doi:10.6041/j.issn.1000-1298.2020.08.023

基于部分亲和场的行走奶牛骨架提取模型

宋怀波[1,2]　李振宇[1,2]　吕帅朝[1,2]　尚钰莹[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:在奶牛关键点预测的基础上,通过点、线重构奶牛骨架结构,可为奶牛跛行检测、发情行为分析、运动量估测等提供重要参考。本研究基于部分亲和场,以养殖场监控摄像头拍摄的视频为原始数据,使用 1 600 幅图像训练了奶牛骨架提取模型,实现了奶牛站立、行走状态下关键点信息和部分亲和场信息的预测,并通过最优匹配连接对奶牛骨架结构进行准确提取。为了验证该模型的性能,采用包含干扰因素的 100 幅单目标奶牛和 100 幅双目标奶牛图像进行了测试。结果表明,该模型对单目标行走奶牛骨架提取的置信度为 78.90%,双目标行走奶牛骨架提取的置信度较单目标下降了 10.96 个百分点。计算了不同关键点相似性(Object keypoint similarity,OKS)下的模型准确率,当 OKS 为 0.75 时,骨架提取准确率为 93.40%,召回率为 94.20%,说明该模型具有较高的准确率。该方法可以提取视频中奶牛骨架,在无遮挡时具有高置信度和低漏检率,当遮挡严重时置信度有所下降。该模型的单目标和双目标图像帧处理速度分别为 3.30、3.20 f/s,基本相同。本研究可为多目标奶牛骨架提取提供参考。

关键词:奶牛;骨架提取模型;部分亲和场;深度学习;关键点相似性

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2020)08-0203-11　　**OSID**:

Skeleton Extraction Model of Walking Dairy Cows Based on Partial Affinity Field

SONG Huaibo[1,2]　LI Zhenyu[1,2]　LÜ Shuaichao[1,2]　SHANG Yuying[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: The skeleton extraction of cows is based on the prediction of key points of cows, which can provide important reference for detection of claudication, analysis of estrus behavior, and estimation of motion of cows through point and line reconstruction of skeleton structure of cows. Based on the partial affinity field, taking the video taken by the monitoring camera of the farm as the original data, totally 1600 images were used to train the cow skeleton extraction model, and the prediction of the key point information and partial affinity field information of the cow in the standing and walking states was realized, and the accurate extraction of the cow skeleton structure through the optimal matching connection was realized. In order to verify the performance of the model, totally 100 images of single cow and 100 images of double cows were tested. The experimental results showed that the model had a 78.90% confidence in the attitude of single target walking cows, and a 10.96 percentage points decrease in the confidence of double target walking cows compared with single target walking cows. In order to test the overall accuracy of the model, the accuracy of the model under different key points similarity OKS was calculated, and the accuracy rate was 93.40% when strict standard OKS was 0.75. Furthermore, the experimental results showed that the method can extract the cow skeleton in the video, which had high confidence and low missing rate when there was no occlusion, and the confidence was decreased when there was serious occlusion. For single target and multi-target detection, the frame processing speed of the model was 3.30 f/s and 3.20 f/s, respectively, and the speed was basically the same, which can lay the foundation for multi-target cow skeleton extraction. The results showed that the model can extract the

收稿日期:2019-11-06　修回日期:2020-02-03

基金项目:陕西省重点产业创新链(群)-农业领域项目(2019ZDLNY02-05)、国家重点研发计划项目(2017YFD0701603)和中央高校基本科研业务费专项资金项目(2452019027)

作者简介:宋怀波(1980—),男,教授,博士,主要从事数字图像处理研究,E-mail:songyangfeifei@163.com

skeleton information of dairy cattle accurately and could be used for the research of lameness and calculation of motion behaviors.

Key words: cows; skeleton extraction model; partial affinity field; deep learning; object keypoint similarity

0 引言

奶牛行走或站立时的姿势、时间与其健康状态密切相关[1-4]。骨架信息可以直观反映奶牛行走或站立时的姿态情况,可为奶牛跛行检测[5-6]、运动量估测[7]、发情行为检测[8]、反刍行为分析[9]等提供重要数据支持。从复杂多变的牛场环境中快速准确地提取奶牛骨架信息,对于提升奶牛养殖业的智能化水平具有重要意义。

骨架提取是运动目标姿态估计的重要手段,大致可分为基于模型、基于特征和基于深度学习 3 种方法。在基于模型的骨架提取及姿态估计研究方面,IVEKOVIC 等[10]基于粒子群优化算法实现了多视图视频序列的人体姿态估计,可从 4 视点视频序列中恢复身体姿势;WANG 等[11]通过初始化人体局部搜索空间,采用树状图模型将检测到的区域整合到人体结构中,最后利用局部特征结合人体整体结构变化来准确推断人体的姿态动作;石祥滨等[12]提出了一种基于连接关系与对称关系的人体姿态估计方法,在图形结构(Pictorial structure, PS)模型的基础上,利用相互连接的部位和对称部位之间的依赖关系实现了无标记静态人体骨架的提取;代钦等[13]提出了一种基于多模整体-星型模型的人体姿态估计方法,该模型可以较好地描述相临近部位之间的分布关系,提高了骨架提取的准确率。此类方法数据量要求低,无需大量训练数据,但模型在部件特征表达上的不足会限制姿态估计的准确度。

在基于特征的骨架提取及姿态估计研究方面,FAN 等[14]提出一种基于图像轮廓的人体姿态估计方法,可以实现复杂背景环境下人体姿态的准确估计;沈春锋等[15]提出一种基于子流形投票统计的残缺人体轮廓姿态估计方法,可以在完成姿态估计的同时修复残缺人体的轮廓;WANG 等[16]提出了一种基于 AdaBoost 算法的人体姿态估计方法,该方法结合自下而上的关节位置分类器和自上而下的整体骨架分类器实现姿态估计;朱珏钰等[17]提出了一种基于二元纯位相匹配滤波器(Binary phase-only filter, BPOF)特征与优化型随机森林的人体姿态估计方法,该方法平均准确率为 95.10%。此类方法需要大量训练数据,一般具有较高的鲁棒性。

在基于深度学习的骨架提取及姿态估计研究方面,PFISTER 等[18]提出了一种基于流动卷积网络的人体姿态估计方法,在 3 个视频姿势估计数据集上取得了较好的效果;WEI 等[19]将卷积网络合并到姿态估计框架中,学习图像特征和与图像相关的空间模型,以进行姿态估计;NEWELL 等[20]使用多尺度特征捕捉人体各关节点的空间位置信息,在 MPII 数据集上的平均准确率为 90.9%;FANG 等[21]利用单镜头多框检测器(Single shot multibox detector, SSD)算法检测人体目标,在 MPII 数据集上的平均精度均值(Mean average precision, mAP)取得了较好的效果;许忠雄等[22]利用双分支深度神经网络同时预测骨点位置和骨点之间的空间关系,并将骨点连接形成图结构,通过图优化实现了多人体姿态估计,平均精度达到 82.60%;CAO 等[23]使用部分亲和场连接身体的各个关键部分,且可以同时学习关键点和部分亲和场,在实现实时性的同时保持了较高的准确性。

