吴振邦, 陈泽锴, 田绪红, 等. 基于 3D 卷积视频分析的猪步态评分方法 [J]. 华南农业大学学报, 2024, 45(5): 743-753.
WU Zhenbang, CHEN Zekai, TIAN Xuhong, et al. A method for pig gait scoring based on 3D convolution video analysis[J]. Journal of South China Agricultural University, 2024, 45(5): 743-753.

基于 3D 卷积视频分析的猪步态评分方法

吴振邦[1]，陈泽锴[1]，田绪红[1]，杨　杰[2,3,4]，尹　令[1,2,4]，张素敏[1,4]
(1 华南农业大学 数学与信息学院, 广东 广州 510642； 2 国家生猪种业工程技术研究中心, 广东 广州 510642；
3 华南农业大学 动物科学学院, 广东 广州 510642； 4 猪禽种业全国重点实验室, 广东 广州 510640)

摘要:【目的】猪肢蹄病是种猪淘汰的重要原因之一，给养殖场带来巨大的经济损失。猪蹄疾病判断通常依赖人工肉眼观察猪只步态进行排查，存在效率低、人力成本高等问题。本文旨在实现自动化猪步态评分，高效判断猪只肢蹄健康状况。【方法】本文提出一种“端到端”的猪步态评分方法，在单头种猪经过测定通道时采集视频，并制作四分制步态数据集。采用深度学习技术分析视频，设计了一种基于 3D 卷积网络的时间注意力模块 (Time attention module，TAM)，有效提取视频帧图像之间的特征信息。将 TAM 与残差结构结合，构建猪步态评分模型 TA3D，对步态视频进行特征提取与步态分类评分。为进一步提升模型性能并实现自动化处理，本文设计了步态关注模块 (Gait focus module，GFM)，能够自动从实时视频流中提取有效信息并合成高质量步态视频，在提高模型性能的同时降低计算成本。【结果】试验结果表明，GFM 可以实时运行，步态视频大小可以减少 90% 以上，显著降低存储成本，TA3D 模型步态评分准确率达到 96.43%。与其他经典的视频分析模型的对比测试结果表明，TA3D 的准确率和推理速度均达到最佳水平。【结论】本文提出的方案可应用于猪只步态自动评分，为猪肢蹄病的自动检测提供参考。

关键词: 图像处理; 深度学习; 猪; 步态评分; 注意力机制; 视频分析; 肢蹄病
中图分类号: TP391.4; S828　　**文献标志码:** A　　**文章编号:** 1001-411X(2024)05-0743-11

A method for pig gait scoring based on 3D convolution video analysis

WU Zhenbang[1], CHEN Zekai[1], TIAN Xuhong[1], YANG Jie[2,3,4], YIN Ling[1,2,4], ZHANG Sumin[1,4]
(1 College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China; 2 National Engineering Research Center for Swine Breeding Industry, Guangzhou 510642, China; 3 College of Animal Science, South China Agricultural University, Guangzhou 510642, China; 4 State Key Laboratory of Swine and Poultry Breeding Industry, Guangzhou 510640, China)

Abstract: 【Objective】Swine limb and hoof disease is one of the significant reasons for culling breeding swine, resulting in substantial economic losses for livestock farms. The diagnosis of swine limb and hoof disease typically relies on manual observation of pig gaits, which consumes high labor costs and has low efficiency. The aim of this study is to achieve automated pig gait scoring, and efficiently determine the health status of swine limb and hoof.【Method】This study proposed an “end-to-end” pig gait scoring method. Videos of individual

收稿日期:2023–11–18　　**网络首发时间:**2024–07–15 12:00:43
首发网址:https://link.cnki.net/urlid/44.1110.S.20240711.1357.002
作者简介:吴振邦，硕士研究生，主要从事深度学习和计算机视觉研究，E-mail: wzb6398@stu.scau.edu.cn；通信作者：尹　令，副教授，博士，主要从事智慧农业研究，E-mail: yin_ling@scau.edu.cn
基金项目:国家自然科学基金 (32172780)

breeding swine passing through designated channels were collected and a four-point gait dataset was created. Deep learning techniques were employed for video analysis. A time attention module (TAM) based on a 3D convolutional neural network was designed to effectively extract feature information between video frame images. By combining TAM with residual structures, the pig gait scoring model TA3D was constructed for feature extraction and gait classification scoring in the gait videos. To further improve model performance and achieve automation, the gait focus module (GFM) was designed. GFM could autonomously extract effective information from real-time video streams to synthesize high-quality gait videos, improving model performance while reducing computational costs. 【Result】The experimental results demonstrated that GFM could operate in real-time and reduced the size of gait videos by over 90%, significantly reducing storage cost, and the gait scoring accuracy of the TA3D model was 96.43%. Moreover, the comparison test results with other classic video analysis models showed that TA3D achieved optimal levels of accuracy and inference speed. 【Conclusion】This paper proposes a solution that can be applied to the automatic scoring of pig gait, providing a reference for the automatic detection of swine limb and hoof disease.

Key words: Image processing; Deep learning; Pig; Gait scoring; Attention mechanism; Video analysis; Limb and hoof disease

猪肢蹄病是由于疾病、营养、管理等因素导致的猪腿部疾病，不同因素导致的发病可能存在遗传和传染的风险[1-4]。根据调查[5]，一般患有肢蹄病的公猪种用年限仅为 28.9 个月，远低于因年龄淘汰的公猪种用时间 (50.7 个月)，且因肢蹄病被淘汰的公猪比例达到了 32.15%，是造成公猪淘汰的主要原因。猪肢蹄病会严重影响猪的健康和养殖效益，造成经济损失，故及时发现存在发病风险的猪并且进行干预十分必要。步态是能够直观反映猪肢蹄健康情况的关键因素，工作人员主要通过观察猪的步态来判断是否存在肢蹄疾病，这样评判存在较高的主观性且工作量较大。因此，开展猪步态分析研究对猪肢蹄病的早期检测具有重要意义。

