

基于 SimCC – ShuffleNetV2 的轻量化奶牛关键点检测方法

宋怀波^{1,2} 华志新^{1,2} 马宝玲^{1,2} 温毓晨^{1,2} 孔祥凤^{1,2} 许兴时^{1,2}

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100;
2. 农业农村部农业物联网重点实验室, 陕西杨凌 712100)

摘要: 针对现有深度学习技术在奶牛关键点检测研究中尚存在网络复杂度高、检测速度慢等问题,提出了一种轻量化奶牛关键点检测模型 SimCC – ShuffleNetV2。在模型中,主干网络采用 ShuffleNetV2 用于特征提取,有利于实现网络的轻量化;检测头采用 SimCC 用于关键点位置预测,SimCC 采取坐标分类的方法使得检测更加简单高效。为了验证模型的性能,本研究设计了奶牛的关键点及骨架结构,并标注了 3 600 幅图像用于模型的训练与测试。试验结果表明,SimCC – ShuffleNetV2 模型的 AP_{50,95} 为 88.07%,浮点运算量为 1.5×10^8 ,参数量为 1.31×10^6 ,检测速度为 10.87 f/s,可以实现奶牛关键点的精确与高效检测。与基于回归的 DeepPose 网络、基于热力图的 HRNet 网络进行了对比试验,结果表明 SimCC – ShuffleNetV2 取得了精度与速度的良好平衡。同时,本研究通过更换不同主干与不同检测头的方式,对比验证了不同模块对模型性能影响,本研究所提出的模型在所有试验中均取得了最佳结果,表明 ShuffleNetV2 与 SimCC 的组合具备良好的关键点检测性能。为了验证模型的有效性,将模型应用于 4 种动作视频中提取骨架序列并将其送入 ST – GCN 网络以实现不同动作的分类,其分类准确率为 84.56%,表明本研究提出的 SimCC – ShuffleNetV2 模型是良好的关键点提取器,可为奶牛行为识别等任务提供关键信息支撑。

关键词: 奶牛; 关键点检测; ShuffleNetV2; SimCC; 动作识别

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2023)10-0275-07 OSID: 

Lightweight Keypoint Detection Method of Dairy Cow Based on
SimCC – ShuffleNetV2

SONG Huaibo^{1,2} HUA Zhixin^{1,2} MA Baoling^{1,2} WEN Yuchen^{1,2} KONG Xiangfeng^{1,2} XU Xingshi^{1,2}

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
2. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China)

Abstract: Cow keypoint detection is important in research fields such as cow body measurement, behavior recognition, and weight estimation. However, existing deep learning methods for cow keypoint detection still suffer problems such as high network complexity and slow detection speed. A lightweight cow keypoint detection model SimCC – ShuffleNetV2 was proposed. In this model, ShuffleNetV2 was used as the backbone for feature extraction to achieve network lightweight. SimCC was used as the head to achieve keypoint position prediction. SimCC adopted a coordinate classification method that was simple and efficient. To validate the effectiveness of the model, cow keypoints and skeleton structures were designed, and 3 600 images were annotated for training and testing. Experimental results showed that the SimCC – ShuffleNetV2 model got an AP_{50,95} of 88.07%, FLOPs of 1.5×10^8 , parameters of 1.31×10^6 , and detection speed of 10.87 f/s, achieving accurate and efficient detection of cow keypoints. Experimental comparisons with the regression-based DeepPose and Heatmap-based HRNet networks demonstrated that SimCC – ShuffleNetV2 got a good balance between accuracy and speed. Moreover, different backbones and detection heads were replaced to verify the influence of different modules on model performance. And the proposed model achieved the best results in all experiments, demonstrating that the combination of ShuffleNetV2 and SimCC had good keypoint detection performance. The model was applied to extract skeleton sequences from four different action videos of cows, and the ST – GCN network was used to classify the four videos, achieving an 84.56% classification accuracy, which

indicated that the proposed SimCC – ShuffleNetV2 model was a good keypoint extractor and could provide key information support for tasks such as cow action recognition.

Key words: dairy cow; keypoint detection; ShuffleNetV2; SimCC; action recognition

0 引言

奶牛关键点指奶牛身体部位重要的关节位置,如头部、肩部、腿关节等^[1]。关键点相连的骨架作为姿态的最直观体现,可以帮助实现奶牛动作行为的非接触识别,从而为姿态表征的疾病预警提供重要技术支持。同时,关键点检测还是牲畜体尺测量与身体部位分割的前置任务,可为奶牛的体重与体况分析提供更为高效的数据支撑。综上,开展奶牛关键点检测技术的研究和应用,将有助于提高养殖效率和健康水平,具有重要研究价值^[2]。

