

DOI: 10.13382/j.jemi.B2104151

基于视觉的四足动物骨架及行走步态特征提取方法^{*}

陈 瑶¹ 张云伟^{1,2,3} 雷金辉^{1,3} 田泽薇¹ 黎 丽¹
(1. 昆明理工大学信息工程与自动化学院 昆明 650500; 2. 昆明理工大学云南省人工智能重点实验室 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 昆明 650500)

摘 要:为解决四足动物行走期间的骨架信息及步态特征提取问题,基于 HRnet 深度神经网络,通过上下文信息增强与多尺度信息融合构建了四足动物骨架的提取模型。在此基础上,建立了基于骨架信息对步态特征进行定量分析的方法。通过在测试数据集上对该模型的有效性进行验证,实验结果表明,该模型具有较好的精度,对四足动物关键点估计的平均相似度为 81.04%,准确率为 92.77%,召回率为 92.75%。基于骨架提取模型,以水牛、羊驼为实验对象,对其行走时的步频特征进行分析计算,实验结果与人工统计结果相比,最大相对误差为 2.73%。通过对一个完整步态周期中水牛和羊驼的髋关节和膝关节角度变化规律进行分析,提取了四足动物行走过程中的关节运动逻辑以及步态顺序。最后,以犀牛为实验对象,验证了方法对拍摄角度的变化具有一定适应性。研究结果可为四足动物运动信息的智能感知提供参考。

关键词: 计算机视觉;深度学习;骨架提取;运动特征

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Method for skeleton and gait parameters extraction of quadrupeds walking based on vision

Chen Yao¹ Zhang Yunwei^{1,2,3} Lei Jinhui^{1,3} Tian Zewei¹ Li Li¹
(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: Due to the need of extracting skeleton information and gait parameters for walking quadruped, a skeleton extraction model of quadruped is proposed by using context information enhancement and multi-scale information fusion based on HRnet, and a quantitative analysis method for gait parameters is established. The validity of the model is verified on the test data set of images. Experiments show that, in the key point estimation of quadruped, the model achieves good performance with the mean similarity of 81.04%, the accuracy of 92.77%, and the recall rate of 92.75%. Based on the skeleton extraction model, the frequency of buffalo and alpaca were analyzed and calculated. Compared with the manual statistical results, the maximum relative error was 2.73%. Through analyzing the angle variations of buffalo's and alpaca's hip and knee joints during a complete gait cycle respectively, the joint motion logic and gait sequence of walking quadruped can be extracted automatically. Finally, taking the rhinoceros as a sample, it is demonstrated that the proposed method can work adaptively in a range of different shooting angle of images. The results can provide a reference for intelligent perception of quadruped motion information.

Keywords: computer vision; deep learning; skeleton extraction; motion features

0 引言

开展对四足动物运动信息的智能感知研究可为动物行为检测、个体身份识别、步态分析、四足仿生机器人等方面的研究提供基础数据支持^[1-7]。而通过视频图像自动识别和提取四足动物在复杂自然环境下行走时的骨架信息,是获取四足动物运动静态参数和动态参数的一种有效方式。

目前,针对生物体进行骨架提取的方法可分为传统方法^[8-9]和基于深度学习^[10-14]的方法。传统方法一般采用图结构模型,此类方法数据量要求低,无需大量训练数据,但过于依赖人工设定的特征,使得模型姿态估计的鲁棒性较低。相较而言,基于深度学习的方法主要利用卷积神经网络自动地获取图像中丰富的语义信息以及多尺度的上下文信息,摆脱了传统方法对人工设计特征的依赖,在实际应用中也取得了较好的效果。Tsohev 等^[11]采用多阶段回归的思想将深度学习引入人体姿态估计领域,该方法通过前馈神经网络提取关键点大致的位置,再利用逐层优化直接回归出关键点的坐标。Fang 等^[12]采用多人检测和单人人体姿态估计的思路进行多人姿态估计,在 MPII 数据集上取得了较好的效果。Newell 等^[13]提出 Hourglass 网络用于人体姿态估计,该网络通过多尺度特征捕捉人体各关键点的空间坐标,取得了较好的效果。Cao 等^[14]使用部分亲和场连接关键点,为多人姿态估计提供了一种全新的思路。在动物信息感知方面,其研究对象主要为人工饲养动物,关于野生动物的研究相对较少。研究方法主要有接触式传感器的方式和非接触式的方式。Daou 等^[15]采用解剖、视频影像以及压电传感器等方式对乌龟静态参数和动态参数进行分析,并用于仿生机构的研究。Chapinal 等^[16]使用 5 个 3D 加速度仪对奶牛行走速度进行评估。但接触式传感器的方式容易使动物产生应激反应,损害动物福利,数据的采集易受到传感器晃动、脱落等因素的影响。非接触式的方式逐渐受到更多研究人员的关注。钱志辉等^[17]采用光学三维运动捕捉技术对德国牧羊犬前肢各关节的运动学特征进行分析。Zhu 等^[18]使用微软 Kinect 设备获取生猪深度图像数据,在提取目标轮廓参数的基础上对生猪体重进行估计。刘波等^[19]基于深度图像匹配生猪骨架端点,用于提取生猪的步频特征。但开放环境中干扰较多,非接触式动物信息智能感知方式的鲁棒性仍有待加强。

综上所述,目前已有较多学者展开对人体关键点估计的研究,并取得了较好的研究成果,但与人体姿态估计相比,四足动物骨架提取存在如下差异和问题:1) 四足动物四肢形体具有较高的相似性,运动情况复杂,对实现前后左右四肢的准确区分提出了更高的要求;2) 部分四足

动物皮毛具有相似的颜色,给关键点的定位造成负面的影响;3) 复杂的自然环境遮挡、色彩干扰以及四足动物的自身遮挡情况严重;4) 个体差异、拍摄角度、距离等因素导致拍摄的同一主体在尺度上存在多样性,给四足动物骨架的准确提取带来了困难。因此,如何通过计算机视觉技术,自动提取开放环境中四足动物的骨架信息,并对运动步态参数进行定量的识别与分析,还有待解决。针对这一问题,本文基于 HRnet^[20] 深度神经网络构建了四足动物骨架的提取模型,给出了利用骨架信息对步态特征进行定量分析的方法,并通过提取出的行走骨架对侧影图像中水牛、羊驼的步频特征及角度特征进行了分析。

