山东农业大学学报(自然科学版),2026,57(4):750-762 VOL.57 NO.4 2026
Journal of Shandong Agricultural University (Natural Science Edition) doi:10.3969/j.issn.1000-2324.2026.04.016

复杂养殖环境下基于骨架特征的牛体行为识别

刘春暖[1],刘 祥[2],王新明[1],芦 旭[1],苏晓林[3],张 亮[1*]
1. 山东农业大学信息科学与工程学院,山东 泰安 271018
2. 山东省农业农村厅,山东 济南 250013
3. 山东省泰安第六中学, 山东 泰安 271000

摘 要: 牛的行为能够直观地反映其生理和健康状态,在复杂养殖环境中准确识别牛体行为对实现智能养殖至关重要。现有牛体行为识别方法受噪声敏感、牛体运动特性影响,导致在实际养殖环境中牛体行为识别的准确性不足。因此,本文提出一种适用于复杂养殖环境的基于骨架特征的牛体行为识别方法,该方法通过提升行为特征对复杂背景噪声及遮挡的抗干扰能力,提高其在实际养殖环境下的牛体行为识别效果。首先,本文基于HRNet对低分辨率特征利用不足,设计了多尺度特征融合(Multi-scale Feature Fusion,MSFF)模块,增强了HRNet的多尺度表征能力,提升了复杂背景下牛体骨架特征提取的准确性;其次,考虑到牛体行为较为迟缓的特点,构建三维热图体(3D heatmap volume)作为输入,减少了噪声干扰,并保留了全局动态特征;最后,针对牛体行为类间相似性,在PoseConv3D行为识别网络的残差层引入帧间运动激励(Inter-frame Motion Excitation,IME)模块,增强不同行为的判别特征,提升类间相似牛体行为的区分能力。实验结果表明,所设计方法在骨架特征提取任务和行为识别任务中分别取得92.3%(AP^{50})和93.9%(Top1_acc)的效果,相比原有方法分别提升了0.8%和4.7%,整体提升6.4%。本文所设计的方法能在遮挡等复杂养殖环境下更准确地提取牛体骨架特征,对行为的判别性特征提取能力更强,为智能养殖中的牛体行为识别提供了有效的解决方案。

关键词: 骨架特征; HRNet; 行为识别; PoseConv3D; 牛体
中图法分类号: S126 **文献标识码:** A **文章编号:** 1000-2324(2026)04-0750-13

Skeleton-based Cattle Behavior Recognition in Complex Farming Environments

LIU Chun-nuan[1], LIU Xiang[2], WANG Xin-ming[1], LU Xu[1], SU Xiao-lin[3], ZHANG Liang[1*]
1. *College of Information Science and Engineering/Shandong Agricultural University, Tai'an 271018, China*
2. *Shandong Provincial Department of Agriculture and Rural Affairs, Jinan 100083, China*
3. *Shandong Province Tai'an Sixth Middle School, Tai'an 271000, China*

Abstract: The behavior of cattle can directly reflect their physiological and health status,and accurately identifying cattle behavior in complex farming environments is crucial for achieving smart livestock farming. Existing cattle behavior recognition methods are susceptible to noise and affected by cattle motion characteristics, resulting in insufficient recognition accuracy in practical farming environments. To address these issues, this paper proposes a skeleton-based cattle behavior recognition approach suitable for complex farming environments, which enhances the robustness of behavioral features against complex background noise and occlusion, thereby improving recognition performance under practical farming conditions. First, to address the insufficient utilization of low-resolution features in HRNet, this paper designs a Multi-scale Feature Fusion (MSFF) module, which enhances the multi-scale representation capability of HRNet and improves the accuracy of cattle skeleton feature extraction under complex backgrounds. Second, considering the relatively slow movement characteristics of cattle, a three-dimensional heatmap volume (3D heatmap volume) is constructed as input to reduce noise interference while preserving global dynamic features. Finally, aiming at the inter-class similarity of cattle behaviors, an Inter-frame Motion Excitation (IME) module is introduced into the residual layers of the PoseConv3D behavior recognition network, enhancing the discriminative features of different behaviors and improving the distinction capability of inter-class similar cattle behaviors. The experimental results show that the designed method achieves a

收稿日期: 2025-10-09 **修回日期:** 2026-06-20
基金项目: 国家自然科学基金项目(62202281)
第1作者简介: 刘春暖(1999-),女,硕士研究生,研究方向:人工智能方法。E-mail:lchunnuan@163.com
***通讯作者:** Author for correspondence. E-mail:zliang@sdau.edu.cn

performance of 92.3% (AP^{50}) and 93.9% (Top1_acc) in skeleton feature extraction and behavior recognition tasks, respectively, representing improvements of 0.8% and 4.7% compared to the original method, with an overall enhancement of 6.4%. The method proposed in this paper can more accurately extract skeleton features of cattle in complex farming environments such as occlusion, and exhibits stronger discriminative capability for behavior feature extraction, providing an effective solution for cattle behavior recognition in intelligent livestock farming.

Keywords: Skeleton feature; HRNet; action recognition; PoseConv3D; cattle

随着人们生活质量的显著提高，对牛肉与牛奶制品等优质食品的需求也越来越大。为满足这一需求，自2014年中央实施基础母牛扩群增量项目以来，我国肉牛和奶牛存栏数量逐年增加。数据显示，2018年我国牛总存栏达到732.5万头，位居世界第三[1]；2020年进口牛总数量超过30万头，交易金额超过6万亿美元[2]。而随着牛养殖规模的持续扩大，对牛群健康与养殖效率的精细化管理需求也愈发迫切。牛的行为能够直观地反映其生理和健康状态[3,4]，因此，准确快速地识别牛体行为不仅是实现健康状态智能感知的关键，也对疾病预防、优化养殖管理及提高牛场生产效益具有重要意义[5-7]。

目前，牛体行为识别的研究主要分为两大类：基于接触的方法和非接触的方法。基于接触的行为识别方法使用穿戴式传感器采集牛体行为数据并进行分析[8,9]。然而，这类穿戴式设备不仅易损坏且成本较高，还可能引起牛只应激反应，影响牛只福利[10,11]。随着视频采集技术和计算机视觉技术的发展，研究人员开始利用图像和视频数据进行非接触的牛体行为识别[12-14]。与接触的识别方法相比，非接触的识别方法具有成本效益优势，并能避免引发应激反应。

