

田泽薇, 张云伟, 陈 瑶. 基于双路径网络的四足动物运动参数提取方法[J]. 江苏农业学报, 2022, 38(2): 403-413.  
doi: 10.3969/j.issn.1000-4440.2022.02.014

## 基于双路径网络的四足动物运动参数提取方法

田泽薇<sup>1</sup>, 张云伟<sup>1,2,3</sup>, 陈 瑶<sup>1</sup>

(1. 昆明理工大学信息工程与自动化学院, 云南 昆明 650500; 2. 昆明理工大学云南省人工智能重点实验室, 云南 昆明 650500; 3. 昆明理工大学云南省计算机技术应用重点实验室, 云南 昆明 650500)

**摘要:** 为了实现畜牧业集约化、规模化、精准化管理中及时分析动物的运动姿态, 针对不同环境中动物关节点识别准确率较低的问题, 提出了一种双路径网络(Dual path net, Dpnet)特征提取方法。首先, 在HRnet(High resolution net)网络输出分支后面添加混合空洞卷积层(Hybrid dilated convolution, HDC)和并行注意力模块(Parallel attention module, PAM)以识别关键点坐标。然后, 利用模型生成的热力图与本研究设定的精确圈进一步精准定位关键点位置, 通过斜率法区分并匹配动物四肢的关节点归属, 从而获取正确的骨架结构。最后, 模型通过骨架结构提取并分析了不同速度下动物的同侧步距、步频、关节角度与步态占空比4种运动参数。试验结果表明, 该方法中模型识别分类动物5类关键点的精确率为91.1%, 召回率为91.0%, 5类关键点的平均相似度为87.0%, 可有效监测农场或牧场犬类、牛群、羊群等个体的健康状况。

**关键词:** 四足动物; Dpnet; 斜率法; 骨架结构; 运动参数

**中图分类号:** TP391.4 **文献标识码:** A **文章编号:** 1000-4440(2022)02-0403-11

## Extraction of quadruped motion parameters based on dual path network

TIAN Ze-wei<sup>1</sup>, ZHANG Yun-wei<sup>1,2,3</sup>, CHEN Yao<sup>1</sup>

(1. College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China; 3. Yunnan Key Laboratory of Computer Technology Application, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** In order to achieve more intensive, large-scale and precise management of the animal husbandry, timely analysis of animal movement posture is a key link in animal breeding. Aiming at the problem of low recognition accuracy of animal joint points in different environments, a feature extraction method based on dual path network (Dpnet) was proposed. Firstly, the hybrid dilated convolution (HDC) and the parallel attention module (PAM) were added after the output branches of the high-resolution network (HRnet) to identify the coordinates of key points. Secondly, the heat map generated by the model and the precise circle set in this study were used to further accurately locate the positions of key points, the slope method was used to distinguish and match the joint points of the animal limbs, so as to obtain the correct skeleton structure. Finally, four motion parameters of animals (ipsilateral stride, stride frequency, joint angle and gait duty cycle) were extracted and analyzed by the model at different speeds through the skeleton structure. The experimental results showed

收稿日期: 2021-06-19

基金项目: 国家自然科学基金项目(51365019)

作者简介: 田泽薇(1997-), 女, 黑龙江大庆人, 硕士研究生, 主要从事计算机视觉、智能化研究。(E-mail) 1537341112@qq.com

通讯作者: 张云伟, (E-mail) zhangyunwei72@gmail.com

that the accuracy rate of the model to identify and classify the key points of the animal was 91.1%, the recall rate was 91.0%, and the average similarity of the five key points was 87.0%, indicating that this method could effectively monitor the health status of dogs, cattle, sheep and other individuals on farms or ranches.

**Key words:** quadruped; dual path network (Dpnet); slope method; skeleton structure; motion parameters

在大型农场犬类活动时或牧场放牧时为了能够及时发现动物的健康隐患,通常有员工负责定期查看动物的活动情况,但人工观察具有主观性且实时性不高,并且具有人力消耗过大和效率较低的缺点,故精确识别动物的步态状况对于农场和牧场养殖有重要意义。

在动物个体或行为识别中大致分为基于模型和基于非模型的 2 种方法。在基于模型的方法中, Nasirahmadi 等<sup>[1]</sup>和 Chen 等<sup>[2]</sup>构建椭圆和矩形模型区分猪的行为,利用模型质点之间的距离判断各种行为。朱家骥等<sup>[3]</sup>把人体星状图模型引入到动物上,采用频谱分析方法计算求得猪的前肢频率参数。在基于非模型的方法中, Ding 等<sup>[4]</sup>定义了一个包括角度特征和距离特征的 219 维向量,将规则学习方法与 Bagging 和随机子空间方法一起使用,提高关节点识别准确率。Jiang 等<sup>[5]</sup>提出了 FLYOLOv3 网络,通过标记母牛的身体部分边界作为样本,测试不同关键部位。Juan 等<sup>[6]</sup>提出了一种基于卷积神经网络的系统来改善总体身体状况评分(BCS)的自动估计方法,对奶牛健康进行整体评价。王浩等<sup>[7]</sup>通过改进的 Faster-CNN 识别猪圈内位置,采用 PNPool 方法判断猪所在区域。

在动物步态分析方面,钱建轩<sup>[8]</sup>用微软 Kinect 相机采集猪只行走的深度图像,二值化提取猪骨架并对端点进行排序,使用最近邻分类器识别猪正常步态和跛行步态。张满囤等<sup>[9]</sup>对奶牛视频序列进行前景提取,细化步态序列得到奶牛骨架图,然后对生成的步态能量图提取特征。Song 等<sup>[10]</sup>通过 NBSM 模型分割奶牛目标区域,再利用 LCCCT 模型拟合头颈部的直线斜率,用最近邻分类技术对奶牛跛行程度进行分类。Gutierrez-Galan 等<sup>[11]</sup>建立了嵌入式前馈神经网络,通过无线网络和智能项圈分类马的三种步态。Zhao 等<sup>[12]</sup>通过形状匹配方法计算奶牛四肢距离从而分析步态顺序、速度等参数,使用决策树分类器将 3 种跛行类别用箱型图表示。在动物行为识别方面,张苏楠等<sup>[13]</sup>利用帧差法获取生猪移动像素,再对一次性多类物体检测器(SSD)进行改进,检测剧烈运动的个体。Yang 等<sup>[14]</sup>基于光流帧中提取的空间时间信息对猪只在不同体型下的饲养行为进行识别。