1 材料与骨架提取方法

1.1 四足动物图像采集

实验数据采集于云南省昆明市动物园,拍摄对象为主要是处于站立或行走的四足动物。利用三脚架对拍摄设备进行固定,拍摄设备距地高度约在 1~1.7 m,距离拍摄对象约在 5~20 m,拍摄设备镜头光圈 f/1.8,等效焦距 26 mm,对焦模式为自动对焦。采集实验对象包括犀牛、羊驼、水牛和老虎,共计 12 只实验对象,部分实验数据如图 1 所示,共采集到不同角度、光照条件下的视频 33 段,视频格式为 mp4,分辨率为 1 280 pixel(水平)×720 pixel(垂直),帧速率为 30 fps,对每段视频进行随机视频帧抽取,得到不同实验对象的站立图像 800 幅,其中 640 幅用于模型的训练集,160 幅用作测试集。



图1 数据采集

Fig. 1 Data acquisition

考虑到四足动物的运动信息主要集中在四肢关节以及头部区域,本文对四足动物骨架关键点的具体定义如图 2(a) 所示,编号 0~13 分别为嘴、脖子、前左髋关节、前左膝关节、前左蹄、前右髋关节、前右膝关节、前右蹄、后左髋关节、后左膝关节、后左蹄、后右髋关节、后右膝关节、后右蹄。关键点坐标包括未标注点、有遮挡标注点以及未遮挡标注点 3 种情况,分别使用 0、1、2 标签进行区分,其中将未标记点位置信息置零。采用开源工具 Labelme 对实验数据集进行关键点和目标框标注。为消除个人主观认知带来的误差保证关键点标注的准确性,本研究对同一关键点采用多人标注的方法,对同一张数据集图片,邀请 10 位人员进行标注,并取 10 个标注点的

期望值作为最终标注的关键点。对标注数据在原图上进行标注关键点可视化显示,如图 2(b)所示,以检查数据标注的准确性,图 2(b)中标签 5 为有遮挡关键点。对标注数据进行 $\pm 30^\circ$ 范围内的随机旋转、水平翻转、0.8~1.25 范围内的缩放、亮度、对比度变换进行数据增强。

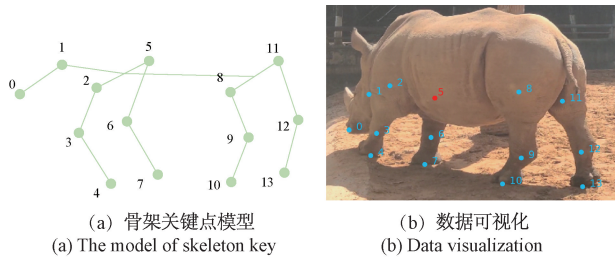


图 2 关键点标注
Fig. 2 Key point annotation points

1.2 特征提取网络

在四足动物关键点估计任务中,由于四肢相似度较高,皮毛颜色接近,自然环境中干扰较多,对特征提取网络的定位精度和语义信息表达能力提出了更高的要求。关键点估计的输出结果即为图片中关键点的空间坐标,如果输出不具备较高的定位精度,则无法实现对关键点坐标的准确预测。其次,四足动物的四肢具有较强的相似性,且运动过程中前后两肢交替运行带来了较大干扰,要实现准确有效的关键点分类,就要求网络的输出要具有很高的语义表达能力。基于以上原因,本文采用 HRnet 作为特征提取网络,HRnet 网络主体结构如图 3 所示。主体为单输入四输出结构,利用 4 个不同分辨率的并行子网进行图像特征的提取,子网个数随着网络的深度的增加而逐渐增多,处于上一层的子网分辨率为下一层子网分辨率的 2 倍,上一层的子网通道数为下一层子网通道数的 1/2。网络可分为 4 个阶段,在第一阶段内,输入图像采用 4 个残差单元进行特征提取;往后的每一阶段子网数逐级递增,在每一子网后串联 4 个残差单元进行特征的提取,然后进行特征融合。其中,第 2~4 阶段分别进行 1、4、3 次的特征融合。

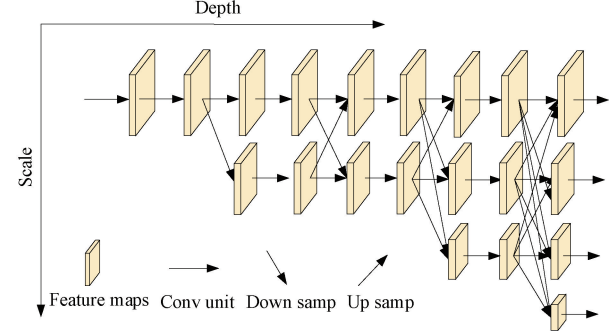


图 3 HRnet 结构
Fig. 3 The structure of HRnet

本文将输入图像的尺寸缩放到 256×256 pixel,HRnet 各子网输出特征图的尺寸和通道数如表 1 所示。

表 1 HRnet 各子网输出特征图的尺寸和通道数
Table 1 Size and channel number of output characteristic diagram of each subnet of HRnet

分枝	通道数	尺寸/pixel
1	32	64×64
2	64	32×32
3	128	16×16
4	256	8×8

1.3 上下文信息增强与多尺度信息融合

目前多种研究均表明,在图像特征提取的过程中,增加感受野有利于特征提取网络获取丰富的上下文信息,对计算机视觉中的多项任务都具有积极作用。在深度神经网络中,对图像进行下采样是增大感受野有效方式,但这种方式会缩小特征图的尺寸,可能导致部分关键点信息无法重建甚至是丢失。显然,这在关键点估计任务中不利的。若通过增大卷积核尺寸来增大感受野则会导致增加参数,加大计算量。空洞卷积能够在参数量不增加的前提下,提高网络的感受野以获得更加丰富的空间信息。空洞卷积如图 4 所示。其通常的做法是根据膨胀系数 r 在原卷积核相邻权值之间插入 $r - 1$ 个 0,得到新的卷积核,计算式为:

$$\begin{cases} F_h = f_h + (f_h - 1) \times (r - 1) \\ F_w = f_w + (f_w - 1) \times (r - 1) \end{cases} \quad (1)$$