现有研究已探索多种基于图像和视频数据的非接触式牛体行为识别方法，如基于RGB[15,16]的识别方法、基于光流[17]的识别方法和基于骨 骼[18]的识别方法。骨骼作为紧凑的动作表示形式，近年来受到了越来越多的关注。与基于RGB和光流的动作识别方法相比，骨架只捕捉动作核心信息，使模型专注于提取行为本身的特 征，不受背景变化、灯光变化等环境干扰。并 且骨架特征更容易与RGB和光流等模态数据进行集成。图卷积网络（Graph Convolutional Network，GCN）将人体骨架序列建模为时空图，被广泛采用[19-21]。ST-GCN通过提取关键点在空间和时间上的特征实现动作识别[22]。Zhao等使用时空图卷积网络来识别动物行为，构建了多物种骨架数据集（Animal-Skeleton）[23]。BlockGCN通过图距离描述物理拓扑的骨骼连接编码方法识别并恢复了先进GCN中被忽视的骨架拓扑[24]。此外，Li等提出结合RGB、光流和骨架特征进行跛行识别，准确率达97.20%[25]。

然而，现有的牛体骨架提取方法多在相对简单的受控环境下进行，在实际牛场中，复杂的养殖环境会对行为识别造成显著干扰。首先，不同栏舍区域的背景纹理、光照条件和拍摄角度差异较大，使得关键点定位的稳定性大幅下降。其次，牛群通常处于高密度饲养状态，牛只之间频繁发生遮挡，尤其是在进食、转身或在狭窄通道中行走时，部分关键点极易被遮挡。再次，牛体毛色差异显著，加上光照反射或阴影变化等因素，骨架提取过程进一步受到干扰。上述因素使得在真实场景下获得稳定、可靠的牛体骨架特征变得十分困难，从而对后续的行为识别提出了更高要求[26,27]。现有方法中，基于骨架特征的牛体行为识别的方法较少，对牛场养殖环境的干扰因素及牛体运动的特性考虑不足。此外，不同牛体行为的特征差异较小，现有方法未能充分利用具有较强判别能力的特征来提升识别精度[28]。

为更有效地提取骨架数据的行为特征并提升识别精度，本文提出了一种基于骨架特征的牛体动作识别方法。与现有方法相比，该方法能够更精准地捕捉牛体运动特征，并提高行为识别的准确性。该方法首先提取牛体骨架信息，并将骨架数据输入3D卷积网络进行行为识别，最终输出行为类别。该方法在实际复杂养殖环境中展现出较强的准确性，能够有效应对多种干扰因素，显著提升识别性能。本文的贡献如下：

（1）针对实际养殖环境复杂，图像分辨率差异大的问题，本文设计了多尺度特征融合（Multi-scale Feature Fusion，MSFF）模块，提升网络对多尺度特征的利用能力，从而提取更加精准的牛体骨架特征。

（2）针对提取的牛体骨架数据噪声较大的

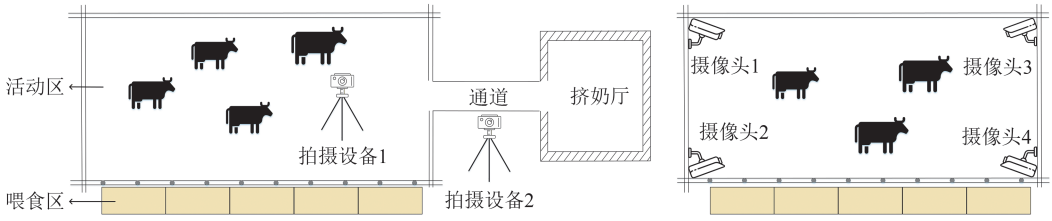
问题，本文提出基于时序均匀抽样方法的三维热图体表示，降低骨架关键点噪声影响，使行为识别模型专注于动作本身特征。

（3）针对不同牛体类间行为特征差异较小的问题，在 PoseConv3D 网络的残差层中融入帧间运动激励（Inter-frame Motion Excitation Module，IME）模块，强化帧间行为差异特征，使模型聚焦于类间行为的判别性特征，提高行为识别准确率。

1 实验数据

原始视频数据集采集自济南市某牛场和泰安市某牛场，数据采集时间为 2023 年 11 月到 2024 年 2 月。为体现数据采集在真实养殖环境下的完备性，本文将数据划分为近景数据和远景数据两类。近景数据包括在牛棚内采集的站立、回头和低头行为视频，以及在通道中采集的站立与行走行为视频，如图 1（a）所示。远景数据则来自牛场监控摄像机的常规拍摄，如图 1（b）所示。近景视频的帧率为 30 帧每秒（Frames Per Second，FPS），分辨率为 1 920×1 080 像素；远景视频的帧率为 30 FPS，分辨率为 1 280×960 像素。

除了视角距离的区分，为进一步提升数据集的多样性和全面性，本文还综合考虑了拍摄牛体时的不同视角（如正面、侧面、背面）、多种光照条件（如强光、暗光、自然光）、不同毛色花纹（如黑白相间、棕色、黑棕相间）、复杂环境因素（如栅栏遮挡、背景杂乱、动物拥挤）以及不同的栏舍结构（采食区、休息区、挤奶通道）。图 2 展示了牛体图像数据集在这些维度下的样例。



(a)近景拍摄场景
(a) Close-shot shooting scene

(b)远景拍摄场景
(b) Long-shot shooting scene

图 1 拍摄场景
Fig. 1 Shooting scene



(a)拍摄角度
(a) Shooting angle

(b)光照条件
(b) Illumination conditions

(c)牛体毛色
(c) Coat color of cattle

(d)环境因素
(d) Environmental factors

(e)栏舍结构
(e) Barn structure

图 2 牛场养殖环境下的牛体图像数据集
Fig. 2 Dataset for cattle images in cattle farming environment

本研究针对站立、行走、回头、低头四种行为开展相关研究，以下为四种行为的具体介绍：

（1）站立：站立行为也是牛的一种休息形式，并且有诸多动作是在站立状态下完成的，如进食、饮水等。

（2）行走：行走是牛常见的行为，该行为能

够直接反映牛的运动量，体现牛的健康状况。并且，在发情当天牛的运动量显著增加，生病牛的活动量会相应地下降。

（3）回头：当患疥螨和湿疹时，牛会频繁出现回头搔痒行为，牛被毛会成片地脱落，导致体重和膘情明显下降。回头行为对于判定疥螨和湿疹等疾病具有重要参考意义。

（4）低头：正常牛会习惯性地前往集中饲料处进食，当牛患异食癖时会舔食和啃咬无营养价值、不应采食的异物，不及时干预会导致新陈代谢障碍，影响牛只福利，在牛场环境下频繁出现低头行为是患异食癖的一种重要表现。