Zheng 等<sup>[15]</sup>建立了检测器 Faster R-CNN 用以识别母猪哺乳期的母性行为。Wu 等<sup>[16]</sup>通过视觉几何群网络(AGG16)提取特征序列,提出 Bi-LSTM 分类模型提取语义信息,识别分类奶牛 5 种行为。Yang 等<sup>[17]</sup>建立了自动识别母猪行为的框架,并基于光流和时空视频技术获取时间和空间信息特征,输入到分层分类器。Guo 等<sup>[18]</sup>提出颜色和纹理特征的背景提取(BSCTF)方法进行奶牛区域检测,再通过光流特征分析法识别奶牛坐骑行为。

目前动物行为识别研究较为成熟,但对于动物具体步态的分析还存在以下问题:自然环境中不同光照、场景变换与拍摄不同角度造成的干扰,以及研究对象本身四肢皮毛的相似性与多目标遮挡带来的关键点定位误差。本研究基于 Dpnet 网络提取四足动物的骨架模型,在此基础上进行空间参数与时间参数分析,以实现步态特征提取,为无人养殖场的监控提供一种有效方法。

## 1 材料与方法

### 1.1 试验数据集

1.1.1 试验数据采集与预处理 使用 Kaggle 网站上公共 dataset 中的 animals-6 图片集做迁移学习,在其中选取象(elephant)、羊(sheep)、马(horse)、奶牛(cow)、狗(dog)、猫(cat)共 6 种四足动物的样本,同时使用金乾象科技的双目相机 KS4A418-D 拍摄猫、犬类与羊奔跑运动状态的视频,视频中图片分辨率为 1 270×720,相机与所拍摄对象基本持平且距离 5~8 m,拍摄角度在 ±270° 之间。在早上 8:00–12:00、下午 13:00–18:00 共拍摄 10 段 1 min 左右的视频,为期 7 d,包含不同的街道环境和路面环境,随机截取并组合成 2 min 视频样本,通过帧分解技术得到 2 920 张原始图像。为增强试验鲁棒性和模型泛化能力,对原始图像做空间翻转并在随机方向旋转 35°、95° 与 155°,进行灰度变换并增加部分方差为 0.02 的高斯噪声使视频样本图像扩张到 4 000 幅,以实现数据增强的目的。将上述 animals-6 样本与视频样本共同组成试验数据集,按照训练集、测试集、验证集 3:1:1 的比例进行分配用于后面网络模型的测试与改进。数据集的组成与分配如表 1 所示。

表 1 试验数据集构成

Table 1 The composition of experimental data set

样本分类	图像数量	所含内容	遮挡率(%)
训练集	10 970	单目标	2
测试集	2 657	单目标	2
验证集	3 657	单目标	2
无遮挡的测试集	500	2 只或多只目标	0
多目标的测试集	500	2 只或多只目标	10~50

1.1.2 数据集标签制作 使用开源工具 labelme 对数据集进行关键点标注,基于四足动物奔跑状态时四肢与脊柱的变化幅度比较大这一特点,本研究共标注与该状态相关联的 14 个点进而对运动参数进行分析。标注的图像如图 1 所示,对应图 2 中 13 个关键点。标注的关键点中 Group ID 分为 3 类:数字 0,没有被遮挡但不在图中显示的点;数字 1,在图中显示但被遮挡住的点;数字 2,在图中既显示又没有被遮挡的点。对于第一类情况对应点会自动置为零,对于后 2 种情况会得到标注点的坐标。关键点按序相连的骨架结构如图 2 所示,标注完成后关键点的信息保存在.json 格式文件中,通过该文件中存取的标签数据进行处理转换为一行数组,完成后续对标注关键点的检测。

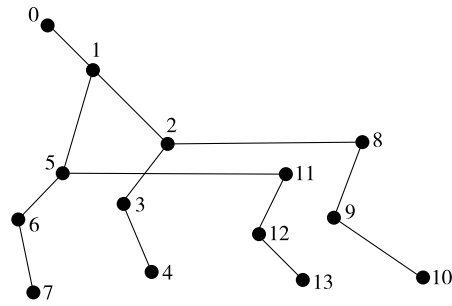


图 1 labelme 标注的数据集示例

Fig.1 Example of data set labeled by labelme

## 1.2 模型结构

1.2.1 HRnet 高分辨率基础网络 基于提取运动目标特征参数需要精准定位关节点的特点,选择自下而上的方法,即先获得四足动物的关节点位置再逐次连接形成骨架模型。其中为了解决网络层数加深伴随下采样后分辨率大幅降低的缺点,以 HRnet<sup>[19]</sup>



关键点 1~13 分别为:嘴、颈、左前髋关节、左前膝关节、左前踝关节、右前髋关节、右前膝关节、右前踝关节、左后髋关节、左后膝关节、左后踝关节、右后髋关节、右后膝关节、右后踝关节。

图 2 四足动物骨架结构

Fig.2 The skeleton structure of tetrapod

作为基础网络,将原始图片预处理为 $256 \times 256$  统一大小输入网络,该网络中包含 4 个模块,每个模块的残差单元从 1 往上依次增加一个,其中模块 3 和模块 4 分别在网络中重复 4 次和 3 次,HRnet 网络各模块信息如表 2 所示。可以看出,高分辨率网络的 4 个模块分辨率逐渐降低的同时通道数逐渐增多,此网络后一支路的输出融合前一对应支路的输入与前面所有输入的特征信息,使得图像特征分辨率始终维持在高水平上,确保模型获取较多的深层信息从而提高最终识别效果。

表 2 HRnet 高分辨率网络各模块信息

Table 2 The information of high resolution network modules

模块	分辨率	通道数
1	1/4	32
2	1/8	64
3	1/16	128
4	1/32	256

1.2.2 Dpnet 双路径网络 基于 HRnet 网络在保持高分辨率的情况下使用常规卷积核  $3 \times 3$  操作提取特征这一设计,重复大量的相同尺寸卷积操作会减弱图片像素点之间的空间关联性。针对 HRnet 模型获取的感受野片面并有固定性的情况,在 HRnet 基础上进行改进,考虑到网络中各层低分辨率逐次交替融合为最终高分辨率的完整性与连贯性,选择在 HRnet 输出的最后 4 个分支后面添加 2 条路径,第 1 条路径设置为 3 层混合空洞卷积,第 2 条路径设置为含有通道和空间 2 种注意力机制的模块。