传统关键点检测方法通过人工提取图像或视频中的关键点特征,如关节点之间的空间位置关系等,再通过机器学习算法进行学习 and 匹配其特征及规律^[3]。JIA 等^[4]通过骨架对奶牛不同身体部位进行划分用以评估其清洁度。该研究在获得奶牛骨架图的过程中采取特征距离变换,细化与特征检测等步骤,然后计算相邻的骨架路径,并与数据库中的模板进行相似度匹配以获得最佳匹配结果,最后确定每个骨骼分支所属的类别后,将身体分为不同部分。结果表明,侧视图与后视图的平均身体部位分割精度分别为 96% 和 91%,为奶牛的分割任务提供了借鉴。但该骨架提取方法受人为设计特征完备性较差、机器学习算法的特征提取和泛化能力较弱的影响,在实际生产过程中难以应用。

随着深度学习的应用,关键点检测技术取得了较多进展。基于卷积神经网络(Convolutional neural networks, CNN)的方法^[5]无需手动设计特征,从整幅图像中学习关键点表征,可以实现端对端的检测,具有良好的特征提取能力和泛化能力,常见的关键点检测网络包括 Hourglass^[6]、DeepPose^[7]、OpenPose^[8]、HRNet^[9]等。诸多研究也将深度学习网络应用于奶牛关键点检测中,RUSSELLO 等^[10]使用 T – LEAP 姿态估计模型,将 LEAP 模型修改为时间序列模型,从图像序列中检测关键点从而预估奶牛姿态,试验结果表明,在奶牛目标上的平均正确关键点的预估比例(Percentage of correct keypoints, PCK)为 93.8%。同时,奶牛关键点检测是实现体尺测量、动作识别等任务的基础^[11]。其中,体尺测量用于评估牲畜体型的大小,体尺关键点检测的准确度直接影响体尺测量的精度。赵宇亮等^[12]为了对猪只的 5 项体尺指标进行非接触式测量,首先在

DeepLabCut 上选取 EfficientNet – b6 模型作为最优主干网络进行猪只体尺关键点检测,其测试集误差为 5.13 像素;在此基础上,通过 3D 坐标转换实现深度图像上关键点坐标的映射,并对离群特征点进行优化,最后提取体尺曲线进行计算。奶牛关键点连接的骨架是姿态的最直观表征,奶牛的姿态数据可用于统计其基本运动行为时间。LI 等^[13]为了准确识别奶牛的基本运动行为(走、站、躺),利用 HRNet 提取奶牛骨架信息,在平行的二维卷积特征中以热图的形式添加对应的奶牛关键点与骨架信息,并选取了 400 个包含该 4 种行为的奶牛视频进行训练与测试,结果表明,经过 5 次交叉验证,最终分类精度为 91.80%。

上述关键点检测研究大多基于手工提取,难以实际应用;或者依赖于 DeepLabCut^[14]平台,泛化性不足,且无法解决多目标间关键点的连接问题。目前基于深度学习的主流算法具备较高的准确率,但其网络复杂程度也较高。为解决上述问题,本研究借鉴 ShuffleNetV2 模型的轻量化性能及 SimCC 所具有的高效简单的关键点坐标分类能力,提出 SimCC – ShuffleNetV2 轻量化模型,以期构建一种实时性强、精度高、鲁棒性强的奶牛关键点检测方法,为奶牛动作识别等研究奠定基础。

1 材料与方法

1.1 材料

本研究视频数据采集于陕西省杨凌科元克隆有限公司,采用 DS – 2DM1 – 714 型圆顶摄像机(海康威视)进行拍摄,摄像机速率为 25 f/s,分辨率为 704 像素 × 480 像素。如图 1 所示,摄像机架设于背靠牛棚的长走廊外,目标与摄像机之间的围栏遮挡

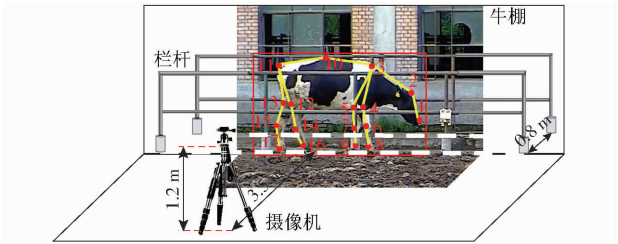


图 1 数据采集与标注方式

Fig. 1 Data collection and annotation methods

1. 鼻子 2. 额头 3. 颈根 4. 右前腿根 5. 左前腿根 6. 右前膝 7. 左前膝 8. 右前蹄 9. 左前蹄 10. 背部 11. 尾根 12. 右后腿根 13. 左后腿根 14. 右后膝 15. 左后膝 16. 右后蹄 17. 左后蹄

了一些关键点,会对骨架提取任务造成一定的干扰。

如图 1 所示,本研究以最大的动物姿态数据集 AP-10k 为参考^[15],设计了奶牛关键点及骨架结构。从视频中截取 1 800 幅图像用于关键点检测任务,为了有效模拟奶牛姿态的变化,采用镜像翻转的方式扩充数据集至 3 600 幅。标注奶牛的边界框与关键点后将其保存为 COCO 格式,并按照 6:2:2 的比例分为训练集、验证集和测试集。