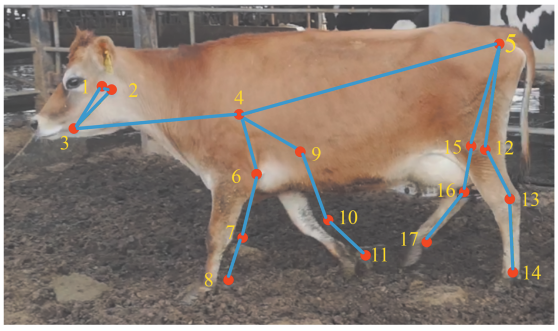
四种行为的具体定义如表 1 所示。其中，回头行为的整个过程定义为头部开始转动为起始，头部转动到露出另一侧的眼睛为止。此外，在站立行为状态下，牛只头部会有一定程度的晃动，与回头和低头行为较为相似。

表 1 牛体行为定义

Table 1 Definitions of the four cattle behaviors

行为类别 Behavior categories	行为定义 Behavior definitions
站立 Standing	牛体不发生明显位移，四肢可能有抬起、放下的动作，头部可能会有轻微晃动。
行走 Walking	在视频中持续保持行走状态。四肢规律地交替前进，并且头颈部规律的起伏。
回头 Turn head	牛体侧拍状态下，满足站立行为的条件，且头部转回贴近身体，露出另一侧的眼睛即判定为回头。
低头 Lower head	头部贴近地面，四肢保持站立，牛体可能会位移，四肢也可能会移动。

参考最大的动物姿态数据集 APT-36K[29]，以牛体侧视视角设计骨架模型，如图 3 所示。牛体侧视视图更能表征牛体的骨架细节信息。本文中使用关键点和连接线模拟骨架模型中牛体的关节和肢体，进而模拟牛体实际骨架结构。



1-右眼部
1-Right eye area
2-左眼部
2-Left eye area
3-鼻部
3-Nose area
4-颈肩部
4-Neck and shoulder area
5-尾根部
5-Root of tail
6-左前大腿部
6-Left fore thigh
7-左前膝部
7-Left fore knee
8-左前蹄部
8-Left fore hoof
9-右前大腿部
9-Right fore thigh
10-右前膝部
10-Right fore knee
11-右前蹄部
11-Right fore hoof
12-左后大腿部
12-Left hind thigh
13-左后膝部
13-Left hind knee
14-左后蹄部
14-Left hind hoof
15-右后大腿部
15-Right hind thigh
16-右后膝部
16-Right hind knee
17-右后蹄部
17-Right hind hoof

图 3 牛体骨架结构

Fig. 3 Cattle skeleton structure

本文以视频为划分单元，将所有视频样本按 7∶2∶1 的比例划分为训练集、验证集和测试集。随后通过视频分帧截图、数据预处理和数据标注处理流程，共计得到 8 650 张牛体图像，划分过程中确保同一视频不会同时出现在不同数据集中，数据集的样本分布及平衡性统计如表 2 所示。

表 2 数据集统计分析表

Table 2 Statistical analysis table of dataset

行为类别 Behavior category	视频数量 Video count	视频时长/s Duration	视频帧数量 Frame count	近景与远景比例 Shot ratio
站立 Standing	180	753	2 319	3∶4
行走 Walking	180	767	2 331	5∶4
回头 Turn head	160	698	2 096	1∶1
低头 Lower head	140	644	1 904	3∶4

2 研究方法

所提出的方法包括牛体骨架提取和行为识别两个主要步骤，如图 4 所示。首先使用 MSFF-HRNet 骨架提取方法获取牛体骨架数据，然后将骨架数据转化为三维热图体并进行预处理，有效保留关键运动特征，最后 PoseC3D-IME 行为特征提取网络实现对牛体行为的识别。

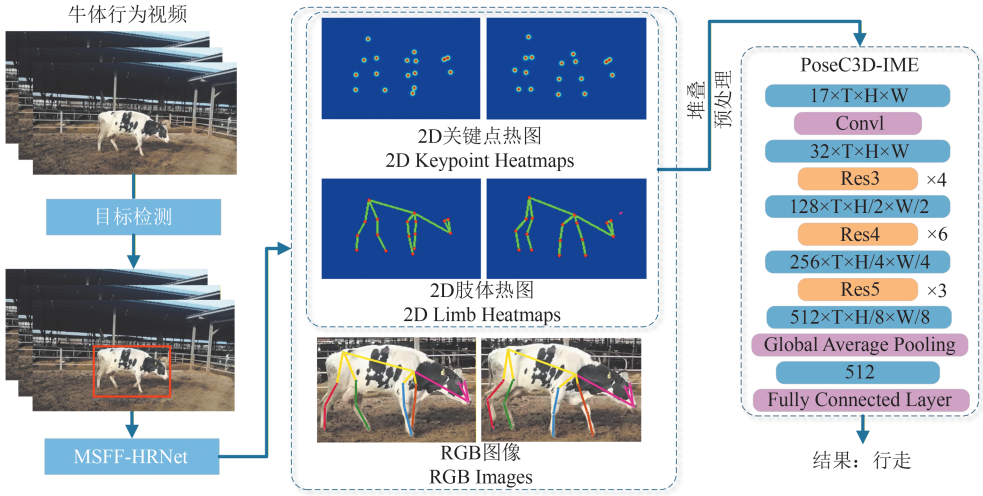


图 4 基于骨架特征的牛体行为识别方法整体流程图

Fig. 4 Overall flowchart of skeleton-based cattle behavior recognition method

2.1 骨架提取方法

本文使用高分辨率网络(High-Resolution Network，HRNet)实现了牛体骨架提取[30]，该网络的核心思路是在整个网络全程保持高分辨率的分支，并通过并行方式持续融合不同尺度的特征。与其他采用串行方式进行特征融合的网络结构相比，HRNet 基于并行方式实现低分辨和高分辨的特征融合，而不是试图从低分辨恢复高分辨，因此预测的热图更精确。

HRNet 在保持高分辨率特征的同时，对低分辨率特征的利用可能不足。在实际养殖场景中，近景与远景的牛体数据存在显著的尺度差异，HRNet 的局限性会进一步加剧。通过优化多尺度特征的融合机制，可以显著增强网络对跨尺度目标的表征能力。Dai 等使用一种多尺度通道注意力模块优化特征融合方式，融合语义和尺度不一致的特征，并通过实验验证了模型性能的提升[31]。

为缓解尺度变化引起的问题，本文针对相邻尺度分辨率的特征融合引入了 MSFF 模块。与传统的求和或拼接式融合方式不同，MSFF 的核心思想是在保证高分辨率特征细节的同时，自适应地整合低分辨率全局语义。将改进后的 HRNet 命名为 MSFF-HRNet，其结构如图 5 所示，该网络保留了原始 HRNet 中 Layer1 和 Basic Block 的基础结构。

与改进前的特征融合方式相比，改进后的结构进行了两次高分辨率特征与低分辨率特征的融合，第一次融合结果作为一个统一的特征表示，作为多尺度通道注意力模块(Multi-scale Channel Attention Module，MS-CAM)的输入，用于权重计算，第二次融合则在 MS-CAM 完成权重分配后，结合全局语义与局部细节，实现对高、低分辨率特征的权重整合。