式中: r 为膨胀系数; f_h 为原始卷积核的高; f_w 为原始卷积核的宽; F_h 为膨胀卷积核的高; F_w 为膨胀卷积核的宽。

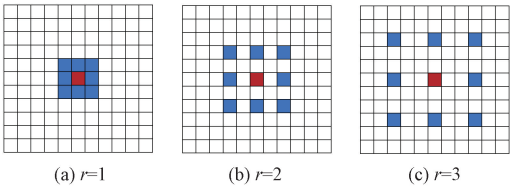


图 4 空洞卷积
Fig. 4 Dilated convolution

为捕获图像中不同尺度的上下文信息,论文利用一个并行的空洞卷积模块 (parallel dilated convolution module, PDCM) 增强网络对上下文信息的感知能力,其结构如图 5 中所示。将 HRnet 网络第 1 阶段提取到的特征图分别输入到膨胀系数分别为 1、3、6、9 的空洞卷积进行特征提取,最后将 4 个特征图进行融合输出。同时,为更好的适应自然环境中不同尺度的图像数据,将 HRnet 输出分别经过 4 次残差模块进行特征的提取和一次的 1×1 卷积调整通道数为 32。最后,将分辨率 1/8、1/16、1/32

特征图采用线性插值放大到 1/4 分辨率大小进行特征融合。

1.4 模型计算步骤

四足动物关键点定位模型结构如图 5 所示,模型计算步骤如下。

1) 输入大小为 256×256 的图像,经过两次步长为 2 的 3×3 卷积,分辨率变为 64×64 ,通道数增大到 64。

2) 将图像特征输入到带有 PDCM 的 HRnet 网络和多尺度融合模块中进行特征提取,得到的特征图分辨率为 64×64 ,通道数减少为 32。

3) 图像特征经过 1×1 卷积将通道数调整为 14,然后,利用上采样和 3×3 卷积得到与原图具有相同分辨率

的关键点热力图,热力图个数与关键点个数一一对应,代表了相应四足动物的关键点位于某一像素区域的概率。设理想状态下四足动物 k 关键点 j 的热力图为 $S_{j,k}^*$,可以根据标注了关键点的二维矩阵计算得到,其计算方法如下:

$$S_{j,k}^* = \exp\left(-\frac{\|P - X_{j,k}\|_2^2}{\sigma^2}\right) \quad (2)$$

式中: $X_{j,k}$ 表示 id 为 k 的四足动物关键点 j 的坐标; σ 表示峰宽; P 表示二维图像中的坐标点。

4) 根据 14 个热力图的通道位置信息及热力图峰值获取四足动物关键点的分类信息及空间位置信息,对关键点有序连接得到四足动物的骨架。

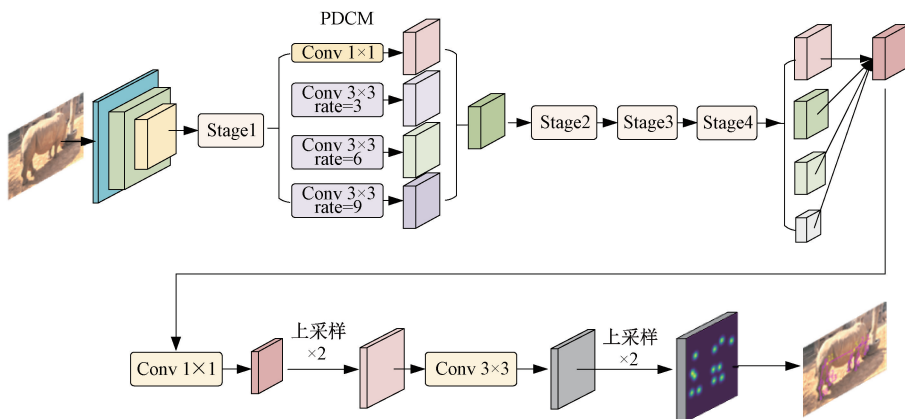


图 5 四足动物骨架提取框架

Fig. 5 Skeleton extraction framework of quadruped

2 行走特征参数提取

2.1 步频特征

定位视频序列中每一帧图像中四足动物的关键点,根据关键点进一步提取四足动物的运动特征参数。为保证运动参数提取的准确性,基于四足动物的侧影图像提取运动特征参数。通过对四足动物行走运动的观察可知,四足动物在行走运动时,存在前肢关键点 4、前肢关键点 7;后肢关键点 10、后肢关键点 13 交替变换。即在一个完整的步态周期内:假设前肢率先开始运动,此时后肢关键点 10、后肢关键点 13 相对距离 $d(P_{10}^i, P_{13}^i)$ 保持不变,前肢关键点 4、前肢关键点 7 相对距离 $d(P_4^i, P_7^i)$ 逐渐减小到极小值然后逐渐增大完成本周期的第一次前肢交替动作。然后前肢关键点 4、前肢关键点 7 相对距离 $d(P_4^i, P_7^i)$ 保持不变,后肢关键点 10、后肢关键点 13 相对距离 $d(P_{10}^i, P_{13}^i)$ 逐渐减小到极小值然后逐渐增大完成本周期的第 1 次后肢交替动作。按照类似的运动规律,接下来前肢和后肢依次进行本周期的第 2