目前,人体姿态估计研究相对成熟,可以准确提取多种环境背景下的人体骨架信息。与人体姿态估计相比,奶牛骨架提取存在如下问题:奶牛四肢具有更高的形体相似度,区分前后肢的难度更大;奶牛皮毛花色对骨架提取也有一定的影响,在奶牛黑色皮毛占比较高时,特征提取准确度会受到负面影响;奶牛养殖的开放式环境色彩与奶牛体色相近,泥土、围栏、墙壁对奶牛的遮挡给骨架提取带来更大的困难。本研究结合深度学习技术,采用基于部分亲和场的行走奶牛骨架提取方法,通过采集的奶牛视频图像提取奶牛骨架信息,以期实现奶牛站立或行走时的姿态估计,为大规模奶牛养殖提供一种有效的监测方法。

1 奶牛骨架提取数据集

1.1 试验数据获取及预处理

试验数据采集于杨凌科元克隆股份有限公司的奶牛养殖场,拍摄对象为处于站立或行走状态的荷斯坦奶牛。采用布置于牛场的 DS-2DM1-714 型圆顶摄像机(海康威视,杭州)的监控视频作为原试验数据,监控摄像头与目标奶牛高度一致且距离目标奶牛约 7 m。

试验共采集 75 头奶牛的 50 段视频,在每段视频中随机选取 20 幅图像,共采集 1 000 幅原始图

像，并通过数据增强扩大样本量。传统的数据增强算法多采用图像亮度调整、改变颜色或对比度、旋转样本、增加噪声等方式增加样本数量。在奶牛骨架提取研究方面，上述方法均无法改变奶牛的姿态。因此，本研究在左右对称的骨架结构中采用镜像的方式左右翻转牛体骨架结构，将样本数量增大为 2 000 幅。数据增强后的样本信息如表 1 所示，其中部分测试集为测试集中具有不同干扰因素的 3 个子集，用于分析模型鲁棒性。

表 1　奶牛图像信息统计

Tab. 1　Statistics of cows images

图像数量	类别	奶牛数量	组成
1 600	训练集	60	单目标与双目标
200	验证集	7	单目标与双目标
200	测试集	8	单目标与双目标
3	部分测试集	6	互不遮挡双目标
9	部分测试集	6	遮挡程度不同的双目标
3	部分测试集	3	单目标
50	多目标测试	7	3 目标与 4 目标

由于所选原始图像均来自养殖场监控摄像头，拍摄视野的宽度约为两头成年奶牛的身长。在 3 目标及以上的情况中，奶牛之间的遮挡过于复杂，给样本标注及骨架提取带来极大困难。因此在目标选择上，2 000 幅图像由单目标和双目标 2 类图像组成，比例约为 1∶1，以保证算法对不同数量奶牛的适用性。为了测试 3 目标及以上奶牛骨架提取效果，本研究在原有数据集外增加了多目标测试集，该测试集由 50 幅 3 目标及 4 目标奶牛图像组成。在遮挡方面，双目标数据集中包含了随机数量的从无遮挡到完全遮挡之间各种遮挡程度的图像，以测试算法对奶牛互相遮挡的鲁棒性。如表 2 所示，为测试该方法的应用性，在采集到的 75 头奶牛的 50 段原始视频中选取 10 段单目标和 10 段双目标奶牛视频作为试验数据。部分测试集 2 包含奶牛擦身而过的场景，部分测试集 3 为不存在奶牛互相遮挡的情况，部分测试集 4 为尺度不同的单目标视频。由表 1 可知，养殖场众多复杂的干扰因素（围栏遮挡、泥土沾染、光照变化、人或工具遮挡等）给骨架提取造成了困难。且奶牛作为一个生命体有着丰富的运动行为，进一步增加了奶牛骨架提取的难度。

表 2　奶牛视频信息统计

Tab. 2　Statistics of cows videos

视频数量	类别	奶牛数量	组成
10	部分测试集 1	1	单目标
5	部分测试集 2	2	遮挡程度不同的双目标
5	部分测试集 3	2	互不遮挡双目标
1	部分测试集 4	1	尺度不同的单目标

1.2　奶牛关键点标签制作

模型训练需要对奶牛图像进行标注。在本研究中，标注信息主要包括图像中奶牛的位置信息、轮廓信息以及由 17 个关键点按指定顺序连接的骨架信息。为了更快捷地获取奶牛的轮廓信息，本研究使用 Labelme 标注工具对原始图像进行标注，分割出牛体区域。如图 1 所示，本研究使用多边形标注奶牛轮廓，标注时使拐点尽可能地接近奶牛的实际轮廓。由于牛尾在牛体姿态监测中影响较小且形状细长多变，因此，在本研究中将牛尾作为遮挡物处理。

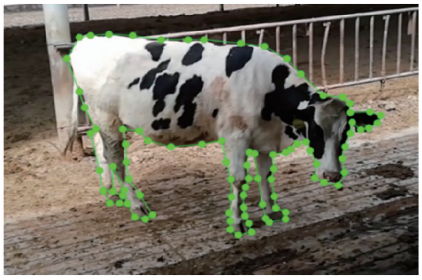


图 1　Labelme 标注奶牛轮廓

Fig. 1　Labeling cows outline using Labelme

奶牛身体区域主要分为头部、背部、腹部、四肢以及尾部 5 个区域。其中背部和腹部区域面积最大，但其运动频率较低且运动形变较小，实际可监测的运动信息偏少。尾部区域过于灵活，易被遮挡且与本研究相关性较小。考虑到这些因素，本研究中着重对奶牛的头部和四肢区域标注关键点。在奶牛的头部区域，对辨识度较高的嘴部、眼睛、耳朵进行标注，在四肢区域，对牛肢的 3 个关节处布置关键点，共计 17 个点，如图 2 所示。其中，编号 1 ~ 17 分别为牛嘴、左眼、右眼、左耳、右耳、前左腿根部、前右腿根部、前左膝、前右膝、前左蹄、前右蹄、后左腿根部、后右腿根部、后左膝、后右膝、后左蹄、后右蹄。