1.2.2.1 混合空洞卷积(HDC)<sup>[20]</sup> 基于第 1 分支与第 2 分支通道数较少且上下文全局信息较多的表现,选择在前 2 个输出分支后面添加混合空洞卷积(HDC)层,前 2 分支分辨率较高故含有的浅层特征较多,因此适合改进卷积操作来丰富感受野提取的背景轮廓等信息。在第 1 条路径中共由 3 层卷积操作组成,假设空洞卷积率(Dilation rate)为  $d$ ,网络中原始卷积核尺寸为  $k \times k$ ,那么空洞卷积核大小  $n$  为

$$n = k + (k-1) \times (d-1) \quad (1)$$

由于空洞卷积在扩大特征信息感受野的同时会出现特征像素读取失去连续性的现象,参与卷积运算的像素点均为膨胀后卷积核的有效权重位置,这样会引起栅格效应,丢失部分有价值的特征信息从而影响到最终网络分类预测的准确性。因此采用不同膨胀系数的空洞卷积并将 3 层卷积操作结合为一条路径,其中为了最大限度地填充丢失部分的像素点从而确保信息表征的全面性, $d$  的选择尤为重要,需满足在当前层的空洞卷积率不大于网络原始卷积核尺寸,即满足条件

$$M_i \leq k \quad (2)$$

$$M_i = \text{Max}[M_{i+1} - 2d_i, M_{i+1} - 2(M_{i+1} - r_i), d_i] \quad (3)$$

式(3)中  $M_i$  为第  $i$  层的最大膨胀系数,  $d_i$  为第  $i$  层的空洞卷积率,  $\text{Max}$  为取 3 层中的最大值。令中间层  $M_2 \leq 3$ , 可得到各层的空洞卷积率  $d_1 = 1, d_2 = 2, d_3 = 5$ 。由此可改进前 2 个分支中普通常规卷积层带来的特征表征能力较弱与信息全局性较差两方面。HDC 层的结构如图 3 所示。

1.2.2.2 并行注意力模块(PAM) 在网络前两分支上下文语义信息得到增强的同时保证感受野的特征提取范围,从而有利于网络全局信息中浅层特征的获取。同时在网络第 3 分支与第 4 分支的输出后面添加 PAM,基于后 2 分支通道数较多所携带的图像细节信息较全面的特点,设置通道注意力机制与空间注意力机制相融合,从而为不同通道的表征和像素的位置分配相应的权重,使得与分类有关的像素点信息在网络训练中更加突出,PAM 的组成图如图 4 所示。

图 4 中后 2 分支的输出特征经过最大池化、平均池化与中值池化的操作后将得到的特征向量进行通道和空间的权值计算。先采用  $1 \times 1$  的卷积调整通道数为 256 并进行数据批标准化,在 2 种注意力机制组合输出的特征向量后降低维度,最后通过 sig-

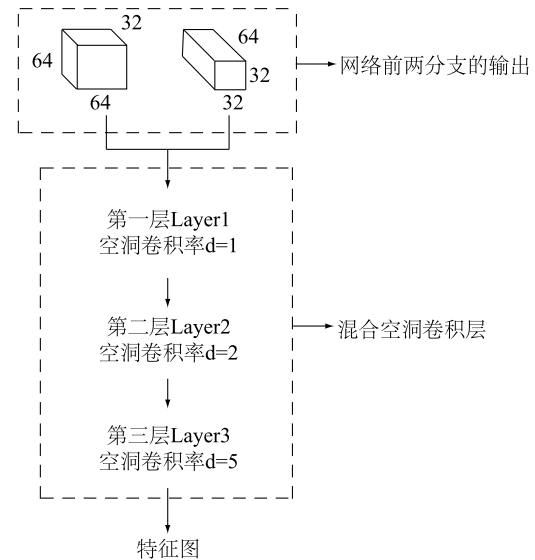


图 3 混合空洞卷积(HDC)层结构图

Fig.3 Layer structure diagram of hybrid dilated convolution (HDC)

moid 激活获取并分配 2 个分支 0~1 的权重值,生成的通道与空间注意力图计算如下:

$$Out_{channel} = W[MLP * (in^{Avgpool} + in^{Maxpool} + in^{Medpool})] \quad (4)$$

$$Out_{spatial} = W[f * (in^{Avgpool}, in^{Maxpool}, in^{Medpool})] \quad (5)$$

$$Out_{branch} = Out_{channel} \otimes in + Out_{spatial} \otimes in \quad (6)$$

式中  $Out_{channel}$  为通道注意力图,  $Out_{spatial}$  为空间注意力图,  $W$  为 sigmoid 激活函数,  $MLP$  为多层感知机(multilayer perception)用以共享权值,  $f$  为  $3 \times 3$  的卷积操作,  $in$  表示后两分支的输出特征,  $Out_{branch}$  为通道权重和空间权重分别与输出特征对位相乘再相加得到的最终分支特征。

第 1 条路径改进了特征表征的固定性,同时第 2 条路径增强了不同特征提取的针对性。2 条路径经过特征融合将低级特征弥补上语义信息同时把高级特征增添上空间信息,从而提高模型输出准确度, Dpnet 网络结构图如图 5 所示。

1.2.3 关键点定位 模型最终输出的热力图是基于关键点高斯响应峰值的位置决定的,因此每个关键点置信度的计算对该点的坐标产生极大的影响。对标注动物的 14 个关节点设置精确圈来约束热力图中的关键点所在位置,即判断热力图中对应每一类关键点的像素位置是否在精确圈内,即满足条件

$$T_k(x_i) = \begin{cases} 1, & \|x_i - x_k\| < r \\ 0, & \text{else} \end{cases} \quad (7)$$

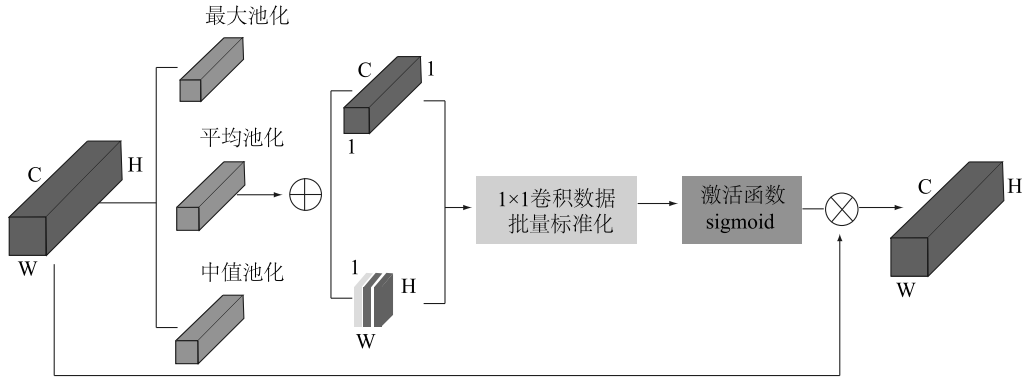


图4 并行注意力模块(PAM)组成图

Fig.4 Composition diagram of parallel attention module (PAM)

