中图法分类号:TP391.4; S828　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-16

论文引用格式:Wang Min, Hou Zhenjie, Qi Fangzheng, Zhang Wenguang. PBVD-13: Construction and validation of a pig behaviour video dataset for 13 behaviours[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260400.(王敏, 侯振杰, 漆方正, 张文广. 面向13种行为的猪行为视频数据集构建和验证(PBVD-13)[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.260400.)[DOI: 10.11834/jig.260400]

面向13种行为的猪行为视频数据集构建和验证(PBVD-13)

王敏[1],侯振杰[1],漆方正[1],张文广[2]

1. 常州大学计算机与人工智能学院, 常州 213100; 2. 内蒙古农业大学乡村振兴研究院,呼和浩特 010010

摘　要:**目的**　猪行为识别是智慧养殖与动物福利评价的重要基础,但现有公开的猪行为视频数据集在行为类别、福利相关行为覆盖及标注粒度等方面仍存在不足。同时,复杂群养环境中的背景干扰、个体遮挡与频繁交互增加了行为相关局部区域和动态变化信息建模的难度。为此,本文构建细粒度猪行为视频数据集,并研究复杂群养环境下猪行为的时空特征建模方法。**方法**　基于公开的真实群养猪视频数据,采用秒级个体行为标注方式构建猪行为视频数据集(Pig Behaviour Video Dataset, PBVD-13),包含13类行为和13,500个行为实例。其中,探索、与玩具互动、争斗和鼻触碰4类为福利相关行为,共包含3513个行为实例,占全部实例的26.02%。在此基础上,提出一种时空感知与增强网络,通过低帧率空间分支增强行为相关区域的空间特征,高帧率运动分支捕获局部动态变化,并结合时序建模实现细粒度猪行为分类。**结果**　PBVD-13上的基准实验表明,不同模型在争斗、鼻触碰、起身和躺卧等复杂细粒度行为上的识别性能差异明显,表明该数据集具有较好的可学习性和一定的任务挑战性。所提出方法在PBVD-13上取得75.92%的均值平均精度(mean average precision, mAP),较最优对比模型SlowFast提高9.41个百分点,并在争斗、起身和躺卧等时序依赖较强的行为上取得更明显的性能提升。进一步在独立的HotPig数据集上进行外部验证,在保持模型主体结构及功能模块不变、仅调整最终分类层输出维度且未针对具体行为类别进行额外针对性参数优化的条件下,取得76.64%的mAP。**结论**　PBVD-13能够为复杂群养环境下多类别猪行为识别及福利相关行为分析提供细粒度数据基础;所提出的时空感知与增强网络能够有效建模行为相关空间区域与动态时序信息,并在不同猪行为数据集上表现出较好的适用性,可为智慧养殖中的猪行为分析与动物福利评价提供技术支持。相关数据集可通过以下地址获取:https://github.com/qiqiqi7777/PBVD-13-dataset。

关键词:猪;视频数据集;行为识别;时空特征;注意力机制

PBVD-13: Construction and validation of a pig behaviour video dataset for 13 behaviours

Wang Min[1], Hou Zhenjie[1], Qi Fangzheng[1], Zhang Wenguang[2]

1. School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213100, China; 2. Rural Revitalization Research Institute, Inner Mongolia Agricultural University, Hohhot 010010, China

Abstract: **Objective** Existing pig behaviour video datasets remain limited in their coverage of behaviour categories, annotation of welfare-related behaviours and annotation granularity. These limitations hinder the training and evaluation of models for multi-class, fine-grained pig behaviour recognition in complex group-housing environments. Furthermore, some datasets are not publicly available, further limiting fair comparisons between methods and the reproducibility of experimen-

收稿日期:2026-07-02;**修回日期**:2026-10-29

© 中国图象图形学报版权所有

tal results. Meanwhile, background interference, occlusion between pigs and frequent interactions in complex group-housing environments make it more difficult to extract behaviour-relevant local regions and temporal dynamics. This hinders fine-grained behaviour recognition, particularly for behaviours that rely on local interaction cues, such as fighting and nose touching, and those involving continuous postural transitions, such as standing up and lying down. To address these issues, this study aims to develop a video dataset of pig behaviour covering multiple basic behaviour classes, fine-grained postures and movements, and welfare-related behaviours. It also proposes a spatiotemporal feature-modelling method for complex group-housing environments to improve fine-grained pig behaviour recognition. **Method** The 13-class Pig Behaviour Video Dataset (PBVD-13) is constructed from publicly available red, green and blue (RGB) videos of pigs in group-housing environments, with individual behaviours annotated at one-second intervals. It comprises 13 behaviour classes, covering basic activities, welfare-related behaviours such as exploration, toy interaction, fighting and nose touching, and fine-grained postural and movement behaviours such as standing up and lying down. The dataset provides a basis for multi-class pig behaviour recognition in complex group-housing environments. To address the difficulty of adequately representing local interactions and continuous motion changes in complex behaviours, a spatiotemporal perception and enhancement network is proposed. It employs a low-frame-rate spatial branch and a high-frame-rate motion branch to extract behaviour-related spatial features and local dynamics, respectively. The spatial branch characterises relatively stable behaviour-related regions, while the motion branch captures local motion changes between consecutive frames. The network further integrates spatial attention, motion feature enhancement and long short-term memory (LSTM) temporal modelling to fuse features across different temporal scales and achieve fine-grained pig behaviour classification. The source footage comprises publicly available recordings of eight pigs housed at a research pig unit at Scotland's Rural College. Twenty videos are selected for fine-grained behaviour annotation. Each complete video serves as the smallest unit for dataset partitioning, and the data are divided into training, validation and test sets in a ratio of 7:2:1 to reduce the risk of information leakage caused by adjacent video frames being assigned to different sets. Behaviour recognition performance is evaluated using average precision (AP) and mean average precision (mAP). Dataset benchmarking, model comparisons, ablation studies and external validation on the independent HotPig dataset are conducted to assess, respectively, the dataset's suitability for different technical approaches, the recognition performance of the proposed method, the contributions of individual modules and the method's applicability across datasets. **Result** PBVD-13 comprises 13 behaviour classes and 13,500 behaviour instances. The four welfare-related behaviour classes comprise 3,513 instances, accounting for 26.02% of all annotated instances. Benchmark experiments show that recognition performance varies considerably across behaviour classes. Complex behaviours, particularly fighting, nose touching, standing up and lying down, remain challenging to recognise, indicating that the dataset supports model training while presenting challenges for behaviour classification. The proposed method achieves a mAP of 75.92% on PBVD-13, outperforming the best-performing comparison model, SlowFast, by 9.41 percentage points. Its AP values for fighting, standing up and lying down reach 60.75%, 52.37% and 41.98%, respectively. Ablation studies show that removing either spatial-attention module, the motion-enhancement module or the LSTM module reduces mAP, indicating that spatial feature enhancement, local motion modelling and temporal information extraction contribute to overall recognition performance. Furthermore, the proposed method achieves an mAP of 76.64% on the independent HotPig dataset, providing preliminary evidence of its applicability across different datasets. Compared with existing pig behaviour datasets examined in this study, PBVD-13 extends the number of behaviour classes to 13. It retains basic activity behaviours while incorporating welfare-related behaviours, including exploration, toy interaction, fighting and nose touching, as well as fine-grained postural and movement behaviours, such as standing up and lying down. The publicly available dataset also supports reproducibility and comparative evaluation in future research. The proposed method also achieves higher AP values than the comparison models for complex behaviours, including fighting, standing up and lying down. These results indicate that dual-branch spatiotemporal feature modelling contributes to the recognition of the 13 pig behaviour classes considered in this study. **Conclusion** This study constructs and publicly releases PBVD-13, a fine-grained video dataset covering 13 pig behaviour classes, and proposes a spatiotemporal perception and enhancement network. The network integrates a low-frame-rate spatial branch, a high-frame-rate motion branch and subsequent temporal modelling to extract behaviour-related spatial features and dynamic information. Benchmarking, model

© 中国图象图形学报版权所有

comparison, ablation and external validation experiments show that PBVD-13 supports the training and evaluation of multi-class pig behaviour recognition methods. The proposed network achieves promising recognition performance under the experimental conditions and demonstrates applicability to the independent HotPig dataset. PBVD-13 provides a data resource for multiclass pig behaviour recognition in complex group-housing environments, welfare-related behaviour analysis and fair comparison of related algorithms. The proposed network provides a methodological basis for jointly modelling local interaction features and continuous postural changes. As the current dataset includes only eight pigs and a limited number of videos, future work should expand the dataset to include more pigs, farming environments and behaviour instances. The model's generalisation should also be evaluated under varying husbandry conditions, individual differences and levels of occlusion to improve its practical applicability. The dataset and source code are available at: https://github.com/qiqiqi7777/PBVD-13-dataset.