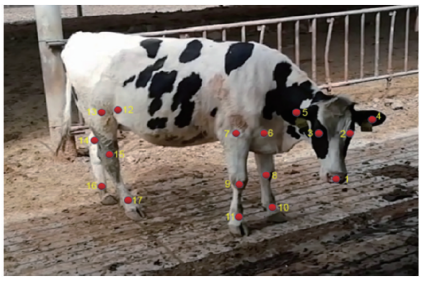


图 2　牛体关键点标定

Fig. 2　Cow key points calibration

在样本图像中，奶牛头部标注了 5 个关键点，四肢区域标注了 12 个关键点。关键点又分为可见、遮挡不可见和未遮挡不可见 3 种情况。在标注时对每个标注点都添加了用以区分以上 3 种情况的标签数据：2 表示可见、1 表示遮挡不可见、0 表示未遮挡不可见。可见时返回当前点坐标信息，遮挡不可见时

返回遮挡物位置信息,未遮挡不可见时将位置信息清零。在本研究中,图像标签采用 COCO2017 数据集的 Object Keypoints 标注类型。为确认标注数据的准确性,将标注信息和与之对应的原始图像结合显示结果如图 3 所示。标注后的图像可见奶牛轮廓信息、实心点代表的可见关键点、环状点代表的被遮挡关键点以及关键点连线的奶牛骨架。

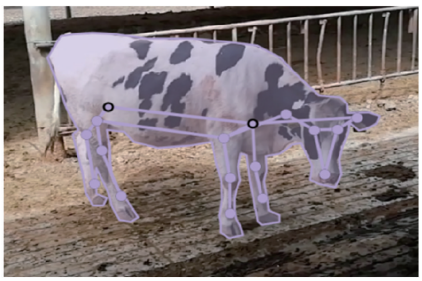


图 3 标注信息可视化

Fig. 3 Annotation information visualization

本研究基于部分亲和场,开展奶牛骨架提取方法研究,为了更快更好地使用现有资源训练卷积神经网络,本研究使用 Windows 10 系统的台式计算机作为开发平台,其硬件配置为 Intel Xeon CPU、E5 - 1620 v4 处理器和 NVIDIA Quadro P2000 显卡。编程语言为 Python3.6,深度学习框架采用 Tensorflow。

2 网络构建

基于部分亲和场的卷积神经网络是基于深度学习的网络结构,本研究选用 VGG - 19 作为前馈网络提取图像特征。为了同时预测奶牛的关键点信息和部分亲和场,本研究采用多阶段二分支网络对前馈网络提取的图像特征进行预测,并通过最优匹配组合关键点和部分亲和场连接出奶牛的骨架结构[23]。由于奶牛体积较大、体表平坦,且皮毛一般为白色和黑色,相比于人体图像特征提取,奶牛图像特征提取对尺度信息更加敏感,提取特征时很容易损失图像细节。因此,本研究在人体姿态估计网络的基础上增加了特征融合。同时提取奶牛图像的浅层和深层特征以减少卷积造成图像细节的损失,在一定程度上可以提高网络的鲁棒性。

2.1 VGG 前馈网络

奶牛四肢的高相似度、皮毛颜色和环境遮挡给奶牛图像特征提取带来较大困难。本研究在以 VGG - 19 为前馈网络的基础上,结合深度特征融合思想,提取奶牛图像的高、低层图像特征以期提高奶牛骨架提取精度。VGG - 19 具有简洁的架构,整个网络使用了同一种卷积核尺寸(3 × 3)和最大池化尺寸(2 × 2),更小的卷积核增加了网络的非线性表达能力,减少了参数量。卷积层低层特征分辨率高,包含更多的位置、细节信息,但其语义性更低,噪声更多;高层特征具有更强的语义信息,分辨率很低,对细节的感知能力较差。如图 4 所示,本研究融合 VGG - 19 第 2、4、8、12 层的图像特征作为前馈网络的特征输出。第 2、4 层提取图像的纹理、颜色、边缘、棱角等低维特征,第 8、12 层提取奶牛语义、整体结构等高维特征。结合深度特征融合思想使前馈网络提取更多的图像特征,提高奶牛骨架提取的准确度。

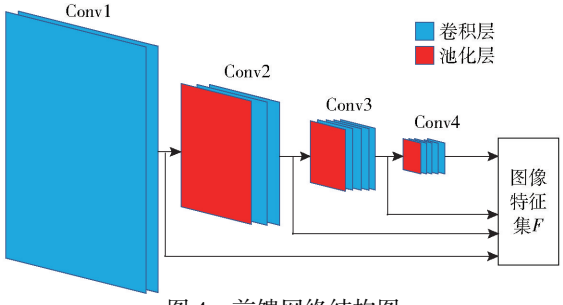


图 4 前馈网络结构图

Fig. 4 Structure diagram of feedforward network

2.2 基于部分亲和场的多阶段二分支网络

传统的姿态估计方法采用从上到下的骨架提取方式,先进行目标检测,再对目标进行骨架提取,耗时较长。为了更快更准确地提取奶牛骨架,本研究采用从下到上的方式,利用全局上下文信息,同时检测奶牛的关键点信息和部分亲和场信息[23]。检测到关键点后,结合部分亲和场推测关键点的连接顺序,展现出奶牛的骨架结构。

2.2.1 二分支网络

为了同时预测关键点信息和部分亲和场信息,将前馈网络输出的图像特征集 F 分别输入到上下分支的卷积神经网络中。二分支网络结构图[23]如图 5 所示,二分支网络的上下分支分别预测关键点信息和部分亲和场信息。

第 1 阶段二分支网络含 3 个 3 × 3 卷积层和 2 个 1 × 1 卷积层,特征图组输入后,上分支网络预测一组关键点信息,下分支网络预测一组部分亲和场信息。在后续每个阶段中含 5 个 7 × 7 卷积层和 2 个 1 × 1 卷积层,输入为前馈网络所提取的图像特征集以及上一阶段的二分支网络预测的关键点信息和部分亲和场。t 阶段对应的上下支网络结构为

