



## 研究论文 ORIGINAL PAPERS

基于骨骼关键点的  
东北虎幼虎行为识别与跟踪刘淑聪<sup>1</sup>, 刘梦雨<sup>1</sup>, 刘丹<sup>2</sup>, 张馨洋<sup>3</sup>, 马光凯<sup>1,4\*</sup>, 姜广顺<sup>4\*</sup>

(1. 东北林业大学计算机与控制工程学院, 哈尔滨 150040;

2. 黑龙江东北虎林园, 哈尔滨 150028;

3. 东北林业大学机电工程学院, 哈尔滨 150040;

4. 国家林业和草原局猫科动物研究中心, 东北林业大学野生动物与自然保护地学院, 哈尔滨 150040)

## 稿件运行过程

收稿日期: 2025-03-17

修回日期: 2025-10-18



## 摘要

关键词: 东北虎幼虎;  
姿态估计;  
行为识别;  
骨骼关键点;  
多目标跟踪;  
深度学习

Keywords: Amur tiger (*Panthera tigris altaica*) cub;  
Pose estimation;  
Behavior recognition;  
Skeletal keypoints;  
Multi-object tracking;  
Deep learning

中图分类号: Q958.1; TP391.41

文献标志码: A

文章编号:

2310-1490(2025)-04-0723-11

DOI: 10.12375/ysdwx.202503015

东北虎 (*Panthera tigris altaica*) 幼虎的行为研究对其保护生物学与发育生态学至关重要。传统人工观察方法效率低下且易受主观因素影响, 亟需发展自动化、客观的技术手段。为此, 提出并验证了一种基于骨骼关键点的深度学习框架, 以实现东北虎幼虎行为的精准识别与跟踪。基于黑龙江东北虎林园和横道河子东北虎林园共计 15 只东北虎幼虎的监控视频, 构建了包含 16 个人工标注关键点的姿态估计数据集及 5 种常见行为的行为识别数据集。利用训练好的高分辨率网络 (HRNet) 生成初始姿态数据, 通过连续关键点序列关联行为标签与个体 ID, 构建了行为识别与跟踪数据集。在该数据集上, 采用多个行为识别网络进行对比实验, 并应用字节跟踪算法 (ByteTrack) 进行多目标个体跟踪。结果表明: 注意力增强自适应图卷积神经网络 (attention-enhanced adaptive graph convolutional neural network, AAGCN) 在幼虎行为识别任务中的准确率最优, 达 76.59%; ByteTrack 在个体跟踪中的多目标跟踪准确率 (multiple object tracking accuracy, MOTA) 高达 92.76%。本方法在东北虎幼虎行为识别与跟踪中表现优异, 为圈养大型猫科 (Felidae) 动物幼崽的行为量化与分析提供了可靠工具, 对野生动物保护和繁育管理具有直接应用价值。

基金项目: 国家重点研发计划课题项目 (2023YFF1305002); 中央高校基本科研业务费专项基金项目 (2572022DS04)

第一作者简介: 刘淑聪 (2000—), 男, 硕士研究生; 主要从事深度学习和人工智能算法研究。E-mail: shucongliu@163.com

\* 通信作者: 马光凯, E-mail: gkmahr@163.com; 姜广顺, E-mail: jgshun@126.com

# Behavior Recognition and Tracking of Amur Tiger Cubs Based on Skeletal Keypoints

LIU Shucong<sup>1</sup>, LIU Mengyu<sup>1</sup>, LIU Dan<sup>2</sup>,  
ZHANG Xinyang<sup>3</sup>, MA Guangkai<sup>1,4\*</sup>, JIANG Guangshun<sup>4\*</sup>

- (1. College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China;  
2. Heilongjiang Siberian Tiger Park, Harbin 150028, China;  
3. College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China;  
4. Feline Research Center of National Forestry and Grassland Administration, College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China)

**Abstract:** Behavioral research on Amur tiger (*Panthera tigris altaica*) cubs is critical to conservation biology and developmental ecology. Traditional manual observation is inefficient and susceptible to observer bias, underscoring the need for automated and objective methods. This study proposes and validates a deep learning framework based on skeletal keypoints for precise behavior recognition and tracking of Amur tiger cubs. Using surveillance videos from 15 cubs at the Heilongjiang Siberian Tiger Park and the Hengdaohezi Siberian Tiger Park, we constructed a dataset comprising a pose-estimation set with 16 manually annotated keypoints and a behavior-recognition set with five common behavior categories. A trained high-resolution network (HRNet) generated initial pose estimates; sequential keypoint series were then linked to behavior labels and individual IDs to construct a behavior recognition and tracking dataset. We benchmarked multiple behavior-recognition networks and applied ByteTrack for multi-object tracking. Results show that the attention-enhanced adaptive graph convolutional neural network (AAGCN) achieved the best behavior recognition accuracy at 76.59%, while ByteTrack reached a multiple object tracking accuracy (MOTA) of 92.76% for individual tracking. The proposed approach performs strongly for behavior recognition and tracking of captive large felid cubs, providing a reliable tool for quantitative behavioral analysis with direct applications to wildlife conservation and breeding management.