Key words: Pig; video dataset; behavioural recognition; spatio-temporal features; attention mechanism

论文引用格式：DOI：10.11834/jig.260400

0 引　言

尽管近年来猪行为识别技术取得了显著进展，但公开可用的猪行为数据集仍然十分有限，难以满足复杂群养环境下多类别行为识别模型的训练与评估需求。数据资源的匮乏不仅限制了模型性能的进一步提升（Bhujel等，2025），也阻碍了不同研究之间的公平比较与技术复现，已成为当前猪行为识别领域的重要瓶颈。

猪是全球养殖规模最大的经济家畜之一，其健康状态与生产效率直接影响畜牧业的经济收益（Xu等，2025）。随着集约化养殖模式的快速发展，高密度饲养引发的健康与福利问题日益受到关注（折帅帅和冯平，2020）。在此背景下，猪只行为监测为评估其健康与福利状态提供了重要依据。其中，饮水、采食、卧、站立和移动等基础生理与活动行为能够反映猪只的日常生活状态，其发生频率、持续时间及活动模式可能随疾病或应激状态而发生变化（Rohan等，2024）。

此外，探索、攻击及社会行为也与猪只的环境适应和福利状况密切相关。已有研究表明，环境丰富化可促进猪只探索，并可能减少攻击、咬尾等不良行为（Godyń等，2019）；在商业化养殖环境中，提供丰富化物品与皮肤损伤减少有关（Smith和Pierdon，2024），而群体组成及环境丰富化条件也会影响散养猪的探索和社会行为（Collarini等，2024）。因此，结合基础行为与福利相关行为开展持续、准确的监测，有助于全面掌握猪只的健康和活动状态，并为动物福利评价提供依据（Berckmans，2017）。

近年来，已有研究陆续构建了猪行为数据集，并推动了相关识别算法的发展。然而，现有公开的猪行为数据集在标注粒度、福利相关行为覆盖及数据公开性等方面仍存在一定局限，主要表现在：标注粒度有限：1）多数数据集仍主要围绕进食、饮水、卧、站立和移动等基础活动状态进行标注，对起身、躺卧、行走等具有更明显姿态变化或连续运动特征的细粒度行为区分不足，难以充分支撑复杂群养场景下的细粒度行为识别。2）福利相关行为覆盖不足：与动物福利评价关系较为密切的行为覆盖仍然有限，尤其是探索、与丰富化设施互动、争斗和鼻触碰等涉及环境利用和社会互动的行为，在现有数据集中尚缺乏较为系统的标注。3）数据公开性不足：部分数据集虽被用于论文实验，但未对外公开，导致算法难以复现与公平比较（Bhujel等，2025）。

除数据集层面的不足外，复杂养殖环境下的猪行为自动识别也面临诸多挑战。随着计算机视觉与深度学习技术的发展，猪行为监测逐渐由人工观察转向基于视频的自动识别（Yang和Xiao，2020），相关技术也应用于猪只识别研究（刘峰等，2023）。

早期研究主要依赖人工设计特征与传统机器学习方法实现行为识别。例如，Viazzi等通过运动历史图像提取平均运动强度与占用指数等特征，并结合线性判别分析实现攻击行为分类（Viazzi等，2014）。Amezcua等利用红外热成像技术检测母猪跛行（Amezcua等，2014），Scolari等则将该技术用于母猪发情检测（Scolari等，2011）。这类方法在特定场景下取得了一定效果，但其对人工特征设计依赖较强，在遮挡、个体交互频繁和行为变化复杂的群养环境中，其特征表达能力与场景适应性仍受到一定限制。

© 中国图象图形学报版权所有

随着深度学习的发展，研究者开始利用卷积神经网络（convolutional neural networks，CNN）、循环神经网络（recurrent neural network，RNN）以及双流网络等方法，对猪行为中的空间外观与时间动态信息进行联合建模。例如，CNN与门控循环单元（gated recurrent unit，GRU）相结合的方法被用于猪只攻击行为识别（Gao等，2023）；InceptionV3与长短期记忆网络（long short-term memory，LSTM）构建的时空模型用于识别猪只对环境丰富化设施的行为（Chen等，2020）；SlowFast双流网络则通过多时间尺度时空建模实现母猪护理行为识别（Li等，2023）。这些研究推动了时空特征建模在猪行为识别中的应用。

近年来，视频动作识别与行为分析领域持续发展，研究重点逐渐由传统时空特征提取向局部区域建模、多尺度特征融合、时序结构建模及多模态表征等方向拓展（Gong和Luo，2025）。在局部区域建模方面，局部动作查询被用于突出视频中的动作相关区域，增强全局上下文与局部动作细节的联合表征（Wang等，2025）；此外，稠密与稀疏骨骼联合表征可引导RGB（red，green，blue）图像中的感兴趣区域（region of interest，ROI）定位，探索利用结构化骨骼信息辅助局部区域定位的多模态识别方法（杜海涛等，2026）。

结合双分支网络、扩张卷积和多尺度特征融合的方法可提升实时动作检测性能（程勇等，2025）；相关研究通过统一框架比较不同的时序动作检测方法，指出时间边界定位与时序结构建模仍是关键问题（Liu等，2025）。在特征表征方面，多模态时空适配器能够增强时空信息建模与知识迁移能力（Chen等，2025），动态多粒度图结构则通过融合不同粒度的骨架信息，减少特征冗余（吴志泽等，2025）。此外，结合语义相关区域选择与时空视频管构建的方法，有助于提升模型对复杂场景的适应能力（Ge等，2026）。然而，领域特定基准研究表明，现有模型在专业领域动作识别中仍存在局限（Yadav等，2026）。

尽管通用视频动作识别方法不断发展，但其研究对象与群养猪只的实际识别场景仍存在差异。在猪行为检测方面，群养环境中的个体遮挡会增加目标感知与行为识别的难度，现有研究还存在公开数据集规模与多样性不足、行为类别覆盖不均衡等问题（Xu等，2025）。从畜禽行为识别的整体研究来看，复杂动态行为识别以及模型的实时性与可扩展性仍是值得关注的问题（Guo等，2026）。群养猪视频中多个体同时出现，而鼻触碰、争斗等交互行为以及起身、躺卧等姿态转换行为又分别依赖局部交互区域和连续运动变化。因此，如何结合群养猪行为特点，有效利用动作识别领域已有的局部区域与时空建模方法，进一步提升细粒度猪行为表征能力，仍值得深入研究。

为此，本文基于苏格兰农村学院（Scotland's Rural College，SRUC）采集的猪群监测视频（Bergamini等，2021），构建了涵盖13类细粒度猪行为的视频数据集（Pig Behaviour Video Dataset，PBVD-13）。该数据集在覆盖基础生理与活动行为的同时，进一步纳入了探索、与玩具互动、争斗和鼻触碰等与福利评价关系较为密切的行为，并采用秒级时空行为标注方式对个体猪及其行为进行标注，为复杂群养环境下多类别猪行为识别提供更加细粒度的数据基础。相比现有猪行为数据集，PBVD-13在行为类别、标注粒度及福利相关行为覆盖方面进行了进一步扩展，并对数据集进行了公开，可为复杂群养环境下的猪行为识别与福利相关行为分析提供数据基础。

在此基础上，为验证所构建数据集的有效性与挑战性，并针对复杂群养环境中背景干扰、个体遮挡及局部运动变化难以充分表征等问题，本文提出一种时空感知与增强网络。该网络通过低帧率空间分支与高帧率运动分支分别建模行为相关区域的空间特征和局部动态变化，并结合时序建模增强行为的时间依赖表征，从而实现复杂群养环境下13类猪行为的联合识别。

综上，本文的主要贡献包括：

1）构建了面向复杂群养环境的猪行为视频数据集PBVD-13，包含13类细粒度猪行为和13500个行为实例，并进一步覆盖多类与福利评价关系较为密切的行为，为多类别猪行为识别与福利相关行为分析提供数据基础。

2）提出时空感知与增强网络，通过双分支分别建模行为相关区域的空间特征与局部运动变化，并结合时序建模增强细粒度行为的时空表征能力。

3）通过系统的对比实验与消融实验，验证了PBVD-13的有效性与挑战性，并进一步分析了所提出时空感知与增强机制在复杂猪行为识别中的作用。

© 中国图象图形学报版权所有

1 数据集

本研究基于公开的猪视频监测数据（Bergamini等，2021），重新构建了包含细粒度行为标签的猪行为数据集。原始数据集于2019年11月5日至12月11日在SRUC的研究猪舍中采集，涉及同一猪栏中的8头背部具有特殊标记的猪。本文从公开数据集中实际获取了1783段MPEG-4 Part 14（MP4）格式视频。猪舍内配备一个三槽饲喂器、两个乳头式饮水器及一个悬挂于猪只可接触高度的塑料丰富化装置，猪栏部分地面采用漏缝地板，并铺设稻草和碎纸，整体环境如图1所示。