四足牲畜的肢蹄疾病自动检测是牲畜饲养过程中的一项难题。近年来传感器技术及人工智能技术的快速发展给牲畜饲养智能化发展带来新的解决方案，其中深度学习技术已经在猪只健康监测中得到了广泛应用[6]。刘波等[7] 使用 Kinect 摄像机采集生猪运动深度图像序列，构建生猪前后肢端点运动模型并且提取步频特征，对异常步态检测具有重要意义。朱家骥等[8] 提出基于星状骨架模型的猪步态分析方法，从图像中提取猪轮廓，由轮廓及质心确定关键轮廓点，以此获得运动规律，并使用频谱分析法计算猪前肢步态频率。李前等[9] 总结了当前奶牛跛行自动识别的主要研究方法并进行分析比对。Zhao 等[10] 提出了一种分析奶牛行走过程中腿部运动的方法，通过分析奶牛腿的摆动来提取步态特征，证明对奶牛跛行退化进行量化是可行的。康熙等[11] 提出基于热红外视频的奶牛跛行运动特征获取与检测的方法，利用弓背曲率对奶牛进行跛行检测，达到 90% 的平均精度。Jiang 等[12] 利用 FLYOLOv3 算法构建奶牛背部位置提取模型，提取奶牛背部曲率数据。最后，使用噪声+双侧长短期记忆模型来预测曲率数据，并匹配奶牛跛行的弯曲特征，从而对奶牛跛行进行分类和检测。Poursaberi 等[13] 提出了一种基于背部姿态分析的奶牛跛行实时检测方法，他们使用图像处理方法获取奶牛背部轮廓并且自动提取出 3 个关键点，根据关键点计算奶牛背部曲率特征，由此评估奶牛跛行程度。

目前对牲畜异常步态的自动检测大部分采用图像处理技术进行图像特征提取和分类。这类方法需提前设定特征，采用图像处理或深度学习方法从图像中找出对应的特征部位，如背、蹄等，再分析统计特定部位的变化构成行为特征，如弓背、步态对称性等，最后使用分类算法如 SVM、KNN 等获得分类结果。而这类方法可能存在 2 个方面问题，一是将视频转化成多张连续图像的集合，专注于单张图像的特定部位特征提取，可能会忽略非特定部位以及视频帧图像之间的特征关系的影响；另一方面这类方法预处理繁琐，手工标注特征消耗时间，人力成本高。针对以上问题，本文提出一种“端到端”的解决方案，即采用基于深度学习的视频分析算法，使用 3D 卷积神经网络 (3D convolutional neural network，3D CNN)[14] 直接对步态视频自动进行特征提取以及自动分类评分。本文提出一种专注于提取步态视频时空特征的注意力模块，将其与

3D CNN 结合，构成基于 3D CNN 的猪步态评分模型 TA3D (Time-attention 3D convolutional network)。为了提高模型准确率与实现方案自动化，本文设计了一个步态关注模块 (Gait focus module，GFM) 用于从视频流中自动提取有效信息并合成高质量的步态视频。

1 试验方法

在当前规模化养猪的背景下，养殖场为了及时调整饲养策略，必须定期对每头生猪进行表型测定以获取其各项身体数据，这为本试验创造了良好的前置条件。本文在猪进入测定栏前布置一条测定通道和相关的拍摄设备，视频采集时尽量在无外力驱赶的前提下让猪只依次通过通道，采集系统自动捕获猪只行进视频并完成步态评分，若发现异常，则会及时发出警告。本文将方案主要分为 2 个部分——合成高质量步态视频和猪步态评分，技术路线如图 1 所示。在合成高质量步态视频部分中，本文设计 GFM 模块，其作用是将目标的步态变化从视频流中自动提取出来合成个体步态视频用于评分；在猪步态评分部分中，本文提出一种基于 3D 卷积的猪步态评分模型，用于对步态视频进行评分。

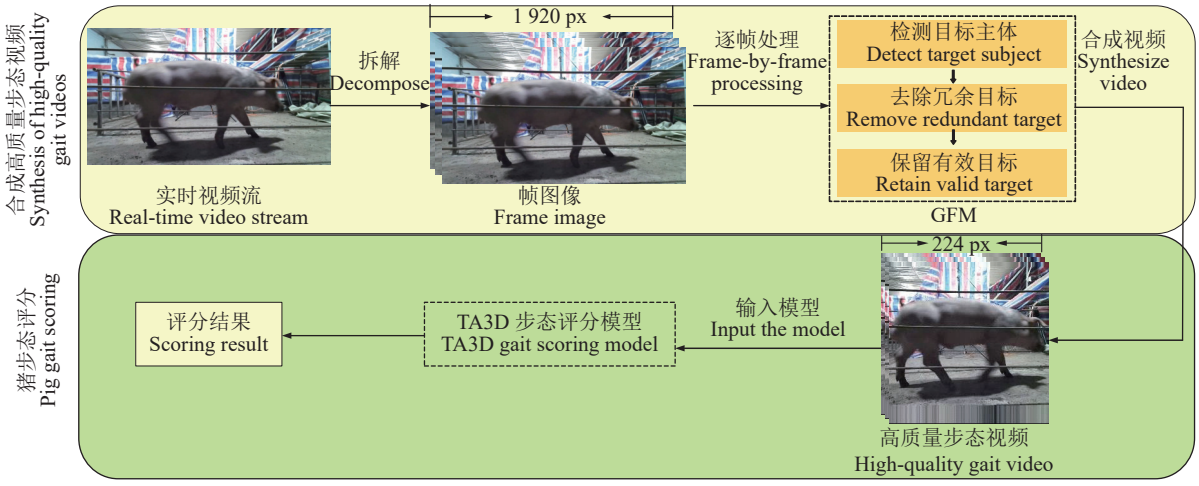


图 1 基于 3D 卷积视频分析的猪步态评分方法流程图

Fig. 1 Flowchart of pig gait scoring method based on 3D convolutional video analysis