$$S^t = \rho^t(F, S^{t-1}, L^{t-1}) \quad (t \geqslant 2) \tag{1}$$

$$L^t = \varphi^t(F, S^{t-1}, L^{t-1}) \quad (t \geqslant 2) \tag{2}$$

式中 ρ^t——t 阶段上分支网络的运算量

φ^t——t 阶段下分支网络的运算量

S^t——t 阶段上分支网络预测结果

L^t——t 阶段下分支网络预测结果

t——二分支网络阶段序号

二分支网络同时预测关键点信息和部分亲和场

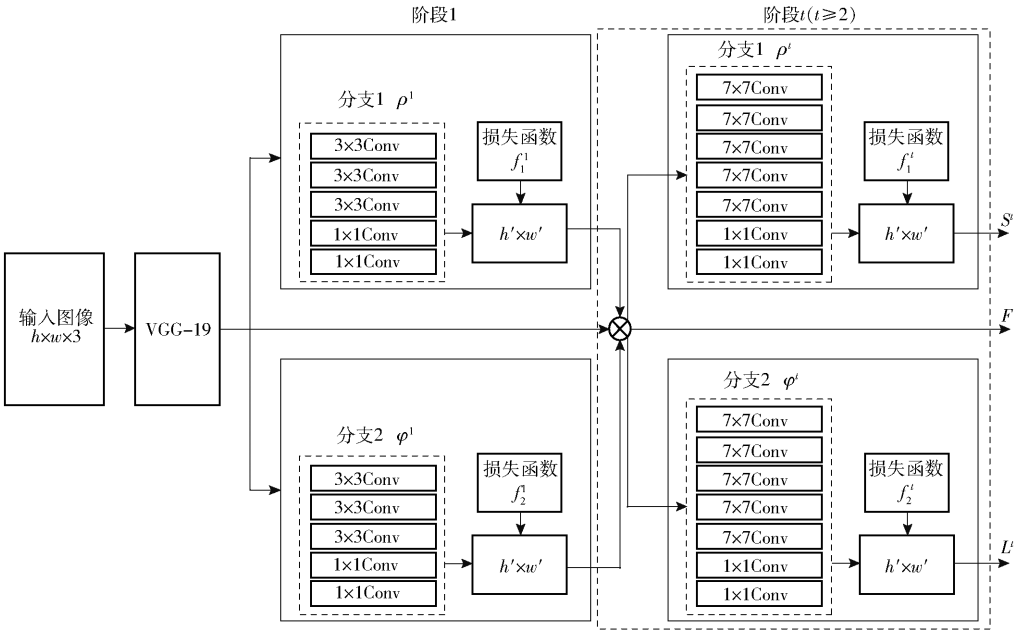


图 5　二分支网络结构图

Fig. 5　Structure diagram of two-branch network

信息，每个阶段的两个分支都有独立的损失函数。本研究将预测值和真实值之间距离平方和的均方误差作为损失函数。

2.2.2　热点图

将 RGB 图像作为原始输入，经过前馈网络提取图像特征集 F。在标注了关键点信息的二维矩阵中得到置信图 $\boldsymbol{S}^*$，每一个置信图代表一个置信度的二维矩阵，而置信度是指某个牛体关键点在各像素区域出现的概率，且其位置唯一。在一般情况下，如果图像中仅有一个目标奶牛且该目标的关键点可被标记，那么在每个关键点的置信图中都会有一个峰值；但如果图像中出现多个目标，那么针对目标 k 的可见关键点 j 会有一个峰值。设目标 k 的置信图为 $\boldsymbol{S}_{j,k}^*$，以 R 为半径的区域中任意一点 P 处对应的置信度为

$$\boldsymbol{S}_{j,k}^*(P)=\exp\left(-\frac{\|\boldsymbol{P}-\boldsymbol{X}_{j,k}\|_2^2}{\sigma^2}\right)\times 100\% \tag{3}$$

式中　$\boldsymbol{X}_{j,k}$——奶牛 k 的关键点 j 的像素坐标

σ——峰宽

$\boldsymbol{P}$——以 R 为半径的区域中的任意一点坐标

当有多个目标时，上分支网络会预测到多个关键点信息，置信度曲线也会出现多个峰值，通过提取高斯曲线的最大值得到每个目标的关键点信息。

2.2.3　部分亲和场

二分支网络的上分支可以预测奶牛的关键点位置，下分支预测部分亲和场存储奶牛肢体的位置和方向信息，利用部分亲和场信息对关键点进行连接。一般情况下，假设只有一个奶牛腿部区域的情况，如图 6 所示。此区域中存在 3 个奶牛关键点，为了便于讨论，只标记了 $X_{j_1,k}$ 和 $X_{j_2,k}$ 两个关键点。P 为图像中任意一点，$\boldsymbol{v}_\perp$ 表示与单位向量 $\boldsymbol{v}$ 垂直方向的向量。

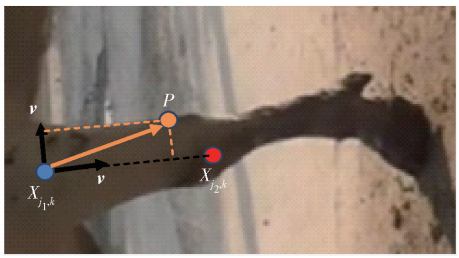


图 6　部分亲和场示意图

Fig. 6　Partial affinity field diagram

其中，$X_{j_1,k}$ 和 $X_{j_2,k}$ 分别对应奶牛 k 的肢体 c 的关键点 j_1 和 j_2 的正确坐标。将 $\boldsymbol{L}_{c,k}^*(P)$ 作为该肢体的部分亲和场。

$$\boldsymbol{L}_{c,k}^*(P)=\begin{cases}\boldsymbol{v} & (\text{点 } P \text{ 不在奶牛 } k \text{ 的肢体 } c \text{ 中})\\ \boldsymbol{O} & (\text{点 } P \text{ 在奶牛 } k \text{ 的肢体 } c \text{ 中})\end{cases} \tag{4}$$

式中 $\boldsymbol{O}$ 为零向量。对任意点 P 在肢体区域内时，$\boldsymbol{L}_{c,k}^*(P)$ 的值就是该点位置的单位向量 $\boldsymbol{v}$。P 在其他位置时，$\boldsymbol{L}_{c,k}^*(P)$ 都为零向量。

但实际奶牛肢体的二维区域由无数个点组成，点集中的任意点 P 需要小于距离阈值，即满足条件

$$\begin{cases}0\leqslant|\boldsymbol{v}(\boldsymbol{P}-\boldsymbol{X}_{j_1,k})|\leqslant l_{c,k}\\ |\boldsymbol{v}_\perp(\boldsymbol{P}-\boldsymbol{X}_{j_1,k})|\leqslant\sigma_l\end{cases} \tag{5}$$

式中　$l_{c,k}$——相邻两关键点间奶牛肢体的长度

σ_l——相邻两关键点间奶牛肢体的宽度

若点 P 处有多个目标 k 的肢体重叠，则其向量场为所有目标向量场的均值。得到奶牛肢体区域中每个

点 P 处的向量,在两个相邻关键点之间的准线上取点 $\boldsymbol{P}(u)$。

$$\boldsymbol{P}(u)=(1-u)\boldsymbol{d}_{j_1}+u\boldsymbol{d}_{j_2} \tag{6}$$

式中 $\boldsymbol{P}(u)$——关节点上的采样点

$\boldsymbol{d}_{j_1}$——关键点 j_1 的预测坐标

$\boldsymbol{d}_{j_2}$——关键点 j_2 的预测坐标

u——$\boldsymbol{d}_{j_1}$ 与 $\boldsymbol{d}_{j_2}$ 的相对距离,取值为 0~1

将每个点 $P(u)$ 位置预测的部分亲和向量 $\boldsymbol{L}(P(u))$ 与从关键点 j_1 指向 j_2 的单位向量相乘后再求和。

连接关键点 j_1 和 j_2 的肢体权重计算式为

$$E=\int_0^1 \boldsymbol{L}_c(P(u))\frac{\boldsymbol{d}_{j_2}-\boldsymbol{d}_{j_1}}{\|\boldsymbol{d}_{j_2}-\boldsymbol{d}_{j_1}\|_2}\mathrm{d}u \tag{7}$$