东北虎(*Panthera tigris altaica*)作为体型最大的猫科(Felidae)动物之一,是国家一级重点保护野生动物<sup>[1]</sup>,同时在IUCN濒危物种红色名录中被评估为濒危(EN)物种<sup>[2]</sup>。东北虎在维持东北亚森林生物系统的稳定性和完整性中发挥着关键作用<sup>[3]</sup>。研究东北虎幼虎的行为模式及其持续时间,不仅有助于揭示成年个体的行为发育轨迹,也为科学保护和人工繁育提供了重要依据。目前,动物行为数据获取仍主要依赖人工观察与标记,不仅耗时耗力,而且容易受观察者主观判断的影响<sup>[4]</sup>。此外,幼虎体型小、行为特征不明显,且常因聚集活动造成严重遮挡,进一步加大了其行为自动识别与个体跟踪的难度,制约了动物行为学研究的精确性与效率,也在一定程度上影响了东北虎保护工作的实施效果。

近年来,随着计算机视觉与人工智能技术的迅

速发展,深度学习在动物行为识别领域取得了巨大进展<sup>[5-7]</sup>。Deng *et al.*<sup>[8]</sup>利用循环神经网络和自注意力机制分支,对骨骼序列的时空信息进行提取,实现了对野生动物飞奔、坐、行走和站立4种行为的识别,准确率达96.8%。Feng *et al.*<sup>[9]</sup>则通过循环神经网络与轻量化卷积神经网络组成双分支结构,提取骨骼序列的时空信息,实现了对野生动物站立、漫步和奔跑3种行为的识别,准确率达95.00%。Lin *et al.*<sup>[10]</sup>利用卷积神经网络提取姿态特征,在自建鸟类行为数据集IMLab-P8-2021上实现了鸟类多种常见行为的识别,其关键点正确百分比(percentage of correct keypoints, PCK)和行为识别的整体准确率(overall accuracy, OA)最高达87.99%和87.81%。Li *et al.*<sup>[11]</sup>采用时空图卷积网络(ST-GCN)充分提取姿态序列中的时空信息,实现了对奶牛跛行的预测,准确

率达97.20%。这些成果表明,深度学习技术在动物行为识别中具有良好的识别性能,其准确性已超越传统学习算法和人工水平。

基于深度学习的方法目前是动物行为识别任务中的重要方法,其中基于骨骼关键点的行为识别主要是基于动物姿态序列进行行为识别和视频理解,与其他模式(如彩色图像/光流)相比,姿态数据紧凑且信息丰富<sup>[12]</sup>,含有多种互补特征信息<sup>[13]</sup>,由于姿态序列仅捕捉动作信息,对背景与光照变化不敏感<sup>[14]</sup>,且不受目标大小或形态变化的影响,能够有效对小物体或密集目标进行行为识别,具有更强的鲁棒性。同时,骨骼关键点的提取计算复杂度较低,适用于计算资源有限的环境,特别适合实时监控和大规模数据处理场景。此外,该方法以行为三要素(姿势-行为-环境)中的“姿势”为切入点,不仅能准确表征个体的运动模式,还可为行为分析提供辅助信息。尽管基于骨骼关键点的深度学习方法在动物行为识别领域展现出显著优势,但在东北虎幼虎群体行为识别的特定场景下,仍缺乏专用数据集和针对性的实验研究。虽然Deng *et al.*<sup>[8]</sup>与Feng *et al.*<sup>[9]</sup>的研究以姿态为基础,但其模型在时空特征融合方面仍较为基础,未能充分挖掘关节间的结构关系与复杂的时空关联,也未能系统融合如关节运动、骨骼长度变化等多维特征,因而在处理行为多变、遮挡严重的幼虎场景时存在局限。这些不足不仅限制了该方法在该场景下的直接应用,也为行为连续性建模及群体交互分析带来了挑战。

为弥补上述研究空白,本研究构建了首个面向东北虎幼虎的骨骼关键点行为识别和跟踪数据集。该数据集涵盖4个独立群体共15只幼虎,其中姿态数据集包含16个关键点,共标注3 092张图中的9 675只幼虎个体姿态;行为数据集包含5种常见行为,共计2 299组行为标签。在此基础上,采用多种

基于骨骼关键点的行为识别网络对幼虎的行为进行识别,并对比了不同多目标跟踪算法在幼虎群体中的个体跟踪效果。实验结果表明,注意力增强自适应图卷积神经网络(attention-enhanced adaptive graph convolutional neural network, AAGCN)和字节跟踪算法(ByteTrack)分别在行为识别和跟踪任务中表现出色,取得了较好的性能。本研究首次成功将基于骨骼关键点的深度学习方法应用于东北虎幼虎群体行为识别中,为其行为发生及持续时间的自动化分析提供了可靠的技术手段。

## 1 研究方法

### 1.1 数据采集

本研究在黑龙江省两个主要东北虎人工繁育基地——哈尔滨市松北新区黑龙江东北虎林园(45°49'00" N, 126°02'00" E)和牡丹江市横道河子东北虎林园(44°47'37" N, 129°06'39" E)开展。两园区饲养环境基本一致,均实行封闭式管理,符合动物福利标准,能有效避免外界干扰。

2019年3—11月,选取两园区内共4只哺乳期雌性东北虎所产的15只幼虎作为研究对象(个体信息详见表1),使用海康威视监控摄像头及硬盘录像机对其进行为期3个月的连续视频监控。虎舍视频样例帧如图1所示,虎舍内设有木制铺板,具有良好的保温、防潮和隔热功能,为幼虎提供了适宜的活动环境。从图1可见,幼虎行为模式多样,但由于笼舍结构限制及幼虎喜好聚集的习性,遮挡情况较为严重。此外,相较于成年虎,幼虎体型较小,行为特征不如成体明显,这给行为识别和统计带来了较大挑战。摄像头安装高度距笼舍地面约3 m,可在不影响幼虎正常活动的前提下,覆盖其绝大部分活动区域。视频以mp4格式录制,帧分辨率为2 560 × 1 440像素,帧率为25 fps,所有视频数据本地存储,总时长达86 h。

表1 研究涉及的东北虎雌虎及其幼虎信息

Table 1 Information of the female Amur tiger and their cubs included in the study