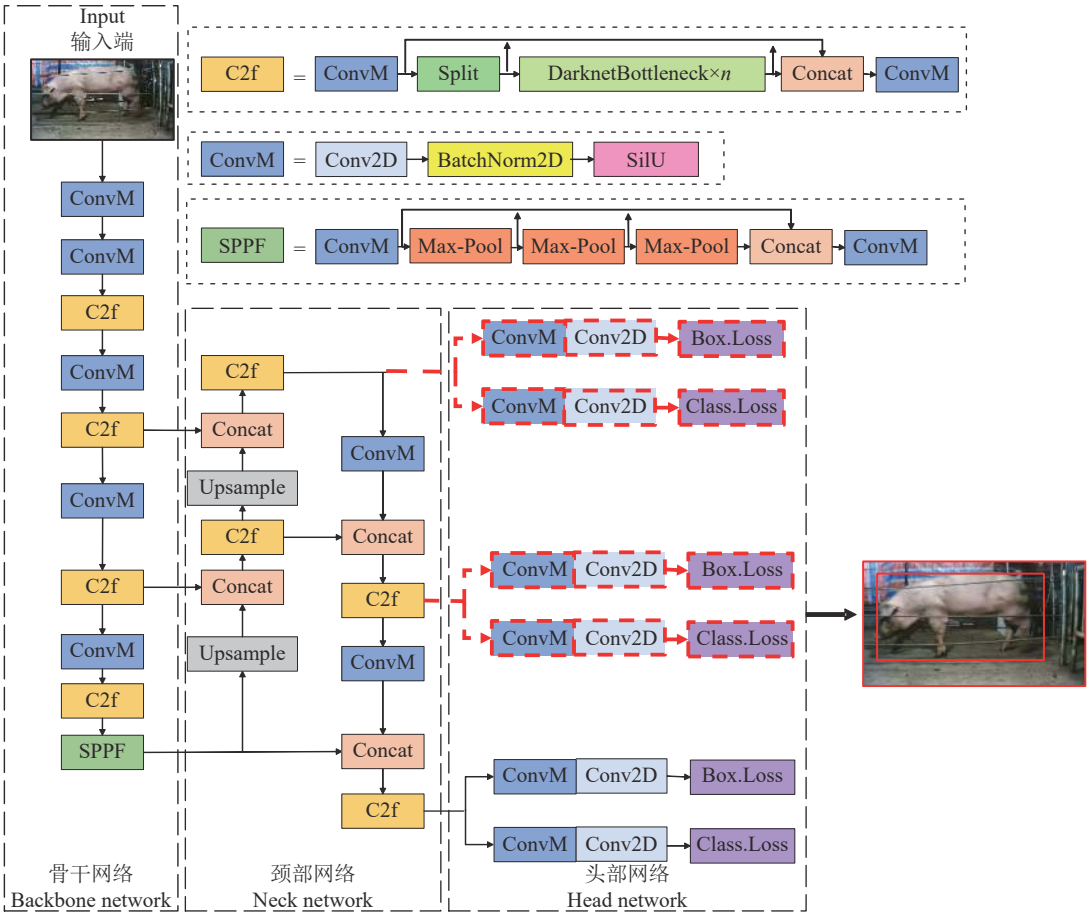
1.1 高质量猪步态视频合成

从实时视频流中采集个体步态视频，获取其步态信息，存在两大问题：一是由于猪的习性或肢蹄病等原因，它们的行走规律存在较大的不确定性，经常出现在通道内移动缓慢甚至停止不前的情况，这将导致源步态视频存在大量的冗余内容；二是采集到的源步态视频中包含大量无效的背景信息。这 2 个问题不仅对预测结果的准确性有影响，还会显著地增加计算成本。为了解决以上问题，本文设计了 GFM 模块，该模块通过以下 3 步来实现高质量步态视频自动提取与合成：检测目标主体、去除冗余目标和保留有效目标。通过这种方法，能够有效提高步态视频的质量和系统实用性。

1.1.1 检测目标主体 为保证试验方法客观性与便捷性，首先要自动识别出待检测目标主体，确定帧图像中目标的位置、大小信息。本方案选择的目标检测器为 YOLOv8，其改进于 YOLOv5，是目前 YOLO 算法系列最新成果，具备高实时性与高准确率等特点。为了满足不同场景的需求，YOLOv8 提供了多种体量的模型，考虑到本文的试验对象特征明显、易于检测，因此选择体量最小的模型 YOLOv8n 作为目标检测器。

YOLOv8 由骨干网络 (Backbone)、颈部网络 (Neck) 和头部网络 (Head)3 个部分组成。其中 Head 作用是根据 Neck 提供的特征信息来预测目标边界框和类别概率。为了检测出不同大小的目标，Head 被设计成拥有 3 种尺度的检测头，用于检测大、中、小 3 种尺度的目标。由于本文试验对象都属于大目标且不存遮挡现象，因此对 YOLOv8n 网络的 Head 进行改进，使其仅保留大目标检测头，在保证检测精度的前提下，尽可能减少计算量，其网络结构如图 2 所示。

1.1.2 去除冗余目标 无外力驱赶时，猪经过通道的速度与其肢蹄状态并不存在强关联关系。例如跛脚猪虽行动缓慢，但健康猪因低头觅食的习性导致行动缓慢的情况也很常见，这使得视频中存在高度重复的冗余帧图像。由于输入网络帧数是固定的，冗余的帧图像不仅会降低视频有效信息密度，还会



去除红色虚线框所示的 2 个检测头作为对 YOLOv8 模型的改进

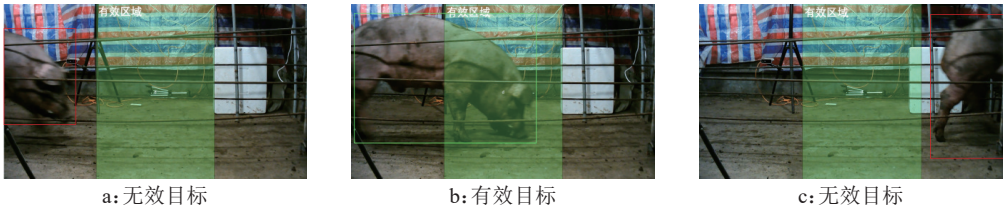
Remove the two detection heads indicated by the red dashed box to improve the YOLOv8 model