由于奶牛动作及姿态变化多样且易受遮挡、光照等因素影响,奶牛关键点检测的准确性不高。为了增强模型的鲁棒性,采集了不同情况的奶牛图像:①正面视角:奶牛朝正面摄像头时,关节位置会产生一定的形变。②关键点缺失:奶牛目标位于图像边缘时,部分关键点会缺失。③躺卧姿态:奶牛在躺卧姿态下,某些关节点会被遮挡。④小目标:拍摄距离较远时会造成奶牛目标在图像中占据的比例较小。训练集和测试集中均包含了各种影响因素下的奶牛图像,且在训练集和测试集中的比例基本保持一致。

1.2 试验平台

试验在 Windows 10 系统下进行,处理器为 Intel (R) Core (TM) i5 - 11400F,图形处理器为 Nvidia RTX2080Ti。深度学习框架为 PyTorch,编程平台为 PyCharm,编程语言为 Python,所有算法均在相同环境下运行。

1.3 总体技术路线

本研究总体技术路线如图 2 所示,首先将图像送入 SimCC - ShuffleNetV2 模型,关键点检测过程类似于编解码,ShuffleNetV2 用于提取并编码 17 个关键点表征。SimCC 用于对水平轴和垂直轴执行坐标分类,最后解码出关键点坐标。为了验证模型的有效性,将 SimCC - ShuffleNetV2 应用于行为识别任务。从 400 段包含 4 种动作(行走、站立、躺卧、跛行)的视频中提取出骨架信息,并将骨架序列送入 ST - GCN (Spatial temporal graph convolutional networks)^[16] 网络训练,ST - GCN 利用时空图卷积来提取骨架序列间的时空信息,并将时空特征融合起来,最后输出动作分类结果。

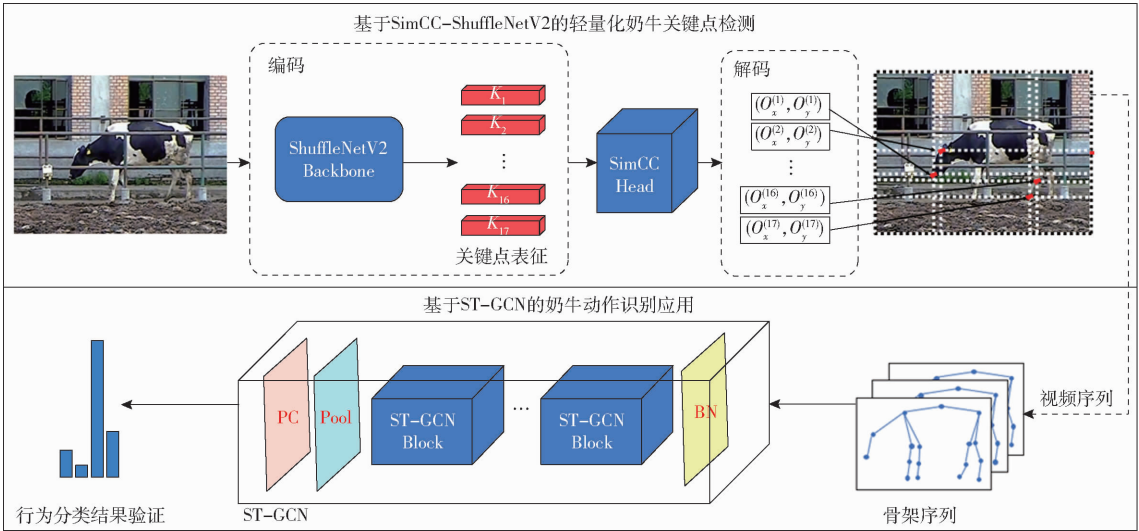


图 2 总体技术路线
Fig. 2 Overview of the proposed method

1.4 ShuffleNetV2 网络结构

ShuffleNetV2^[17] 为主干特征提取网络,结构如图 3 所示,其中 Conv 为卷积操作,DWConv (Depthwise convolution) 为深度卷积。

分组卷积操作让不同的卷积核学习不同的特征,从而提高模型的表达能力。左右分支连接后将输出特征进行通道混洗,从而达到不同通道间信息交换的目的,有利于增加模型的非线性表示能力。上述结构在保持网络准确性的同时,具有更高的计算效率和更小的模型参数,故本研究将其作为主干网络。

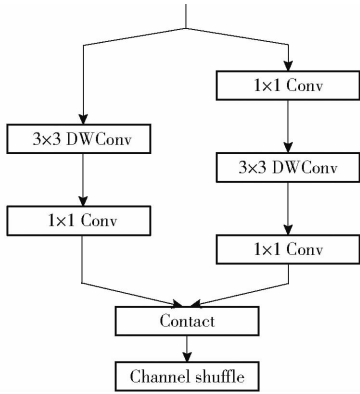


图 3 ShuffleNetV2 结构
Fig. 3 Structure of ShuffleNetV2

1.5 SimCC 结构

SimCC^[18]在网络中作为检测头,SimCC 模块采用坐标分类的思路,将不同坐标值划分为不同类别,从水平与垂直维度将关键点坐标进行分类,从而实现关键点检测,其结构如图 4 所示。

