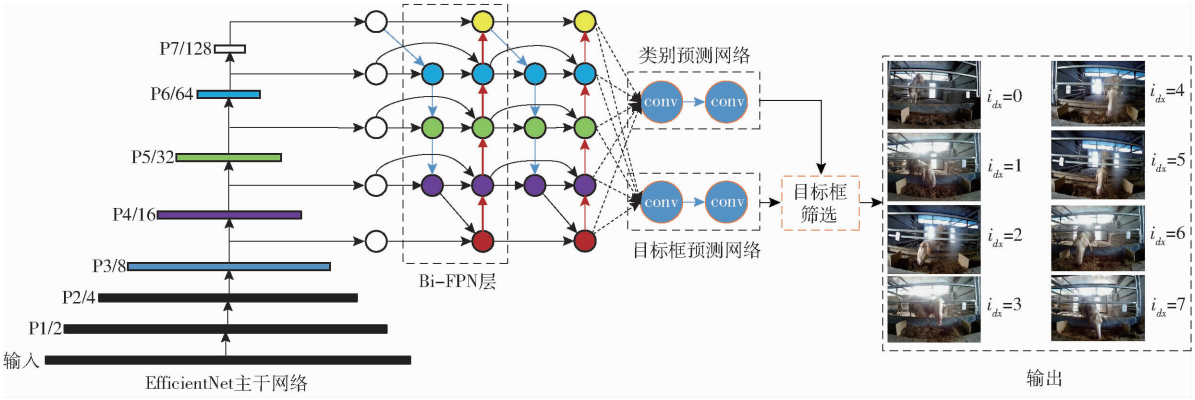


图 2　增加目标框筛选模块的 EfficientDet 结构图

Fig. 2　Architecture diagram of modified EfficientDet with box selection module

图 2 中 i_{dx} 是羊嘴状态编码值，羊嘴状态 S1、S3、PU1、PU2、PU3、PD2、PD3 对应的编码值分别为 1、2、3、4、5、6、7。增加目标框筛选模块后的 EfficientDet 网络仍采用复合尺度参数 φ 对所有模块的深度、宽度和分辨率进行联合缩放，φ 取值范围为 0 ~ 7。EfficientNet 模块的深度、宽度和分辨率缩放公式为

$$D_{\mathrm{EffN}} = \alpha^{\varphi} \tag{1}$$

$$W_{\mathrm{EffN}} = \beta^{\varphi} \tag{2}$$

$$R_{\mathrm{EffN}} = \gamma^{\varphi} \tag{3}$$

其中

$$(\alpha\beta^2\gamma^2)^{\varphi} = 2^{\varphi} \tag{4}$$

式中　D_{EffN}——EfficientNet 模块深度

W_{EffN}——EfficientNet 模块宽度

R_{EffN}——EfficientNet 模块分辨率

α、β、γ——通过小网格搜索确定的不小于 1 的常数

Bi－FPN 的宽度和深度缩放公式为

$$W_{\mathrm{bifpn}} = 64 \times 1.35^{\varphi} \tag{5}$$

$$D_{\mathrm{bifpn}} = 3 + \varphi \tag{6}$$

式中　W_{bifpn}——Bi－FPN 模块宽度

D_{bifpn}——Bi－FPN 模块深度

类别预测网络与目标框预测网络的宽度 W_{pred} 与 W_{bifpn} 相等，深度缩放公式为

$$D_{\mathrm{box}} = D_{\mathrm{class}} = 3 + \varphi/3 \tag{7}$$

式中　D_{box}——目标框预测网络的深度

D_{class}——类别预测网络的深度

输入图像分辨率 R_{input} 的缩放公式为

$$R_{\mathrm{input}} = 512 + 128\varphi \tag{8}$$

式(8)表明，即便 φ 取得最大值 7，R_{input} 仍远小于湖羊采食视频帧图像分辨率。因此，先采用双三次样条插值法在不改变图像宽高比的前提下将湖羊采食视频帧图像分辨率压缩至 $R_{\mathrm{input}} \times \frac{3R_{\mathrm{input}}}{4}$，然后利用在图像上、下方填补空灰色像素（RGB 值为[128，128，128]）的方法将图像调整至满足 EfficientDet 网络对输入图像分辨率的要求。基于改进的 EfficientDet 网络的湖羊嘴部状态检测流程图如图 3 所示。