图1 猪活动数据收集环境

Fig. 1 Pig activity data collection environment

图像数据由安装于距地面约2.5m处的Intel RealSense D435i摄像头采集。原始数据同时包含RGB图像与深度信息，均以1280×720分辨率和6帧/秒（frames per second，fps）采集，本文仅使用其中的RGB视频数据进行后续数据集构建与行为标注。原始采集过程中，每1800个连续采集帧对应约300秒，即约5分钟的实际监测时长。在视频预处理阶段，为统一视频处理流程及秒级标注时间轴，对获取的RGB视频进行重新编码，即将每个原始观测帧重复5次并以30 fps保存为MP4格式。该过程不涉及帧插值，也不产生新的独立观测信息，仅改变视频的编码帧率和帧索引方式，因此视频的有效采样率仍为6 fps。

原始研究从完整数据中选取了12个视频序列，共7200帧进行人工标注，并为每帧中可见猪只提供矩形边界框、个体编号（individual identifier，ID）以及卧、移动、进食、饮水和站立5类行为标签。上述标注主要覆盖猪只的基础活动状态，对探索、玩具互动、争斗和鼻触碰等社会互动及福利相关行为的覆盖较为有限，因此难以满足多类别细粒度猪行为识别与福利相关行为分析的需求。

从现有猪行为数据集的整体情况来看，行为类别规模与福利相关行为覆盖仍存在一定局限。如表1所示，多数数据集主要关注进食、饮水、躺卧和移动等基础行为，对鼻触碰、玩具互动等社会互动及环境探索相关行为支持不足。虽然部分研究已开始引入攻击与探索行为，但相关数据集通常类别数量有限，且部分数据未公开，导致模型复现与跨方法公平比较受到限制。

表1 现有猪行为识别数据集的对比

Table 1 Comparison of existing pig behaviour recognition datasets

数据集（Dataset）	行为类别数	行为类别	福利相关行为	是否公开
Pig Behaviour Recognition Video（Li等，2020）	5	进食、卧、移动、抓挠、爬跨	抓挠、爬跨	否
Edinburgh Pig Behaviour Video（Bergamini等，2021）	5	饮水、进食、卧、移动、站立	—	是
Pig Behaviour Recognition（Hao等，2023）	6	饮水、进食、卧、行走、探索、争斗	探索、争斗	否
Pig Aggressive Behaviour Recognition（Ji等，2023）	2	攻击、非攻击	攻击	是
Pig Behaviour Recognition（Zhang，2025）	6	卧、睡眠、探索、进食、行走、爬跨	探索、爬跨	是
PBVD-13（本文）	13	饮水、进食、卧、坐、站立、移动、行走、探索、与玩具互动、争斗、鼻触碰、起身、躺卧	探索、与玩具互动、争斗、鼻触碰	是

© 中国图象图形学报版权所有

为提升数据集在实际养殖环境中的适用性，本文基于原始RGB视频数据，参考Atomic Visual Actions（AVA）数据集的秒级时空行为标注方式，对原有行为类别进行扩展并重新标注（Gu等，2018）。在保留饮水、进食、卧、移动和站立5类原有基础活动行为的基础上，进一步增加坐、行走、起身和躺卧4类细粒度姿态与运动状态行为，以及探索、与玩具互动、争斗和鼻触碰4类与福利评价关系较为密切的行为，最终构建包含13类猪行为的PBVD-13数据集。其中，探索和与玩具互动主要反映猪只对环境及丰富化设施的探索与利用，争斗和鼻触碰则反映猪只间的社会互动。各类行为的属性及其与福利评价的关系见补充材料表S1。上述4类福利相关行为占13类行为标签的30.77%，共包含3513个行为实例，占全部13500个行为实例的26.02%。

在数据集构建过程中，本文从原始RGB视频中选取20段视频用于细粒度行为标注与后续实验。首先对所选视频进行预处理，并按照视频实际时间轴以1秒为间隔选取标注帧，即每秒选取一个时刻对应的图像帧进行行为标注。随后，对每个标注帧中出现的猪只分别进行逐实例行为标注。

标注过程中，采用VGG Image Annotator工具对目标猪进行矩形边界框标注，并记录对应的猪只编号及行为类别信息（Dutta and Zisserman，2019）。每个行为实例对应标注帧中的一头目标猪，包含边界框坐标、ID以及对应的行为标签。图2展示了行为标注过程及标注界面示例。

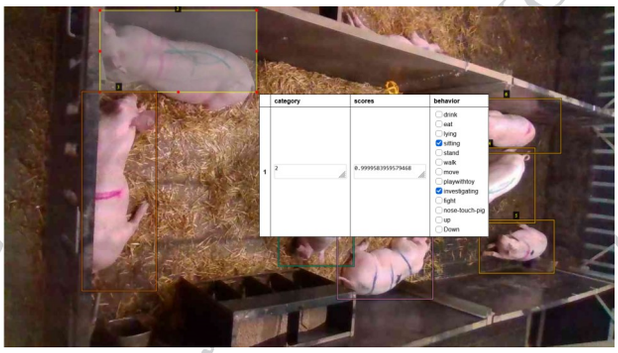


图2　使用VGG Image Annotator标注过程

Fig. 2　Annotation process using VGG Image Annotator

完成行为标注后，数据集中共包含13500个猪行为实例，涉及视频中出现的8头猪个体。为避免同一视频中时间上高度相关的标注帧被划分到不同数据子集中，从而降低潜在的信息泄漏风险，本文采用视频级划分，以完整视频作为最小划分单元。具体而言，将上述20段视频按照7∶2∶1的比例划分为训练集、验证集和测试集，分别包含14、4和2段视频。同一视频中的所有标注帧及其对应行为实例均归属于同一数据子集。最终，训练集、验证集和测试集分别包含9594、2725和1181个行为实例，



图3　猪的13种具体行为

Fig. 3　13 Specific behaviours of pigs

占比分别为71.07%、20.19%和8.75%。具体Video ID及各行为类别分布见补充材料表S2和表S3。

为保证行为定义的一致性与可区分性，本文在标注过程中综合考虑猪只的姿态、运动状态以及与环境或其他个体之间的交互关系。由于猪只在实际养殖环境中可能同时满足多个行为定义，对于存在重叠的行为类别，优先标注语义更具体且能够反映主要行为状态的类别。例如，猪只在进食过程中通常同时处于站立状态，此时优先标注“进食”行为，以表征其与饲喂器之间明确的交互行为。

上述13类行为示例如图3所示，具体定义及对

© 中国图象图形学报版权所有

应行为实例数见表2。

表2　13种猪行为的具体定义及其相应的标注实例数

Table 2　Definitions and numbers of annotated instances for the 13 pig behaviours

行为类型	行为定义	标注实例数
饮水	猪头部接触饮水器并表现出饮水行为	130
进食	猪头部伸入饲喂器并进行采食行为	741
卧	猪身体呈稳定卧姿，腹部或身体侧面与地面接触，躯干基本处于水平状态	873
坐	猪后躯着地，前肢支撑身体，躯干保持抬起状态	3782
站立	猪四肢伸展，身体保持静止站立状态	2527
移动	猪发生短距离位置调整或局部身体移动，但未形成持续的行走过程	1480
行走	猪通过连续步态产生持续、明显的空间位移	234
探索	猪使用鼻部主动嗅探或接触地面、围栏、垫料等周围环境或非特定交互目标	2562
与玩具互动	猪主动接触或推动丰富化装置	56
争斗	两头或多头猪之间发生攻击性身体接触	420
鼻触碰	猪使用鼻部主动接触另一头猪的头部或身体	475
起身	猪由卧姿或坐姿转变为站立状态	106
躺卧	猪由站立状态转变为卧姿或坐姿	114

2　研究方法

本文提出的整体网络结构如图4所示，主要由时空感知网络与时空特征增强网络组成。

首先，对输入视频序列进行不同时间尺度的时序采样，构建具有不同时间分辨率的猪行为帧序列，并输入双分支时空感知网络进行特征提取。其中，慢速分支侧重于学习行为相关的空间语义特征，快速分支则侧重于捕获行为过程中的动态变化信息。在双分支特征提取过程中，通过空间注意力模块强化行为相关关键区域的特征表达，并在快速分支中引入运动特征增强模块，以进一步突出相邻时刻之间的动态变化特征。同时，通过双分支之间的多阶段特征交互与融合，实现不同时间尺度时空信息的有效互补。