即所有点 $P(u)$ 的部分亲和场在 j_1 到 j_2 方向上投影的总和,当 $P(u)$ 位置的部分亲和向量与单位向量 $\boldsymbol{v}$ 同向时,得到权重 E 最大。由于相邻两关键点之间的线段上有无数个点,实际操作中在线段上进行等间隔取样,计算一定数量 $P(u)$ 点的权重 E。

2.2.4 匹配关键点和部分亲和场

匹配关键点和部分亲和场为骨架时,不同目标的关键点可能产生错误连接,如图 7 中连接 $d_{j_1}^1$ 与 $d_{j_2}^2$。



图 7 关键点匹配示意图

Fig. 7 Key point matching diagram

设有 J 种类型的关键点,同一类型 j 的关键点有 N_j 个,通过上分支网络,预测候选点集的集合 D_J 为

$$D_J=\{d_j^m:j\in\{1,2,\cdots,J\},m\in\{1,2,\cdots,N_j\}\} \tag{8}$$

式中 j——关键点类型序号

m——关键点序号

d_j^m——类型 j 的第 m 个关键点

如图 7 所示,假设图中有两个目标,且只考虑图中标注的两头奶牛 c 类型肢体。同一类型的关键点分别有两个,因此共有 4 种连接。寻找正确的连接将成为一个二分图匹配问题,将关键点集 D_{j_1}、D_{j_2} 看作图节点,将关键点的连接看作图边。以同一个节点不允许同时被两条边连接为原则,在所有的边中筛选出满足条件的边。再结合评估部分亲和场的权重 E,找到最大权重的边。用 E_c 表示 c 类肢体权重的和,计算式为

$$\max_{Z_c}E_c=\max_{Z_c}\sum_{m\in D_{j_1}}\sum_{n\in D_{j_2}}E_{mn}z_{j_1,j_2}^{m,n} \tag{9}$$

式中 $z_{j_1,j_2}^{m,n}$——两点之间的连接状态,取 0 或 1

Z_c——c 类肢体的子集

E_{mn}——关键点 $d_{j_1}^m$ 和 $d_{j_2}^n$ 之间部分亲和场权重

D_{j_1}——j_1 类关键点集

D_{j_2}——j_2 类关键点集

两个边连接同一节点需满足

$$\begin{cases}\sum\limits_{n\in D_{j_2}}z_{j_1,j_2}^{m,n}\leqslant 1 & (m\in D_{j_1})\\ \sum\limits_{m\in D_{j_1}}z_{j_1,j_2}^{m,n}\leqslant 1 & (n\in D_{j_2})\end{cases} \tag{10}$$

即没有相同的肢体连接到同一关键点,最后通过 Hungarian 算法找到正确的边。

3 测试结果与分析

本研究以 16∶2∶2的比例将试验样本分为训练集、验证集和测试集(表 1),并在测试集中分别选取 100 幅单目标、100 幅双目标和 50 幅 3 目标及以上的奶牛原始图像。本研究测试奶牛每类关键点的置信度,再通过求均值的方式计算奶牛整体置信度。对于多目标情况,各类关键点置信度取所有目标的均值。为了便于记录,本研究使用骨架中的关键点编号如图 2 所示。

本研究用 OKS 指标评价预测置信度。置信度 1、置信度 2 和置信度 3 分别记录单目标、双目标、3 目标及以上奶牛各类关键点的置信度,其中置信度的计算公式为

$$O_{KSp}=\frac{\sum\limits_i\exp(-d_{p^i}^2/(2S_p^2\sigma_i^2))\delta}{\sum\limits_i\delta}\times 100\% \tag{11}$$

式中 p——目标奶牛 O_{KSp}——置信度

p^i——目标奶牛 p 的关键点 i

$d_{p^i}^2$——标注点与预测关键点之间的欧氏距离

S_p^2——目标 p 在二维图像中的面积

σ_i——关键点 i 的归一化因子,即权重,表示该点对整体的影响程度

δ——关键点可见程度,0 表示不可见,1 表示可见

测试集测试结果如表 3 所示,目标数为 1 时各类关键点置信度不小于 72.00%,且不同类别间置信度差异较大,奶牛整体置信度为 78.90%。目标数为 2 时各类关键点的置信度均有所降低,整体置

信度为 67.94%，表明在预测单目标奶牛骨架信息时的置信度较高。考虑到奶牛间的复杂遮挡及其他干扰因素增多会给提取奶牛骨架造成较大干扰，因此提取到的双目标奶牛骨架信息置信度相对较低。目标数为 3 及以上时，整体置信度为 48.59%，由于数据均来自养殖场监控摄像头，拍摄视野的宽度约为两头成年奶牛的身长。如图 8c 所示，在 3 目标及以上的情况中，奶牛之间的遮挡过于复杂，预测难度增大。试验表明，本研究提取 3 目标及以上奶牛骨架的准确度较低。

表 3　各类关键点的置信度

Tab. 3　Confidence of various key points　　%

关键点序号	1	2	3	4	5	6	7	8	9
置信度 1	83.80	76.56	80.73	82.57	77.82	79.27	80.87	83.72	80.69
置信度 2	70.05	65.08	67.30	78.56	76.43	61.12	73.53	64.09	78.61
置信度 3	49.03	45.66	48.00	51.14	45.98	49.49	53.12	52.48	46.45
关键点序号	10	11	12	13	14	15	16	17	平均值
置信度 1	79.99	76.71	74.09	72.00	79.79	79.90	75.40	77.40	78.90
置信度 2	67.47	74.77	60.70	65.42	66.03	63.03	59.32	63.44	67.94
置信度 3	49.00	48.95	44.98	47.81	50.38	52.55	44.29	46.69	48.59



(a) 单目标骨架



(b) 双目标骨架



(c) 3目标及以上骨架

图 8　部分模型骨架提取测试结果

Fig. 8　Parts of model prediction results

后左腿根部、后右腿根部两类关键点与整体水平相差最大。如图 8a 所示，以奶牛身体侧面 90°为观测角度时，奶牛另一侧后腿根部总处于被遮挡的情况，且相比于其他关键点位置，奶牛后腿根部体积更大且更平坦，在二维图像中表现出的纹理特征也较少一些。在本次研究中，所有原始图像均通过镜像的方式以增加数据集整体的数量，由于奶牛四肢高度的相似性和镜像所造成的左右互换性，在四肢区域的 12 类关键点中，同一位置左右侧关键点的置信度相近，使得后腿根部左右两侧关键点置信度较低。