图 2　YOLOv8 网络结构及改进

Fig. 2　YOLOv8 network architecture and improvements

增加存储和计算成本。因此，本文采用相邻帧目标检测框的交并比 (Intersection over union，IoU) 计算前后两帧内目标的重合度，以此判断帧图像是否冗余。经过反复试验对比，阈值为 0.96 时的效果最好。当前后两帧目标检测框的 IoU 大于 0.96 时，后帧被判定为冗余帧图像，作丢弃处理，随后顺位计算下一帧，重新计算两帧的目标检测框的 IoU，以此类推，直至完成整个视频的检测。

1.1.3　*保留有效目标*　对视频进行逐帧检测时，包含目标的每一帧都会得到目标位置和大小信息。显然，每一帧获得的结果无论位置和大小都会不断变化，并且目标在进入画面和退出画面时只出现头部或臀部，这些都属于不包含步态信息的无效帧。若简单地将检测框统一大小后进行拼接将不仅会导致视频帧的剧烈抖动，而且还会因包含大量无效信息而影响结果和增加计算成本。因此需要给视频输入端添加一个有效区域，尽量降低由于目标大小变化造成的影响。本文将有效区域设为画面中间的三分之一，当目标与有效区域接触时才算是有效目标。判定有效目标的示意图如图 3 所示。当某帧图像被认定为非冗余且有效时，对其目标检测框进行裁切，以长边为长度裁切正方形范围，若超出图像



a：无效目标
a: Invalid target

b：有效目标
b: Valid target

c：无效目标
c: Invalid target

图 3　判定步态有效目标

Fig. 3　Determining valid gait target

范围则使用边缘像素补齐。然后将获得的图像调整为 224 像素×224 像素保存。最终，将所有图像顺序合并成高质量猪步态视频。GFM 合成高质量步态视频的示例图如图 4 所示。根据图 4可见，经过 GFM 处理后，目标主体信息被完整保留，而大部分背景信息被去除。在示例图中，源视频序列共有 7 帧，首尾两帧由于被判定为无效目标而被删除，第 4 帧因为与第 3 帧重合度过高而被认为是冗余目标被删除。

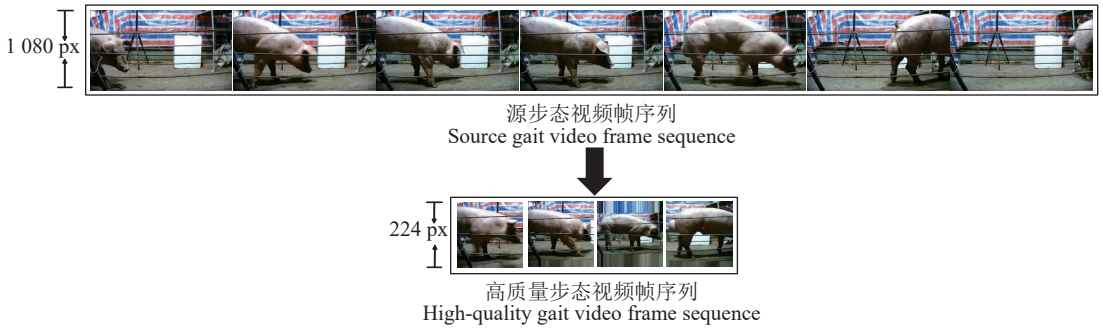
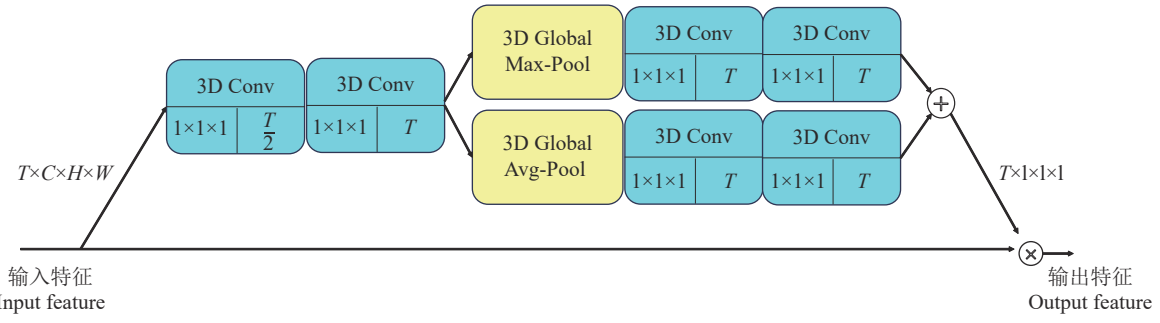