随后，利用数据集中人工标注并归一化后的猪个体目标矩形框，对双分支输出的高层特征在各时间步分别进行感兴趣区域对齐（Region of Interest Align，ROI Align），从而提取与目标猪个体对应的局部时空行为特征。训练和测试阶段均采用相同的标注目标框作为猪个体区域信息。最后，对双分支获得的个体级特征进一步进行空间增强、时序建模与特征融合，并输入分类模块完成行为类别预测。

2.1　时空感知网络

由于群体猪行为识别过程中存在背景干扰强、体交互复杂以及局部运动变化明显等问题，传统视频网络难以同时有效关注行为关键区域及其动态变化信息。因此，本文设计了时空感知网络。

考虑到低帧率采样能够保留行为过程中更加稳定的空间语义结构，并减少相邻帧之间的冗余信息，因此更适合提取行为相关的静态空间特征；而高帧率采样能够保留更加连续的时间动态变化，有利于捕获猪行为中的局部细粒度运动模式。

因此，本文分别构建低帧率空间分支与高帧率运动分支，对行为中的静态空间信息与动态运动信息进行解耦建模，从而增强模型对复杂猪行为的时空表征能力。

2.1.1　低帧率空间分支

低帧率空间分支采用基于残差网络构建的空间特征提取结构，由5个残差块构成，并在相邻残差块之间插入4个帧级空间注意力模块（frame-level spatial attention module，FL-SAM），以增强行为相关区域的空间响应能力，如图5。

© 中国图象图形学报版权所有

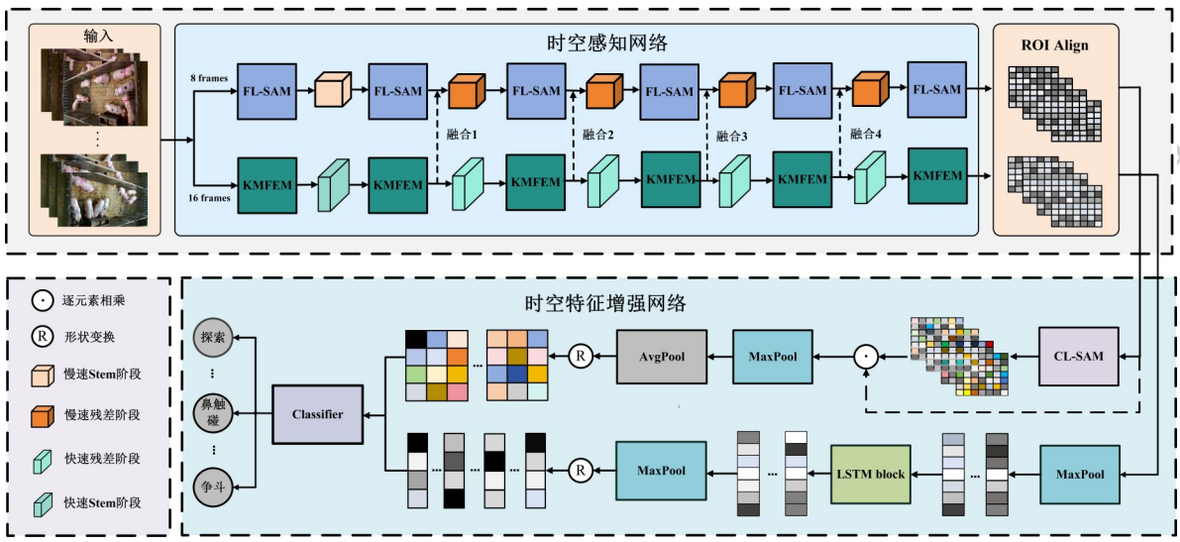


图4　整体网络架构

Fig. 4　The overall framework of our network

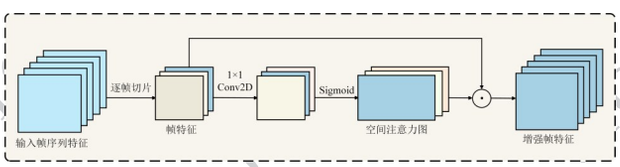


图5　空间特征提取结构

Fig. 5　Spatial Feature Extraction Structure

如图5所示，对于输入特征 $\boldsymbol{X} \in \mathbf{R}^{C \times T \times H \times W}$，式中 C、T、H、W 分别表示通道数、时间维度、特征图高度与宽度。FL-SAM对时间维度上的每一帧特征独立进行空间注意力建模，对于第 t 帧特征，首先通过轻量化分组卷积提取不同局部区域的响应特征，并结合空间注意力机制生成注意力权重 $\boldsymbol{A}$，以减少不同身体区域之间的特征干扰。随后生成的空间注意力权重通过逐元素加权作用于原始特征，从而突出行为关键区域并抑制背景干扰，得到输出 $\tilde{\boldsymbol{X}}$，其计算过程如下：

$$\boldsymbol{A} = \sigma\left(\mathrm{Conv}_{1\times1}^{(2)}\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}^{(1)}(\boldsymbol{X})\right)\right)\right)\right) \tag{1}$$

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} \odot \boldsymbol{A} \tag{2}$$

式中，$\sigma(\cdot)$ 表示Sigmoid函数，$\odot$ 表示逐元素乘法。由于低帧率分支主要负责提取稳定的静态空间信息，因此FL-SAM更关注行为相关区域的空间响应。该机制能够自适应聚焦于嘴部、四肢等关键区域，抑制背景与个体干扰，以轻量化方式增强低帧率分支的空间特征表达能力，并为后续关键运动特征增强模块（Key Motion Feature Enhancement Module，KMFEM）提供更具判别性的空间先验。

2.1.2　高帧率运动分支

高帧率运动分支采用帧间差分残差网络，并在不同特征提取阶段引入KMFEM，以增强模型对局部运动变化的建模能力。其中，KMFEM中的运动特征增强结构如图6所示。

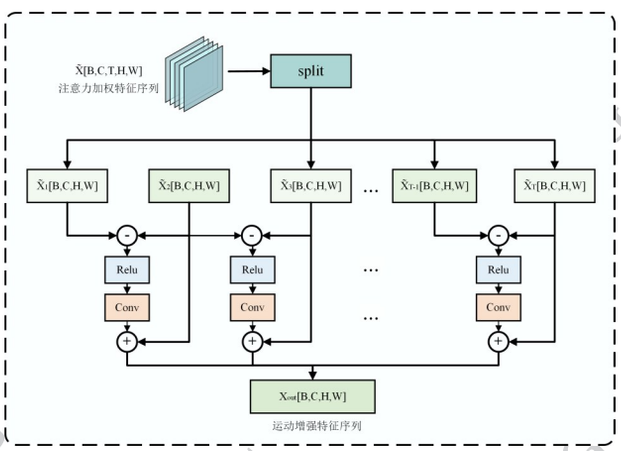


图6　KMFEM中的运动特征增强结构

Fig. 6　Motion feature enhancement structure in KMFEM

该分支保留输入序列中的全部帧进行特征提取，相较于低帧率空间分支具有更高的时间采样密度，以保留更加连续的动态变化信息。在不同阶段的特征提取后引入KMFEM模块，KMFEM主要通过空间注意力增强与运动特征增强两个过程提升高帧率特征表达，其中FL-SAM的空间注意力计算过程已在2.1.1节中给出，图6重点展示其运动特征增强过程。对于相邻两帧特征图：$\boldsymbol{X}_t, \boldsymbol{X}_{t+1} \in \mathbf{R}^{C \times H \times W}$，

© 中国图象图形学报版权所有

KMFEM首先通过共享参数的FL-SAM生成对应的空间注意力权重$\boldsymbol{A}_t$与$\boldsymbol{A}_{t+1}$，并得到加权后的特征表示$\tilde{\boldsymbol{X}}_t$与$\tilde{\boldsymbol{X}}_{t+1}$。随后，对加权后的相邻帧特征进行差分计算：

$$\boldsymbol{M}_t = \mathrm{ReLU}\left(\tilde{\boldsymbol{X}}_{t+1} - \tilde{\boldsymbol{X}}_t\right) \tag{3}$$

该操作仅保留正向运动变化，从而突出行为相关区域的动态响应并抑制静态背景噪声。随后，通过卷积进一步聚合差分运动信息。为避免运动差分过程造成原始语义信息丢失，引入残差连接，将卷积聚合后的运动差分特征与对应的当前帧特征进行残差融合，最终输出为：

$$\boldsymbol{X}_{out,t+1} = \tilde{\boldsymbol{X}}_{t+1} + \mathrm{Conv}\left(\boldsymbol{M}_t\right) \tag{4}$$

通过上述残差融合方式，在增强相邻时刻动态变化信息的同时保留原始空间语义特征。经过多个KMFEM模块逐层增强后，高帧率分支输出运动增强特征图，其中包含更加显著的行为相关动态变化信息，从而增强模型对猪行为中细粒度运动模式的表征能力，并降低背景运动与非目标干扰对行为识别的影响。