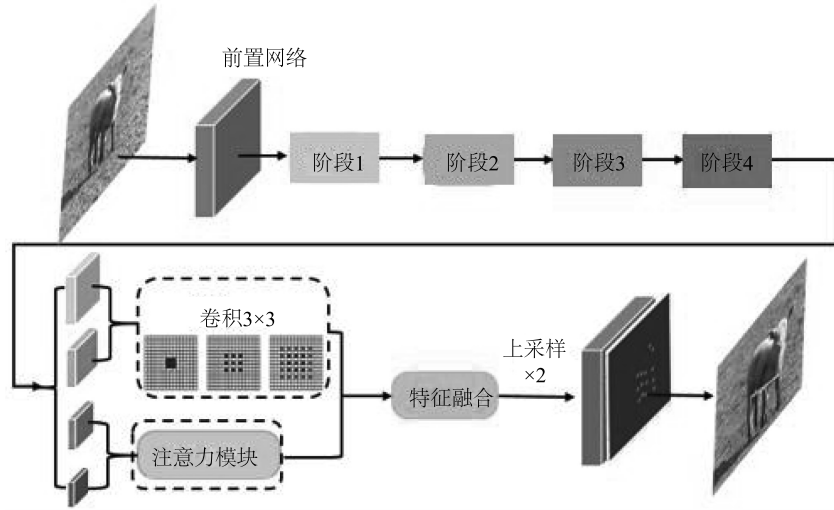


图5 Dpnet 双路径网络结构图

Fig.5 Structure diagram of dual path network (Dpnet)

式中  $x_i$  表示热力图上第  $i$  个关键点的像素位置坐标,  $x_k$  表示属于第  $k$  类的关节点实际应在的位置坐标,  $r$  设置精确圈的半径大小为 20 像素, 若生成热力图上的关节点与实际位置坐标的距离在  $r=20$  以内, 则第  $i$  个像素符合精确圈范围。将检测出的关键点高斯响应最大值与精确圈结合生成定位准确度更高的热力图, 计算式为

$$l_k(x_i) = \sum_{j=1}^n \frac{1}{\pi r^2} G(x_j^k + \|x_i - x_k\|) T_k(x_i) \quad (8)$$

式中  $j=(1, 2, \dots, n)$  为生成热力图上关键点所在像素位置的序号,  $x_j^k$  为  $j$  像素位置属于第  $k$  类关节点的高斯响应峰值,  $G(x_j^k + \|x_i - x_k\|)$  为双线性插值的核,  $\frac{1}{\pi r^2}$  的相乘系数为确保以  $r$  为半径的精确圈外

label 是 0。通过计算每类关节点所在像素位置 2 个方向的插值聚合为更精准的坐标, 过程如图 6 所示。

### 1.3 四肢区分与匹配

在一个运动周期中, 由于四足动物四肢的有序交替摆动特征, 会产生关节坐标丢失的现象, 即前双腿、后双腿、同侧双腿间有关节相互遮挡的现象。模型由此提取出的骨架结构会形成空间交叉的误判, 从而影响后续运动特征参数的提取。故本研究提出一种斜率法来解决四肢同侧区分与异侧匹配的问题。

假设  $S_i^j (i=1, 2, 3, 4)$  表示第  $i$  拍步态时第 1 肢踝关节与图像原点之间的斜率, 随着一个步态周期的完成,  $S_1^i, S_2^i, S_3^i, S_4^i$  四肢的斜率也完成一次周期变化。以动物四肢编号按左前肢 LF(left forelimb) - 右

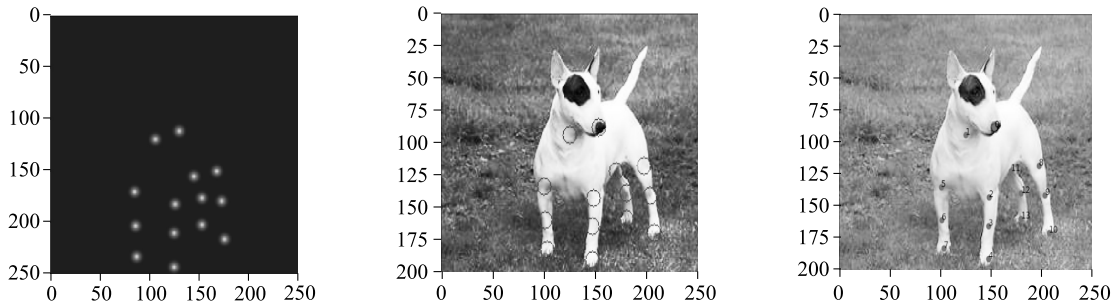


图 6 动物关节点定位过程图

Fig.6 Process map of animal joint point positioning

前肢 RF(right forelimb) -左后肢 LH(left hindlimb) - 右后肢 RH(right hindlimb) 的顺序为前提分析, 对应的四肢为  $L_1^i, L_2^i, L_3^i, L_4^i$ 。具体分析包括以下步骤:

Step1: when  $d_{\min}(L_1^i, L_3^i) \&\& d_{\min}(L_1^i, L_4^i)$

if  $|S_1^i - S_3^i| < |S_1^i - S_4^i|$ , then  $L_1^i$  and  $L_3^i$  are on the same side

else,  $L_1^i$  and  $L_3^i$  are on the same side

Step2: if  $|S_1^i - S_2^i| \rightarrow 0$ , then  $L_1^i = LF$  and  $L_2^i = RF$ ,  $L_3^i = LH$  and  $L_4^i = RH$

else,  $L_1^i = LF$ , and  $L_4^i = RF$ ,  $L_3^i = LH$  and  $L_2^i = RH$

if  $|S_3^i - S_4^i| \rightarrow 0$ , then  $L_3^i = LH$  and  $L_4^i = RH$ ,  $L_1^i = LF$  and  $L_2^i = RF$

else,  $L_3^i = LH$ , and  $L_2^i = RH$ ,  $L_1^i = LF$  and  $L_4^i = RF$

其中 step1 是判断四肢中的两肢是否同侧匹配, step2 则在 step1 基础上区分同侧两肢的前后关系, 完成四肢的归属。通过分析四肢踝关节-原点的斜率变化对缺失的关节坐标分配到相应四肢, 可以正确区分同异侧肢并匹配四肢关节点, 得到的关键点匹配效果如图 7 所示。