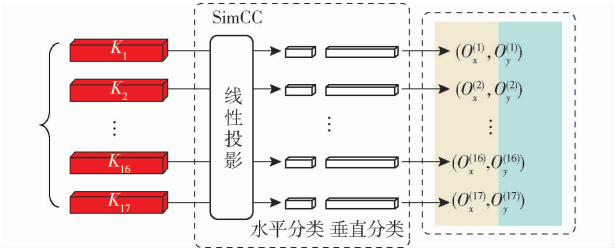


图 4 SimCC 结构
Fig. 4 Structure of SimCC

在主干网络提取特征时,关键点由 (n, H', W') 展平至 $(n, H' \times W')$,即输出 n 个关键点的一维向量。再通过线性投影将坐标编码为 n 个 SimCC 表征,表征后的坐标可以通过两条独立的一维向量来描述:

$$p' = (\text{round}(x^p k), \text{round}(y^p k)) \tag{1}$$

式中 round ——线性投影
 (x^p, y^p) ——表征前的坐标
 p' ——表征后的坐标
 k ——缩放因子

经过线性投影后,输出的特征图维数为 $W \times k$ 维和 $H \times k$ 维。缩放因子 $k(k > 1)$ 的作用是使一维向量长度超过图像边长,关键点定位精度增强到亚像素级别。在坐标解码过程中,给定的第 p 个关键点表征输入至仅一个线性层的水平和垂直坐标分类器。需将坐标还原到图像尺度,要将分类的最大概率所在位置除以缩放因子,计算公式为

$$\begin{cases} O_x^{(i)} = \frac{\text{argmax}(O_x(i))}{k} \\ O_y^{(i)} = \frac{\text{argmax}(O_y(i))}{k} \end{cases} \tag{2}$$

式中 $(O_x^{(i)}, O_y^{(i)})$ ——还原到图像尺度后的坐标
 $(O_x(i), O_y(i))$ ——解码前的位置坐标
 argmax ——取最大值操作

与直接基于像素点回归的模型不同,SimCC 将定位问题转化为分类问题,可以避免回归模型存在的训练难度高、容易受到噪声和异常值影响等问题,具有更高的精度和鲁棒性。基于热力图的关键点检测方法通过显式地渲染高斯热图,让模型输出目标概率分布。由于特征提取时需要进行多次卷积和下采样操作,通常导致生成的热力图尺寸小于图像原尺寸,因此将坐标映射回原图时会存在不可避免的量化误差。而 SimCC 采用两个方向上、长度大于原

图像尺寸的一维向量对关键点进行表征,坐标表征为数值,精度不受缩放影响,避免了量化误差。

1.6 关键点检测评价指标

OKS(Object keypoint similarity)^[19]是一种用于衡量关键点预测值与真实值间相似度的度量指标。计算 OKS 时要首先提取关键点预测值与真实值间的特征向量,然后使用欧氏距离来计算两个特征向量的相似度,基于这种相似度计算,可以进行关键点的匹配和识别。 OKS_p 的计算公式为

$$OKS_p = \frac{\sum_i \exp\left(\frac{-d_{pi}^2}{2s_p^2\sigma_i^2}\right)\delta(v_{pi} > 0)}{\sum_i \delta(v_{pi} > 0)} \tag{3}$$

式中 OKS_p ——第 p 个目标的 OKS
 δ ——用于选取可见点的计算函数
 p_i ——第 p 个目标的第 i 个关键点
 d_{pi} ——关键点预测值与真实值间的欧氏距离
 v_{pi} ——关键点的可见性
 s_p ——目标边界框的面积
 σ_i ——第 i 个关键点标注值和实际值间的标准偏差

模型的准确度采用平均精度 (Average precision, AP) 来衡量,AP 值越高,说明模型的准确度越高。OKS 类似于目标检测中的交并比 (IOU)^[20],用于计算关键点检测的 AP。当 OKS 大于给定阈值 T 时,检测结果视为真阳性,精度 (Precision, P) 指检测结果真阳性的比例,召回率 (Recall, R) 指所有真实标注中的关键点被检测到的比例,然后计算 PR 曲线与坐标轴所围面积即为 AP。其中 AP_{50} 表示 OKS 阈值为 0.50 时的 AP, $AP_{50,95}$ 表示在 OKS 阈值从 0.50 到 0.95 之间,步长为 0.05 时的平均 AP。