在置信度 2 中，双目标奶牛的各类关键点置信度均低于置信度 1，且整体置信度下降为 67.94%。如图 8b 所示，两头奶牛并行时存在不同程度的互相遮挡，且被遮挡部分的预测结果较差。当遮挡面积约 70% 时，只能预测到镜头方向奶牛的骨架信息。

在本研究中，计算了测试集的 OKS 值，即所有目标奶牛的 OKS 值。从 0.50 ~ 0.95 以 0.05 等间隔设置 OKS 值，计算不同 OKS 值下的准确率与召回率，结果如表 4 所示。

OKS 值越高，准确率和召回率越低，当 OKS 取 0.5，即宽松标准(loose metric)时，准确率为 97.30%，

表 4 不同 OKS 值的模型准确率和召回率

Tab. 4 Model accuracy and recall rate of different OKS value %

OKS 值	准确率	召回率
0.50	97.30	97.00
0.55	97.30	97.00
0.60	97.30	97.00
0.65	95.80	95.90
0.70	93.40	94.20
0.75	93.40	94.20
0.80	86.90	84.90
0.85	83.50	83.20
0.90	83.50	83.20
0.95	77.00	77.90

召回率为 97.00%。当 OKS 取 0.75 时,准确率为 93.40%,召回率为 94.20%,表明本研究算法提取奶牛骨架具有较高的准确率。本试验使用测试集图像的部分原始视频测试该模型的检测速度,在只有单目标时为 3.30 f/s,双目标时为 3.20 f/s。

4 奶牛骨架提取的影响因素分析

由于奶牛养殖环境复杂多变,由牛场监控摄像头得到的奶牛图像存在拍摄视野过小、奶牛被围栏遮挡、牛体之间复杂多变的遮挡、光照引起的图像亮度变化等各种干扰因素[24-26]。为研究算法的鲁棒性,对影响骨架提取的干扰因素进行分析。

4.1 奶牛数量对预测置信度的影响

为了避免奶牛花色形状、亮度、遮挡程度等其他因素对测试造成干扰,本研究从测试集的双目标子集中随机挑选出互不遮挡的 3 幅双目标奶牛图像,测试出每头奶牛骨架的置信度后,将双目标图像裁剪为单目标奶牛,再对 6 幅只有单头奶牛的图像进行测试,测试结果如表 5 所示。

表 5 不同奶牛数量时的置信度

Tab. 5 Confidence in different cow quantities %

目标	单目标	双目标
奶牛 1	80.854	80.473
奶牛 2	68.812	68.358
奶牛 3	73.241	72.931
奶牛 4	81.121	77.941
奶牛 5	72.502	73.210
奶牛 6	67.518	66.137

结果表明,3 幅双目标奶牛中,置信度差值最大为 3.18 个百分点,最小为 0.31 个百分点,平均差异为 1.1 个百分点。数据表明,奶牛数量增加对骨架置信度的影响较小,证明本文算法适用于多目标奶牛的骨架提取。

4.2 双目标奶牛间的遮挡程度对置信度的影响

为了测试奶牛之间遮挡对骨架提取的影响,本研究从测试集挑选出 3 组奶牛图像,定义为图像 1、2、3。当遮挡超过 50% 时,目标奶牛被遮挡的关键点数量过多,可预测到的关键点数量过少,无法连接出骨架信息,且在二维图像中奶牛形体不规则,准确估计遮挡面积难度较大。本研究的筛选条件为相同两头奶牛不遮挡、遮挡约 25% 和遮挡约 50% 共 3 等遮挡程度。被遮挡奶牛置信度测试结果如图 9 所示。

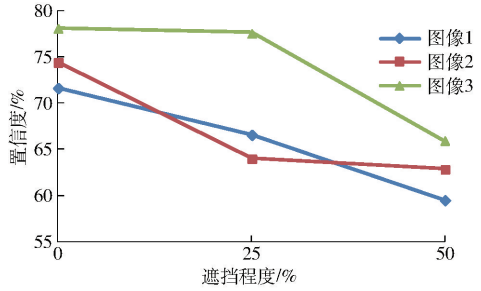


图 9 不同遮挡时的置信度

Fig. 9 Confidence under different occlusions

测试结果表明,被遮挡约 25% 时 3 组奶牛平均置信度下降了 5.3 个百分点,被遮挡约 50% 时较未遮挡置信度平均下降了 12 个百分点。这是由于当待测目标奶牛被其他奶牛遮挡时,网络优先预测其他奶牛的关键点。目标奶牛被遮挡部分关键点未被预测,无法对剩余关键点的预测提供有用信息,使整体的骨架提取置信度与被遮挡面积近似成线性关系。

4.3 光照对预测置信度的影响

在测试集中,随机采集 3 幅不同奶牛图像,分别记为牛 41、牛 62、牛 77。为了保证测试结果不被光照以外的因素干扰,本研究通过对同一幅图像调整亮度的方式模拟光线强度对图像的影响。当原图像亮度减小超过 50% 时,监控摄像头将开启夜视功能。因此本研究保持图像对比度、饱和度不变,分别增大 25%、50% 亮度和减小 25%、50% 亮度。测试结果如图 10 所示。

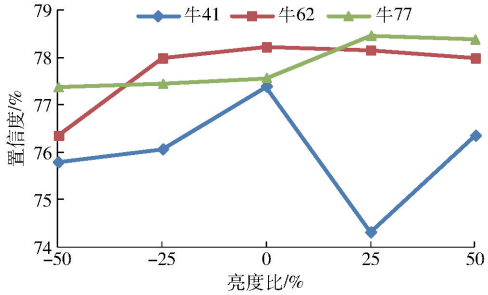


图 10 不同亮度时的置信度

Fig. 10 Confidence at different brightness values

由图 10 可知，牛 41、牛 62、牛 77 3 组图像改变亮度后置信度方差分别为 0.000 392、0.000 195、0.000 089。测试结果表明，在光线强度满足摄像头未开启夜视功能的情况下对奶牛的骨架提取影响较小。充分说明本研究方法在实际奶牛养殖中的有效性。

5　奶牛骨架提取应用试验

5.1　试验测试结果

在实际应用中，原始数据多以视频的形式进行存储。为了进一步验证该方法的实际应用性，本研究在采集到的 75 头奶牛的 50 段原始视频中选取 10 段单目标和 10 段双目标奶牛视频作为试验数据。奶牛数量增加对骨架置信度的影响较小，但奶牛间的互相遮挡会对骨架提取造成干扰。为了不失一般性，试验中选取的 10 段双目标测试视频中有 5 段视频包含奶牛擦身而过的运动场景，5 段视频为不存在奶牛互相遮挡的情况。其中由于奶牛行走速度的差异，将 20 段视频的时长剪切至 10 ~ 12 s。