次交替动作。前两肢和后肢分别进行两次交替运动后,四足动物四肢依次向前移动,一个完整的步态周期结束。本文对步态周期定义为两前肢或者两后肢在相邻周期内第 1 次交替变换两肢相对距离为极小值时的时间差,为消除侧影图像中,左右关键点误判的问题。对第 i 帧图像中前肢蹄部关键点 P_4^i, P_7^i ,后肢关键点坐标 P_{10}^i, P_{13}^i 分别按照横坐标 x 值的大小作升序重新排列。理想状态下,序列图像中两前肢和两后肢的相对距离变化情况如图 6 所示,则步频特征 f 计算如式(3)所示。

$$f = \frac{1}{[(f_9 - f_1) + (f_{11} - f_3)]/2f_{ps}} \quad (3)$$

式中: f_{ps} 表示视频帧速率; f_i 表示视频帧序号,其中 $i = 1, 2, 3, \dots$ 。

2.2 角度特征

四足动物行走过程中,在一个步态周期内,四肢按照一定的步态顺序交替前移同时配合各关节角度的有序变换,以实现自身空间位置的平稳移动。为探究四肢关节角度的变化规律,定量分析四足动物的运动方式,本文在准确提取四足动物步频特征的基础上,分析在一个完整

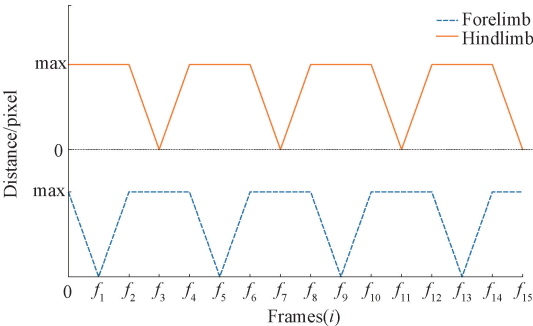


图 6 理想相对距离变化曲线

Fig. 6 Variation curve of ideal relative distance

步态周期内四足动物膝关节和髋关节的角度变化情况。对膝关节角度 θ_1 定义为向量 $\vec{P_k P_a}$ 和向量 $\vec{P_k P_h}$ 的夹角，髋关节角度 θ_2 定义为向量 $\vec{P_h P_k}$ 与速度向量 $\vec{v}$ 的夹角。其中， P_a, P_k, P_h 表示同一条腿上踝关节、膝关节和髋关节的坐标，速度向量可以根据嘴关键点 0 确定。角度特征示意如图 7 所示。

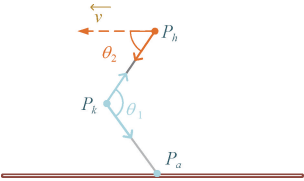


图 7 关节角度示意图

Fig. 7 Schematic diagram of joint angle

可以直接获得四肢关节点的坐标，对属于同一肢的踝关节、膝关节和髋关节可以实现准确的匹配，但在侧影图像中，在前肢或后肢的前后关系发生交换后则可能存在左右侧关系错误匹配的情况。因此，为消除左右关键点误判给计算带来的误差，需要对关键点的所属关系进行重新匹配。针对此种情况本文建立了一种基于特定交换规则的分配方法。对第 i 帧图像中右前肢、左前肢、右后肢、左后肢的踝关节、膝关节和髋关节算法检测到的 3 个关键点坐标依次用 $\mathbf{g}_{Fr}^i = [P_2^i \ P_3^i \ P_4^i]$ 、 $\mathbf{g}_{Fl}^i = [P_5^i \ P_6^i \ P_7^i]$ 、 $\mathbf{g}_{Hr}^i = [P_8^i \ P_9^i \ P_{10}^i]$ 、 $\mathbf{g}_{Hl}^i = [P_{11}^i \ P_{12}^i \ P_{13}^i]$ ，其中 $i = 1, 2, 3, \dots$ 。按照 $P_4^i, P_7^i, P_{11}^i, P_{13}^i$ 横坐标 x 值的大小对 $\mathbf{g}_{Fr}^i, \mathbf{g}_{Fl}^i, \mathbf{g}_{Hr}^i, \mathbf{g}_{Hl}^i$ 作升序排列。假设步态周期内第 1 帧图像中从左至右四肢的位置关系依次为右前肢、左前肢、右后肢、左后肢，则直接将 $\mathbf{g}_{Fr}^i, \mathbf{g}_{Fl}^i, \mathbf{g}_{Hr}^i, \mathbf{g}_{Hl}^i$ 4 组数据依次分配给四足。接着在相对距离 $d(P_4^i, P_7^i)$ 变为局部极小时，对 $\mathbf{g}_{Fr}^i, \mathbf{g}_{Fl}^i$ 排列位置进行交换。同理，采用类似的方法对 $\mathbf{g}_{Hr}^i, \mathbf{g}_{Hl}^i$ 的位置关系进行处理。即可保证图像序列中四肢关键点的正确匹配。

3 实验与分析

本实验在 Windows 操作系统上进行，CPU 为英特尔酷睿 i5 9400F (2.9 GHz)，GPU 为英伟达 GTX 1660Super (6 G 显存)，深度学习框架为 Tensorflow1.11，软件平台为 Python&Pycharm，初始学习率设置为 0.001，batch_size = 2，epochs = 30，采用 Adam 方法进行训练优化。由于目前还没有针对四足动物键点检测的大规模公开数据集。考虑到关键点估计任务都是在图像中定位出关键点的空间位置，而与检测目标本身的运动模式关联性较小。为提高模型的泛化能力，加快损失值的收敛，获取较优的模型初始权值。本文采用迁移学习的思想，利用人体关键点数据集 MPII 对模型首先进行 30 次的迭代训练。

3.1 四足动物骨架提取

考虑到部分关节的相似性以及检测的难易程度，本文将 14 个关键点划分为 5 个大类，即嘴、脖子、髋关节、膝关节、踝关节，分别包含 1、1、4、4、4 个关键点。采用 O_{ks} 评价指标分别考查算法对四足动物每一类关键点预测的相似度。

$$O_{ks} = \frac{\sum_i \exp \{ -d_i^2 / 2S^2 \sigma_i^2 \} \delta(\vartheta_i > 0)}{\sum_i \delta(\vartheta_i > 0)} \times 100\% \quad (4)$$