此外,本研究所提出的 SimCC - ShuffleNetV2 为轻量级模型,因此浮点运算量 (Floating point operations, FLOPs)、参数量 (Params) 与检测速度作为重要的轻量化评估指标也纳入评价。浮点运算量的计算方式是将各层参数量乘以输入数据的维度,然后对所有层的结果求和。参数量表示模型中的参数数量,通常用于衡量模型的复杂度和容量。

2 结果与分析

2.1 SimCC - ShuffleNetV2 训练结果

本研究中共训练 200 轮次,网络训练过程损失值与 AP_{50} 的变化曲线如图 5 所示。

SimCC - ShuffleNetV2 的网络参数空间较小且层数较低,训练过程中更容易找到全局或局部最优

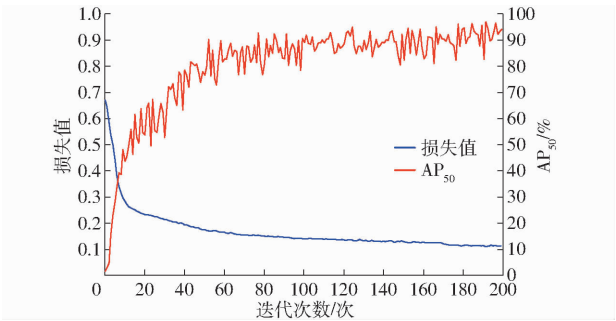


图 5 训练过程

Fig. 5 Training process

解,故曲线收敛较快,当迭代次数达到 50 次左右时,模型学习效果已经达到饱和,关键点损失值稳定在 0.15 左右,AP₅₀稳定在 93% 左右。

2.2 关键点检测效果

SimCC - ShuffleNetV2 在测试集上的 AP_{50,95} 为 88.07%, AP₅₀ 为 97.76%, 参数量为 1.31×10^6 , 浮点运算量为 1.5×10^8 , 检测速度为 10.87 f/s。不同情况下关键点可视化效果如图 6 所示,标示出 17 个关键点并绘制出骨架。奶牛朝向摄像头位于正面视角时,关节位置会产生一定的变化,训练集中此类图像较少,但检测效果同样良好,表明模型的泛化能力较强。奶牛在躺卧姿态下,有些关节点会被遮挡,模型检测时会预测被遮挡关键点的位置,虽然结果有一定偏离,但整体效果依然能准确反映奶牛姿态。在远距离情况下,此时奶牛目标较小,模型依然能准确标示出关键点位置,表明模型对小目标同样敏感。针对上述情况,一般采用尺度变换和多尺度融合技术来提高模型检测的准确性,而 SimCC - ShuffleNetV2 对正面视角、躺卧姿态与远距离情况下亦能准确检测,表明模型具备良好的性能。