在视频中，每帧图像提取到的奶牛骨架置信度存在一定差异。这是由于一系列奶牛的连续动作中，某些动作的奶牛轮廓清晰、关键点部位全部可见，骨架提取准确度也相对较高；相反，在部分情景中，牛体被严重遮挡，大部分关键部位处于不可见状态，骨架的整体置信度会降低。如图 11 所示，在双目标情景中，内侧奶牛被外侧奶牛遮挡严重，大部分关键点的位置都需要预测，其整体置信度较低。在本试验中，当奶牛骨架整体置信度低于 60% 时，将其视为不可信，在测试时不画出其骨架的点线模型，同时认为该帧图像为漏帧状态。



(a) 关键点　(b) 部分亲和场　(c) 骨架提取结果

图 11　遮挡程度超过 50% 的情况

Fig. 11　Situation of serious occlusion

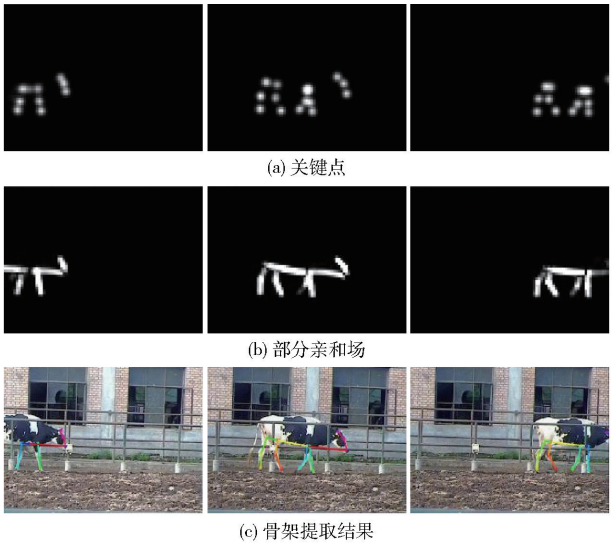
本试验中，使用帧间差分法每 3 帧取一次图像进行测试。同时使用漏帧率和置信度评测该方法在视频中的骨架提取效果。其中漏帧率为漏帧数与总帧数的比值，测试结果如表 6 所示。

表 6　视频骨架提取置信度与漏帧率

Tab. 6　Confidence and missing frame rate of video skeleton extraction　%

视频类别	置信度	漏帧率
单目标	78.30	3.13
双目标无遮挡	76.15	7.08
双目标有遮挡	62.46	27.59

其中，单目标视频和双目标无遮挡视频的骨架提取置信度相对较高，漏帧率较低，具有较高的实际应用性。部分测试结果如图 12 所示，基于部分亲和场的奶牛骨架提取模型同时预测奶牛关键点和部分亲和场，通过匹配上下支网络预测结果提取奶牛骨架；双目标有遮挡视频奶牛骨架提取置信度为 62.46%，明显低于单目标和双目标无遮挡情况。同时，由于遮挡的影响，整体置信度低于阈值的帧数较多，漏帧率为 27.59%，明显高于另外 2 种情况，实际应用性较低。



(a) 关键点

(b) 部分亲和场

(c) 骨架提取结果

图 12　视频骨架提取的部分结果

Fig. 12　Partial results of video skeleton extraction

5.2　不同尺度对奶牛骨架提取的影响

在拍摄时，调整奶牛与摄像机之间的距离，图像中奶牛的尺度会有所变化。为验证在一定范围内改变图像中奶牛尺度并不会影响网络提取奶牛骨架，本研究使用一段有明显距离变化的单目标奶牛视频为原始数据，采用跳帧的方式将其分为一系列奶牛图像，通过提取不同位置的奶牛骨架并可视化结果来测试网络对目标尺度的鲁棒性。

用奶牛在图像中所占面积百分比来度量拍摄点到奶牛的距离，通过计算奶牛在不同占比时的整体置信度来验证不同尺度对奶牛骨架提取的影响。7 个不同尺度的骨架提取结果如图 13 所示。

奶牛面积占比约为 50% 时，预测置信度达到最大，为 80.7%。增大奶牛面积占比，置信度在较小范围内波动。奶牛面积占比减小时，置信度呈下降趋势。在奶牛面积占比约为 20% 时，置信度下降至 65.3%。若继续减小奶牛面积占比，则置信度最终将低于 60%，变为不可信结果。其中部分测试的可视化结果如图 14 所示。

由于卷积神经网络通过卷积、池化等方式提取图像特征，随着网络深度的增加，图像细节信息会逐

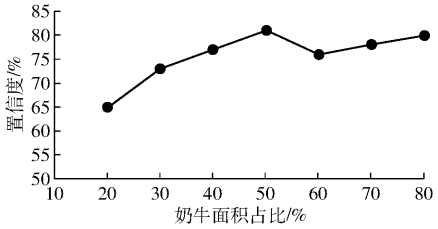


图 13 不同尺度奶牛骨架的提取结果

Fig. 13 Results of skeleton extraction in different scales of dairy cattle



(a) 面积占比约20% (b) 面积占比约50% (c) 面积占比约80%

图 14 部分测试的可视化结果

Fig. 14 Visualization of partial results

渐丢失。本研究通过特征融合同时提取卷积神经网络的深层语义特征和浅层细节特征,在一定程度上可以减少图像的细节丢失。但拍摄距离过远时,图像模糊导致细节丢失严重,则该网络会无法获得可靠结果。

6 结论

(1)采用本文方法可以较准确地提取单目标奶牛行走或站立时的骨架信息。在无遮挡的情况下,可以较准确地提取多目标奶牛行走的骨架信息。该方法鲁棒性较高,在一定范围内提取精度不会被亮度等因素干扰,说明采用本文方法在未开夜视功能时提取行走奶牛骨架信息是可行的。

(2)在提取单目标奶牛骨架时,平均置信度为 78.90%。在提取双目标奶牛骨架时,奶牛相互遮挡使得置信度下降为 67.94%,在双目标以上时,置信度为 48.59%。研究表明,单目标奶牛的骨架提取置信度最高;双目标时,奶牛之间存在遮挡,骨架置信度有所降低;3 目标及以上时,遮挡概率增大,对骨架置信度的影响进一步加大。

(3)在无遮挡或遮挡度较低时,采用本文方法可以较准确地提取多目标奶牛骨架信息;奶牛互相遮挡时,在一定范围内随着遮挡程度的增大,骨架置信度逐渐降低,说明本文方法对奶牛之间遮挡的干扰尚有待改善。

(4)采用本文方法可以提取视频中奶牛骨架,在无遮挡时具有高置信度和低漏检率,遮挡严重时置信度有所下降。在实际应用时,奶牛大多处于可见状态,被遮挡严重的情形很少。因此,本文方法可以用于提取奶牛骨架、监测奶牛信息,为奶牛智能养殖研究提供一种可靠方法。

参 考 文 献

[1] KING M T M, LEBLANC S J, PAJOR E A, et al. Cow-level associations of lameness, behavior, and milk yield of cows milked in automated systems[J]. Journal of Dairy Science, 2017, 100(6):4818 – 4828.