与传统多尺度特征融合的方法相比，MSFF 模块能够自适应地融合高低分辨率特征，通过注意力机制动态调整各尺度特征的贡献比例，从而提升跨尺度特征的表征能力。同时，该模块兼顾局部细节与全局语义，并通过加权融合实现两者的平衡。

假设 M 为 MS-CAM 模块，MSFF 特征融合结构对输入特征的处理过程可以使用式(1)表示：

$$z = M(x \uplus y) \otimes x + (1 - M(x \uplus y)) \otimes y \quad (1)$$

式(1)中 x 代表低分辨率的特征，y 代表高分辨率的特征，z 代表融合后的特征，$\uplus$ 代表初始化特征融合，本文中采用逐元素求和。

MS-CAM 为双分支结构的模块，双分支用于分别指导两种分辨率进行特征融合。上侧分

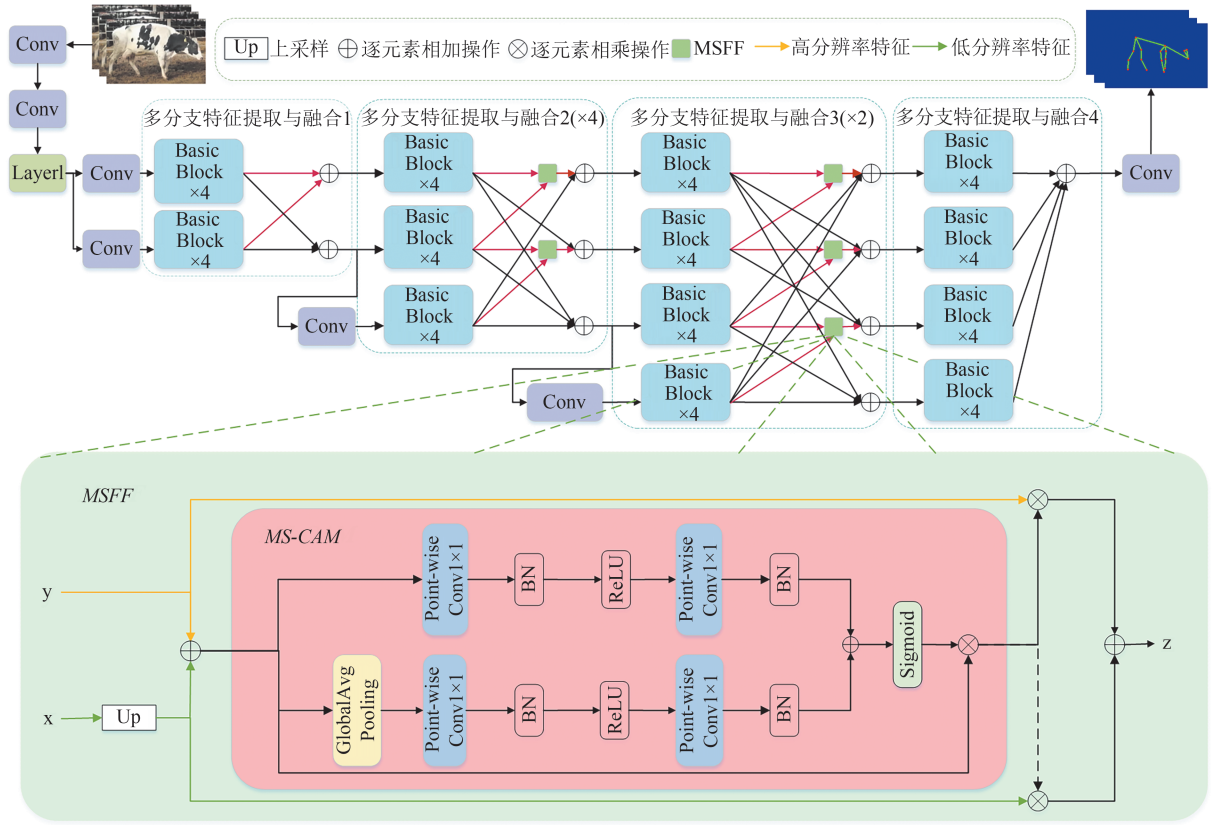


图 5 MSFF-HRNet网络结构

Fig. 5 MSFF-HRNet network structure

支使用点卷积，实现每个空间位置上通道间点对点的交互，提取局部上下文信息特征；下侧分支首先通过池化操作进行特征降维，再利用点卷积得到全局上下文特征，最后将两者相加经过激活函数得到最终的通道特征权重。

2.2　行为识别方法

PoseConv3D 是一种基于 3D 卷积网络构建的基于骨架特征的动作识别方法[32]，在通过不同方法获得的输入骨架上具有良好的通用性。当前主流方法的输入数据格式一般为骨架关键点的坐标，这导致模型对骨架关键点的噪声比较敏感，鲁棒性较差。

本文使用的 PoseConv3D 方法与现有基于骨架特征的动作识别方法的不同之处在于，该方法使用 3D 热图体作为数据输入，降低骨架关键点噪声对行为识别的影响，并针对牛体行为较为迟缓的特点，采用均匀采样，保留全局动态特征，减少信息冗余。三维热图体的生成与处理过程如图 6 所示，主要包括生成热图、主题中心裁剪和均匀采样。

首先从不同的动作视频中提取牛体的二维骨架连续序列，将其生成二维牛体骨架关键点（或肢体）热图，然后将所有的二维热图沿时间维度堆叠形成三维热图体，其数据格式为 $K \times T \times H \times W$，其中 K 表示骨架关键点个数，$H$ 和 W 分别表示帧高度和宽度，T 表示帧数。

为了减少三维热图体的信息冗余，在输入特征提取网络之前还对三维热图体进行以下处理：

（1）主体中心裁剪（Subjects Centered Cropping），当牛体区域只在一个小区域时，将整个热图送入模型会包含较多的冗余信息。为了处理这种情况，本文利用以主题为中心裁剪的方法，首先找到所有骨架关键点的最小边界框。然后，根据最小边界框裁剪所有的帧。

（2）均匀抽样（Uniform Sampling），比如对于总计包含 96 帧的三维热图体，使用均匀抽样的方法，先将三维热图体均匀分成 32 份，每份包含 2 帧。然后，在每份的 2 帧中随机抽取 1 帧。最后将抽取的 32 帧重新组合成三维热图体。