图 4　高质量步态视频合成

Fig. 4　High-quality gait video synthesis

1.2　猪步态评分网络模型设计

1.2.1　时间注意力模块设计　众多基于 3D 卷积网络的研究[15-16] 表明，3D 卷积不仅能够有效提取到图像空间维度上的特征，也能提取到图像序列之间时间维度上的特征，是一种十分适合处理视频的方法。受卷积块注意力模块 (Convolutional block attention module，CBAM)[17] 的启发，本文提出一种基于 3D 卷积网络的时间注意力模块 (Time attention module, TAM)。CBAM 专注于提取图像的通道特征和空间特征，而视频是由多帧图像组成，因此本文设计了一种 TAM，使网络关注到视频帧之间的特征关系，更有利于网络理解视频。TAM 采取“压缩–激励”的模式[18]，“压缩”即通过全局平均池化层或全局最大池化层将输入特征图的空间维度压缩，得到一个全局描述；“激励”即将全局描述映射到一个较小的维度，然后再通过一个激励函数产生注意力权重。TAM 结构如图 5 所示，3D 卷积中特征图的数据格式通常为 $C\times T\times H\times W$，其中 C 表示通道数，T 表示时间维度，H 和 W 分别表示高度和宽度。为了加强模块在时间维度上进行特征提取能力，TAM 首先对输入特征图进行临时的维度变换，将 C 与 T 进行交换，获得 $T\times C\times H\times W$ 的数据格式，以便于 3D 卷积在时间维度 T 间融合特征信息。TAM 通过多个 1×1×1 卷积操作[19]，在降低模型复杂度的同时，融合时间维度上的特征，提高模型的泛化能力。随后通过 3D 全局平均池化和 3D 全局最大池化操作分别获得特征图在时间维度上的特征分布和最显著特征。将两者结合后使用 Relu 函数进行激活，得到输入特征在时间维度上的权重向量，最后将该权重向量与输入特征图进行逐元相乘并还原为 $C\times T\times H\times W$ 的格式。TAM 结构能够精准地提取和加权视频序列中的关键特征，为模



T、C、H 和 W 分别表示特征图的时间维度、通道数、高度和宽度；3D Conv 表示 3D 卷积层；3D Global Max-Pool 表示 3D 全局最大池化层；3D Global Avg-Pool 表示 3D 全局平均池化层

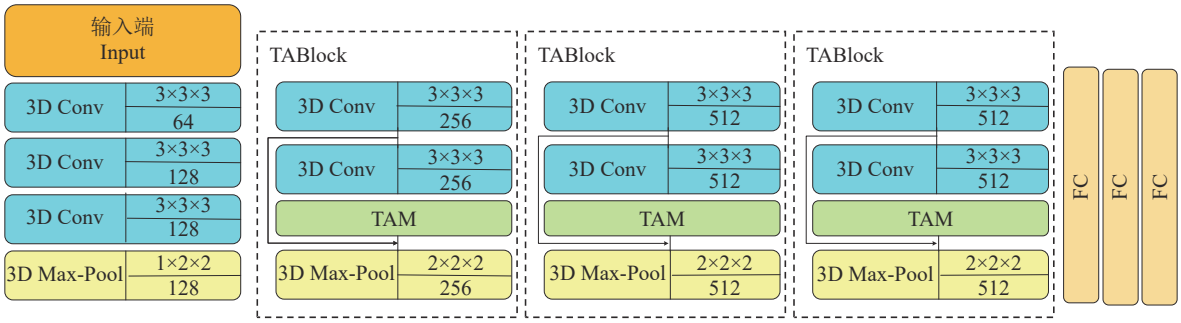
T, C, H and W represent the time dimension, number of channels, height, and width of the feature map, respectively; 3D Conv represents a 3D convolutional layer; 3D Global Max-Pool represents a 3D global maximum pooling layer; 3D Global Avg-Pool represents a 3D global average pooling layer

图 5　TAM 模块结构

Fig. 5　Structure of TAM module

型提供了对动态时序信息的强化处理能力，从而对应用场景中的复杂时空动态变化具有更强的建模能力。

1.2.2　TA3D 网络结构整体设计　TA3D 是本文提出的一个基于 3D 卷积的端到端的视频分析网络模型，用于直接对猪步态视频进行特征提取并自动评分。其中，TAM 作为 TA3D 的关键组件，能够关注视频中的时间维度，帮助网络更好地捕捉和理解动态时空变化，从而提高网络对复杂场景下步态变化的建模能力。为了提高网络的泛化能力，输入网络的帧图像有 50% 的概率被镜像翻转，TA3D 网络的整体结构如图 6 所示。



3D Conv 表示 3D 卷积层；3D Max-Pool 表示 3D 最大池化层；FC 表示全连接层；TAM 表示时间注意力模块；TABlock 表示残差结构

3D Conv represents 3D convolutional layer; 3D Max-Pool represents 3D maximum pooling layer; FC represents fully connected layer; TAM represents time attention module; TABlock represents residual structure

图 6　TA3D 网络结构

Fig. 6　TA3D network architecture

网络引入了残差结构[20]，残差结构将有效解决梯度消失问题，达到加深网络和避免网络丢失特征信息的目的，有效提高网络的特征表达能力。本文残差结构 TABlock 由 TAM 与 3D 卷积层组成，将输入特征与输出特征进行逐元相加保证特征信息不丢失，最后使用 Relu 函数进行激活。网络最后使用 3 个全连接层将提取到的高级特征映射到具体的类别上，并且在 3 个全连接层中使用 Dropout 操作[21]。Dropout 采用正则化技术用于减少深度神经网络的过拟合现象，它在训练过程中将随机丢弃一部分神经元输出，来降低神经网络的复杂性和容量，从而提高模型的泛化能力。

网络使用交叉熵损失函数 (Cross entropy loss function, $\mathrm{Loss_{CE}}$) 计算训练时预测结果与真实值之间的差异，其计算公式如下：

$$\mathrm{Loss_{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{M} y_{ic}\ln\left(p_{ic}\right), \tag{1}$$

式中，N表示该批次的样本量，M表示类别数量，如果样本i的真实类别等于c则y_{ic}为 1，否则y_{ic}为 0。p_{ic}表示样本i属于类别c的预测概率。

2　数据与试验

2.1　数据采集

视频数据于 2022 年 12 月至 2023 年 2 月采集自广东省清远温氏种猪科技有限公司养殖场，试验对象为日龄在 180 d 左右的杜洛克猪和长白猪。测定通道长约 2.0 m，宽约 0.6 m，高约 0.6 m，摄像机布设于通道侧方，高度约 0.4 m。采集数据时，让猪在自然状态下通过通道并对其进行录像，通过时间约为 2~9 s。采用的摄像机为通用的光学摄像机，采集帧率为 25 帧/s，分辨率为 1 920 像素×1 080 像素。本文共采集 162 头猪，每头猪采集 2~3 次，共采集 406 段源步态视频。