[2] POURSABERI 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.

[3] 何东健,孟凡昌,赵凯旋,等. 基于视频分析的犊牛基本行为识别[J/OL]. 农业机械学报,2016,47(9):294 – 300.
HE Dongjian, MENG Fanchang, ZHAO Kaixuan, et al. Recognition of calf basic behaviors based on video analysis[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(9): 294 – 300. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20160940&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.09.040. (in Chinese)

[4] 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/OL]. 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/OL]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(Supp.):290 – 296. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=2019s045&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.S0.045. (in Chinese)

[5] 宋怀波,阴旭强,吴峁华,等. 基于自适应无参核密度估计算法的运动奶牛目标检测[J/OL]. 农业机械学报,2019,50(5):196 – 204.
SONG Huaibo, YIN Xuqiang, WU Dihua, et al. Detection of moving cows based on adaptive kernel density estimation algorithm [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019,50(5):196 – 204. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20190523&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.023. (in Chinese)

[6] 宋怀波,姜波,吴倩,等. 基于头颈部轮廓拟合直线斜率特征的奶牛跛行检测方法[J]. 农业工程学报,2018,34(15):190 – 199.
SONG Huaibo, JIANG Bo, WU Qian, et al. Detection of dairy cow lameness based on fitting line slope feature of head and neck outline[J]. Transactions of the CSAE,2018,34(15):190 – 199. (in Chinese)

[7] 顾静秋,王志海,高荣华,等. 基于融合图像与运动量的奶牛行为识别方法[J/OL]. 农业机械学报,2017,48(6):145 – 151.
GU Jingqiu, WANG Zhihai, GAO Ronghua, et al. Recognition method of cow behavior based on combination of image and

activities[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(6): 145 - 151. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20170619&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.06.019. (in Chinese)

[8] 刘忠超,何东健. 基于卷积神经网络的奶牛发情行为识别方法[J/OL]. 农业机械学报,2019,50(7):186 - 193.
LIU Zhongchao, HE Dongjian. Recognition method of cow estrus behavior based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 186 - 193. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20190719&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.07.019. (in Chinese)

[9] 宋怀波,李通,姜波,等. 基于 Horn - Schunck 光流法的多目标反刍奶牛嘴部自动监测[J]. 农业工程学报,2018,34(10):163 - 171.
SONG Huaibo, LI Tong, JIANG Bo, et al. Automatic detection of multi-target ruminate cow mouths based on Horn - Schunck optical flow algorithm[J]. Transactions of the CSAE, 2018, 34(10): 163 - 171. (in Chinese)

[10] IVEKOVIC S, TRUCCO E. Human body pose estimation with PSO[C]// Proceedings of IEEE International Conference on Evolutionary Computation. Canada, 2006.

[11] WANG L, CHENG L, THI T H, et al. Human action recognition from boosted pose estimation[C]// Proceedings of Digital Image Computing: Techniques and Applications. Australia, 2010.

[12] 石祥滨,丁肖,代钦,等. 基于连接关系与对称关系的人体姿态估计方法[C]//第十四届中国虚拟现实大会. 沈阳,2014.
SHI Xiangbin, DING Xiao, DAI Qin, et al. Pose estimation based on connection and symmetry[C]// Proceedings of the 14th China Virtual Reality Conference. Shenyang, 2014. (in Chinese)

[13] 代钦,石祥滨,乔建忠,等. 基于整体-星型模型的人体姿态估计[J]. 小型微型计算机系统,2018,39(12):2687 - 2693.
DAI Qin, SHI Xiangbin, QIAO Jianzhong, et al. Human pose estimation with global constellated model[J]. Journal of Chinese Mini - Micro Computer Systems, 2018, 39(12): 2687 - 2693. (in Chinese)

[14] FAN Bin, WANG Zengfu. Pose estimation of human body based on silhouette images[C]// Proceedings of Information Acquisition. China, 2004.

[15] 沈春锋,林学阎,史元春. 从残缺的人体轮廓进行姿态估计[J]. 清华大学学报(自然科学版),2008, 48(10): 1659 - 1662.
SHEN Chunfeng, LIN Xueyan, SHI Yuanchun. Human pose estimaties from corrupt silhouettes[J]. Journal of Tsinghua University(Science and Technology), 2008, 48(10): 1659 - 1662. (in Chinese)

[16] WANG S, AI H, YAMASHITA T, et al. Combined top-down/bottom-up human articulated pose estimation using AdaBoost learning[C]// Proceedings of the 20th International Conference on Pattern Recognition. Turkey, 2010.

[17] 朱珏钰,曹亚微,周书仁,等. 基于随机森林深度特征选择的人体姿态估计[J]. 计算机工程与应用,2017,53(2):172 - 176.
ZHU Jueyu, CAO Yawei, ZHOU Shuren, et al. Human pose estimation based on random forest depth feature selection[J]. Computer Engineering and Applications, 2017, 53(2): 172 - 176. (in Chinese)

[18] PFISTER T, CHARLES J, ZISSERMAN A. Flowing convnets for human pose estimation in videos[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, 2015.

[19] WEI S, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016.

[20] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]// Proceedings of the Computer Vision(ECCV). Amsterdam, 2016.

[21] FANG H, XIE S, TAI Y, et al. RMPE: regional multi-person pose estimation[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Italy, 2017.

[22] 许忠雄,张睿哲,石晓军,等. 深度学习实时多人姿态估计与跟踪[J]. 中国电子科学研究院学报,2018,13(4):491 - 496.
XU Zhongxiong, ZHANG Ruizhe, SHI Xiaojun, et al. Real-time pose estimation and tracking base on deep learning[J]. Journal of China Academy of Electronics and Information Technology, 2018, 13(4): 491 - 496. (in Chinese)

[23] 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). America, 2017.

[24] WU Dihua, WU Qian, YIN Xuqiang, et al. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector[J]. Biosystems Engineering, 2020, 189: 150 - 163.

[25] JIANG B, SONG H B, HE D J. Lameness detection of dairy cows based on a double normal background statistical model[J]. Computers and Electronics in Agriculture, 2019, 158: 140 - 149.

[26] JIANG B, WU Q, YIN X Q. FLYOLOv3 deep learning for key parts of dairy cow body detection[J]. Computers and Electronics in Agriculture, 2019, 166: 1 - 8.