雌虎编号 Female tiger ID	生产日期 Parturition date	幼虎数量/只 No. of cubs	观察起始日期 Observation start	观察结束日期 Observation end	所在园区 Location
1522	2019-05-24	4	2019-05-24	2019-08-30	哈尔滨东北虎林园
1132	2019-07-03	3	2019-07-03	2019-09-27	哈尔滨东北虎林园
1532	2019-06-16	4	2019-06-24	2019-07-13	横道河子东北虎林园
泰山 Taishan	2019-06-19	4	2019-06-22	2019-07-15	横道河子东北虎林园



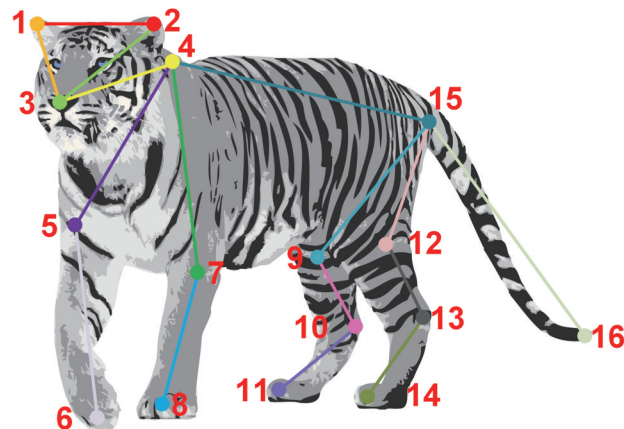
图1 虎舍监控视频样例帧

Figure 1 Sample frames from the surveillance video in the tiger enclosure

### 1.2 数据集的制作

为实现研究目标,构建了一个基于骨骼关键点的东北虎幼虎多目标行为识别和跟踪数据集。构建流程主要包括姿态估计、行为识别与个体跟踪3部分。首先,定义了16个覆盖虎体主要关节部位的关键点(具体定义与骨架连接方式见图2),并对视频中清晰可见的个体进行人工标注,共标注3092张图中的9675个幼虎个体姿态,形成姿态估计数据集。在此基础上,采用已训练的VarifocalNet-R50目标检测网络<sup>[15]</sup>获取每只幼虎的检测框,并利用高分辨率网络(high-resolution network, HRNet)<sup>[16]</sup>进一步生成各

个体的骨骼关键点数据。随后,利用ByteTrack<sup>[17]</sup>多目标跟踪算法,以目标检测框和关键点数据作为输入,对视频序列中的幼虎进行跟踪,并对跟踪结果进行人工校对,从而获取每个个体的姿态时间序列,制作跟踪数据集。为建立行为识别数据集,本研究依据东北虎行为谱<sup>[18]</sup>,选取了“走、卧息、蹲坐、站立、玩耍”5种常见行为(具体定义与示例见表2)。对视频数据按每10min采样一次,每次提取100帧行为序列的规则进行处理,并由专业人员对可识别的行为进行人工精确标注,最终完成2299组行为标签的标注。各行为类别的样本数量统计见表2。



1. 右耳; 2. 左耳; 3. 鼻子; 4. 颈部; 5. 右肩; 6. 右前爪; 7. 左肩; 8. 左前爪; 9. 右臀; 10. 右膝盖; 11. 右后爪; 12. 左臀; 13. 左膝盖; 14. 左后爪; 15. 尾根; 16. 尾尖。

1. Right ear; 2. Left ear; 3. Nose; 4. Neck; 5. Right shoulder; 6. Right forepaw; 7. Left shoulder; 8. Left forepaw; 9. Right hip; 10. Right knee; 11. Right hind paw; 12. Left hip; 13. Left knee; 14. Left hind paw; 15. Tail base; 16. Tail tip.

图2 东北虎关键点与姿态骨架示意图

Figure 2 Schematic diagram of keypoints and pose skeleton of the Amur tiger

表2 东北虎幼虎的5种常见行为定义及其数据分布

Table 2 Definitions and data distribution of five common behaviors in Amur tiger cubs

行为 Behavior	定义 <sup>[18]</sup> Definition <sup>[18]</sup>	图示 Illustration	样本数量/组 Sample size
走 Walking	通过四肢的摆动使身体较慢地向前移动		372
卧息 Lying	卧于地上休息		1 137
蹲坐 Sitting	前肢伸直支撑身体,后肢蜷于身下,整个臀部接触基底		224
站立 Standing	四肢接触基底并支撑身体,静止不动		257
玩耍 Playing	个体间在一起进行轻松愉快的活动		309

### 1.3 目标检测模型

为了给关键点检测模型和后续跟踪模型提供准确的目标框,采用在密集和小目标检测中表现优异的VarifocalNet-R50算法,来实现东北虎幼虎个体的多目标检测。VarifocalNet-R50是一种基于深度学习的高效目标检测模型,其利用50层残差网络(residual network 50, ResNet-50)<sup>[19]</sup>提取图像特征,并通过变焦焦点机制对不同尺度的目标进行优化,从而提高了小物体和密集目标的检测能力。这一特性使其适应于东北虎幼虎这一应用场景,为后续的关键点检测、行为识别和多目标跟踪任务奠定了可靠的基础。

### 1.4 基于HRNet的关键点检测

在VarifocalNet-R50提取的目标检测框基础上,利用HRNet对东北虎幼虎进行骨骼关键点检测,共提取16个关键点坐标。HRNet是一种专为高精度关键点检测任务设计的高分辨率深度学习网络。与传统的低分辨率特征提取方法不同,HRNet首先通过高分辨率网络提取图像的空间特征;随后通过多分支结构融合不同尺度的特征,逐步增强关键点检测的精确度;最终基于融合后的特征实现关键点的定位与分类,从而获得幼虎的关键点坐标。本研究应用HRNet为东北虎幼虎行为识别提供了准确的关键点数据,为后续行为识别任务奠定了基础。