2.2　数据集构建

本文视频数据由 3 位畜牧专家共同评定并给出评分。畜牧专家根据猪行走的整体情况进行评分，评分标准如表 1 所示。专家们按照四分制进行评分，对应 4 个步态级别，评分越低表示肢蹄问题越严重，数据集的最终构成如表 2 所示。将这些视频按照 8∶2 的比例进行切分构建步态数据集。此外，从这些视频中随机抽取了 3 082 张图像，使用 Labelimg 工具进行标注，按照 8∶2 的比例构建目标

表 1　四分制步态评分标准

Table 1　Four point gait socring scale

步态评分 Gait scoring	步态级别 Gait level	状态描述 State description
1	不及格	猪只行走困难，背部扭曲，难以承重
2	及格	猪只行走犹豫，步幅较短，略有弓背
3	良好	猪只行走正常，步幅较长
4	优秀	猪只行走稳健，步态对称，步频稳定

表 2 数据集中不同步态级别数据的占比

Table 2 The proportion of different gait levels in the dataset

步态级别 Gait level	数量 Quantity	占比/% Percentage
不及格 Fail	71	17.5
及格 Pass	58	14.3
良好 Good	101	24.9
优秀 Excellent	176	43.3

检测数据集。

2.3 试验平台及训练参数

本文试验均在 Ubuntu22.04 系统上完成, 处理器为 AMD Ryzen Threadripper PRO 3945WX, 内存 32 G, 显卡型号为 NVIDIA GeForce RTX 3090。使用 Pytorch 深度学习框架构建所有模型并进行训练和验证, 所有参与训练的模型均达到最佳效果, 训练过程完全收敛。

YOLOv8n 网络模型训练使用 AdamW 优化器, Batch-size 为 533, 共迭代 150 回合。TA3D 网络模型使用 SGD 优化器, Batch-size 为 12, 共迭代 80 回合。

2.4 评价指标

YOLOv8n 模型需从视频中检测出猪只主体, 使用精确率 (Precision) 和召回率 (Recall) 作为评价指标。计算公式如下:

$$\text{Precision} = \frac{T_P}{T_P + F_P} \times 100\%, \tag{2}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \times 100\%, \tag{3}$$

式中, T_P 表示将正类预测为正类的样本数目, F_P 表示将负类预测为正类的样本数目, F_N 表示将正类预测为负类的样本数目, 本文将预测框与真实框的 IoU 大于 0.5 的情况认定为正类, 其余为负类。

TA3D 网络模型需对步态视频进行分类评分, 使用精确率、召回率和准确率 (Accuracy) 等作为评价指标, 精确率和召回率的计算公式与式 (2) 和式 (3) 相同, 准确率计算公式如下:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + F_N + T_N + F_P} \times 100\%, \tag{4}$$

式中, T_P 表示正确预测为该步态评分的样本数目, T_N 表示正确预测为其他步态评分的样本数目, F_P 表示其他步态评分被错误预测为该步态评分的样本数目, F_N 表示该步态评分被错误预测为其他步态评分的样本数量。总之, T_P 和 T_N 表示预测正确的样本数目, F_N 和 F_P 表示预测错误的样本数目。

2.5 试验结果与分析

2.5.1 猪只检测试验结果 为了验证本文使用改进 YOLOv8n 模型的方法, 对包含不同检测头的 YOLOv8n 模型进行测试, 在本文数据集上的测试结果如表 3 所示。分析表 3 可知, 3 种模型均在本文的数据集上有着优越的性能, 精确率和召回率基本一致且都在 99% 以上。在保持高精度的情况下, YOLOv8-1h 的浮点运算量分别比 YOLOv8n 和 YOLOv8n-2h 减少了 24.39% 和 10.14%, 帧率提高了 17.3% 和 8.5%, 检测帧率远远超过视频帧率 (25 帧/s)。由此验证了本文中仅使用单个大尺度检测头的 YOLOv8n 模型在有同样检测精度的同时有效降低了浮点运算量。

表 3 改进 YOLOv8n 试验结果

Table 3 Experimental results of improved YOLOv8n

模型[1)] Model	精确率/% Precision	召回率/% Recall	帧率/(帧·s^{-1}) Frames per second	浮点运算量/(×10^9) Floating point operations
YOLOv8n	99.8	99.7	185	8.2
YOLOv8n-2h	99.7	99.8	200	6.9
YOLOv8n-1h	99.8	99.7	217	6.2

1)YOLOv8n-2h表示YOLOv8n保留大、中2种尺度的检测头; YOLOv8n-1h表示YOLOv8n只保留大尺度检测头

1)YOLOv8n-2h represents YOLOv8n keeps two detection heads with large and medium scales; YOLOv8n-1h represents YOLOv8n keeps only one detection head with large scale

2.5.2 GFM 试验结果 本文设计 GFM 模块用于从视频流中自动提取有效信息合成高质量步态视频, 实现方案的自动化。GFM 通过计算前后帧目标交并比, 有效去除冗余目标, 并且在保留目标主体等有效信息的同时消除无效背景信息。视频分辨率从 1920 像素×1 080 像素降低为 224 像素×224 像素, 帧图像像素数量直接减少 97%, 最终获得包含高密度信息的高质量步态视频。本文统计了使用 GFM 前后视频时长对比和视频大小对比, 结果如图 7 所示。根据图中数据可知, 在使用 GFM 合成