为实现低帧率空间特征与高帧率运动特征的有效融合，首先对高帧率分支特征进行时间维对齐。设低帧率与高帧率分支的输出特征分别为$\boldsymbol{F}_s$和$\boldsymbol{F}_f$，其时间长度分别为T_s和T_f，且满足$T_f = \alpha T_s$，式中α为两分支的时间尺度倍率。采用时间维步长为α的三维卷积对高帧率特征$\boldsymbol{F}_f$进行降采样，并依次经过批归一化和ReLU激活，使其时间长度由T_f对齐至T_s。随后，将时间对齐后的高帧率特征与低帧率特征沿通道维进行拼接，得到时空联合特征$\boldsymbol{F}_{sf}$：

$$\tilde{\boldsymbol{F}}_f = ReLU\left(BN\left(Conv3D_{s_t=\alpha}\left(\boldsymbol{F}_f\right)\right)\right) \tag{5}$$

$$\boldsymbol{F}_{sf} = Concat_C\left(\boldsymbol{F}_s, \tilde{\boldsymbol{F}}_f\right) \tag{6}$$

式中，$\tilde{\boldsymbol{F}}_f$表示时间对齐后的高帧率特征，$\mathrm{Concat_C}()$表示沿通道维进行拼接。本文高、低帧率分支的时间长度分别为16和8，因此设置α为2。经过上述多阶段特征交互后，低帧率分支和高帧率分支分别获得增强后的空间语义特征与动态运动特征。

随后，利用归一化后的猪个体目标矩形框，在两个分支的各时间步特征图上分别进行ROI Align，以提取目标猪个体对应的局部特征，并作为后续空间增强与时序建模的输入。

2.2 时空特征增强网络

在时空感知网络提取关键空间特征与关键运动变化特征之后，本文进一步构建了时空特征增强网络，用于对行为特征进行空间增强与时间建模。如图4所示，该网络由低帧率空间增强分支与高帧率时序建模分支组成。

其中，低帧率分支引入通道级空间注意力模块（Channel-Level Spatial Attention Module，CL-SAM），用于增强行为相关的空间语义信息；高帧率分支则利用LSTM对行为变化过程中的时间依赖关系进行建模，最终融合两条分支特征完成行为分类。

2.2.1 低帧率空间增强分支

在低帧率分支中，传统卷积特征图中不同通道对行为分类的重要性存在明显差异，部分通道包含背景与冗余信息，容易削弱关键行为区域的表达能力。因此，本文在ROI Align提取得到的个体猪空间特征基础上，引入CL-SAM对特征进行通道级增强。

对于输入特征图$\boldsymbol{F} \in \mathbf{R}^{T\times C\times H\times W}$，首先利用Sigmoid函数生成特征注意力权重，随后将此权重与原始特征逐元素相乘得到增强特征$\boldsymbol{F}'$，具体流程见公式：

$$\boldsymbol{F}' = \sigma(\boldsymbol{F}) \odot \boldsymbol{F} \tag{7}$$

式中，$\sigma(\cdot)$表示Sigmoid激活函数，其输出范围为(0, 1)，用于生成输入特征对应的注意力权重；$\odot$表示逐元素乘法。通过该方式，网络能够增强行为相关特征响应，同时抑制无关背景信息干扰。

2.2.2 高帧率时序建模分支

在高帧率分支中，仅依赖空间特征难以有效区分“起身”与“躺卧”等具有相似空间结构但时间变化方向相反的行为。因此，本文利用LSTM对行为变化过程进行时序建模，建模过程如图7所示。

首先，为了聚焦于每帧中对行为变化最敏感的区域，我们对特征图进行最大池化操作，得到一组关键帧序列表示$\boldsymbol{X} = \left[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\right]$，式中$\boldsymbol{x}_t \in \mathbf{R}^{256}$表示第$t$帧对应的运动特征向量，$T$表示时间序列长度。随后，将时间序列输入LSTM进行动态建模，具体计算流程见下述公式：

$$\boldsymbol{h}_t = \mathrm{LSTM}\left(\boldsymbol{x}_t, \left(\boldsymbol{h}_{t-1}, \boldsymbol{c}_{t-1}\right)\right) \tag{8}$$

式中，$\boldsymbol{h}_t$与$\boldsymbol{c}_t$分别表示隐藏状态与记忆单元状态。通过LSTM，网络能够学习猪行为在时间维度上的动态变化规律，从而增强模型对时序敏感行为的表达

© 中国图象图形学报版权所有

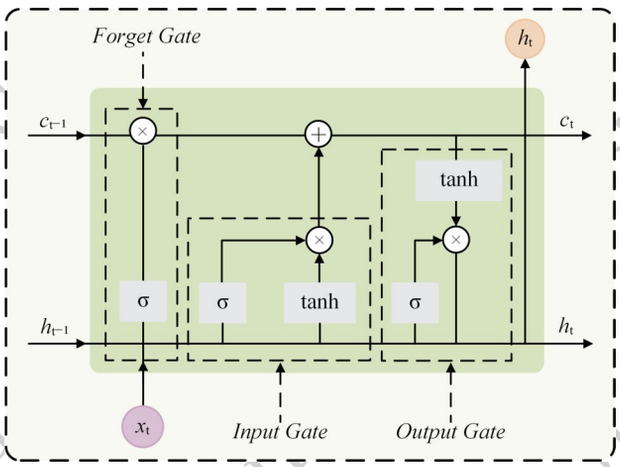


图 7　利用LSTM进行时序特征提取

Fig. 7　LSTM-based temporal feature extraction

能力。

最终，将低帧率分支得到的空间增强特征与高帧率分支得到的时序运动特征进行拼接融合，并输入全连接层完成行为分类。通过空间增强与时间建模的联合优化，所提出的时空特征增强网络能够有效提升模型对复杂猪行为的时空表达能力。

2.3　评价指标

本研究采用平均精度（average precision，AP）和均值平均精度（mean average precision，mAP）作为主要评价指标，用于衡量模型对不同猪行为类别的识别性能。对于每一行为类别，根据模型输出的类别预测置信度计算Precision-Recall曲线，并据此获得该类别的AP；mAP为全部行为类别AP的平均值，用于评价模型在13类猪行为上的整体识别性能。对于3.1节中的端到端行为检测方法，其AP、mAP沿用检测任务的评价定义，其中样本匹配同时考虑目标定位与行为类别预测结果。

3　验证与评估

3.1　数据集有效性与挑战性验证

为验证PBVD-13对不同类型猪行为识别方法的适用性，并评估数据集中复杂行为的识别难度，本文选取两类具有代表性的技术路线进行基准验证。其中，ST_TransNeXt（Hao等，2025）和基于MobileNetV2-Autoformer的猪攻击行为识别模型（Pu等，2025）主要利用视频时空特征进行行为类别识别；YOLOv8-EAPNet（Liu等，2025）和DM-GD-YOLO（Mao等，2024）则采用端到端检测框架，在完成猪只目标定位的同时进行行为类别预测。通过不同技术路线在PBVD-13上的实验结果，分析该数据集对不同方法的适用性以及各类猪行为的识别难度。其中，视频行为识别方法的AP、mAP反映行为分类性能，而端到端检测方法的评价结果同时受到目标定位与行为类别预测的影响，因此本节主要用于分析PBVD-13对不同技术路线的适用性及类别识别难度。

如表3所示，各方法在PBVD-13上的整体mAP分布在61.46%至68.63%，表明该数据集具有较好的可学习性。从类别结果来看，“进食”“站立”等行为的AP普遍超过90%，而“争斗”“鼻触碰”“躺卧”等复杂行为的识别性能明显降低，部分结果不足40%。这种明显的类别性能梯度表明，PBVD-13在能够被现有方法有效学习的同时，仍保留了对局部交互和动态状态转换行为的较高区分难度，具有一定的任务挑战性。

表3　不同代表性方法在PBVD-13数据集上的基准结果

Table 3　Benchmark results of representative methods on the PBVD-13 dataset

行为类别	视频行为识别AP(%)		端到端行为检测AP(%)	
	ST_TransNeXt	MobileNetV2-Autoformer	YOLOv8-EAPNet	DM-GD-YOLO
饮水	85.10	88.41	91.70	91.51
进食	95.37	93.15	97.21	98.57
卧	78.24	79.73	79.50	79.99
坐	79.00	83.55	81.71	85.53
站立	92.50	94.21	93.05	95.37
移动	63.30	88.55	82.43	83.64
行走	71.59	83.30	77.53	82.39
探索	59.71	67.17	66.75	68.71
与玩具互动	88.61	80.42	79.97	80.44
争斗	19.11	37.82	38.42	39.55
鼻触碰	20.71	29.31	19.30	20.19
起身	27.37	48.10	29.91	29.00
躺卧	18.31	18.53	17.28	17.85
分类mAP	61.46	**68.63**	—	—
检测mAP	—	—	65.70	**67.13**