为了更好地捕捉牛体类间行为的具有判别

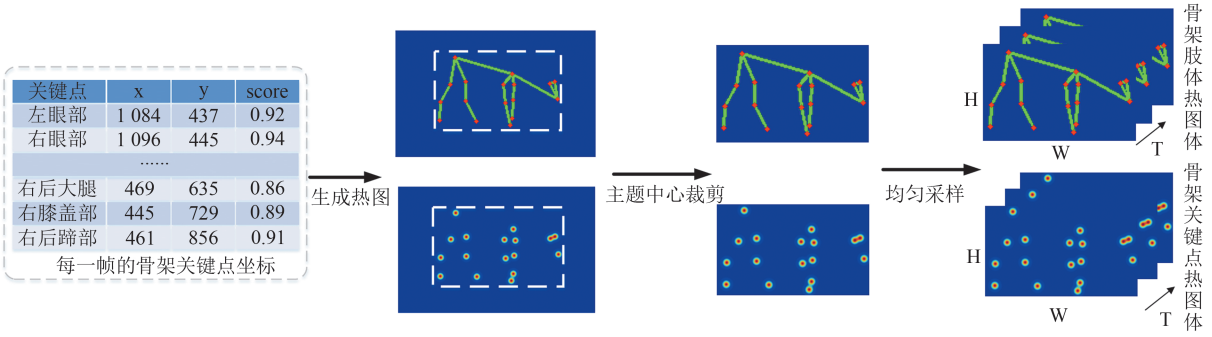


图 6 三维热图体的生成与处理

Fig. 6 Generation and processing of 3D heatmap volume

性的运动特征，本文在 PoseConv3D 主干网络的 Res4 和 Res5 两个残差层中添加 IME 模块，其结构如图 7 所示。利用该模块生成运动特征掩码，然后将掩码与输入特征融合后输出，增强 PoseConv3D 网络对判别性特征的提取能力，使其能更好地识别牛体行为中细微动作。将使用帧间运动激励模块重新设计残差层的 PoseConv 3D 网络命名为 PoseC3D-IME 网络。

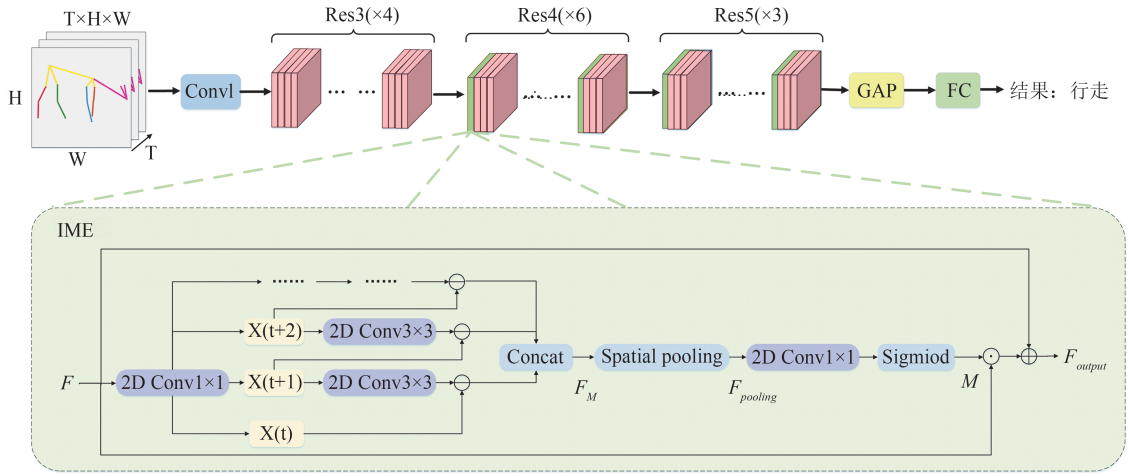


图 7 PoseC3D-IME 的网络结构

Fig. 7 PoseC3D-IME network structure

对于输入特征 F，首先使用 2D 卷积进行通道数的压缩，然后沿着时间维度分割，得到代表每一帧的运动特征 F_r，根据式(2)获取帧间运动特征 F_m。

$$F_m = K*F_r[:,t+1,:,:,:] - F_r[:,t,:,:,:] \quad (2)$$

式(2)中 K 代表 3×3 的 2D 卷积，* 代表卷积操作，3×3 卷积提供了适度的感受野，既能捕捉小范围的空间变化，又不会引入过大的计算量。

将 F_m 沿时间维度将运动特征连接，并将 0 填充到最后一个元素，得到帧间运动特征集合，其计算公式如式(3)所示。

$$F_M = [F_m(1),F_m(2),...,F_m(t-1),0] \quad (3)$$

式(3)中 F_M 为帧间运动特征集合，$F_m(1)$、$F_m(2)\cdots F_m(t-1)$ 为各时刻运动特征。

接着对 F_M 进行平均池化操作，计算过程如式(4)所示。

$$F_{pooling} = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_M[:,:,:,i,j] \quad (4)$$

式(4)中 H 和 W 分别表示 F_M 的高和宽。

将 F_{output} 送入 1×1 大小的 2D 卷积层，再使用 Sigmoid 激活函数得到通道特征掩模 M，模块的最后输出结果利用式(5)计算得到。

$$F_{output} = F + F*M \quad (5)$$

3 实验与结果分析

3.1 实验设置

实验所用的服务器设备配置为 Intel(R) Xeon(R) GOLD 5318Y 处理器，基本频率为

2.1 GHz。GPU 为 NVIDIA GeForce RTX 3090，显存 24 GB。深度学习框架为 PyTorch1.8，编程语言为 Python3.8。所有的实验都是在同一环境下进行的。

MSFF-HRNet 的总 epoch 设为 210。在输入网络之前，图像尺寸均重塑为 256×256 像素。为了在训练过程中动态调整学习率以增强模型的收敛性和最终性能，实验过程采用自适应矩估计（Adam）优化器，使用线性 warm up 热身训练策略调整学习率进行训练，前 500 个 iters 进行热身训练，初始学习率为 0.001，从 170 到 210 epoch 衰减学习率，初始学习率设置为 0.000 5。

PoseC3D-IME 的训练 epoch 数为 240，采用随机梯度下降（SGD）作为优化器，初始学习率为 0.2，动量（Momentum）设置为 0.9，使用带动量的 SGD 优化器加快模型的收敛速度。权重衰减（Weight_decay）设置为 0.000 3，权重衰减能够调节模型复杂度对损失函数的影响，防止模型在训练过程中产生过拟合。使用余弦退火算法（Cosine Annealing）调节学习率。输入尺寸为 32×17×56×56，记为 $T\times K\times H\times W$，其中 K 为骨架关键点个数，H 和 W 分别为热图的高度和宽度，T 表示帧数。批处理大小为 16，即每次训练迭代处理 16 个热图序列。

3.2　评价指标

3.2.1 骨架提取模型性能评估　物体关键点相似度（Object key point similarity，OKS）是常用的骨骼关键点检测算法的评估指标。其计算公式如式（6）所示：

$$OKS_p=\frac{\sum_i \exp\left\{-\frac{d_{p^i}^2}{2S_p^2\sigma_i^2}\right\}\delta(v_{p^i}>1)}{\sum_i \delta(v_{p^i}>1)} \tag{6}$$