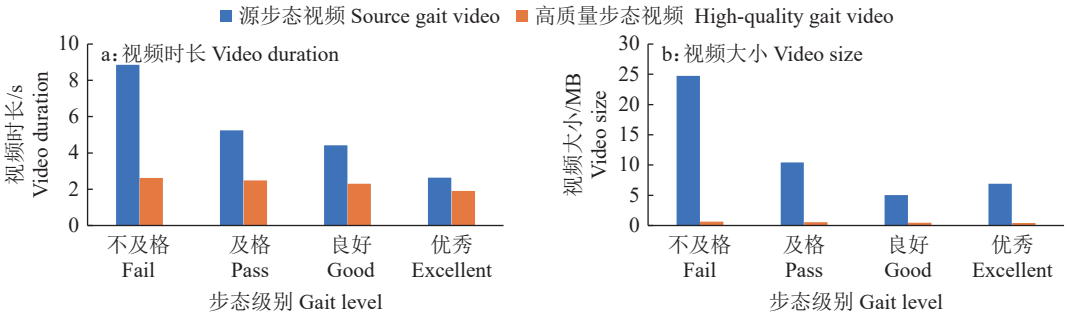


图 7　GFM 数据处理前后对比

Fig. 7　Comparison of GFM data before and after processing

高质量步态视频后，各级别步态视频时长分别缩短了 70.40%、52.48%、47.96% 和 27.38%，各级别步态视频大小分别减少了 97.53%、95.01%、91.04% 和 94.04%。由此可知，GFM 有效缩短了视频长度，使视频时长相对统一并且大幅提高了视频中有效信息的密度。另外，GFM 使得视频大小减少 90% 以上，极大地减小了内存占用，有效降低了存储成本。

2.5.3　步态评分试验结果　本文随机选择了 4 个级别共 84 段步态视频对步态评分模型 TA3D 进行测试，并将预测结果与真实级别进行对比，获得评分测试结果混淆矩阵如表 4 所示。根据表内数据可知，本文方案在 84 段测试视频上的预测正确样本的数目为 81，准确率达到 96.43%。进一步分析发现，预测错误的样本全部为“及格”级别。“及格”级别共计 12 个样本，其中 9 个样本正确预测为 2 分步态，2 个样本错误预测为 3 分步态，1 个样本错误预测为 4 分步态。“及格”级别召回率为 75%，其余召回率均达到 100%。

表 4　评分测试结果混淆矩阵

Table 4　Confusion matrix of score testing results

真实步态级别 Real gait level	不同预测评分的样本数 Sample size with different predicted score				召回率/% Recall
	1分 1 point	2分 2 points	3分 3 points	4分 4 points	
不及格 Fail	15				100.00
及格 Pass		9	2	1	75.00
良好 Good			21		100.00
优秀 Excellent				36	100.00
精确率/% Precision	100.00	100.00	91.30	97.30	

为了进一步验证 TA3D 的性能以及 GFM 的作用，在相同条件下将其与近年来 4 种优秀的视频分析模型[22-25] 进行对比。试验结果如表 5 所示。根据表格数据可知，TA3D 与 GFM 结合后在测试集上的准确率达到 96.43%，与不使用 GFM 的情况相比提高 2.38 个百分点。对于基础模型，TA3D 的准确率与 I3D、Slowfast 和 TANet 相比分别提高了 11.91、2.29 和 1.19 个百分点，与 VideoSwin 持平。在加入 GFM 后，TA3D 的准确率与 I3D、Slowfast 和 TANet 相比分别提高了 4.76、9.53 和 3.57 个百分点。这 4 种视频分析模型的实现原理各有不同：VideoSwin 基于 Swin Transform[26] 架构，I3D、Slowfast 和 TA3D 基于 3D 卷积网络，TANet 基于 2D 卷积网络。其中，TANet 与本文的

表 5　各视频分析模型对比

Table 5　Comparison of various video analyzing models

模型 Model	准确率/% Accuracy	单视频平均推理时间/s Average inference time per video
I3D	82.14	2.02
Slowfast	91.67	2.12
VideoSwin	94.05	1.28
TANet	92.86	1.24
TA3D	94.05	1.18
I3D+GFM	91.67	0.40
Slowfast+GFM	86.90	0.44
VideoSwin+GFM	96.43	0.43
TANet+GFM	92.86	0.26
TA3D+GFM	96.43	0.26

TA3D 在特征提取模块的设计上呈现相似之处: 两者均采用了“压缩–激励”的特征提取方式。同时, 两者在设计思路和对输入特征图的处理维度上存在差异。对于两者的特征提取模块, TANet 设计的时序自适应模块 (Temporal adaptive module) 凭借其双分支设计, 动态生成视频相关的自适应卷积核, 同时兼顾时间和空间维度的特征提取能力, 能够有效捕获视频中的时空信息。相较之下, 本文设计的 TAM 对特征图的时间维度处理更为激进, 更适用于处理时长较短、动作较为激烈的时间敏感型视频, 例如本文中的步态视频。从结果上看, TA3D 与 VideoSwin 均为效果最佳的模型, 但是在使用 GFM 后, TA3D 的推理时间仅为 VideoSwin 的 60.47%, 因此可以认为与 GFM 结合后的 TA3D 为最佳模型。从表格中可以发现, GFM 在识别效果上并非对所有不同原理的视频分析算法有积极作用: 对于 I3D, GFM 使其准确率提升 9.62 个百分点; 对于 Slowfast, GFM 则起着相反的作用, 使其准确率降低了 4.77 个百分点。

为了进一步分析模型对不同级别步态的识别情况, 本文继续统计了各模型在使用 GFM 前后对各级别步态识别的精确率与召回率, 结果分别如表 6 和表 7 所示。精确率表示测试集中预测各类别正确的比率, 分析各模型的精确率对比可知, 各模型对于“良好”和“优秀”步态有着较好的识别效果, 并且 GFM 对于“不及格”步态有较明显的积极作用。使用 GFM 后, TA3D 对于“不及格”步态的精确率分别提高了 16.67 个百分点, 达到了 100.00%, 说明 GFM 有效地提取到了这个级别步态的特征。召回率表示测试集中各类别被正确检出的比率, 根据各模型的召回率对比可知, “及格”步态召回率明显低于其他级别的步态, 这意味着识别错误的样本主要为“及格”级别, 这与表 4 得到的结论相同。