式中： d_i 表示预测关键点与真实关键点之间的欧氏距离； ϑ_i 表示关键点是否检测到，当其为 0 时，表示该关键点不可见； S 表示目标尺度大小； σ_i 表示关键点 i 的归一化因子。

检测结果如表 2 所示，算法检测关键点平均相似度为 81.04%。其中，在脖子、髋关节的关键点上，预测关键点与实际关键点相似度较低，从测试数据上来看，可能是由于脖子、髋关节关键点周围存在相似的皮毛以及实验对象自身遮挡所导致，而鼻子、膝关节、踝关节特征相对明显，关键点的定位相对而言更加准确。

表 2 各类关键点测试结果

Table 2 Results of various key points (%)						
关键点	嘴	脖子	髋关节	膝关节	踝关节	平均值
水牛	82.37	77.68	78.53	81.90	83.32	80.76
犀牛	81.97	78.39	77.96	81.59	83.47	80.68
羊驼	82.38	79.01	78.55	82.13	83.67	81.15
虎	84.74	78.96	78.16	82.10	83.90	81.57
平均值	82.87	78.51	78.30	81.93	83.59	81.04

然而， O_{ks} 相似度评价指标只从与标记点在距离上的偏离程度来评价模型的优劣，相似度较高的关键点并不一定适用于步态参数的提取，这对于后续步态参数的提取工作不利的。另外、不可见关键点未参与计算，对模型性能的测试还不够全面。本文采用以下方法进一步测试

模型对关键点估计的准确性。

1)利用目标框长度对图像进行归一化。2)设某一关键点 10 个标注点的期望值为 0,求期望值与标注点间的距离最大为 r_i ,则以 0 为圆心, r_i 为半径范围内的区域为关节区域。3)如果模型预测点刚好落在该区域内,则认为关键点被正确预测,如图 8 所示。测得模型的准确率召回率如表 3 所示。

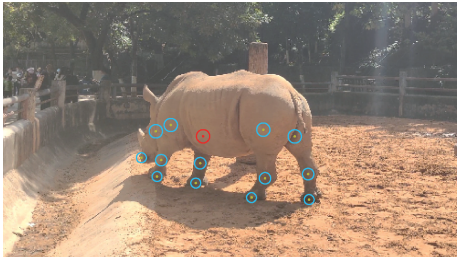


图 8 关节区域示意图

Fig. 8 Schematic diagram of joint area

表 3 各类关键点准确率

Table 3 The accuracy rate of key points (%)					
Subjects	水牛	犀牛	羊驼	虎	平均值
准确率	92.03	91.59	93.89	93.57	92.77
召回率	91.98	91.50	94.01	93.50	92.75

从表 3 可知:模型对水牛、犀牛、羊驼、老虎关键点定位的准确率分别为:92.03%、91.59%、93.89%、93.57%。召回率率分别为:91.98%、91.50%、94.01%、93.50%。

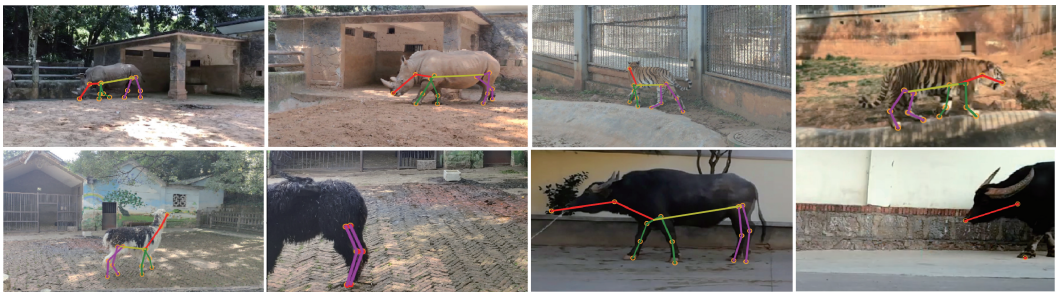


图 9 测试结果可视化

Fig. 9 Visualization of prediction results

3.2 步频特征

从采集数据中挑选水牛、羊驼的行走侧影图像序列进行实验。按照前文所述,依次分别计算两前肢关键点 4、7 和两后肢关键点 11、13 的相对距离,得到相对距离变化曲线如图 10、11 所示。前后肢相对距离呈周期性变化,相对距离在增大到局部极大后,基本维持不变,实验结果与实际场景基本保持一致,从侧面验证了论文方法对关键点估计的准确性。相对距离局部极小值间距大小

表明本文方法在四足动物关键点定位上具有较高的可靠性和准确性。

为更好地验证本文方法的有效性,分别在 Hourglass 和 HRnet 上进行四足动物关键的估计实验。由表 4 可知,在所述实验平台上,本文方法相较于 Hourglass 的参数量有所增加,导致处理速度稍慢。但关键点估计的准确率却提高了 13.7%,说明所提方法在四足动物关键点的估计任务中取得了较好的效果。此外,本文方法在参数量和运算速度与 HRnet 几乎保持一致的前提下,准确率提高了 0.98%,说明并行空洞卷积模块和多尺度特征融合的方式在四足动物关键点的估计任务中具有积极的作用。

表 4 实验结果对比

Table 4 Comparison of experimental results			
方法	参数	速度/ f_{ps}	准确率/%
Hourglass	25 547 760	11.92	79.07
HRnet	28 601 408	10.73	91.79
Ours	28 685 504	10.57	92.77

有序连接各关键点的到四足动物骨架如图 9 所示,可以看出论文方法对光照强度、目标尺度的改变具有一定的适应能力,在关键点轻微遮挡的情况下表现出了较好的准确性,对目标不完整的情况具有一定的判断能力。但当环境中存在相近的色彩时,则有可能导致关键点误判,另外,对于遮挡比较严重或者身体缺失面积较多的情况,模型则无法检测出关键点位置。