式（6）中 p 代表真值中的某个牛的全值，p^i 代表某个牛的第 i 个关键点，$d_{p^i}^2$ 代表关键点，p^i 代表真实坐标与预测坐标的欧氏距离，v_{p^i} 代表关键点的可见性，$v_{p^i}=0$ 表示未标记，$v_{p^i}=1$ 表示标记但不可见，$v_{p^i}=2$ 表示标记且可见，S_p^2 代表关键点所属目标的尺度因子，该值为标注时目标所占的像素面积，σ_i^2 代表骨架关键点的归一化因子，反映当前关键点的标注难度，其值越大表示关键点越难标注，δ 代表筛选出可见骨架关键点的函数。

为了判定 OKS 值为多少时预测结果有效，使用平均精度（Average Precision，AP）来计算，AP 值表示正确预测的关键点个数占总预测个数的百分比。AP 值的计算公式如式（7）所示，设定阈值 t，仅当预测关键点的 OKS 值超过 t 时判定该骨架关键点检测成功。通过计算 OKS 分别为 0.5、0.55、0.6、0.65、0.7、0.75、0.8、0.85、0.9、0.95 的 AP 值再求平均值，即为平均精度均值（Mean Average Precision，mAP），AP^m 用来评估中等尺寸的牛体关键点的检测性能。

$$AP=\frac{T}{L} \tag{7}$$

式（7）中 T 代表所有满足阈值 t 条件的 OKS 值个数，S 代表预测的所有 OKS 值个数。

此外，本文还使用了平均召回率（Average Recall，AR）的概念，AR 值表示正确检测的样本个数占总标注真值个数的百分比。AR 的计算公式如式（8）所示。设定阈值 t，当 OKS 值大于 t 时判定为预测成功。参考平均准确率的计算方式，平均召回率均值（Mean Average Recall，mAR）为 OKS 值为 0.5、0.55、0.6、0.65、0.7、0.75、0.8、0.85、0.9、0.95 时得到的所有 AR 的平均值。

$$AR=\frac{T}{L} \tag{8}$$

式（8）中 T 代表所有满足阈值 t 条件的 OKS 值个数，L 代表所有标注 OKS 值的个数。

3.2.2 行为识别模型性能评估　Top-1 准确率（Top1_accuracy，Top1_acc）是一种用于量化动作识别算法性能的评估指标，该评价指标计算模型给出的所有预测结果中与给定视频的真实标签一致的样本占总样本数的百分比。该评价指标的计算过程如式（9）所示。

$$Top1_acc=\frac{\sum_i^N \delta(class_i^{ture}=rank_1(class_i^{pred}))}{N} \tag{9}$$

式（9）中 δ 代表判断识别结果的条件，条件成立取 1，否则取 0，$class_i^{true}$ 代表第 i 类的真实标注类别，$class_i^{pred}$ 代表第 i 个行为的预测行为类别，N 代表样本总量。

3.3　结果与分析

3.3.1 骨架提取结果　使用本文所提出的骨架提取模型进行测试，最终 AP^{50} 为 92.3%。训练后的模型输出图像中牛体的骨架关键点及其关节连接线。提取的牛体骨架关键点和关节连线效果如图 8 所示。



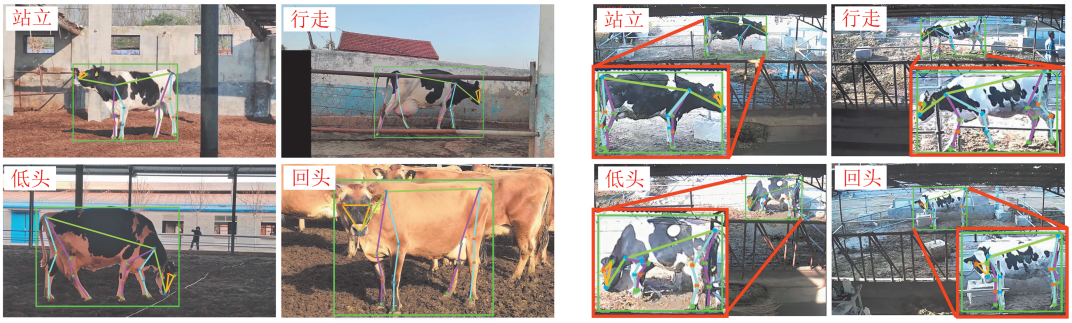
图 8 不同条件下骨架关键点的检测效果

Fig. 8 Skeletal keypoint detection effects in different conditions

但是，由于牛场实际环境比较复杂，并且牛体关节存在被遮挡的情况，出现了一些偏差和错误。通过标记更多场景下的样本进行训练，可以改善检测效果。此外，在连续的动作视频序列中，得到的牛体骨架质量在可接受范围内，对动作识别的影响较小。

3.3.2 *行为识别结果*　PoseC3D-IME 的 Top1_acc 为 93.9%。为了更直观地展示本文行为识别的效果，绘制了牛体站立、行走、低头和回头四种典型动作的可视化结果如图 9 所示。绿色矩形为牛只边界框。这些点和连接线分别代表牛体骨架的关键点和肢体，视频画面左上角的红色文本表示动作识别结果。为了更直观地展示远景下的行为识别可视化结果，图中将检测到的牛只边界框和牛体骨架进行了放大展示。

3.3.3 *不同骨架关键点提取算法的比较*　使用自建牛体骨架提取数据集在堆叠沙漏网络、残差网络、轻量级网络和高分辨率网络模型上进行训练并测试了不同模型的性能表现，具体评估指标结果如表 3 所示。



(a)近景下的行为识别可视化结果
(a) Visualization results of behavior recognition under close shot

(b)远景下的行为识别可视化结果
(b) Visualization results of behavior recognition under long shots

图 9 行为识别可视化结果

Fig. 9 Visualization results of the behavior recognition

由实验结果可知，MSFF-HRNet 骨架提取网络在评估指标上领先于其他方法。基于堆叠沙漏网络 Hourglass[33]的方法由于其特征融合方式在进行上采样和下采样操作时只能利用当前阶段上采样（或下采样）的信息，虽然指标相差不大，但是指标与 HRNet 相比低了接近 5%，说明在网络结构中始终保持高分辨率分支对牛体骨架提取任务有较大的意义。残差网络 ResNet[34]在牛体骨架提取