HRNet适用于东北虎幼虎骨骼关键点检测的原因在于其在关键点检测任务中的高精度和强鲁棒性。相比于基于回归的关键点检测方法,HRNet采用热图回归(heatmap regression)方式,能够更精准地定位关键点位置,这对于东北虎幼虎这种目标较小、特征细节较为复杂的对象尤为重要。此外,HRNet的多尺度融合结构能够在不同分辨率下保持关键点检测的稳定性,使其在处理不同姿态、不同尺度的幼虎图像时仍能保持较高的检测精度。

### 1.5 基于骨骼关键点的行为识别

采用基于骨骼关键点的方法对东北虎幼虎进行行为识别,其整体流程如图3所示。该方法通常采用图卷积网络(graph convolutional network, GCN)处

理骨骼关键点这类非欧几里得结构数据。GCN通过聚合相邻节点的特征,有效融合全局信息与局部特征,从而增强模型的表达能力,并具备较强的空间不变性,能够应对节点位置和结构的变化。与时序卷积网络(temporal convolutional network, TCN)结合的GCN能够准确识别个体行为。此外,基于骨骼关键点的行为识别方法能够从骨架数据中提取多维度特征,包括骨骼特征、关节特征、关节运动特征和骨骼运动特征等,从而全面描述幼虎的行为模式。最终,通过时空图卷积网络(spatio-temporal graph convolutional network, STGCN)对这些特征进行预测并融合,准确识别个体行为并输出识别结果。

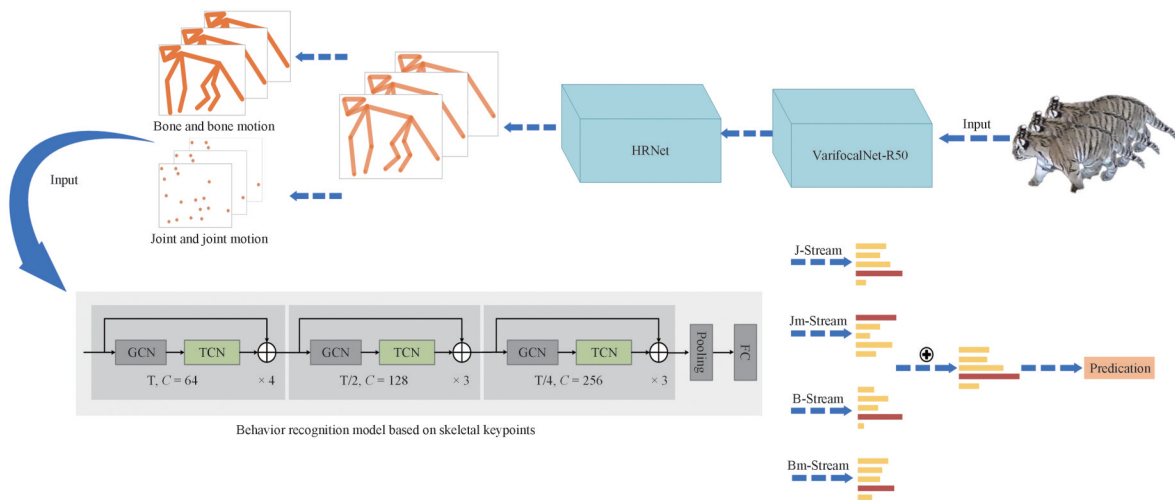


图3 基于骨骼关键点的东北虎幼虎行为识别总体流程

Figure 3 Overall flowchart of behavior recognition for Amur tiger cubs based on skeletal keypoints

为比较不同模型的性能,采用了5种当前流行的基于骨骼关键点的行为识别模型,包括自适应图卷积网络(adaptive graph convolutional network, AGCN)<sup>[20]</sup>、AAGCN<sup>[21]</sup>、通道拓扑精化图卷积网络(channel-wise topology refinement graph convolutional network, CTRGCN)<sup>[22]</sup>、STGCN<sup>[13]</sup>和STGCN++<sup>[12]</sup>。通过综合识别用时及识别准确率筛选性能最佳的模型,以支撑东北虎幼虎行为识别的实际应用需求。

### 1.6 跟踪模型

在多数东北虎幼虎场景中,个体分布密集且遮挡频繁,传统跟踪方法难以有效应对由此产生的低置信度检测问题,常出现跟踪稳定性差和轨迹碎片化等现象。ByteTrack算法充分利用了从高分到低分检测框,通过两阶段匹配策略,先将高置信度检

测框与现有轨迹匹配,再利用低置信度检测框补充未匹配轨迹,并为未匹配的高置信度检测框生成新轨迹。这一方法显著减轻了检测缺失和轨迹碎片化问题,增强了多目标跟踪的鲁棒性。

为准确实现东北虎幼虎的多目标行为识别及个体行为统计分析,采用能有效保留并关联低置信度检测框的ByteTrack算法,将其应用于幼虎群体的多目标跟踪中。为进一步评估不同跟踪模型的性能,将ByteTrack与另外两种主流跟踪算法,即基于深度关联度量的跟踪算法(SORT with deep association metric, DeepSORT)<sup>[23]</sup>和以观测为中心的跟踪算法(observation-centric SORT, OCSORT)<sup>[24]</sup>进行对比,旨在选择性能最佳的模型,以支持东北虎幼虎的行为识别和统计分析。