## 2 结果与分析

使用 Windows 10 操作系统进行试验的配置与优化, 试验硬件条件为 AMD 锐龙 R5 3700x CPU 处理器和 NVIDIA RTX2080TI 显卡。软件条件为 Python3.5, 框架为 Tensorflow1.4, 编辑器为 Pycharm。对网络模型设置 100 个训练周期, batch\_size = 30, 网络初始学习率为  $1 \times 10^{-3}$ , 网络训练采用 Adam 优化方法。

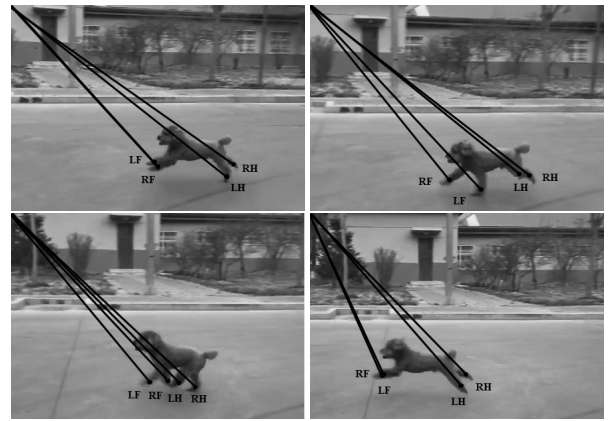


图 7 动物四肢的区分与匹配结果

Fig.7 Discrimination and matching results of animal limbs

### 2.1 模型预测分类结果

为进一步验证模型分类性能, 将设定网络输出的 14 个关键点概括划分为 5 大类, 分别是嘴、颈、髋关节、膝关节、踝关节, 并在测试集中随机选出 800 张单目标、100 张多目标无遮挡、100 张多目标有遮挡的图片。通过 3 个评价指标 [关键点相似度 (OKS)、精确率、召回率] 来分析每类关键点的预测值与真值之间的相似度和准确度, 并通过混淆矩阵可视化模型分类正确与错误的图片数量。其中 3 个评价指标的计算式为

$$OKS = \frac{\sum_i \exp\left(\frac{-d_i^2}{2s^2\sigma_i^2}\right)\delta}{\sum_i \delta} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

式(9)中  $d_i$  为模型预测点与实际关键点间的欧



氏距离,  $s$  为四足动物的图像所占像素面积,  $\sigma_i$  为关键点预测值与真实值的标准差;  $\delta$  的作用是保证标注了的关键点才能参与计算,  $\delta=0$  表示关键点不可见未遮挡,  $\delta=1$  表示关键点不可见被遮挡,  $\delta=2$  表示关键点可见未遮挡。式(10)和(11)中  $TP$  为实例中在属于 5 种类别的情况下预测类别正确的数量,  $FP$  为把不属于某一类的实例预测为该类的数量,  $FN$  为把属于某一类的实例预测为其他类的数量。模型分类预测的混淆矩阵如图 8 所示。

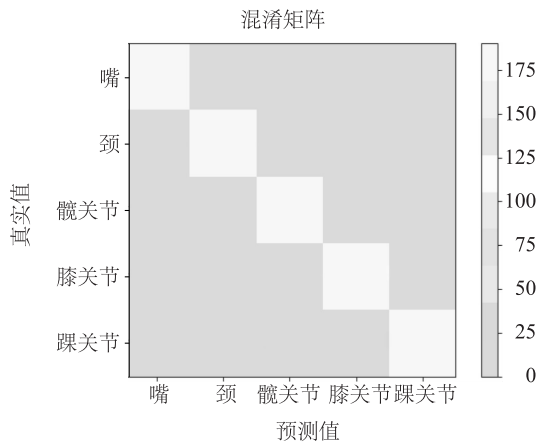


图 8 双路径网络 Dpnet 分类结果混淆矩阵

Fig.8 Confusion matrix of Dpnet classification results

通过混淆矩阵对模型最后 softmax 归一化输出的结果进行统计与分析, 由图 8 可看出在矩阵中对角线上的数值最大且颜色最深, 对角线上的颜色越深说明分类结果越准确, 体现了模型预测的有效性。对比各类关键点分类的精确率与召回率(表 3)可看出对于图片中特征明显度较高的关节位置分类精确率较高, 如标注嘴部和颈部的关键点由于  $\pm 270^\circ$  间两类关键点始终处于较易于观察的情况, 故模型预测的效果也比较好; 而髋关节、膝关节这类关键点易受多目标的遮挡影响和背景环境的干扰, 故模型识别准确度较低。由此得到的骨架图如图 9 所示。

表 3 双路径网络 Dpnet 测试分类的准确度

Table 3 Accuracy of Dpnet test classification

关键点	评价指标		
	关键点相似度	精确率(%)	召回率(%)
嘴	0.86	91.6	93.0
颈	0.87	91.8	90.5
髋关节	0.89	88.1	88.1
膝关节	0.88	89.9	89.1
踝关节	0.85	93.9	94.9
平均值	0.87	91.1	91.0