基本相等,说明实验对象行走速度基本不变。

对相对距离曲线进行三次样条插值并计算局部极小值,得到水牛、羊驼的平均步态周期分别为 42 帧和 39 帧。为检验论文方法的准确性,采用人工方法对图像序列中实验对象的步态周期进行统计,得到各自的步频特征如表 5 所示,本文方法与人工统计结果最大误差为 2.73%,说明论文方法可以用于四足动物步频特征的提取。

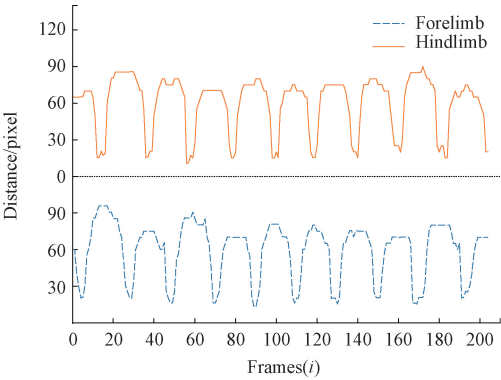


图 10 水牛相对距离变化曲线
Fig. 10 The relative distance curve of buffalo

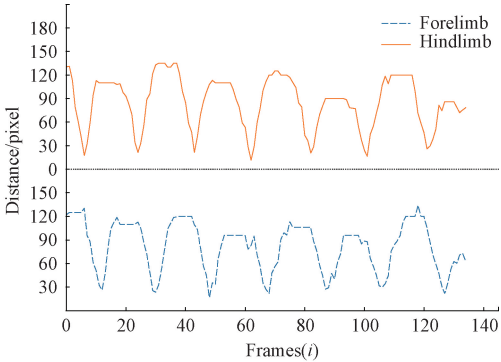


图 11 羊驼相对距离变化曲线
Fig. 11 The curve of relative distance of alpaca

表 5 频率值
Table 5 The value of frequency

方法	水牛	羊驼
本文 f/Hz	0.71	0.77
人工统计 f/Hz	0.73	0.79
误差/%	2.73	2.53

3.3 角度特征

在一个完整的步态周期内,分别计算图像序列中水牛和羊驼四肢的膝关节角度 θ_1 和髋关节角度 θ_2 ,实验结果如图 12、13 所示。从二者实验结果来看,四足动物在行走过程中前肢和后肢情况存在一定的差异,在某一步态周期内,前肢膝关节角度在多数情况下保持在 180° 左右,即前肢此时几乎处于完全伸直的状态,在此过程中,髋关节角度逐渐增大。结合实际图像序列分析可知,此时前肢始终与地面接触,即前肢处于支撑相,前肢膝关节完全张开,以更好的支撑四足动物身体的重量。同时,前肢绕髋关节向后旋转,使得四足动物身体向前移动。当前肢髋关节角度旋转到一定角度后,此时前肢膝关节角度变小,前肢弯曲抬起,同时髋关节向前旋转,前肢向前

迈进。此时前肢离开地面,即前肢处于运动相;而后肢膝关节则在整个步态周期内变化范围较小,其中水牛膝关节角度约为在 160° ,羊驼膝关节角度约为 150° 。后肢髋关节角度采用与前肢类似的起伏变化,后肢处于支撑相时,髋关节角度增大,推动身体向前运动,当增大到一定角度后,髋关节角度收缩,后肢向前移动。总体而言,在四足动物前行过程中,各个关节在一定时间差上有序变化形成了一套严密的运动逻辑,使得四足动物个体可以平稳移动。

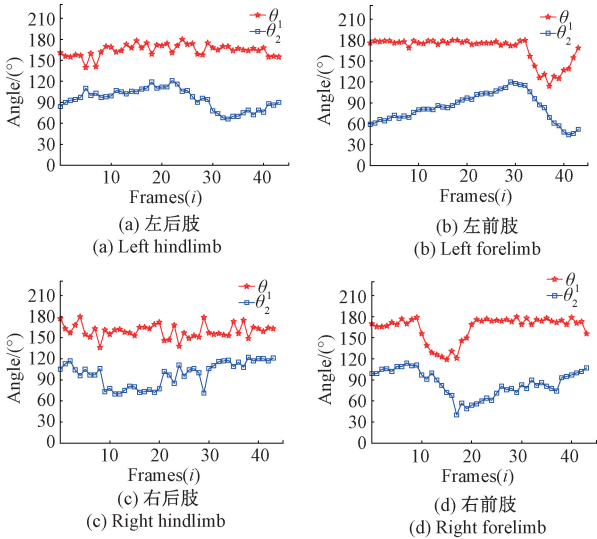


图 12 水牛四肢膝关节和髋关节角度
Fig. 12 The angle of knee joint and hip joint of buffalo

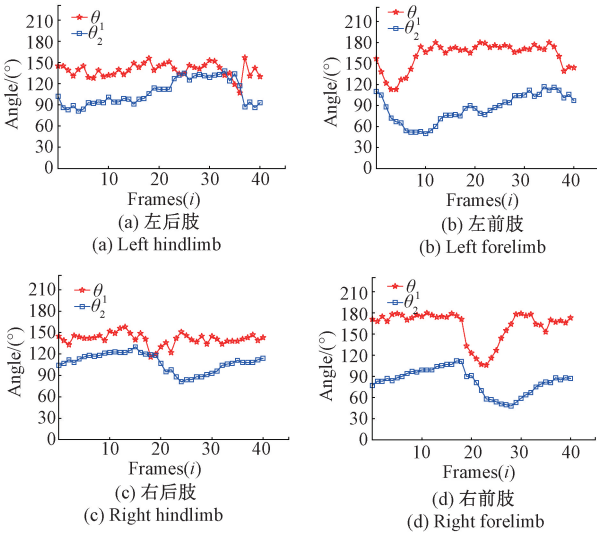


图 13 羊驼四肢膝关节和髋关节角度
Fig. 13 The angle of knee joint and hip joint of alpaca

基于上述分析,论文将实验对象各个髋关节角度绘制到同一图中,进一步分析实验对象的摆腿顺序,结果如

图 14、15 所示。根据对角度特征的分析可知,髋关节角度整体变化趋势减小时表示对应肢体向前迈进,则羊驼的迈腿顺序为左前→右后→右前→左后→左前循环进行;水牛的迈腿顺序为右后→右前→左后→左前→右后循环进行。其共同的特点表现为同侧后肢运动后同侧前肢运动,前肢运动后接着对侧后肢运动然后对侧前肢运动。同时,在两腿切换过程中还存在双腿共同运动的情况,这与实际观察结果是一致的。