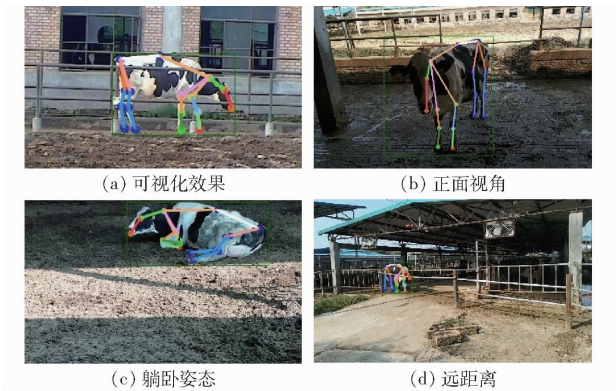


图 6 不同情况下的关键点检测效果

Fig. 6 Keypoint detection effects in different situations

2.3 关键点误检分析

如图 7 所示,虚线框所指的缺失关键点被错误检测在非奶牛身体部位。模型检测效果较差的可能原因在于,本研究中数据集中包含的此类图像较少,

模型可能会过拟合,缺失关键点被误认为是图像中某些类似的局部特征。此外,当奶牛关键点缺失时造成的姿态约束条件、关节角度等先验信息缺失也可能会导致模型误判。

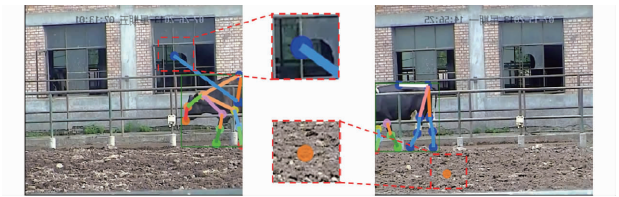


图 7 关键点误检

Fig. 7 Error detection of keypoint

3 结果与讨论

3.1 不同关键点检测算法性能比较

为了评估 SimCC - ShuffleNetV2 模型对奶牛关键点检测的性能,在相同条件下,分别基于 DeepPose、HRNet 两种经典关键点检测算法对同一数据集进行训练,其中 DeepPose 为直接基于回归的模型,HRNet 为基于热力图的模型。采用 AP_{50,95}、浮点运算量、参数量与检测速度对训练完成后的模型进行评估。3 种检测算法的性能指标如表 1 所示。对比其结果可知,SimCC - ShuffleNetV2 的 AP 分别比 DeepPose 和 HRNet 提高 23.65、2.16 个百分点。检测速度比 DeepPose 减少 7.44 f/s,比 HRNet 提高 4.78 f/s。其参数量分别比 DeepPose 和 HRNet 减少 2.23×10^7 和 2.72×10^7 , 浮点运算量分别减少 8.93×10^9 、 1.718×10^{10} 。DeepPose 直接基于回归方式无需过多后处理且主干网络为结构简单的 Alexnet^[21],故检测速度较快。但同时 Alexnet 的学习能力有限,而且直接在图像像素点中回归坐标较为困难,故精度较低。HRNet 能一直保持高分辨率的表征,生成的热力图尺寸也为高分辨率,预测的关键点在空间上更精确。但由于其复杂的网络结构带来了巨大计算量从而导致检测速度较慢。而 SimCC - ShuffleNetV2 在拥有最高准确度的同时有较快的检测速度,实现了精度与速度的良好平衡。

表 1 不同关键点检测算法性能比较

Tab. 1 Comparison of different keypoint detection models

算法	平均精度/ %	浮点运 算量	参数量	速度/ (f·s ⁻¹)
DeepPose	64.42	9.080×10^9	2.36×10^7	18.31
HRNet	85.91	1.733×10^{10}	2.85×10^7	6.09
本文算法	88.07	1.500×10^8	1.31×10^6	10.87

3.2 不同主干网络性能比较

主干网络是用于特征提取的主要组件,为了对比不同主干网络对模型性能的影响,保持 SimCC 为

检测头,分别测试以 Res - 50^[22]、ResNeSt - 50^[23]、MobileNetV2^[24]作为主干网络的模型性能。如表 2 所示,ShuffleNetV2 为主干网络时的 AP 分别比 Res - 50、ResNeSt - 50 和 MobileNetV2 提高 1.96、3.87、3.16 个百分点,检测速度分别提高 5.14、4.19、3.05 f/s;参数量分别比 Res - 50、ResNeSt - 50 和 MobileNetV2 减少 3.54×10^7 、 3.74×10^7 、 9.8×10^5 ,浮点运算量分别少 5.36×10^9 、 6.63×10^9 、 1.60×10^8 。ShuffleNetV2 为主干网络时,AP 最高,浮点运算量与参数量最小,检测速度最快,说明 ShuffleNetV2 能在保持与更大网络相同精度的前提下,还能有更小的模型体积和浮点运算量。且对比同为轻量化网络的 MobileNetV2,同样有更出色的性能与速度。

表 2 不同主干网络性能比较

Tab. 2 Comparison of different backbones

主干网络	平均精度/ %	浮点运 算量	参数量	速度/ (f·s ⁻¹)
Res - 50	86.11	5.51×10^9	3.67×10^7	5.73
ResNeSt - 50	84.20	6.78×10^9	3.87×10^7	6.68
MobileNetV2	84.91	3.10×10^8	2.29×10^6	7.82
ShuffleNetV2	88.07	1.50×10^8	1.31×10^6	10.87

3.3 不同检测头性能比较

为了评估不同检测头对模型性能的影响,保持 ShuffleNetV2 为主干网络,对比测试了以热力图为检测头的模型。其结果如表 3 所示,SimCC 作检测头时比以热力图作为检测头时的 AP 提高 2.03 个百分点,检测速度提高 3.97 f/s;参数量减少 6.24×10^6 ,浮点运算量减少 1.22×10^9 。此外,在 SimCC 检测头的浮点运算量为 3×10^6 ,参数量为 5.6×10^4 ,相较于整体模型,其浮点运算量与参数量可忽略不计。而基于热力图的检测头的浮点运算量为 1.24×10^9 ,在整个模型中占比为 90.8%;参数量为 6.29×10^6 ,在整个模型中占比为 83.3%,基于热力图的检测头占据了网络大部分的浮点运算量与参数量,导致模型检测速度较低。

表 3 不同检测头的性能比较

Tab. 3 Comparison with Heatmap-based method

检测头	平均精度/ %	浮点运 算量	参数量	速度/ (f·s ⁻¹)
Heatmap	86.04	1.37×10^9	7.55×10^6	6.90
SimCC	88.07	1.50×10^8	1.31×10^6	10.87

如图 8 所示,热力图表征图像中每一个关键点的概率分布。热力图由图像进行卷积和池化操作得到,需要大量的计算成本和内存开销^[25]。SimCC 作为检测头无需高分辨的热力图即可实现良好的检测