© 中国图象图形学报版权所有

综合来看，PBVD-13能够支持不同技术路线的猪行为分析方法开展训练与评价，同时多种方法在复杂细粒度行为上均表现出一定的识别困难，说明该数据集既具有较好的可学习性，也能够对局部交互和动态状态转换等复杂行为表现出较高的识别挑战，可为复杂群养环境下猪行为分析研究提供实验基准。

3.2 对比实验

为验证本文提出时空感知与增强网络的有效性，本节选取ACTION-Net（Wang等，2021）、TSM（Lin等，2019）、I3D（Carreira等，2017）以及SlowFast（Feichtenhofer等，2019）作为对比模型。在统一的行为分类设置下，各方法采用相同的视频级训练、验证和测试集划分，训练与测试阶段均使用相同的人工标注猪个体目标框进行个体区域特征提取，以保证不同方法在一致实验条件下进行比较。

如表4所示，本文方法在13类猪行为识别任务上取得了75.92%的mAP，较最优对比模型SlowFast提高9.41个百分点。其性能优势在时序依赖较强的复杂行为上更加明显，其中争斗、起身和躺卧的AP分别达到60.75%、52.37%和41.98%，较对应类别的最佳对比结果分别提高23.42、14.01和21.36个百分点。结果表明，本文方法能够更有效地表征复杂行为中的局部动态变化与时序特征。

为进一步评价模型的复杂度与实时性，表5给出了各方法的参数量（Params）、运算量（GFLOPs）和推理速度（FPS）。本文模型的参数量为34.67 M，与SlowFast的33.95 M较为接近；受额外时空特征建模模块影响，其计算量增加至70.95 GFLOPs，推理速度为23.40 FPS，但mAP达到75.92%，较SlowFast提高9.41个百分点。结果表明，本文模型在增加一定计算开销的同时获得了更明显的识别性能提升，体现了较好的精度-效率权衡。

3.3 外部数据集验证

为进一步验证所提方法在不同猪行为数据集上的适用性，本文在HotPig（Bonneau de Beaufort等，2025）独立数据集上开展额外验证。实验中除根据HotPig的6类行为将最终分类层输出维度相应调整外，其余网络结构及各功能模块均与PBVD-13实验保持一致，未针对HotPig中的具体行为类别进行额外的针对性参数优化。

实验结果表明（表6），本文方法在HotPig上的mAP达到76.64%。结果说明，所提方法的有效性并非仅局限于PBVD-13，在不同数据来源和行为类别设置下仍具有较好的识别性能。

表4 不同模型在PBVD-13数据集上的行为识别性能对比

Table 4 Performance comparison of different models for pig behaviour recognition on the PBVD-13 dataset

Behaviour	ACTION-Net	TSM	I3D	SlowFast	Ours
饮水	71.50%	96.7%	70.4%	81.28%	**81.63%**
进食	**98.60%**	59.52%	96.57%	95.06%	95.75%
卧	77.21%	77.15%	91.71%	74.37%	**96.67%**
坐	79.37%	81.71%	89.03%	82.45%	**94.23%**
站立	92.25%	**93.06%**	92.63%	92.24%	92.29%
移动	73.65%	80.41%	80.42%	81.75%	**82.77%**
行走	77.33%	78.76%	81.26%	82.28%	**90.13%**
探索	62.90%	64.51%	61.29%	66.28%	**72.48%**
与玩具互动	**86.70%**	78.93%	19.43%	77.07%	85.33%
争斗	29.11%	30.22%	37.33%	36.92%	**60.75%**
鼻触碰	18.71%	11.92%	22.08%	36.53%	**40.00%**
起身	25.34%	25.18%	33.52%	38.36%	**52.37%**
躺卧	17.21%	14.72%	20.62%	19.54%	**41.98%**
mAP	62.30%	60.98%	61.24%	66.51%	**75.92%**

表5 不同模型的复杂度与推理效率对比

Table 5 Comparison of model complexity and inference efficiency

Method	Params(M)	GFLOPs	FPS	mAP (%)
I3D	27.25	28.41	70.37	61.24
TSM	23.53	32.70	73.97	60.98
ACTION-Net	27.75	34.43	29.54	62.30
SlowFast	33.95	50.38	30.66	66.51
Ours	34.67	70.95	23.40	75.92

3.4 消融实验

3.4.1 时空模块有效性分析

本次实验分别对时空感知网络中的FL-SAM和KMFEM模块，以及时空特征增强网络中的CL-SAM与LSTM模块进行消融实验，并与完整模型进行对比分析。实验采用控制变量法，除输入方案外，其余超参数均保持一致，实验结果如表7所示。

消融实验结果表明，各模块对模型性能提升均具有重要作用。移除时空感知网络中的FL-SAM模

© 中国图象图形学报版权所有

表6　所提方法在HotPig外部数据集上的验证结果

Table 6　Validation results of the proposed method on the HotPig dataset

数据集	行为类别	AP(%)
HotPig	drink	77.47
	eat	84.79
	lying	93.17
	move	75.13
	social	34.31
	stand	94.95
	mAP	**76.64**

表7　不同模块的消融实验结果

Table 7　Ablation study of different modules

Behaviour	Ours	No FL-SAM	No KMFEM	No CL-SAM	No LSTM
饮水	81.63%	77.42%	77.55%	76.37%	**84.45%**
进食	95.75%	96.31%	**97.15%**	96.76%	95.97%
卧	**96.67%**	94.49%	94.85%	93.83%	95.41%
坐	**94.23%**	92.15%	92.59%	90.82%	90.61%
站立	92.29%	**94.93%**	94.69%	94.02%	94.79%
移动	**82.77%**	80.14%	81.31%	80.18%	80.94%
行走	**90.13%**	86.37%	81.44%	83.39%	84.02%
探索	72.48%	65.34%	**75.54%**	69.47%	71.83%
与玩具互动	85.33%	83.54%	82.55%	86.20%	**96.67%**
争斗	60.75%	**64.53%**	56.79%	44.84%	50.25%
鼻触碰	**40%**	33.11%	36.45%	33.43%	28.63%
起身	52.37%	44.99%	49.67%	**55.20%**	39.37%
躺卧	**41.98%**	40.30%	25.92%	36.42%	23.29%
mAP	**75.92%**	73.36%	72.81%	72.38%	72.02%

块后，模型整体mAP由75.92%降至73.36%。其中，“探索”“饮水”与“鼻触碰”等行为下降较为明显，分别下降至65.34%、77.42%和33.11%。这是由于上述行为更加依赖头部与局部交互区域的空间特征，而FL-SAM能够增强模型对关键局部区域的关注，从而提高低层空间特征的表达能力。

移除KMFEM模块后，模型mAP下降至72.81%。其中，“行走”、“躺卧”等依赖动态变化特征的行为下降较为明显，分别降至81.44%和25.92%。说明KMFEM能够通过帧间差分增强关键区域的运动变化响应，从而提升模型对短时动态特征的建模能力。

移除时空特征增强网络中的CL-SAM模块后，模型mAP下降至72.38%。其中，“争斗”“行走”与“饮水”等行为下降较为明显，尤其“争斗”行为由60.75%降至44.84%。这是由于CL-SAM能够进一步强化高层语义空间特征，增强模型对行为主体及其环境关系的表达能力，从而提升复杂行为的判别能力。

移除LSTM后，模型mAP下降至72.02%。其中，“起身”与“躺卧”行为下降最为明显，分别降至39.37%和23.29%。这是由于上述行为具有明显的时序依赖特征，单帧空间信息难以完整描述其动作演化过程，而LSTM能够通过建模长时间序列信息，有效提升模型对动态行为变化的理解能力。

3.4.2　模块协同作用分析

最后，本文进一步分析了不同模块组合对模型性能的影响，实验结果如表8所示。结果表明，包含LSTM的组合整体性能更优，说明时序建模对行为识别具有重要作用。其中，KMFEM与LSTM的组合取得了较高的mAP，表明短时运动特征与长时序依赖之间具有良好的互补性。

表8　时空模块组合实验结果

Table 8　Experimental results of spatio-temporal module combinations

FL-SAM	KMFEM	CL-SAM	LSTM	mAP
✓	✓			70.11%
✓		✓		69.54%
✓			✓	72.19%
	✓	✓		69.7%
	✓		✓	71.87%

此外，包含FL-SAM的组合整体表现优于未使用FL-SAM的组合结果，说明低层关键区域增强能够有效提升空间特征表达能力。实验还发现，在部分组合中，即使未引入CL-SAM，模型仍能够保持较高性能。这可能是由于ROI操作后，随着网络深度增加，主体特征中已隐式包含部分环境上下文信息，从而在一定程度上缓解了高层空间注意力缺失带来的影响。总体来看，各模块之间形成了有效的时空互补关系，共同提升了模型性能。