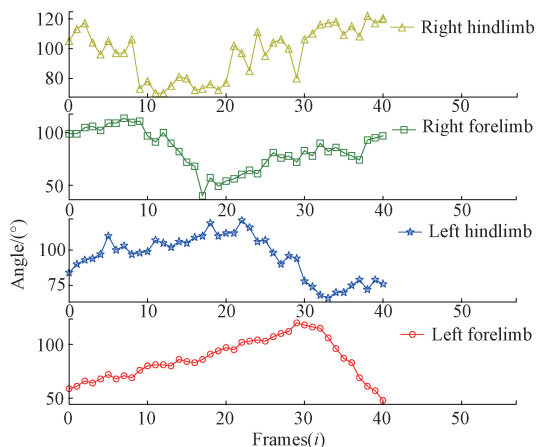


图 14 水牛四肢髋关节角度

Fig. 14 Hip joint angle of buffalo limbs

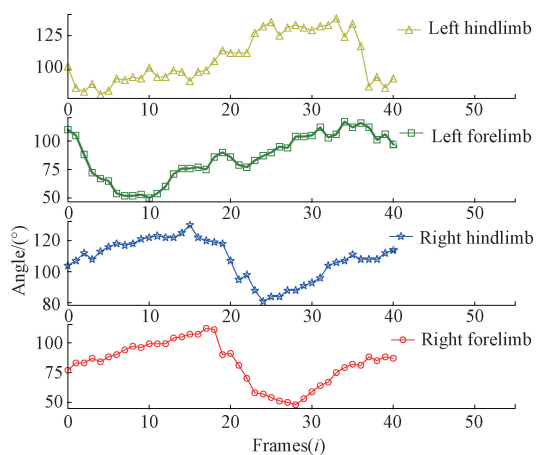


图 15 羊驼四肢髋关节角度

Fig. 15 Hip joint angle of alpaca limbs

3.4 角度变化对四足动物骨架提取的影响

为探究拍摄角度变化或四足动物四肢多次交替运动情况下,本文方法对四足动物骨架提取的准确率。本文以犀牛为实验对象进行实验,选取拍摄角度发生明显变化的测试数据进行骨架提取。论文对拍摄角度的定义为犀牛运动方向与相机光轴之间的夹角,其中运动方向远离图像平面 XY 时为正,运动方向接近图像平面 XY 时为负,相机光轴垂直于图像平面 XY ,面向拍摄目标的一向

为正,反之为负。

实验结果如图 16 所示,当拍摄角度为 $\pm 90^\circ$ 左右时准确率约为 88.4%,当拍摄角度为 $\pm 45^\circ$ 左右时准确率约为 91.9%,当拍摄角度为 $\pm 30^\circ$ 左右时准确率约为 90.1%。

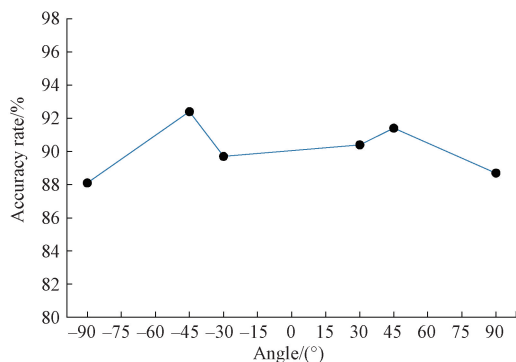


图 16 不同角度下关键点估计准确率

Fig. 16 The accuracy of key point estimation from different angles

对不同角度的犀牛骨架提取结果进行可视化如图 17 所示,本文方法存在对关键点位置错误估计的情况,主要表现为:在不同的拍摄角度下,前后或左右同类关键点(如左前肢和右前肢的踝关节)在二维图像中不易区分,造成属于不同肢体的同类关键点在位置接近或位置错误的地方具有较高的响应,算法无法估计出准确的关键点坐标。在对动物步态特征参数计算时可能形成噪声干扰。

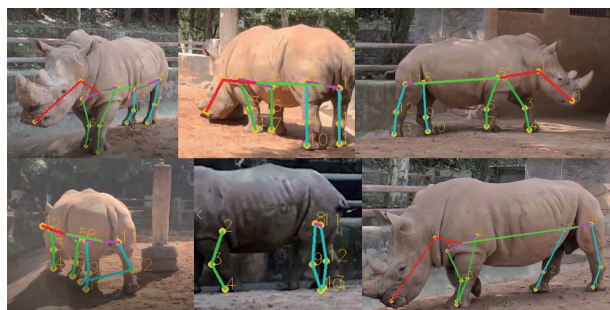


图 17 不同角度下骨架提取

Fig. 17 The skeleton extraction of different angles

4 结 论

本文在 HRnet 深度神经网络的基础上构建了四足动物骨架的提取模型,其基本思路为利用并行空洞卷积增强网络对上下文信息的感知能力,并结合多尺度的信息融合进行图像特征提取,然后通过上采样将特征图恢复到原图大小进行热力图的预测,实现了对四足动物骨架信息的自动提取。实验结果表明该模型对关键点估计的