表 3 不同骨架提取方法的对比
Table 3 Comparison of different skeleton extraction methods

模型 Model	mAP	AP^{50}	AP^{m}	mAR
SHN[33]	70.3	90.5	53.7	71.6
ResNet[34]	71.4	90.7	56.5	73.1
MobileNet[35]	66.7	88.6	50.3	68.7
YOLO-Pose[36]	70.1	88.7	55.0	74.5
HRNet[30]	73.5	91.5	58.6	76.2
MSFF-HRNet	**75.6**	**92.3**	**68.5**	**79.7**

任务中有较好的性能表现。轻量级网络MobileNet[35]由于采用了较多的下采样操作，参数量相较于其他网络结构明显偏低。YOLO-Pose[36]更注重实时性，牺牲了一些精度，在实际的复杂养殖环境中会做出一些错误的判断。总体来说，MSFF-HRNet在保留高分辨率特征的同时，充分融合来自低分辨率分支的全局语义信息。通过跨分辨率特征的双阶段融合机制，模型在特征层实现了空间细节与语义上下文的互补，从而显著增强了对不同尺度牛体关键点的检测能力，将APm从58.6%（HRNet）提升至68.5%，实现了更高质量的骨架关键点提取。

3.3.4 不同行为识别算法的比较　由于GCN的方法与骨架数据具有较高的适配性，因此被广泛应用于基于骨架数据的动作识别任务中。为了验证以三维热体图作为行为识别模型输入的有效性和PoseC3D-IME的性能，本文与PoseC3D方法和主流的GCN方法（ST-GCN，BlockGCN）进行对比实验，结果如表4所示。PoseC3D-IME以93.9%的高准确率优于ST-GCN、BlockGCN和PoseC3D。这说明基于三维热图体方法对骨架关键点的噪声有更好的包容性，能更好地关注到行为本身特征，而PoseC3D-IME在仅较PoseC3D略微增加参数量和计算量（FLOPs）的情况下，仍显著提升了识别性能，特别是对回头行为的识别，证明了帧间运动激励模块能够增强识别牛体行为中细微动作的能力。

表 4 不同动作识别方法的对比
Table 4 Comparison of different action recognition methods

主干网络 Backbone network	输入格式 Input hormat	Top1_acc /%					参数量/M Parameter/M	FLOPs/G
		站立 Standing	行走 Walking	低头 Lower head	回头 Turn head	所有行为 All behaviors		
ST-GCN[22]	关键点坐标	88.8	86.1	87.5	82.1	86.4	3.1M	1.03G
BlockGCN[24]	关键点坐标	92.0	90.1	88.5	85.4	90.2	1.3M	0.44G
PoseConv3D[32]	三维热图体	94.4	91.7	90.6	85.7	90.9	0.52M	0.31G
PoseC3D-IME	三维热图体	**94.4**	**94.4**	**93.8**	**92.9**	**93.9**	0.54M	0.39G

3.3.5 消融实验　本小节针对各组件的有效性进行了实验验证，分别评估了MSFF模块和IME模块对牛体行为识别性能的影响。具体评估指标如表5所示，实验结果表明，这两个模块均对模型性能有显著提升。每次改变一个组件，都使用相同的超参数和训练设置重新训练。

表 5 MSFF和IME的消融实验
Table 5 Ablation Studies on MSFF and IME

ID	MSFF	IME	Top1_acc /%				
			站立 Standing	行走 Walking	低头 Lower head	回头 Turn head	所有行为 All behaviors
1	×	×	90.0	88.5	85.6	84.5	87.5
2	√	×	94.4	91.7	90.6	85.7	90.9
3	×	√	92.3	89.6	87.8	85.2	89.2
4	√	√	**94.4**	**94.4**	**93.8**	**92.9**	**93.9**

为了验证MSFF模块的有效性，我们对比了使用MSFF模块和不使用MSFF模块的模型性能。实验结果显示，添加MSFF模块（ID 2）后，模型在四种行为（站立、行走、低头、回头）上的识别准确率均有显著提升。整体平均准确率从87.5%提升至90.9%。这验证了MS-CAM能够有效地

引导高低分辨率特征的融合，确保全局和局部上下文信息都能得到充分提取。而MSFF模块通过其两步特征融合过程，优化了多分辨率特征的整合，增强了模型提取精确骨架信息的能力。

为了验证IME模块的有效性，比较了完整版本（含IME模块）的方法与剔除 IME 模块的方法在性能上的差异。实验结果显示，IME 模块使模型能够更好地利用类间行为的判别性特征，尤其是在低头和回头等动作幅度较小、特征差异不明显的行为识别中，准确率分别提升了 3.2% 和 7.2%。这进一步证明了 IME 模块在增强时间维度特征表达能力方面的有效性，为解决牛体类间行为相似难以区分的问题提供了重要支持。

3.3.6 性能衰减规律 为验证模型在复杂场景下的鲁棒性，本文对不同遮挡比例和光照条件下的识别性能进行了评估。遮挡通过随机区域掩蔽实现，光照干扰通过降低视频帧亮度模拟。表 6 展示了不同干扰强度下的识别性能。

表 6 干扰强度-性能衰减结果

Table 6 Interference intensity-performance degradation results

干扰类型 Disturbance type	干扰强度 Disturbance intensity	Top1_acc /%	性能衰减/% Performance degradation
对照组	无遮挡+自然光	93.9%	0
遮挡	20%遮挡	92.5%	1.4%
	30%遮挡	91.7%	2.2%
	50%遮挡	89.4%	4.5%
光照强度	强光	92.7%	1.2%
	暗光	91%	2.9%

从结果可见，随着遮挡比例或光照减弱程度的增加，模型准确率呈平缓下降趋势。当遮挡比例达到 50% 时，准确率仍保持在 89.4%，仅下降约 4.5%。这表明所提出的模型在严重遮挡与低光照环境下仍能维持较高识别精度，验证了其在真实养殖场景中的鲁棒性。

4 结论

本文验证了所提出的牛体行为识别方法提升了在实际养殖环境中的准确性，主要结论如下：

（1）本文通过 MSFF-HRNet 骨架提取模型完成牛体骨架提取，结合 PoseC3D-IME 行为识别模型实现牛体行为识别，最终在所有行为类别上的 Top1_acc 达到 93.9%，相比于原始模型提高了 6.4%，取得了良好的识别效果。

（2）MSFF-HRNet 模型在牛体骨架提取任务中的 AP^{50} 相比于改进前提升了 0.8%，这一结果验证了所提出的特征融合方法能够有效提升牛体骨架提取模型的性能，在遮挡以及低分辨率场景中仍保持良好效果。

（3）实验证明，以三维热图体作为行为识别模型的数据输入，其效果优于以骨架关键点作为行为识别模型的数据输入的方法。同时，本文所提出的 IME 模块增强了牛体类间行为的判别能力，在低头和回头等动作幅度较小、特征差异不明显的行为识别中，Top1_acc 分别提升了 3.2% 和 7.2%，在所有行为类别中提高了 4.7%，验证了该方法在复杂养殖环境下的实用性与鲁棒性优势。

尽管本文提出的牛体行为识别方法在复杂养殖环境中表现出较高的准确性，但仍存在一定的局限性。模型的泛化能力受到训练数据多样性的制约，尽管数据采集涵盖了近景、远景、多视角、多光照等条件，但数据主要来源于特定牛场环境，不同养殖场在栏舍结构、管理方式、牛品种等差异下，仍可能对模型效果产生影响。未来工作中，可引入多视角视觉融合、时序上下文建模以及遮挡恢复机制，进一步提升模型在真实养殖环境中的鲁棒性和适用性，为实现智能化牛场养殖管理提供更可靠的技术支持。

参考文献

[1] 许尚忠. 中国牛产业发展概况及展望[J]. 畜牧产业, 2021(07):35-40.