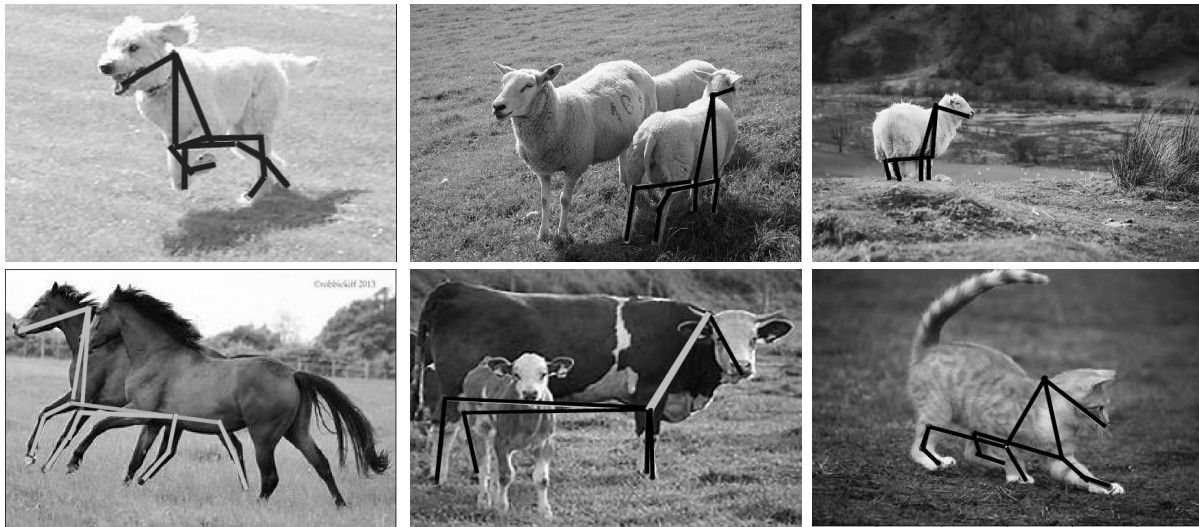


图 9 试验动物骨架图的可视化

Fig.9 Visualization of skeleton diagram of experimental animal

本研究模型最终的参数总量为 28.69 M, 参数利用率为 98%。为验证双路径网络的优越性, 本研究将只设置四分支的 HDC 层或 AM 注意力模块与两大分支加权融合的方法作对比, 结果如表 4

所示。

## 2.2 四足动物步态特征参数

选取猫、狗、羊分别以 20 km/h、30 km/h、40 km/h 速度奔跑的运动视频进行测试。

表 4 不同方法试验结果对比

Table 4 Comparison results of different methods

方法	关键点相似度	精确率 (%)	召回率 (%)
混合空洞卷积(HDC)	0.84	89.4	89.6
并行注意力模块(PAM)	0.86	90.2	89.9
双路径网络(Dpnet)	0.87	91.1	91.0

$$d = \frac{1}{4} \sqrt{[(x_{f_{i+1}}^{LF} - x_{f_i}^{LF})^2 + (y_{f_{i+1}}^{LF} - y_{f_i}^{LF})^2 + \otimes + (x_{f_{i+1}}^{RH} - x_{f_i}^{RH})^2 + (y_{f_{i+1}}^{RH} - y_{f_i}^{RH})^2]} * \frac{h}{H} \quad (12)$$

式中  $\frac{h}{H}$  为图片像素与实际距离的转换系数,将动物四肢在一个周期内前进的距离求和并取平均值,得到的结果如表 5 所示。由表 5 可知猫在奔跑过程中的步距小于犬的同侧前进距离,由于猫的体型较小且四肢较短,运动起来时步态轻盈故与犬身体的延展性相比较弱,对于其他大型四足动物如羊等同侧步距参数较犬类等小中型动物要多出 1 m 左右。

2.2.2 步幅频率参数 四足动物的步频可通过同侧肢踝关节的相对距离变化来计算。本研究选取 2

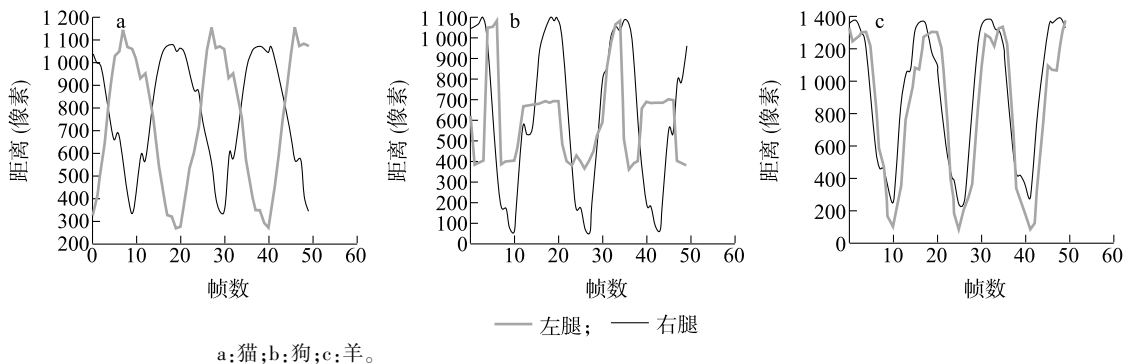
2.2.1 同侧步距参数 在模型中获取研究对象一个步态周期内 4 个角点的坐标位置后,通过定位每一肢踝关节离图像原点距离最远和最近时的 2 关节坐标欧氏距离变化,计算四肢踝关节相对本身迈步距离的平均值,假设  $(x_{f_i}^{LF}, y_{f_i}^{LF})$ 、 $(x_{f_{i+1}}^{LF}, y_{f_{i+1}}^{LF})$  分别为第  $f_i$  帧和第  $f_{i+1}$  帧图像中动物左前肢踝关节坐标,则步距计算式为:

对踝关节的标注点,分别是 4、10 与 7、13,分析一个周期内动物四肢同侧运动一次的帧数范围,得到四足动物同侧双腿距离变化(图 10)。

表 5 不同动物同侧步距参数对比

Table 5 Comparison of ipsilateral stride parameters of different animals

四足动物	左前腿步距 (m)	右前腿步距 (m)	左后腿步距 (m)	右后腿步距 (m)	平均步距 (m)
猫	2.58	2.63	2.57	2.66	2.61
狗	4.18	4.10	4.11	4.09	4.12
羊	5.44	5.46	5.41	5.40	5.43



a: 猫; b: 狗; c: 羊。

图 10 四足动物同侧双腿距离变化

Fig.10 Changes of distance between the legs on the same side of quadruped

通过图 10(a) 可看出速度在 20 km/h 时猫的运动步态接近于对角小跑的运动形式,从同侧两肢距离变化曲线上能看出四肢的交替规律性,猫同侧步距的相邻帧数范围在 9~29 和 19~40 两段。由图 10(b) 可以发现狗运动过程中左侧双腿跨距先达到最大,接着变为最小后又增大到原距离的 2/3 处,而右侧双腿距离先增至最大后缩至最小符合正常奔跑情况,由此可见狗在 10~20 帧时可能有上台阶或遇到坑洼跳跃的情况,狗同侧的两段步态周期帧数分别为 7~32 和 10~28。由图 10(c) 可见羊同侧两肢的