平均相似度为 81.04%, 准确率为 92.77%, 召回率为 92.75%, 具有较好的精度。在此基础上, 基于骨架信息建立了对步态特征进行定量分析的方法, 并以水牛、羊驼为实验对象, 对其行走时的步频特征进行分析计算, 结果与人工统计结果相比, 最大相对误差为 2.73%。最后, 论文对一个完整步态周期中水牛和羊驼的髌关节和膝关节角度变化规律进行分析, 准确提取了行走过程中的关节运动逻辑以及步态顺序。研究结果为四足动物骨架的自动提取提供了一种思路, 可用于四足动物步态参数的提取、行为分析、动物个体身份的识别、仿生机器人运动分析与控制等方面的研究。

由于目前尚未发现与四足动物关键点估计相关的大型公开数据集, 实验中的数据集规模较小, 模型的实际泛化能力还有待进一步的提升。同时, 本文目前所做工作主要是针对典型四足动物行走运动, 后续研究将继续改进完善, 针对其他运动模式开展研究。

参考文献

- [1] 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.
- [2] KIM C H, SHIN H C, LEE H H, et al. Trotting gait analysis of a lizard using motion capture [C]. International Conference on Control, Automation and Systems, IEEE, 2013: 1247-1251.
- [3] THORUP V M, MUNKSGAARD L, ROBERT P E, et al. Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms [J]. Animal, 2015, 9(10): 1704-1712.
- [4] MATERTENS W, JURGEN V, BAERT J, et al. Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: The GAITWISE system [J]. Biosystems Engineering, 2015, 110(1): 29-39.
- [5] ANDREW W, GREATWOOD C, BURGHARDT T. Visual localisation and individual identification of holstein friesland cattle via deep learning [C]. IEEE International Conference on Computer Vision Workshop, 2017: 2850-2859.
- [6] FRICKE P M, CARVALHO P D, GIORDANO J O, et al. Expression and detection of estrus in dairy cows: The role of new technologies. [J]. Animal, 2014, 8(5): 134-143.
- [7] 杨秋妹, 肖德琴, 张根兴. 猪只饮水行为机器视觉自动识别 [J]. 农业机械学报, 2018, 49(6): 232-238.
YANG Q M, XIAO D Q, ZHANG G X. Automatic pig drinking behavior recognition with machine vision [J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(6): 232-238.
- [8] 郭晓峰, 王耀南, 周显恩, 等. 基于目标轮廓与骨架特征的棋子识别算法 [J]. 电子测量与仪器学报, 2018, 32(9): 142-149.
GUO X F, WANG Y N, ZHOU X EN, et al. Chess piece recognition algorithm based on target contour and skeleton features [J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(9): 142-149.
- [9] 徐超, 肖潇, 骆燕, 等. 基于距离变换的新型骨架提取方法 [J]. 仪器仪表学报, 2012, 33(12): 2851-2856.
XU CH, XIAO X, LUO Y, et al. New skeleton extraction method based on distance transformation [J]. Chinese Journal of Scientific Instrument, 2012, 33(12): 2851-2856.
- [10] 游伟, 王雪. 人为骨架特征识别边缘计算方法研究 [J]. 仪器仪表学报, 2020, 41(10): 156-164.
YU W, WANG X. Study on the edge computing method for skeleton-based human action feature recognition [J]. Chinese Journal of Scientific Instrument, 2020, 41(10): 156-164.
- [11] TSOHEV A, SZEGEDY. DeepPose: Human pose estimation via deep neural networks [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1653-1660.
- [12] FANG H, XIE S, TAI Y, et al. RMPE: Regional multi-person pose estimation [C]. IEEE International Conference on Computer Vision (ICCV), 2017: 2353-2362.
- [13] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation [C]. European Conference on Computer Vision, Springer, 2016: 833-851.
- [14] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields [C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1302-1310.
- [15] DAOU H E, LIBOUREL P A, RENOUS S, et al. Methods and experimental protocols to design a simulated bio-mimetic quadruped robot [J]. International Journal of Advanced Robotic Systems, 2013, 10(5): 1-8.
- [16] CHAPINAL N, PASSILLE A M D, PASTELL M, et al. Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle [J]. Journal of Dairy Science, 2011, 94(6): 2895-2901.
- [17] 钱志辉, 苗怀彬, 任雷, 等. 基于多种步态的德国牧羊犬下肢关节角 [J]. 吉林大学学报(工学版), 2015,

45(6): 1857-1862.

QIAN ZH H, MIAO H B, REN L, et al. Lower limb joint angles of German shepherd dog during foot-ground contact in different gait patterns [J]. Journal of Jilin University (Engineering and Technology Edition), 2015, 45 (6): 1857-1862.

- [18] ZHU Q, REN J, BARCLAY D, et al. Automatic animal detection from kinect sensed images for livestock monitoring and assessment [C]. IEEE International Conference on Computer & Information Technology Ubiquitous Computing & Communications Dependable, 2015: 1154-1157.

- [19] 刘波, 朱伟兴, 杨建军, 等. 基于深度图像和生猪骨架端点分析的生猪步频特征提取[J]. 农业工程学报, 2014(10):131-137.

LIU B, ZHU W X, YANG J J, et al. Extracting of pig gait frequency feature based on depth image and pig skeleton endpoints analysis [J]. Transactions of the Chinese Society for Agricultural Engineering, 2014(10): 131-137.

- [20] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C].

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 5686-5696.

作者简介



陈瑶, 2018 年于华东交通大学获得学士学位, 现为昆明理工大学硕士研究生, 主要研究方向为计算机视觉。

E-mail: 2586313633@qq.com

Chen Yao received his B. Sc. degree from East China Jiaotong University in 2018.

Now he is a M. Sc. candidate at Kunming University of Science and Technology. His main research interest includes computer vision.



张云伟 (通信作者), 2007 年于上海交通大学获得博士学位, 现为昆明理工大学教授, 硕士生导师, 主要研究方向主为机器人学、机器视觉、传感检测技术。

E-mail: zhangyunwei72@gmail.com

Zhang Yunwei (Corresponding author) received his Ph. D. degree from Shanghai Jiaotong University in 2007. Now he is a professor and M. Sc. supervisor at Kunming University of Science and Technology. His main research interests include robotics, machine vision and sensor detection technology.