© 中国图象图形学报版权所有

3.5 特征可视化分析

为进一步验证时空感知网络对关键行为区域的特征提取能力，本文对FL-SAM与KMFEM输出的特征图进行了可视化分析，如图8所示。

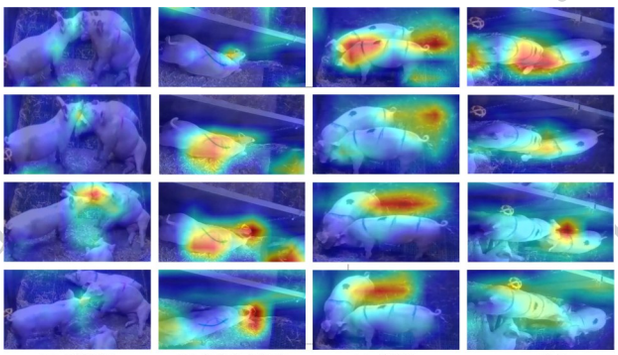


图8 FL-SAM空间响应与KMFEM运动响应可视化结果

Fig. 8 Visualization results of spatial responses in FL-SAM and motion responses in KMFEM ((a) nose touching; (b) toy interaction; (c) motion; (d) fighting)

图8(a)与(b)分别展示了FL-SAM在“鼻触碰”与“与玩具互动”行为中的空间响应结果，可以观察到模型能够有效聚焦于猪头部及交互区域。图8(c)与(d)展示了KMFEM在“移动”与“争斗”行为中的运动响应结果，模型能够突出行为相关的动态变化区域，同时有效抑制背景噪声干扰。

为进一步分析空间注意力机制对高层语义特征的增强效果，本文对CL-SAM输出的通道响应进行了可视化分析，如图9所示。图中展示了经空间注意力残差网络提取的256维特征图中，激活均值大于1e-1的88个通道的可视化结果。

结果表明，经过空间注意力增强后，仅少量通道对行为相关区域产生较强响应，说明CL-SAM能够有效抑制冗余背景信息，并增强关键行为区域的语义表达能力。

为验证LSTM对行为时序变化的建模能力，本文进一步对“躺卧”与“起身”行为的时序特征进行了可视化分析，如图10所示。

结果表明，LSTM能够有效捕获行为变化过程中的动态时序信息，对具有明显状态转换特征的行为具有较好的时序表达能力。

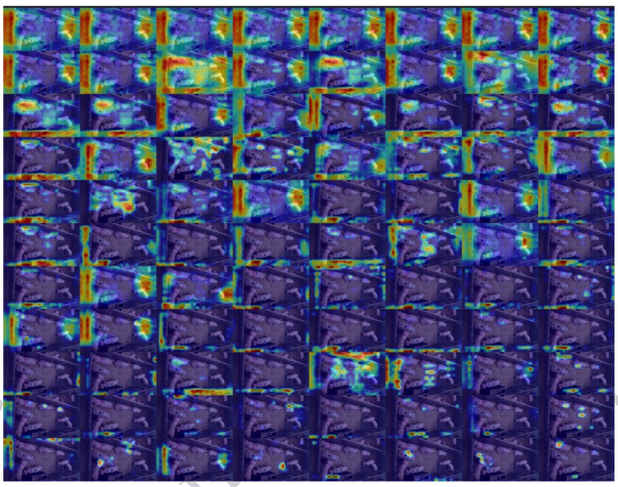


图9 CL-SAM通道响应可视化结果

Fig. 9 Visualization results of channel responses in CL-SAM

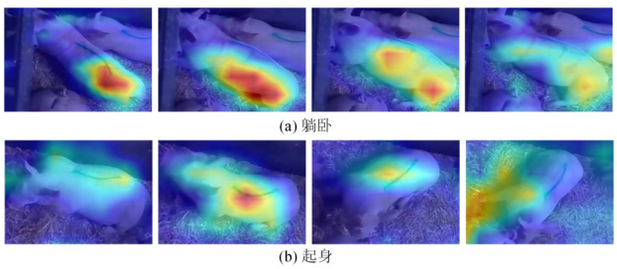


图10 LSTM时序特征响应可视化结果

Fig. 10 Visualization results of temporal feature responses in LSTM ((a) Lie Down; (b) Stand Up)

4 结 论

本文面向复杂群养环境下细粒度猪行为识别与福利相关行为分析需求，构建了包含13类猪行为和13500个行为实例的PBVD-13视频数据集。在保留饮水、进食、卧、移动和站立等基础行为的基础上，进一步扩展了坐、行走、起身、躺卧等细粒度姿态与运动行为，以及探索、与玩具互动、争斗和鼻触碰4类福利相关行为，从而提升了数据集在行为类别、标注粒度及福利相关行为覆盖方面的丰富性。

实验结果表明，PBVD-13具有较好的可学习性，同时在争斗、鼻触碰、起身和躺卧等复杂细粒度行为上具有较高的识别挑战性，可为复杂群养环境下多类别猪行为识别提供有效的实验基准。在此基础上，本文提出的时空感知与增强网络在PBVD-13上取得75.92%的mAP，并在独立HotPig数据集上取得76.64%的mAP，进一步验证了所提方法在不同

© 中国图象图形学报版权所有

猪行为数据集上的适用性。未来将进一步扩展不同养殖场景下的数据规模和行为类别,为猪行为识别与动物福利分析提供更加丰富的数据基础。

参考文献(References)

Amezcua R, Walsh S, Luimes P H and Friendship R M. 2014. Infrared thermography to evaluate lameness in pregnant sows. Canadian Veterinary Journal , 55(3): 268-272

Berckmans D. 2017. General introduction to precision livestock farming. Animal Frontiers , 7(1): 6-11 [DOI:10.2527/af.2017.0102]

Bergamini L, Pini S, Simoni A, Vezzani R, Calderara S, D'Eath R B, et al. 2021. Extracting accurate long-term behavior changes from a large pig dataset// Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Volume 5. SCITEPRESS: 524-533[DOI: 10.5220/0010288405240533]

Bhujel A, Wang Y, Lu Y, Morris D and Dangol M. 2025. A systematic survey of public computer vision datasets for precision livestock farming. Computers and Electronics in Agriculture, 229: 109718 [DOI:10.1016/j.compag.2024.109718]

Bonneau de Beaufort L, Xavier C, Renaudeau D, Largouët C and Gondret F. 2026. Data Paper: HotPig, a behavioural dataset of pigs under heat stress. Animal - Open Space, 5: 100112 [DOI: 10.1016/j.anopes.2025.100112]

Carreira J and Zisserman A. 2017. Quo vadis, action recognition? A new model and the kinetics dataset//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE: 6292-6303 [DOI: 10.1109/CVPR.2017.502]

Chen C, Zhu W X, Steibel J, Siegford J, Wurtz K, Han J J, et al. 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. Computers and Electronics in Agriculture, 169: 105166 [DOI:10.1016/j.compag.2019.105166]

Chen H X, Huang Z Z, Hong Y, Wang Y S, Lyu Z C, Xu Z E, et al. 2025. Efficient transfer learning for video-language foundation models//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE: 29129-29138 [DOI:10.1109/CVPR52734.2025.02712]

Cheng Y, Gao Y Y, Wang J, Yang L, Xu X L, Cheng Y, et al. 2025. Combining dilated convolution and multiscale fusion temporal action detection. Journal of Image and Graphics, 30(02): 0406-0420 (程勇, 高园元, 王军, 杨玲, 许小龙, 程遥, 等. 2025. 结合扩张卷积与多尺度融合的实时时空动作检测. 中国图象图形学报, 30(02): 0406-0420) [DOI:10.11834/jig.240098]

Collarini E, Capponcelli L, Pierdomenico A, Cordoni G and Norscia I. 2024. Balancing welfare and habitat damage in pigs (Sus scrofa) under extensive farming: an ethological approach for determining the effects of group composition and environmental enrichment. Frontiers in Ethology, 3: 1450585 [DOI: 10.3389/fetho.2024.1450585]

Du H T, Zhao F, Liu H Q and Tang Y. 2026. Dense-Sparse Skeleton Joint Representation-Guided RGB Image ROI Localization for Multimodal Action Recognition [J/OL]. Journal of Image and Graphics, 1-16 (杜海涛, 赵凤, 刘汉强, 唐垚. 2026. 稠密-稀疏骨骼联合表征引导RGB图像ROI定位的多模态行为识别 [J/OL]. 中国图象图形学报, 1-16 [DOI:10.11834/jig.260177]