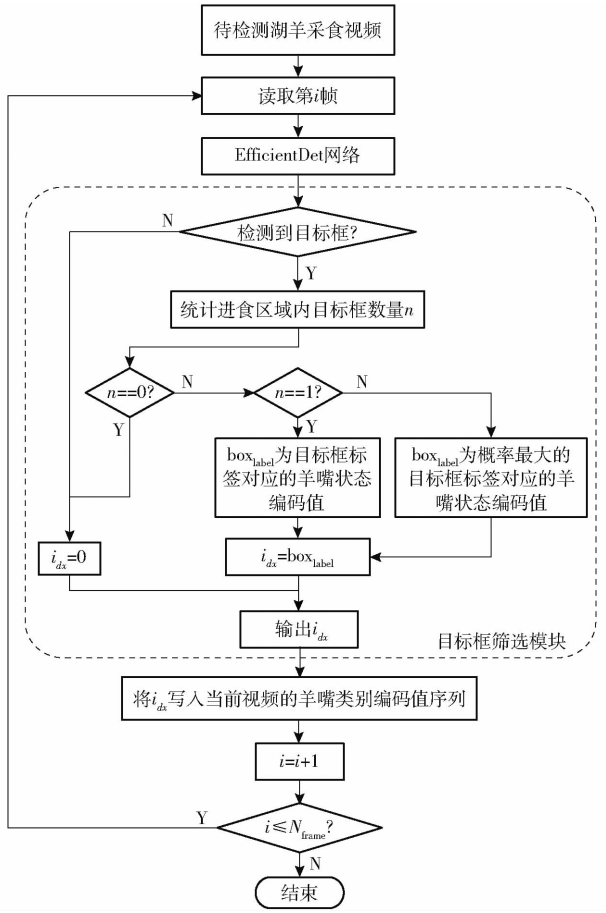


图 3　湖羊嘴部状态检测流程图

Fig. 3　Mouth state detection flow chart

图 3 中，N_{frame} 为待检测的采食视频的总帧数。视频被逐帧送入 EfficientDet 网络，将所得到的类别及目标框检测结果送入筛选模块。首先剔除采食区域外的目标框，然后根据进食区域内目标框数量 n，分 3 种情况确定进食区域内的单个羊嘴状态目标框：若 n 为 0，则网络未在采食区内检测到羊嘴，本帧羊嘴状态类别编码值 i_{dx} 赋值为 0；若 n 为 1，则将类别预测网络输出的羊嘴状态标签对应的编码值赋给 i_{dx}；若 n 大于 1，选择类别预测网络输出的目标框标签最大概率对应的羊嘴状态类别编码值赋给 i_{dx}。采食视频经过改进的 EfficientDet 网络处理后，得到一个与帧顺序对应的 i_{dx} 序列。

1.2.2 羊嘴一次上下颌张合视频段分割

采用 i_{dx} 序列分割方法自动分割羊嘴一次上下颌张合动作对应的连续视频。引入正则表达式(Regular expression)过滤符合羊嘴上下颌一次张合动作所对应的 i_{dx} 序列片段模式，进而将视频分割为多个与上下颌张合相对应的 i_{dx} 序列片段。表 1 数据表明，羊侧脸面对相机时，无论抬头还是低头，其上下颌的一次张合对应的 i_{dx} 均为 1 和 2；湖羊正脸面对相机时，其抬头、低头姿态下短时咀嚼对应的嘴部状态类别存在差异。因此，本文将短时咀嚼动作分为 3 类：侧面短时咀嚼、正面抬头短时咀嚼和正面低头短时咀嚼，各类短时咀嚼行为对应的单次上下颌张合的 i_{dx} 序列片段如表 2 所示。

表 2 一次上下颌张合对应的湖羊嘴状态序列及 i_{dx} 序列

Tab. 2 Sequences of mouth state label and i_{dx} corresponding to single opening-closing of jaw