距离同步变化且相差较小,与猫和狗的步距相比多了大约 300 个像素,由此可见奔跑速度在 40 km/h 时动物的身体基本延展开且前两肢与后两肢同步开始运动,羊的同侧两肢距离周期分别为 10~15 帧与 11~16 帧。假设  $f_i$  为当前踝关节距离最小时对应的帧数,则  $f_{i+8}$  为踝关节第 2 次距离最小时对应的帧数,故步频的计算式为:

$$f = \frac{1}{(f_{i+8}^L - f_i^L) + (f_{i+8}^R - f_i^R) * \frac{1}{2F}} \quad (13)$$



式中 $f_{i+8}^l - f_i^l$ 为左侧两肢摆动一个周期内的帧数间隔, $f_{i+8}^r - f_i^r$ 为右侧两肢摆动一个周期内的帧数间隔, $F$ 为摄像机帧率(1 s 30 帧)。故由前一周期和相邻后一周期的平均帧数时长与帧率可计算得到四足动物的步频参数(表 6)。

表 6 四足动物步频参数对比

Table 6 Comparison of quadruped stride frequency parameters

四足动物	上一个周期的帧数	下一周期的帧数	步频(s)
猫	20	21	1.47
狗	25	18	1.39
羊	5	5	5.88

2.2.3 关节角度参数 选取标注的关键点 3 即左前肢膝关节点,将 3 种动物在连续运动周期内膝关节的变化进行对比分析(图 11)。膝关节角度( $\theta$ )计算式为:

$$\theta = \frac{\overrightarrow{C_k C_h} \times \overrightarrow{C_k C_a}}{|\overrightarrow{C_k C_h}| \times |\overrightarrow{C_k C_a}|} \quad (14)$$

式中,向量 $\overrightarrow{C_k C_h}$ 、 $\overrightarrow{C_k C_a}$ 分别是膝关节指向髋关节、踝关节的向量。通过式(14)计算得到动物的关节角度变化(图 12)。

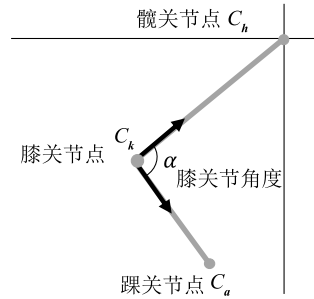
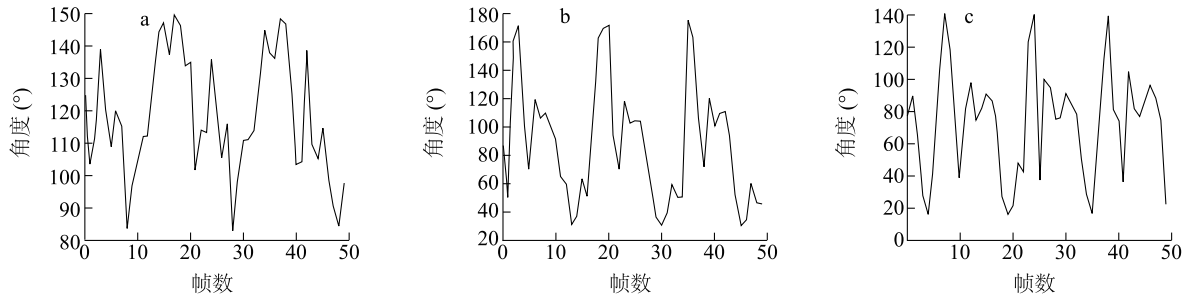


图 11 动物膝关节向量图

Fig.11 Knee joint vector diagram of quadruped



a:猫;b:狗;c:羊。

图 12 四足动物膝关节角度变化图

Fig.12 Changes of knee joint angle of quadruped

由图 12 可知在 20 km/h 时猫的左前膝关节先由 100°变为局部最大 140°,体现向前跨步的过程,接着在第 8 帧降为 85°以下,此时为膝关节弯曲度最大即后肢向前接地且身体压缩的过程,第 15 帧时角度增到最大 148°左右,该阶段是前腿由弯曲到迈进即身体拉伸的过程,到第 19 帧时角度变小体现后脚向前跃进且前脚恢复初始状态的过程。在 30 km/h 和 40 km/h 的运动状态下动物四肢弯曲的程度比较接近,其中羊的最大膝关节角度较犬类小约 30°且后半周期运动帧数比犬类多 2 帧左右,这是由于在跑步步态中羊的体型较大故需弯曲膝关节以获取落地后的减震平衡同时受羊四肢较长的影响,而相比于犬类,由于其体态轻盈故运动速度恒定时膝关节可以完全展开至 180°,即前肢完全伸展以给后

肢足够的动力迈进从而完成奔跑动作。以上运动过程与实际观察效果是一致的,可见本研究算法是可行的。

2.2.4 步态占空比参数 在一个步态周期序列内,包括支撑相和摆动相 2 种序列相位,2 种相位按照研究对象速度的不同占空比也不同。假设 $t_i$ 、 $t_{i+4}$ ( $i=1,2,3,4$ )分别为第  $i$  肢腿的抬起与落地时刻,则一个八拍周期序列  $T$  为

$$t_1=0, t_5=\beta T \quad (15)$$

$$t_2=f, t_6=f+\beta T \quad (16)$$

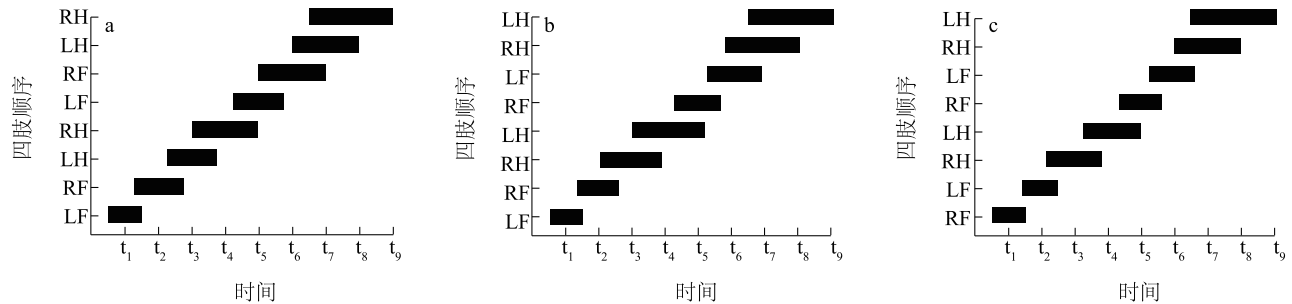
$$t_3=p, t_7=p+\beta T \quad (17)$$

$$t_4=p+h, t_8=p+h+\beta T \quad (18)$$

式中 $f$ 为前双腿分别落地的时间差, $h$ 为后双腿分别落地的时间差, $p$ 为同侧双腿分别落地的时

间差,  $\beta$  为负荷系数, 即一个周期序列中四肢交替着地时间总和与周期总时间的比值。由式 (15) 至式

(18) 计算得到四足动物步态占空比如图 13 所示。将图 13 量化得到表 7。



a: 猫; b: 狗; c: 羊。  $t_1 \sim t_8$  分别为四足动物在一个八拍序列周期内占用的时间。LF: 左前肢; RF: 右前肢; LH: 左后肢; RH: 右后肢。