Dutta A and Zisserman A. 2019. The VIA annotation software for images, audio and video// Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: Association for Computing Machinery: 2276-2279 [DOI: 10.1145/3343031.3350535]

Feichtenhofer C, Fan H, Malik J and He K. 2019. SlowFast networks for video recognition//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE: 5907-5915[DOI:10.1109/ICCV.2019.00599]

Gao Y, Yan K, Dai B, Sun H, Yin Y, Liu R, et al. 2023. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatio-temporal attention mechanism. Computers and Electronics in Agriculture, 205: 107606 [DOI:10.1016/j.compag.2022.107606]

Ge W X, Qu H Y, Yan R, Xie G S, Yao Y Z, Shu X, et al. 2026. Condensed test-time adaptation of VLMs for action recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Denver, CO, USA: IEEE: 16977-16987

Godyń D, Nowicki J and Herbut P. 2019. Effects of environmental enrichment on pig welfare—A review. Animals, 9(6): 383 [DOI: 10.3390/ani9060383]

Gong P and Luo X D. 2025. A survey of video action recognition based on deep learning. Knowledge-Based Systems, 320: 113594 [DOI: 10.1016/j.knosys.2025.113594]

Gu C H, Sun C, Ross D A, Vondrick C, Pantofaru C, Li Y Q, et al. 2018. AVA: A video dataset of spatio-temporally localized atomic visual actions//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 6047-6056 [DOI: 10.1109/CVPR.2018.00633]

Guo J J, Xu Y D, Zhang B B, Zhou P, Luo S W, Zhuo Y H, et al. 2026. A survey on deep learning-based livestock and poultry behavior recognition: Methods, dataset and prospects. Engineering Applications of Artificial Intelligence, 181(4): 115608 [DOI: 10.1016/j.engappai.2026.115608]

Hao W, Shu H, Hu X, Han M and Li F. 2025. ST_TransNeXt: A novel pig behavior recognition model//Advances in Brain Inspired Cognitive Systems. BICS 2024. Lecture Notes in Computer Science, vol 15498. Singapore: Springer: 23 [DOI: 10.1007/978-981-96-2885-

© 中国图象图形学报版权所有

8_23］

Hao W，Zhang K，Zhang L，Han M，Hao W，Li F，et al. 2023. TSML：A new pig behavior recognition method based on two-stream mutual learning network. Sensors，23（11）：5092［DOI：10.3390/s23115092］

Ji H，Teng G，Yu J，Wen Y，Deng H and Zhuang Y. 2023. Efficient aggressive behavior recognition of pigs based on temporal shift module. Animals，13（13）：2078［DOI：10.3390/ani13132078］

Li B，Xu W J，Chen T M，Cheng J H and Shen M X. 2023. Recognition of fine-grained sow nursing behavior based on the SlowFast and hidden Markov models. Computers and Electronics in Agriculture，210：107938［DOI：10.1016/j.compag.2023.107938］

Li D，Zhang K，Li Z and Chen Y. 2020. A spatiotemporal convolutional network for multi-behavior recognition of pigs. Sensors，20（8）：2381［DOI：10.3390/s20082381］

Lin J，Gan C and Han S. 2019. TSM：Temporal shift module for efficient video understanding//Proceedings of the IEEE/CVF International Conference on Computer Vision（ICCV）. Seoul，South Korea：IEEE：2083-2091［DOI：10.1109/ICCV.2019.00217］

Liu F，Wu W J，Liu X L，Wang X R，Fang Y P，Li G L，et al. 2023. Progress of computer vision and deep learning methods for pig's identity and behavior recognition. Journal of Huazhong Agricultural University，42（3）：47-56（刘峰，吴文杰，刘小磊，王欣然，方亚平，李国亮，等. 2023. 计算机视觉与深度学习在猪只识别中的研究进展. 华中农业大学学报，42（3）：47-56）［DOI：10.13300/j.cnki.hnlkxb.2023.03.006］

Liu J，Yan Y Q，Yang Y S，Hao Y H，Chen B F，Yang M K，et al. 2025. Pig recognition based on YOLOv8-EAPNet. INMATEH-Agricultural Engineering，77（3）：676-688［DOI：10.35633/inmateh-77-55］

Liu S M，Zhao C，Zohra F，Soldan M，Pardo A，Xu M M，et al. 2025. OpenTAD：a unified framework and comprehensive study of temporal action detection //2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. IEEE：2616-2626［DOI：10.1109/CVPRW67362.2025.00247］

Mao R，Shen D Z，Wang R Q，Cui Y M，Hu Y F，Li M，et al. 2024. An integrated gather-and-distribute mechanism and attention-enhanced deformable convolution model for pig behavior recognition. Animals，14（9）：1316［DOI：10.3390/ani14091316］

Pu Y，Zhao Y Q，Ma H and Wang J X. 2025. A lightweight pig aggressive behavior recognition model by effective integration of spatio-temporal features. Animals，15（8）：1159［DOI：10.3390/ani15081159］

Rohan A，Rafaq M S，Hasan M J，Asghar F，Dottorini T and Bashir A K. 2024. Application of deep learning for livestock behaviour recognition：a systematic literature review. Computers and Electronics in Agriculture，224：109115［DOI：10.1016/j. compag. 2024. 109115］.

Scolari S C，Clark S G，Knox R V and Tamassia M A. 2011. Vulvar skin temperature changes significantly during estrus in swine as determined by digital infrared thermography. Journal of Swine Health and Production，19（3）：151-155［DOI：10.54846/jshap/685］

She S S and Feng P. 2020. Research progress on welfare improvement of intensive pig breeding. Journal of Yulin University，30（6）：44-48（折帅帅，冯平. 2020. 猪集约化养殖的福利改善研究进展. 榆林学院学报，30（6）：44-48）［DOI：10.16752/ki. jylu. 2020. 06.009］

Smith K C and Pierdon M K. 2024. Utilization of enrichment objects by growing pigs in a commercial facility and the impact on behavior and skin lesions. Applied Animal Behaviour Science，272：106181［DOI：10.1016/j.applanim.2024.106181］

Viazzi S，Ismayilova G，Oczak M，Sonoda L T，Fels M and Guarino M，et al. 2014. Image feature extraction for classification of aggressive interactions among pigs. Computers and Electronics in Agriculture，104：57-62［DOI：10.1016/j.compag.2014.03.010］

Wang M M，Huang Z Y，Kong X J，Shen G J，Dai G and Wang J D，et al. 2025. Action Detail Matters：Refining video recognition with local action queries//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. IEEE：19132-19142［DOI：10.1109/CVPR52734.2025.01782］

Wang Z W，She Q and Smolic A. 2021. ACTION-Net：Multipath excitation for action recognition//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. IEEE：13209-13218［DOI：10.1109/CVPR46437.2021.01301］

Wu Z Z，Chen X，Xu T，Nian F D，Wang X F and Li T. 2025. Dynamic multi-granularity graph convolutional networks for skeleton-based action recognition. Journal of Image and Graphics，30（8）：2822-2834（吴志泽，陈鑫，徐童，年福东，王晓峰，李腾. 2025. 基于动态多粒度图卷积网络的人体骨架行为识别. 中国图象图形学报，30（8）：2822-2834）［DOI：10.11834/jig.240352］

Xu J Y，Ying Y B，Wu D H，Hu Y L and Cui D. 2025. Recent advances in pig behavior detection based on information perception technology. Computers and Electronics in Agriculture，235：110327［DOI：10.1016/j.compag.2025.110327］

Yadav T，Salehi M，Park J S，Ramanujan V，Hajishirzi H and Choi Y，et al. 2026. VideoNet：A large-scale dataset for domain-specific action recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. IEEE：12881-12891

Yang Q M and Xiao D Q. 2020. A review of video-based pig behavior recognition. Applied Animal Behaviour Science，233：105146［DOI：10.1016/j.applanim.2020.105146］

Zhang N N. 2025. Pig Behavior Recognition Dataset［Dataset］. Mendeley Data，V1.［DOI：10.17632/y5b89bsv69.1］

作者简介

王敏，女，硕士研究生，研究方向为计算机视觉，深度学习。
E-mail：S24150812044@smail.cczu.edu.cn

© 中国图象图形学报版权所有

侯振杰,通信作者,男,教授,主要研究方向为计算机视觉,深度学习。E-mail:houzj@cczu.edu.cn
漆方正,男,硕士研究生,主要研究方向为计算机视觉,深度学习。E-mail: qifangzheng777@gmail.com
张文广,男,教授,主要研究方向家畜遗传育种与数量基因组学。E-mail: atcgnmbi@aliyun.com

© 中国图象图形学报版权所有