湖羊头姿态	湖羊嘴状态序列	i_{dx} 序列片段	正则表达式
侧脸面对相机	多帧张嘴、一帧闭嘴	1…2	1 +2
抬头正脸面对相机	多帧张嘴、多帧上下颌错开、一帧闭嘴	3…4…5	3 +4 +5
	多帧张嘴、一帧闭嘴	3…5	3 +5
	多帧上下颌错开、一帧闭嘴	4…5	4 +5
低头正脸面对相机	多帧上下颌错开、一帧闭嘴	6…7	6 +7

表 2 中，由于湖羊侧脸面对相机短时咀嚼和低头正脸面对相机短时咀嚼各仅存在一种嘴部状态序列，所以各自只能匹配一种 i_{dx} 序列模式。分别采用正则表达式“1 +2”和“6 +7”描述湖羊侧脸面对相机和低头正脸面对相机时一次上下颌张合对应的 i_{dx} 序列片段匹配模式，其中，“ +”代表其前面的字符必须至少出现一次。抬头正脸面对相机短时咀嚼存在 3 种嘴部状态序列，其原因是湖羊抬头正脸面对相机短时咀嚼可能会出现以下 2 种特殊情况：湖羊一口咬住过多的饲料使部分嘴部被遮挡，不能完整地提取出张嘴状态的特征，导致未识别到张嘴状态，对应的湖羊嘴状态编码值序列片段是 4…5；湖羊为了将嘴边的饲料全部拢到嘴里，上下颌基本不错开，以大幅度张嘴状态为主，此时对应的湖羊嘴状态编码值序列片段为 3…5。因此，湖羊抬头正脸面对相机的一次上下颌张合对应的 i_{dx} 序列片段匹配模式有 3 类，分别是“3 +5”、“4 +5”及“3 +4 +5”。

利用表 2 所示各正则表达式处理 i_{dx} 序列，提取符合要求的 i_{dx} 序列片段并分行存储在 i_{dx} 序列片段文本中，即 i_{dx} 序列片段文本文件中的每一行对应一个 i_{dx} 序列片段。

1.2.3 短时咀嚼行为识别

湖羊的一次鸣叫同样对应着湖羊上下颌的 1 次张合，其 i_{dx} 序列片段也符合 1…2 或 3…5 的模式。统计分析数据集中鸣叫及短时咀嚼视频段的帧数，结果表明鸣叫行为持续帧数为(49 ± 1.2)帧，远大于短时咀嚼持续帧数((25 ± 5.8)帧)。因此，引入 i_{dx} 序列片段长度阈值区分鸣叫行为和短时咀嚼行为，记该阈值为 *th_lenidxseq* 且设其值为 42。利用 EfficientDet 网络、正则表达式及阈值 *th_lenidxseq* 实现短时咀嚼和鸣叫行为分类的流程如图 4 所示。

图 4 中，j 为 i_{dx} 序列片段文本的行编号，N_{chew} 和 D_{chew} 分别为视频中提取到的短时咀嚼次数和短时咀嚼持续时间，$l_{en(idxseq_j)}$ 为 i_{dx} 序列片段文本第 j 行的序列片段长度。

2 结果与分析

2.1 湖羊嘴状态检测方法测试与结果分析

湖羊嘴状态检测方法测试硬件环境如下：处理器为 Intel Core i9 - 7920X 12 核，2.9 GHz，64 GB 内存，显卡为 GTX NVIDIA TITAN Xp，12 GB。统计 1.1 节获取的湖羊嘴状态数据集的羊嘴状态数量如图 5 所示。

在 Pytorch 深度学习开发平台上实现检测网络，网络优化算法选用带动量和权重衰减的随机梯度下降，动量参数和权重衰减系数分别设为 0.9 和 0.000 5，学习率设为 0.000 1，迭代 125 次。式(1) ~(7)表明，φ 越大，主干网络、Bi - FPN 及目标框预测网络和类别预测网络的分辨率、深度及宽度的特征扩大比例越高，整个网络的检测精确率也越高，但同时计算量也