## 2 结果

### 2.1 实验设置

实验平台配置英特尔中央处理器和英伟达 GeForce RTX 3090 显卡,以保证实验过程具备较高的计算性能和结果可比性。所有实验代码均采用 Python 编写,并在 PyTorch 框架下运行。在姿态估计实验中,将东北虎幼虎姿态估计数据集按照 8:2 的比例划分为训练集和验证集,二者互不重叠,其中训练集包含 2 474 张图像,验证集包含 618 张图像。模型训练共进行 350 轮迭代,使用自适应矩估计优化器 (adaptive moment estimation optimizer, Adam)<sup>[25]</sup> 进行优化,初始学习率设为 0.000 01,并通过 1 000 步的热启动 (warm-up) 逐渐增至 0.001。从第 110 轮开始,采用余弦退火 (cosine annealing) 策略,在 240 轮内将学习率逐步降低至 0.000 05,最终在第 350 轮完成训练。这种学习率调整策略有助于在训练后期进一步优化模型参数,从而提升模型性能。在基于骨骼关键点的行为识别实验中,将行为识别数据集按 8:2 的比例划分为训练集与验证集,两者互不重叠。其中,训练集包含 1 839 组行为样本,验证集包含 460 组行为样本。为应对行为类别不平衡并抑制过拟合,实验中对损失函数设置了类别权重。具体而言,根据 5 种行为的样本比例,类别权重分别设置为 3.051 3、1.000 0、5.080 4、4.428 4 和 3.681 2。模型训练随机数种子统一设置为 1 655 740 202,输入姿态估计序列的长度均为 100 帧,并对训练过程进行 50 轮迭代。采用随机梯度下降优化器 (stochastic gradient descent optimizer, SGD)<sup>[26]</sup>,将学习率从 0.000 01 开始经过 2 轮热启动到达初始学习率 0.01,训练从第 3 轮起采用余弦退火策略,在之后的 47 轮内逐步将学习率降低至 0.000 001,于第 50 轮完成训练。

### 2.2 评价指标

为评估姿态估计模型的性能,采用平均绝对欧几里得距离 (mean absolute Euclidean distance, MAED) 作为评价指标,并采用准确率 (accuracy,  $A_{cc}$ ) 来衡量东北虎幼虎行为识别模型的性能。准确率指模型正确预测的样本数量占总样本数量的比例,计算公式为

$$A_{cc} = \frac{\sum_{i=1}^N \mathbb{I}(y_i = \hat{y}_i)}{N} \times 100\%。 \quad (1)$$

式中: $N$  为样本总数; $y_i$  为第  $i$  个样本的真实标签; $\hat{y}_i$  为模型对第  $i$  个样本的预测标签; $\mathbb{I}(\cdot)$  为指示函数,当内部条件为真时,取值为 1,否则为 0。

为评估跟踪模块的性能,使用多目标跟踪准确率 (multiple object tracking accuracy, MOTA) 作为评价指标,其计算公式为

$$M_{OTA} = 1 - \frac{\sum_t (F_{Nt} + F_{Pt} + I_{DSwt})}{\sum_t G_{Tt}}。 \quad (2)$$

式中: $t$  为视频中第  $t$  帧影像; $F_{Pt}$  (false positive) 为第  $t$  帧中的假阳性; $F_{Nt}$  (false negative) 为第  $t$  帧中的假阴性; $I_{DSwt}$  (ID switch) 为第  $t$  帧中身份切换的次数; $G_{Tt}$  (ground truth) 为第  $t$  帧中所有真实边界框的数量。

### 2.3 关键点检测实验结果

为了评估 HRNet 关键点检测算法的性能,将其与简单坐标分类 (simple coordinate classification, SimCC)<sup>[27]</sup>、分布感知的关键点表征 (distribution-aware coordinate representation of keypoints, DARK)<sup>[28]</sup>、简单基线模型 (simplebaseline)<sup>[29]</sup> 和基于视觉变换器的姿态估计模型 (vision transformer for pose estimation, ViTPose)<sup>[30]</sup> 等关键点检测模型进行对比实验。所有对比方法均在 VarifocalNet-R50 目标检测结果的基础上进行,以确保实验条件一致。

由表 3 可知,SimCC 方法的 MAED 为 46.89 像素,性能显著低于其他方法,表明其难以在东北虎幼虎场景中实现准确的关键点定位。DARK、Simple-Baseline 和 ViTPose 方法的 MAED 分别为 18.60、19.94、19.37 像素,性能相近且均优于 SimCC,显示出一定关键点检测能力。HRNet 的 MAED 为 17.55 像素,表现最佳,表明 HRNet 能够更准确地定位东北虎幼虎的骨骼关键点,展现出在高细节复杂度场景下的优越建模能力,为后续的行为识别和分析提供更加可靠的数据支持。

### 2.4 基于骨骼关键点的行为识别实验结果

为评估不同模型在东北虎幼虎行为识别任务中的性能,对比了 AGCN、AAGCN、CTRGCN、STGCN 和 STGCN++ 五种基于骨骼关键点的行为识别模型。所有模型均使用 HRNet 获取的骨骼关键点数据,实验结果如表 4 所示。

表3 关键点检测实验结果

Table 3 Keypoint detection experimental results 像素

方法 Methods	平均绝对欧几里得距离 MAED
SimCC	46.89
DARK	18.60
SimpleBaseline	19.94
ViTPose	19.37
HRNet	17.55

表4 基于骨骼关键点的行为识别实验结果

Table 4 Experimental results of behavior recognition based on skeletal keypoints %

模型 Model	准确率 Accuracy
AGCN	76.15
AAGCN	76.59
CTRGCN	74.62
STGCN	70.46
STGCN++	75.93

从表4可知,AAGCN模型在基于骨骼关键点的东北虎幼虎行为识别任务中性能最优,准确率达76.59%,高于其他对比模型。该结果表明,AAGCN在基于骨骼关键点的行为识别中具有较强的表征能力,能够更有效地捕捉东北虎幼虎的行为模式,体现出其在复杂动物行为识别任务中的适用性与先进性。