效果,将关键点检测问题转换为分类问题从而降低了计算成本和内存开销,可大幅提高模型的计算速度。



图 8 热力图表征
Fig. 8 Heatmap representation

3.4 基于 ST - GCN 的奶牛动作识别应用

为了验证 SimCC - ShuffleNetV2 的有效性,将模型用于奶牛运动视频中提取出骨架序列,并送入 ST - GCN 网络以实现动作识别。人工筛选了 400 段视频作为动作识别的数据,每段时长约为 (10 ± 4) s,如图 9 所示,动作行为包含躺卧、站立、行走、跛行(每种动作 100 段视频)。



图 9 不同动作行为
Fig. 9 Different actions

如图 10 所示,ST - GCN 是基于骨架序列的动作识别方法,其中 Attention(Spatial attention)为空间注意力机制,GCN(Graph convolutional network)为图卷积网络,TCN(Temporal convolutional network)为时间卷积网络。该模型使用图卷积神经网络处理骨架视频序列,然后使用 CNN 提取特征,最后通过分类器选取具有最高概率的类别。分类准确率是衡量 ST - GCN 效果的主要指标,即正确预测视频样本占视频样本总数的百分比。

如图 11 所示,动作分类结果标示在视频左上

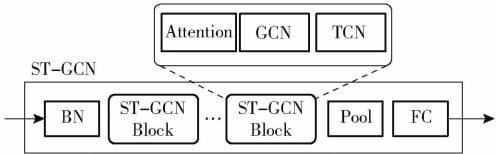


图 10 ST - GCN 结构

Fig. 10 Structure of ST - GCN

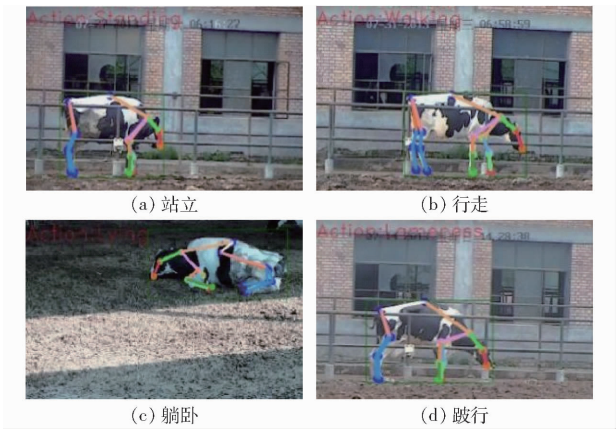


图 11 动作识别效果

Fig. 11 Action detection effects

角,ST - GCN 能对不同的动作进行有效的区分。同时 SimCC - ShuffleNetV2 提取的骨架序列送入训练后,ST - GCN 在测试集上的分类准确率为 84.56%。表明 SimCC - ShuffleNetV2 能良好地表征奶牛的姿态以供 ST - GCN 网络学习。

4 结论

- (1) ShuffleNetV2 作为轻量化网络,其特有结构能保持与大模型相当精度的同时更加轻量,有利于实现高效的特征提取,更加适合于实际应用场景。
- (2) SimCC 将关键点检测问题转换为分类问题,从而降低了问题的复杂度,更加简单和高效。相较于基于回归的方式,模型精度更高;相较于基于热力图的方式,模型能在保持相当精度的情况下更加轻量。
- (3) ST - GCN 的分类准确率达到 84.56%,表明 SimCC - ShuffleNetV2 是良好的关键点提取器,在动作识别任务中有良好的应用前景。

参 考 文 献

[1] POURSAHERI A, BAHR C, PLUK A, et al. Real-time automatic lameness detection based on back posture extraction in dairy cattle: shape analysis of cow with image processing techniques[J]. Computers and Electronics in Agriculture, 2010, 74(1): 110 - 119.

[2] 宋怀波,李振宇,吕帅朝,等. 基于部分亲和场的行走奶牛骨架提取模型[J]. 农业机械学报, 2020, 51(8): 203 - 213. SONG Huaibo, LI Zhenyu, LÜ Shuaichao, et al. Skeleton extraction model of walking dairy cows based on partial affinity field [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 203 - 213. (in Chinese)

[3] 邓益依,罗健欣,金凤林. 基于深度学习的人体姿态估计方法综述[J]. 计算机工程与应用,2019,55(19): 22 - 42. DENG Yinong, LUO Jianxin, JIN Fenglin. Overview of human pose estimation methods based on deep learning[J]. Computer Engineering and Applications, 2019, 55(19):22 - 42. (in Chinese)

[4] JIA N, KOOTSTRA G, KOERKAMP P G, et al. Segmentation of body parts of cows in RGB-depth images based on template matching[J]. Computers and Electronics in Agriculture, 2021, 180: 105897.

[5] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. Deepcut: a deeper, stronger, and faster multi-person pose estimation model[C]//Computer Vision-ECCV 2016. Springer International Publishing, 2016: 34 - 50.

[6] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]// Computer Vision-ECCV 2016. Springer International Publishing, 2016: 483 - 499.

[7] TOSHEV A, SZEGEDY C. DeepPose: human pose estimation via deep neural networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1653 - 1660.