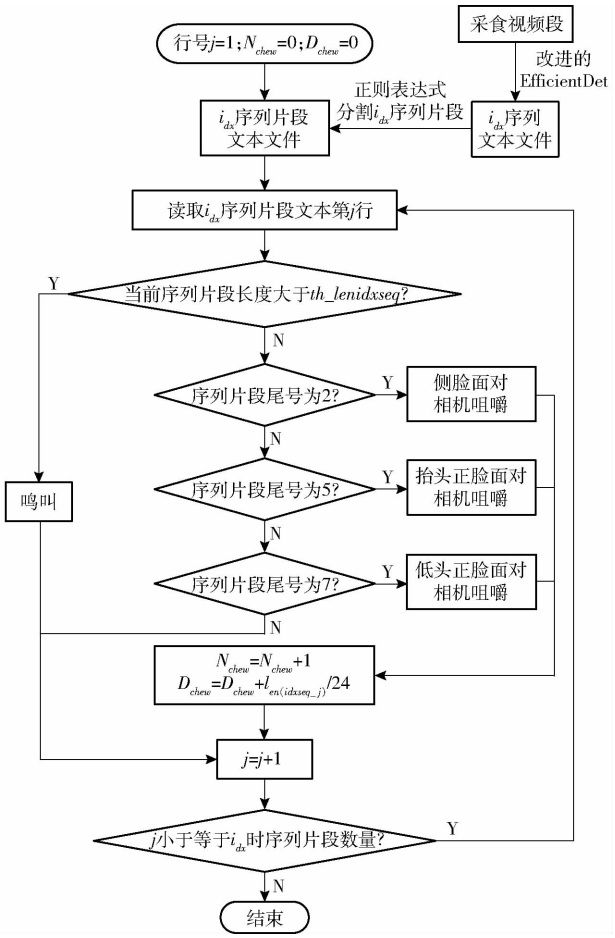


图 4　短时咀嚼与鸣叫行为分类流程图

Fig. 4　Flow chart of short term chewing and chirping classification

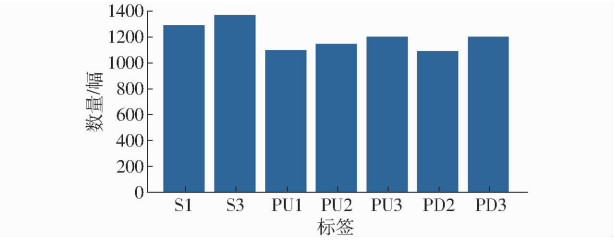


图 5　羊嘴状态检测模型构建数据集组成

Fig. 5　Dataset composition for sheep mouth status detection model establishment

会相应增大。受限于硬件性能，本文仅测试比较 φ 取值 0～4 的网络性能，并与 YOLO v5、SSD 两种主流单阶段检测模型进行性能对比。

将图 5 所示各类羊嘴状态图像按照 8∶2的比例划分为羊嘴状态检测模型构建所需的训练集和验证集。把训练集图像分别输入到 EfficientDet－D0～D4、YOLO v5 和 SSD 以训练检测模型，用验证集图像测试各模型性能。采用精确率（P）、召回率（R）、均值平均精确率（mAP）以及传输速率（FPS）评价模型性能。利用验证集图像测试各模型性能，结果如表 3 所示。

表 3 数据表明，EfficientDet 网络的羊嘴状态检测精确率 P 随着 φ 的增大而增大，EfficientDet－D0 的精确率与 YOLO v5 相当且优于 SSD。在召回率方面，YOLO v5 的召回率最高，SSD 的召回率最低，EfficientDet 网络模型的召回率均在 97% 附近。EfficientDet 网络模型的均值平均精确率均高于 SSD 网络模型，且远优于 YOLO v5 网络。EfficientDet 网络模型检测速度随 φ 的增大而降低，其中，EfficientDet－D2 的传输速率与 YOLO v5 相当，EfficientDet－D4 和 SSD 的传输速率均低于 10 f/s。

表 3　羊嘴状态检测模型性能对比

Tab. 3　Performance comparison of different models for sheep mouth state detection