2.5 跟踪模型实验结果

为验证 ByteTrack 跟踪算法在东北虎幼虎场景下的性能,选取5个随机视频片段,将其与 DeepSORT 和 OCSORT 两种主流多目标跟踪算法进行对比,并对评估结果进行平均处理。实验结果显示,ByteTrack 在本研究数据集上表现最佳(图4)。在身份切换(ID switches, IDSW)方面,ByteTrack 在5个视频片段中的平均身份切换次数为17.6,显著低于 DeepSORT(58.6)和 OCSORT(67.8),显示出其在保持目标身份一致性方面的明显优势。此外,ByteTrack 的 MOTA 最高达到了92.76%,进一步表明其在跟踪精度和稳定性方面的优越性,能更好满足对每只东北虎幼虎个体进行准确跟踪和行为统计的需求。

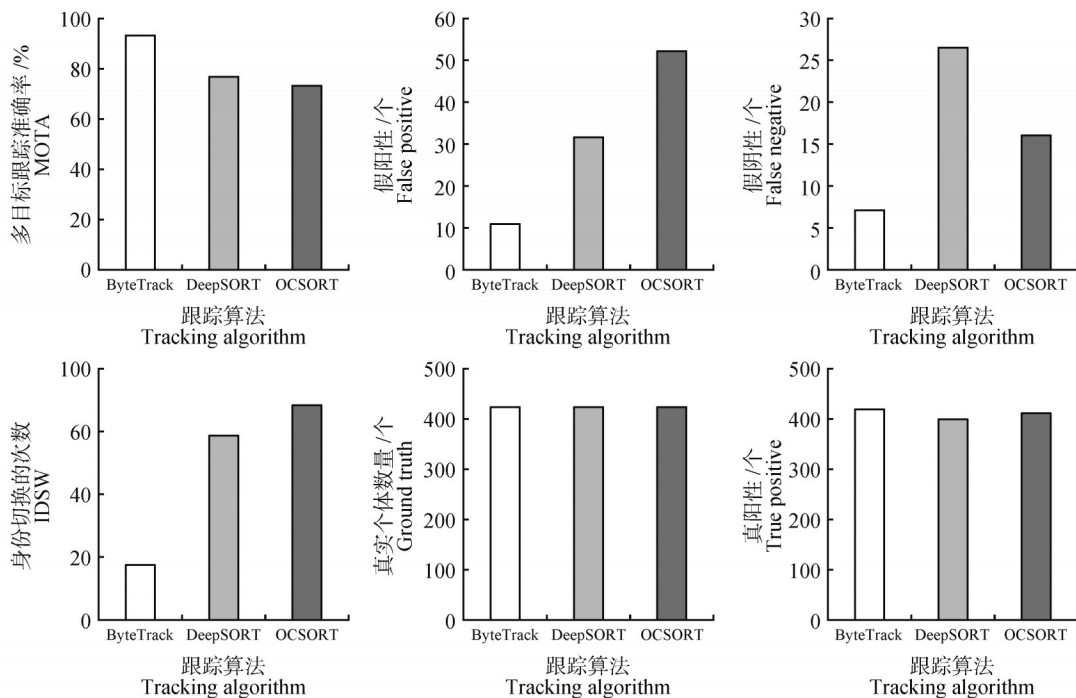


图4 跟踪比较实验中各算法的评估指标

Figure 4 Evaluation metrics of different methods in the tracking comparison experiment

2.6 行为持续时间的统计分析

本研究将基于骨骼关键点的行为识别模型与多目标跟踪算法相结合,实现了多只东北虎幼虎个体

行为的自动化统计分析。通过这一方法,对包含3只东北虎幼虎的样例视频进行识别与统计,计算每只幼虎5种常见行为的持续时间,并与人工标注

结果进行对比,计算相对误差。结果表明,自动化统计与人工标注之间的差异总体较小(表5),验证了该方法的整体有效性。然而,由于各类别样本数量存在不均衡(如“卧息”1 137组,而“蹲坐”仅224组),模型对不同行为的识别精度存在一定差异。其中,“卧息”行为在3只幼虎中的相对误差均低于2%,识别稳定性最高;“走”行为的相对误差均低于5%,说明模型在样本量充足的常见运动行为中具有较高的可

靠性;而“蹲坐”和“站立”等中等样本量行为的误差约为2%~9%,识别精度略有波动;“玩耍”行为由于姿态变化复杂且样本量相对较少,相对误差均在5%~11%,识别性能相对较低。这表明模型在多数行为类别上具有稳定的统计性能,但在样本稀少、姿态多变的行为类别上仍有提升空间。未来可通过数据增强或代价敏感学习等策略进一步优化模型在不平衡数据下的表现。

表5 东北虎幼虎行为时长自动统计与人工标注误差

Table 5 Errors between automated and manual duration annotation for Amur tiger cub behaviors

虎ID Tiger ID	行为 Behavior	自动识别 /s Automatic recognition	人工标注 /s Manual annotation	相对误差 /% Relative error
幼虎1 Cub 1	走	78	80	2.50
	卧息	271	276	1.81
	蹲坐	41	40	2.50
	站立	49	48	2.08
	玩耍	53	48	10.42
幼虎2 Cub 2	走	84	88	4.55
	卧息	219	222	1.35
	蹲坐	26	24	8.33
	站立	19	20	5.00
	玩耍	74	70	5.71
幼虎3 Cub 3	走	48	50	4.00
	卧息	106	108	1.85
	蹲坐	51	52	1.92
	站立	12	11	9.09
	玩耍	49	45	8.89