[8] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291 - 7299.

[9] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693 - 5703.

[10] RUSSELLO H, RIK V, KOOTSTRA G. T-LEAP: occlusion-robust pose estimation of walking cows using temporal information [J]. Computers and Electronics in Agriculture, 2021, 192: 106559.

[11] LI C, LEE G H. From synthetic to real: unsupervised domain adaptation for animal pose estimation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 1482 - 1491.

[12] 赵宇亮,曾繁国,贾楠,等. 基于 DeepLabCut 算法的猪只体尺快速测量方法研究[J].农业机械学报,2023,54(2):249 - 255. ZHAO Yuliang, ZENG Fanguo, JIA Nan, et al. Rapid measurements of pig body size based on DeepLabCut algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(2):249 - 255. (in Chinese)

[13] LI Z, SONG L, DUAN Y, et al. Basic motion behaviour recognition of dairy cows based on skeleton and hybrid convolution algorithms[J]. Computers and Electronics in Agriculture, 2022, 196: 106889.

measurements by the four-electrode method[J]. Soil Science Society of America Journal, 1980(6):1216 – 1221.

[28] NADLER A. Matrix nonuniformity effect on measurement of apparent bulk soil electrical conductivity[J]. Soil Science, 1988, 145(3): 157 – 162.

[29] 徐志闻, 刘亚斌, 胡夏嵩, 等. 基于水分和原位电导率的西宁盆地盐渍土含盐量估算模型[J]. 农业工程学报, 2019, 35(5): 148 – 154.

XU Zhiwen, LIU Yabin, HU Xiasong, et al. Salt content estimation model of saline soil in Xining Basin based on water content and in-situ electrical conductivity[J]. Transactions of the CSAE, 2019,35(5): 148 – 154. (in Chinese)

[30] 曹巧红, 龚元石. 土壤电导率对时域反射仪测定土壤水分的影响[J]. 土壤学报, 2001, 38(4): 483 – 490.

CAO Qiaohong, GONG Yuanshi. Effect of soil bulk electrical conductivity on time domain reflectometry measurement of soil water content [J]. Acta Pedologica Sinica, 2001,38(4): 483 – 490. (in Chinese)

[31] ALJOUANI B, SANCHEZ-ESPIGARES J A, WESSOLEK G. Estimating pore water electrical conductivity of sandy soil from time domain reflectometry records using a time-varying dynamic linear model[J]. Sensors, 2018, 18(12): 4403.

[32] MIRZAKHANINAFCHI H, MANI I, HASAN M, et al. Development of prediction models for soil nitrogen management based on electrical conductivity and moisture content[J]. Sensors, 2022, 22(18): 6728.

[33] 毛任钊, 田魁祥, 松本聪, 等. 盐渍土盐分指标及其与化学组成的关系[J]. 土壤, 1997, 29(6): 326 – 330.

MAO Renzhao, TIAN Kuixiang, MATSUMOTO Cong, et al. Relationship between salinity index and chemical composition of saline soil[J]. Soils, 1997,29(6): 326 – 330. (in Chinese)

[34] 刘广明, 杨劲松. 土壤含盐量与土壤电导率及水分含量关系的试验研究[J]. 土壤通报, 2001, 32(增刊 1): 85 – 87.

LIU Guangming, YANG Jinsong. Experimental study on the relationship between soil salt content and soil electrical conductivity and water content[J]. Chinese Journal of Soil Science, 2001, 32(Supp. 1): 85 – 87. (in Chinese)

(上接第 281 页)

[14] MATHIS A, MAMIDANNA P, CURY K M, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning[J]. Nature Neuroscience, 2018, 21(9): 1281 – 1289.

[15] YU H, XU Y, ZHANG J, et al. Ap-10k: a benchmark for animal pose estimation in the wild[J]. arXiv preprint arXiv:2108.12617, 2021.

[16] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2018.

[17] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116 – 131.

[18] LI Y, YANG S, LIU P, et al. SimCC: a simple coordinate classification perspective for human pose estimation [C] // Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 89 – 106.

[19] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context [C] // Computer Vision-ECCV 2014: 13th European Conference. Springer International Publishing, 2014: 740 – 755.

[20] REN S, HE K, GIRSHICK R, et al. Faster R – CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137 – 1149.

[21] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84 – 90.

[22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.

[23] ZHANG H, WU C, ZHANG Z, et al. Resnest: split-attention networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2736 – 2746.

[24] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510 – 4520.

[25] 张飞宇, 王美丽, 王正超. 引入 Transformer 和尺度融合的动物骨骼关键点检测模型构建[J]. 农业工程学报, 2021, 37(23): 179 – 185.

ZHANG Feiyu, WANG Meili, WANG Zhengchao. Construction of the animal skeletons keypoint detection model based on Transformer and scale fusion[J]. Transactions of the CSAE, 2021, 37(23): 179 – 185. (in Chinese)