网络模型	精确率/%	召回率/%	均值平均精确率/%	传输速率/(f·s^{-1})	训练时间/h
YOLO v5	93.50	98.57	65.35	22.58	30
SSD	91.14	96.19	93.49	8.14	36
EfficientDet－D0	93.43	96.85	97.74	36.98	18
EfficientDet－D1	95.64	97.73	98.84	28.18	20
EfficientDet－D2	95.71	97.90	98.79	22.89	27
EfficientDet－D3	96.41	97.47	98.74	15.78	36
EfficientDet－D4	96.58	97.31	98.89	7.56	48

综合考虑检测模型的 P、R、mAP 及 FPS，选择带目标框筛选模块的 EfficientDet－D1 网络模型检测结果作为后续短时咀嚼行为识别的输入。为测试 EfficientDet－D1 网络模型在检测 7 种不同羊嘴状态时的性能表现，使用训练好的带目标框筛选模块的 EfficientDet－D1 网络模型检测一段时长 1 min 的采食视频，结果如图 6a 所示。

图 6a 中每条横虚线下面是视频帧实际类别，横

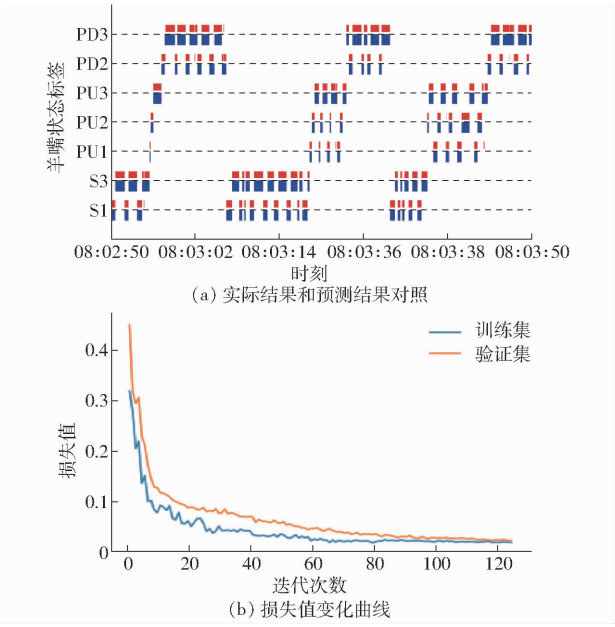


图 6　不同类别羊嘴状态的检测结果

Fig. 6　Comparison of detecting results of sheep mouth status using modified EfficientDet－D1

虚线上面是模型预测的类别结果。图 6a 表明,改进的 EfficientDet - D1 模型可以准确检测出湖羊嘴部状态,且保证了各个状态准确检测时间的连续性。图 6b 所示的模型训练集和验证集的损失函数曲线表明,经过约 100 轮迭代后,验证集的损失函数曲线开始逼近训练集损失函数曲线,且两条曲线均在[0, 0.1]之间收敛。

2.2 短时咀嚼行为识别正确率分析

从 1.1 节获得的视频库(除去构建训练集、验证集的视频)中随机选择 2 段鸣叫视频和 30 段采食视频,通过 Windows Media Player 软件播放视频,人工观察、记录视频中的短时咀嚼行为发生的次数和持续时间分别为 1 400 次和 1 335.22 s。使用带目标框筛选模块的 EfficientDet - D1 网络结合正则表达式的行为识别方法自动提取短时咀嚼行为发生次数及持续时间分别为 1 280 次和 1 213.89 s。这表明,改进的 EfficientDet - D1 网络结合正则表达式的短时咀嚼行为识别方法能在存在鸣叫行为干扰的情况下,短时咀嚼行为发生次数自动识别正确率为 91.42%,算法自动提取到的短时咀嚼行为持续时间与人工提取结果平均误差为 9.15 s,平均正确率为 90.85%。