图 13 动物在一个周期内的步态占空比图

Fig.13 The gait duty cycle diagram of quadruped in one cycle

表 7 四足动物在一个周期内的步态占空比参数

Table 7 The gait duty cycle parameters of quadruped in one cycle

四足动物	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	$\beta$
猫	0	0.5T	0.5T	1.0T	0.5T	1.0T	1.0T	1.5T	0.5
狗	0	0.3T	0.9T	1.2T	0.4T	0.7T	1.3T	1.6T	0.4
羊	0	0.1T	0.7T	0.8T	0.3T	0.4T	1.0T	1.1T	0.3

$t_1 \sim t_8$  分别为四足动物在一个八拍序列周期内占用的时间,  $\beta$  为负荷系数, T 为一个完整序列周期。

由图 13 可知 3 种动物的步态顺序随周期序列循环更替。猫的四肢步态顺序为 LF→RF→LH→RH, 在  $t_2$  和  $t_3$  时 RF 和 LH 两肢几乎同时运动, 通过对角线前进使身体处于一个稳定的三角裕态中。狗的步态顺序为 LF→RF→RH→LH, 在犬类运动过程中犬前肢的髋关节处实质是肩带肌, 前肢处于摆动最大相位时肩带肌完全拉伸有助于后肢膝关节弯曲, 完成下一周期序列的运动。羊的步态顺序为 RF→LF→RH→LH, 与狗相比两者的初始摆动相不同, 羊在  $t_2$  和  $t_3$  时先 LF 作摆动相, 再 RH 作摆动相, 这种顺序有助于羊在奔跑时后肢接替前肢的位置, 使得支撑相的时间最短, 故  $\beta$  最小。通过分析占空比参数验证了本研究算法与实际场合的一致性。

### 3 结论

本研究选用猫、狗、羊 3 种四足动物作为研究对象, 得到的结论如下: (1) 提出了 Dpnet 双路径网络, 通过添加的 HDC 与 PAM 路径提高模型在不同场景下动物关键点识别的准确率, 使用 1 000 张测试集验证各类关键点平均相似度为 0.87, 识别精确率达到 91.1%。(2) 为了获取正确且精度更高的骨架结构, 本研究将热力图与精确圈结合限制关键点像素位

置, 同时提出斜率法实现四肢关节点的区分与归属。(3) 通过模型提取的骨架结构计算动物的运动参数, 包括步距、步频、关节角度变化与步态占空比, 试验证明添加双路径模型的识别精确率分别比并行注意力模块的单路径和混合空洞卷积的单路径提高了 0.9 个百分点与 1.7 个百分点, 验证了模型的有效性, 故本方法可用于大型农场、牧场动物的智能监测工作中。

### 参考文献:

- [1] NASIRAHMADI A, EDWARDS S A, STURM B. Implementation of machine vision for detecting behaviour of cattle and pigs[J]. Livestock Science, 2017, 202: 25-38.
- [2] CHEN C, ZHU W, MA C, et al. Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs[J]. Computers and Electronics in Agriculture, 2017, 142: 380-387.
- [3] 朱家骥, 朱伟兴. 基于星状骨架模型的猪步态分析[J]. 江苏农业科学, 2015, 43(12): 453-457.
- [4] DING W, HU B, LIU H, et al. Human posture recognition based on multiple features and rule learning[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(11): 2529-2540.
- [5] JIANG B, WU Q, YIN X, et al. FLYOLOv3 deep learning for key parts of dairy cow body detection[J]. Computers and Electronics

- in Agriculture, 2019, 166:104982.
- [6] ALVAREZ J R, ARROQUI M, MANGUDO P, et al. Body condition estimation on cows from depth images using Convolutional Neural Networks[J]. Computers and Electronics in Agriculture, 2018, 155:12-22.
- [7] 王浩, 曾雅琼, 裴宏亮, 等. 改进 Faster R-CNN 的群养猪只圈内位置识别与应用[J]. 农业工程学报, 2020, 36(21):201-209.
- [8] 钱建轩. 基于骨架分析和步态能量图的猪的步态识别[D]. 镇江: 江苏大学, 2018.
- [9] 张满囤, 王萌萌, 刘天鹤, 等. 基于骨架能量图的奶牛步态识别[J]. 江苏农业科学, 2020, 48(19):257-262.
- [10] SONG H B, JIANG B, WU Q, et al. Detection of dairy cow lameness based on fitting line slope feature of head and neck outline[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(15):190-199.
- [11] GUTIERREZ-GALAN D, DOMINGUEZ-MORALES J P, CEREZUELA-ESCUDEO E, et al. Embedded neural network for real-time animal behavior classification[J]. Neurocomputing, 2017, 272(jan.10):17-26.
- [12] ZHAO K, BEWLEY J M, HE D, et al. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique[J]. Computers and Electronics in Agriculture, 2018, 148:226-236.
- [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.
- [14] YANG A, HUANG H, ZHU X, et al. Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features[J]. Biosystems Engineering, 2018, 175:133-145.
- [15] ZHENG C, ZHU X M, YANG X F, et al. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. Computers and Electronics in Agriculture, 2018, 147:51-63.
- [16] WU D, WANG Y, HAN M, et al. Using a CNN-LSTM for basic behaviors detection of a single dairy cow in a complex environment[J]. Computers and Electronics in Agriculture, 2021, 182:1-12.
- [17] YANG A, HUANG H, ZHENG B, et al. An automatic recognition framework for sow daily behaviours based on motion and image analyses[J]. Biosystems Engineering, 2020, 192:56-71.
- [18] GUO Y Y, ZHANG Z, HE D J, et al. Detection of cow mounting behavior using region geometry and optical flow characteristics[J]. Computers & Electronics in Agriculture, 2019, 163(8):34-45.
- [19] SUN K, XIAO B, LIU D. Deep high-resolution representation learning for human pose estimation[C]//IEEE. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, USA:IEEE, 2019:5693-5703.
- [20] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.

(责任编辑:张震林)