[2] 汤文洁,苏惟真,姜 冰. 中国牛产业国际贸易回顾[J]. 乳品与人类,2023(03):3-10.

[3] Chen C, Zhu W, Norton T. Behaviour recognition of pigs and cattle: Journey from computer vision to

deep learning[J]. Computers and Electronics in Agriculture, 2021,187:106255.

[4] Zhang L, Guo W, Lv C, et al. Advancements in artificial intelligence technology for improving animal welfare: Current applications and research progress[J]. Animal Research and One Health, 2024,2(1):93-109.

[5] 张宏鸣,孙 扬,赵春平,等.反刍家畜典型行为监测与生理状况识别方法研究综述[J].农业机械学报,2023,54(03):1-21.

[6] Shu H, Bindelle J, Guo L, et al. Determining the onset of heat stress in a dairy herd based on automated behaviour recognition[J]. Biosystems Engineering, 2023,226:238-251.

[7] Zhang Y, Zhang Y, Gao M, et al. Multimodal behavior recognition for dairy cow digital twin construction under incomplete modalities: A modality mapping completion network approach[J]. Artificial Intelligence in Agriculture, 2025,15(3):459-469.

[8] 程国栋,吴建寨,邢丽玮,等.基于IMU的细粒度奶牛行为判别[J].中国农业大学学报,2022,27(04):179-186.

[9] Wang L, Arablouei R, Alvarenga F A P, et al. Classifying animal behavior from accelerometry data via recurrent neural networks[J]. Computers and Electronics in Agriculture,2023,206:107647.

[10] Camargo V A, Pajor E A, Bayat S, et al. Predicting main behaviors of beef bulls from accelerometer data: A machine learning framework[J]. Smart Agricultural Technology,2024,9:100683.

[11] Liang H T, Hsu S W, Hsu J T, et al. An IMU-based machine learning approach for daily bBehavior pattern recognition in dairy cows[J]. Smart Agricultural Technology,2024,9:100539.

[12] Nasir M F, Fuentes A, Han S, et al. Multi-camera fusion and bird-eye view location mapping for deep learning-based cattle behavior monitoring[J]. Artificial Intelligence in Agriculture,2025,15:724-743.

[13] Tassinari P, Bovo M, Benni S, et al. A computer vision approach based on deep learning for the detection of dairy cows in free stall barn[J]. Computers and Electronics in Agriculture,2021,182:106030.

[14] 李恩泽,王克俭,司永胜,等.基于改进ConvNeXt的奶牛行为识别方法[J].农业机械学报,2024,55(05):282-289+404.

[15] 曹天一,李宝山,王月明,等.基于起卧频率异常的牛产犊自动预警[J].农业工程学报,2022,38(S1):201-208.

[16] 付辰伏,任力生,王 芳.基于改进YOLO v8的牛只行为识别与跟踪方法[J].农业机械学报,2024,55(05):290-301.

[17] Jiang B, Yin X, Song H, et al. Single-stream long-term optical flow convolution nNetwork for action recognition of lameness dairy cow[J]. Computers and Electronics in Agriculture,2020,175:105536.

[18] Chi H G, Ha M H, Chi S, et al. InfoGCN: Representation learning for human skeleton-based action recognition[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE,2022:20154-20164.

[19] Cai J, Jiang N, Han X, et al. JOLO-GCN: Mining joint-centered light-weight information for skeleton-based action recognition[C]. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2021: 2734-2743.

[20] Chen Y, Zhang Z, Yuan C, et al. Channel-wise topology refinement graph convolution for skeleton-based action recognition[C].2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE,2021:13339-13348.

[21] Gupta P, Thatipelli A, Aggarwal A, et al. Quo vadis, skeleton action recognition?[J]. International Journal of Computer Vision,2021,129(7):2097-2112.

[22] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2018,32(1).

[23] Zhao Y, Feng L, Tang J, et al. Automatically recognizing four-legged animal behaviors to enhance welfare using spatial temporal graph convolutional networks[J]. Applied Animal Behaviour Science, 2022,249:105594.

[24] Zhou Y, Yan X, Cheng Z Q, et al. BlockGCN: Redefine topology awareness for skeleton-based action recognition[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE,2024:2049-2058.

[25] Li Y, Shu H, Bindelle J, et al. Classification and analysis of multiple cattle unitary behaviors and movements based on machine learning methods[J]. Animals,2022,12(9):1060.

[26] Liu H, Reibman A R, Boerman J P. Video analytic system for detecting cow structure[J]. Computers and Electronics in Agriculture,2020,178:105761.

[27] 张宏鸣,李永恒,周利香,等.基于改进YOLO v3的肉牛多目标骨架提取方法[J].农业机械学报,2022,53(03):285-293.

[28] Hua Z, Wang Z, Xu X, et al. An effective PoseC3D model for typical action recognition of dairy cows based on skeleton features[J]. Computers and Electronics in Agriculture,2023,212:108152.

[29] Yang Y, Yang J, Xu Y, et al. APT-36K: A large-scale benchmark for animal pose estimation and tracking [C]. Koyejo S, Mohamed S, Agarwal A, et al (Eds.). Advances in Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates, Inc., 2022, 35:17301-17313.

[30] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [C].2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE,2019:5686-5696.

[31] Dai Y, Gieseke F, Oehmcke S, et al. Attentional feature fusion[C]. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE,2021:3559-3568.

[32] Duan H, Zhao Y, Chen K, et al. Revisiting skeleton-based action recognition[C].2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE,2022:2959-2968.

[33] Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]. Leibe B, Matas J, Sebe N, Welling M (Eds.. Computer Vision-ECCV 2016. Lecture Notes in Computer Science, Vol. 9912. Cham: Springer International Publishing, 2016:483-499.

[34] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE,2016:770-778.

[35] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:4510-4520.

[36] Maji D, Nagori S, Mathew M, et al. YOLO-Pose: Enhancing YOLO for multi person pose estimation using object keypoint similarity loss[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, LA, USA: IEEE,2022:2636-2645.