误差产生的原因主要有 2 方面:一是基于改进的 EfficientDet - D1 网络不能够以 100% 的精确率检测羊嘴状态。少量图像帧中羊嘴状态的错误检测结果可能导致算法识别的短时咀嚼行为的提前或延迟结束,也可能导致算法未能准确识别短时咀嚼的起始帧或结束帧,进而导致少量完整的咀嚼行为未被正确识别,使累积咀嚼次数和时长的统计出现误差。二是作为参照标准的短时咀嚼次数和咀嚼时长通过人工观察视频方法提取,不可避免地带有一定的主观性以及误差。

本文方法能够自动识别每栏饲养一只湖羊条件下的湖羊短时咀嚼行为。若将本文方法应用到每栏饲养多只湖羊的场景,则需研究羊嘴状态检测网络输出多目标框的筛选方法,然后解决同一羊嘴目标在连续视频帧中标识与跟踪问题。在此基础上,可以实现单栏多只湖羊的总短时咀嚼次数及时间的统计,可用于单栏多只湖羊采食量的估算。

3 结论

(1)将从视频流中自动识别湖羊短时咀嚼行为的问题转换为羊嘴状态类别编码值序列的自动分类问题,提出了一种基于改进 EfficientDet 网络和正则表达式的湖羊短时咀嚼行为自动识别方法。首先,在 EfficientDet 网络基础上增加目标框筛选模块,实现羊嘴状态自动检测;然后,利用正则表达式提取连续视频帧羊嘴状态类别编码值序列;最后利用编码值序列的结构特征和长度特征构建短时咀嚼行为识别规则,避免鸣叫行为对短时咀嚼行为识别的干扰。

(2)利用本文方法对舍饲环境下采集的湖羊采食视频自动识别短时咀嚼行为的测试结果表明,改进的 EfficientDet - D1 网络在保证检测速度的同时,对羊嘴状态检测精确率达到 95.64%,均值平均精确率为 98.84%,传输速率为 28.18 f/s,优于 YOLO v5 和 SSD 网络。对短时咀嚼次数自动识别正确率和持续时间平均正确率分别为 91.42% 和 90.85%,实现了湖羊短时咀嚼行为的自动检测。本文方法将基于视频的湖羊短时咀嚼行为识别问题转换为羊嘴状态编码值序列分类问题,降低了分类模型的复杂度,为湖羊短时咀嚼行为的自动识别提供了一种新的研究思路。

参 考 文 献

[1] 张玉荣. 湖羊养殖技术要点[J]. 农村新技术, 2017(1):24 - 26.
ZHANG Yurong. Key points of Hu sheep breeding technology[J]. New Rural Technology, 2017(1):24 - 26. (in Chinese)

[2] 任贵福. 舍饲羊多发病原因及防治[J]. 种植与养殖,2005(7):42.
REN Guifu. Causes and prevention of frequentness of sheep in house feeding[J]. Planting and Breeding, 2005(7):42. (in Chinese)

[3] 任晓惠,刘刚,张淼,等. 基于支持向量机分类模型的奶牛行为识别方法[J]. 农业机械学报,2019,50(增刊): 290 - 296.
REN Xiaohui, LIU Gang, ZHANG Miao, et al. Dairy cattle's behavior recognition method based on support vector machine classification model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 290 - 296. (in Chinese)

[4] 王奎,武佩,宣传忠,等. 放牧家畜牧食信息监测的研究进展[J]. 南京农业大学学报,2020,43(3): 403 - 413.
WANG Kui, WU Pei, XUAN Chuanzhong, et al. Progress in monitoring the grazing information of livestock[J]. Journal of Nanjing Agricultural University, 2020, 43(3): 403 - 413. (in Chinese)

[5] GIOVANETTI V, DECANDIA M, MOLLE G, et al. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer[J]. Livestock Science, 2017, 196: 42 - 48.

[6] CHELOTTI J O, VANRELL S R, MILONE D H, et al. A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle[J]. Computers and Electronics in Agriculture, 2016, 127: 64 - 75.

[7] ROMBACH M, MÜNGER A, NIEDERHAUSER J, et al. Evaluation and validation of an automatic jaw movement recorder (RumiWatch) for ingestive and rumination behaviors of dairy cows during grazing and supplementation[J]. Journal of Dairy Science, 2018, 101: 2463 - 2475.

(下转第 426 页)