### 3 讨论

精准监测东北虎幼虎群体的行为模式及其持续时间,对理解其发育规律、评估其福利水平具有重要意义。传统行为统计方法主要依赖人工标注,不仅耗时耗力、主观性强且具有入侵性,干扰动物的正常生活<sup>[9,31]</sup>。因此,开发一种能够自动识别和统计东北虎幼虎行为的技术方法,实现高效、非侵入式自动化监测,已成为一个亟待解决的问题。

Li *et al.*<sup>[5]</sup>使用计算机视觉技术可自动识别狮(*Panthera leo*)和狼(*Canis lupus*)等野生动物的行为,但其方法基于彩色图像,易受环境和光照变化干扰,且忽略了能提供多种互补特征的姿态信息。相比之

下,本研究采用姿态数据进行行为识别,数据更加紧凑且信息丰富,对背景和光照变化具有更好的鲁棒性。Deng *et al.*<sup>[8]</sup>和Feng *et al.*<sup>[9]</sup>虽基于姿态进行了野生动物行为识别,但未能充分利用姿态序列中的多维度特征信息,且主要针对成年个体,未针对幼崽目标小、行为特征不明显的问题进行优化。本研究通过融合姿态的5种特征信息,充分挖掘各特征之间的互补性,并在东北虎幼虎行为识别任务中验证了其有效性。

本研究采用基于骨骼关键点的行为识别方法,实现了对东北虎幼虎视频的自动行为识别,并结合多目标跟踪算法,完成了对多只个体行为的持续性统计。显著提高了行为识别和统计的效率与准确

性。在实验过程中,结合了先进的目标检测、关键点检测和跟踪算法,验证了其在东北虎幼虎场景下的适用性。所获取的目标框和骨骼关键点信息不仅极大丰富了可用数据,为东北虎幼虎行为分析提供了更深入的研究视角,也为其保护工作提供了更为精准的技术支持。

本研究的实验结果表明,基于骨骼关键点的行为识别方法能有效捕捉东北虎幼虎的动作特征。通过提取个体骨骼关键点进行行为分析,在一定程度上提高了行为识别的准确性。然而,该方法在实际应用中对训练样本的数量有较高依赖。此外,对采集的图像进行关键点标注等数据处理时可能消耗大量时间与人力。尽管本研究采用训练完成的 VarifocalNet-R50 和 HRNet 算法以较高精度获取了东北虎幼虎的骨骼关键点数据,但在处理复杂场景或个体行为快速变化时,模型的鲁棒性和稳定性仍有待提升。这些因素也可能影响后续行为识别与跟踪统计的准确性。

为进一步完善本研究方法,未来工作将围绕以下方面展开:(1)扩充数据集,增加更多东北虎幼虎个体的行为样本,提升数据的多样性和代表性;(2)优化行为识别、目标检测和关键点检测模型增强对复杂背景与遮挡情况的适应能力,并提升对幼虎特定行为的识别精度;(3)拓展应用场景,基于本方法的有效性与实用性,将其集成至自动化监控系统中,服务于东北虎幼虎群体的长期行为监测和科学保护。

#### 4 结论

本研究构建了首个基于骨骼关键点的东北虎幼虎行为识别与跟踪数据集,为该领域的定量研究提供了宝贵资源。在此基础上,采用基于骨骼关键点的行为识别方法,结合目标检测和跟踪算法,致力于自动化统计东北虎幼虎群体的行为。实验结果表明,在行为识别方面,AAGCN 能够最有效地从骨骼关键点序列中提取特征,在幼虎行为识别任务中的准确率达到 76.59%,性能优于其他对比模型。在个体跟踪方面,ByteTrack 在幼虎个体跟踪任务中的 MOTA 达到 92.76%,并能显著减少身份切换次数,为实现个体级别的持续行为监测提供了可靠支撑。在应用验证方面,将行为识别与跟踪算法结合,对样例视频进行自动化行为统计的结果与人工标注高度

吻合,多数行为时长的相对误差控制在 5% 以内,证明了该方法在宏观行为谱分析中的实用价值。

尽管本方法在复杂场景下的鲁棒性以及快速变化行为的识别精度仍有提升空间,但本研究成功证实了基于骨骼关键点的解决方案在东北虎幼虎行为监测中的巨大潜力。未来工作将集中于优化模型、扩充数据集,并推动该技术向自动化监控系统中转化,从而为野生动物保护与繁育管理提供更强大的技术工具。

#### 参考文献:

- [1] 国家重点保护野生动物名录(2021年2月1日修订)[J]. 野生动物学报, 2021, 42(2): 605-640.  
List of national key protected wild animals in China (revised on February 1, 2021) [J]. Chinese Journal of Wildlife, 2021, 42(2): 605-640.
- [2] GOODRICH J, WIBISONO H, MIQUELLE D, et al. *Panthera tigris* [J/OL]. The IUCN Red List of Threatened Species, 2022: e.T15955A214862019 [2025-01-14]. <https://dx.doi.org/10.2305/IUCN.UK.2022-1.RLTS.T15955A214862019.en>.
- [3] ALIBHAI S K, GU J Y, JEWELL Z C, et al. 'I know the tiger by his paw': A non-invasive footprint identification technique for monitoring individual Amur tigers (*Panthera tigris altaica*) in snow [J]. Ecological Informatics, 2023, 73: 101947.
- [4] 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.
- [5] LI W N, SWETHA S, SHAH M. Wildlife action recognition using deep learning [EB/OL]. TechRxiv (2025-10-27) [2025-10-30]. <https://www.techrxiv.org/users/43067/articles/1347384-wild-life-action-recognition-using-deep-learning>.
- [6] SWARUP P, CHEN P, HOU R, et al. Giant panda behaviour recognition using images [J]. Global Ecology and Conservation, 2021, 26: e01510.
- [7] 马光凯, 张静, 戴文锐, 等. 基于Transformer的东北虎体侧条纹个体识别[J]. 野生动物学报, 2024, 45(4): 734-743.  
MA G K, ZHANG J, DAI W R, et al. Body stripes individual identification of Amur tigers based on transformer [J]. Chinese Journal of Wildlife, 2024, 45(4): 734-743.
- [8] DENG S C, TANG G Z, MEI L. Wild mammal behavior recognition based on gated transformer network [C]//2022 International Conference on Cyber-Physical Social Intelligence (ICCSI), November 18-21, 2022. Nanjing: IEEE, 2022: 739-743.
- [9] FENG L Q, ZHAO Y Q, SUN Y C, et al. Action recognition using a spatial-temporal network for wild felines [J]. Animals, 2021, 11(2): 485.
- [10] LIN C W, HONG S D, LIN M X, et al. Bird posture recognition based on target keypoints estimation in dual-task convolutional neural networks [J]. Ecological Indicators, 2022, 135:

- 108506.
- [11] LI Z Y, ZHANG Q R, LV S C, *et al.* Fusion of RGB, optical flow and skeleton features for the detection of lameness in dairy cows [J]. *Biosystems Engineering*, 2022, 218: 62–77.
- [12] DUAN H D, WANG J Q, CHEN K, *et al.* PYSKL: towards good practices for skeleton action recognition [C]//Proceedings of the 30th ACM International Conference on Multimedia, October 10–14, 2022. Lisboa: ACM, 2022: 7351–7354.
- [13] YAN S J, XIONG Y J, LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition [EB/OL]. arXiv (2018-01-25) [2025-01-07]. <https://doi.org/10.48550/arXiv.1801.07455>.
- [14] DUAN H D, ZHAO Y, CHEN K, *et al.* Revisiting skeleton-based action recognition [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18–24, 2022. New Orleans: IEEE, 2022: 2969–2978.
- [15] ZHANG H Y, WANG Y, DAYOUB F, *et al.* VarifocalNet: An IoU-aware dense object detector [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20–25, 2021. Nashville: IEEE, 2021: 8514–8523.
- [16] 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), June 15–20, 2019. Long Beach: IEEE, 2020: 5693–5703.
- [17] ZHANG Y F, SUN P Z, JIANG Y, *et al.* ByteTrack: multi-object tracking by associating every detection box [C]//AVIDAN S, BROSTOW G, CISSÉ M, *et al.* Computer Vision-ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, proceedings, part XIII. Cham: Springer, 2022: 1–21.
- [18] 乔征磊, 韦钦国. 东北虎描述性全行为谱的构建 [J]. *黑龙江畜牧兽医*, 2015(9): 207–209.  
QIAO Z L, WEI Q G. Construction of descriptive behavioral ethogram of Amur tiger [J]. *Heilongjiang Animal Science and Veterinary Medicine*, 2015(9): 207–209.
- [19] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27–30, 2016. Las Vegas: IEEE, 2016: 770–778.
- [20] SHI L, ZHANG Y F, CHENG J, *et al.* Two-stream adaptive graph convolutional networks for skeleton-based action recognition [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15–20, 2019. Long Beach: IEEE, 2020: 12018–12027.
- [21] SHI L, ZHANG Y F, CHENG J, *et al.* Skeleton-based action recognition with multi-stream adaptive graph convolutional networks [J]. *IEEE Transactions on Image Processing*, 2020, 29: 9532–9545.
- [22] CHEN Y X, ZHANG Z Q, YUAN C F, *et al.* Channel-wise topology refinement graph convolution for skeleton-based action recognition [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10–17, 2021. Montreal: IEEE, 2022: 13359–13368.
- [23] WOJKE N, BEWLEY A, PAULUS D. Simple online and real-time tracking with a deep association metric [C]//2017 IEEE International Conference on Image Processing (ICIP), September 17–20, 2017. Beijing: IEEE, 2018: 3645–3649.
- [24] CAO J K, PANG J M, WENG X S, *et al.* Observation-centric SORT: Rethinking SORT for robust multi-object tracking [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17–24, 2023. Vancouver: IEEE, 2023: 9686–9696.
- [25] KINGMA D P, BA J. Adam: A method for stochastic optimization [EB/OL]. arXiv (2017-01-30) [2025-01-14]. <https://doi.org/10.48550/arXiv.1412.6980>.
- [26] BOTTOU L. Large-scale machine learning with stochastic gradient descent [C]//LECHEVALLIER Y, SAPORTA G. Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics, Paris France, August 22–27, 2010. Heidelberg: Springer, 2010: 177–186.
- [27] LI Y J, YANG S, LIU P D, *et al.* SimCC: A simple coordinate classification perspective for human pose estimation [C]//AVIDAN S, BROSTOW G, CISSÉ M, *et al.* Computer Vision-ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, proceedings, part VI. Cham: Springer, 2022: 89–106.
- [28] ZHANG F, ZHU X T, DAI H B, *et al.* Distribution-aware coordinate representation for human pose estimation [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13–19, 2020. Seattle, IEEE, 2020: 7093–7102.
- [29] XIAO B, WU H P, WEI Y C. Simple baselines for human pose estimation and tracking [C]//Computer Vision-ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, proceedings, part VI. Cham: Springer, 2018: 472–487.
- [30] XU Y F, ZHANG J, ZHANG Q M, *et al.* ViTPose: Simple vision transformer baselines for human pose estimation [EB/OL]. arXiv (2022-04-26) [2025-01-14]. <https://arxiv.org/abs/2204.12484>.
- [31] NASIRI A, YODER J, ZHAO Y, *et al.* Pose estimation-based lameness recognition in broiler using CNN-LSTM network [J]. *Computers and Electronics in Agriculture*, 2022, 197: 106931.