表 6 使用 GFM 前各视频分析模型对比

Table 6 Comparison of various video analyzing models before using GFM

模型 Model	精确率/% Precision				召回率/% Recall			
	不及格 Fail	及格 Pass	良好 Good	优秀 Excellent	不及格 Fail	及格 Pass	良好 Good	优秀 Excellent
I3D	57.14	76.92	95.24	93.10	80.00	83.33	95.24	75.00
Slowfast	71.43	100.00	100.00	96.88	100.00	83.33	100.00	86.11
VideoSwin	93.33	88.89	91.30	97.30	93.33	66.67	100.00	100.00
TANet	82.35	90.91	95.45	97.06	93.33	83.33	100.00	91.67
TA3D	83.33	100.00	95.45	97.06	100.00	83.33	100.00	91.67
平均值 Average	77.52	91.34	95.49	96.28	93.33	80.00	99.05	88.89

表 7 使用 GFM 后各视频分析模型对比

Table 7 Comparison of various video analyzing models after using GFM

模型 Model	精确率/% Precision				召回率/% Recall			
	不及格 Fail	及格 Pass	良好 Good	优秀 Excellent	不及格 Fail	及格 Pass	良好 Good	优秀 Excellent
I3D+GFM	88.24	80.00	95.00	94.59	100.00	66.67	90.48	97.22
Slowfast+GFM	66.67	100.00	91.30	94.44	93.33	33.33	100.00	94.00
VideoSwin+GFM	93.75	91.67	95.00	100.00	100.00	91.67	90.48	100.00
TANet+GFM	83.33	100.00	91.30	97.06	100.00	75.00	100.00	91.67
TA3D+GFM	100.00	100.00	91.30	97.30	100.00	75.00	100.00	100.00
平均值 Average	86.40	94.33	92.78	96.68	98.67	68.33	96.19	96.50

本文继续探究了各个模型运行效率以及 GFM 对它们运行效率的影响, 图 8 展示了各个模型单回合平均训练时间和单视频平均推理时间, 以及 GFM 对他们的影响。根据图 8 可知, 使用 GFM 后

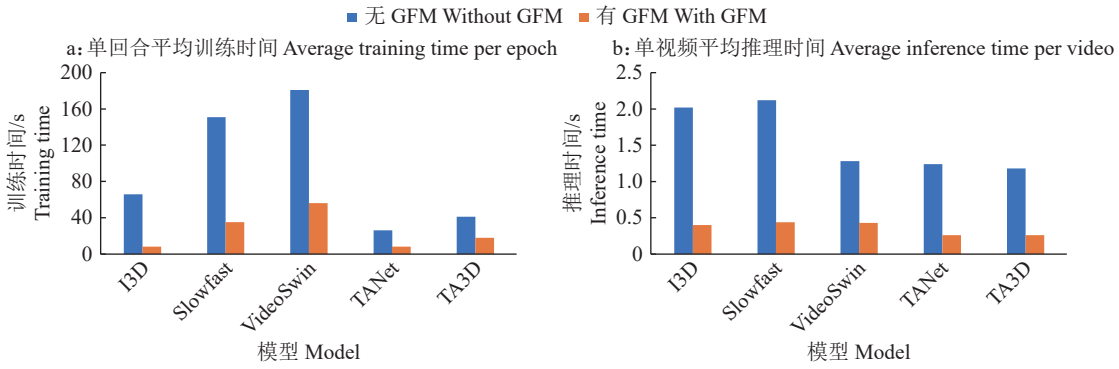


图 8　各模型训练及推理时间对比

Fig. 8　Comparison of training and inference times for various models

I3D、Slowfast、VideoSwin、TANet 和 TA3D 的单回合平均训练时间分别降低了 87.9%、76.8%、69.0%、69.2% 和 56.0%，单视频平均推理时间分别降低了 80.2%、79.2%、66.4%、79.0% 和 78.0%。由此证明，GFM 在消耗少量算力的情况下，有效地从源步态视频中提取到了重要信息，大幅降低了输入帧图像的分辨率，以此显著降低了计算成本。

试验结果表明，各模型识别错误的样本主要为“及格”步态，分析出现这种情况的可能原因如下：

1) “及格”步态与相邻 2 类步态的区别特征不明显。“及格”步态同时包含了“不及格”步态和“良好”步态的部分特征，这对模型的特征提取过程造成了阻碍。处于这个步态级别的猪是判定的主要矛盾点，在进行人工判定时也常常出现不一致的情况。

2) “及格”步态数据较少，与其他级别步态样本数量比例不均衡。试验数据采集过程中，我们尽量保证各个级别的样本均衡采集，而实际中处于“及格”步态级别的猪只明显少于其他级别。“及格”步态级别处于“健康”与“非健康”的中间界限，是一种不稳定的级别。处于这个级别的猪只已经存在一定的肢蹄损伤，可能随时向相邻的步态级别转化，因此处于这个级别的猪只数量明显少于其他级别。

3　结论

在猪只饲养过程中，及时发现存在肢蹄病的猪只能够有效地避免经济损失，而步态是能够直观反映猪只肢蹄情况的关键因素。为了解决该问题，本文引入基于深度学习的视频分析技术对猪只进行步态评分，所取得的结论如下：

1) 本文设计了一种步态关注模块 (GFM)，该模块可以自动从视频流中剥离目标，并且通过删除无效信息和冗余目标来合成高质量的步态视频。在保证识别效果的情况下，有效地降低存储和计算成本。试验证明，GFM 在不降低识别效果前提下，将数据存储成本大幅减少 90% 以上，并显著减少了各模型的训练时间和推理时间，具有重要的实际应用意义。

2) 本文采用了“压缩−激励”的设计思路，构建了一种时间注意力模块 (TAM)，并将其与 3D 卷积和残差结构相结合，形成了一种“端到端”的视频分析模型 TA3D 用于猪步态评分。TA3D 采用线性结构的简约设计，使其具有较快的推理速度；TAM 在时间维度上提取更深层次的特征，而残差结构的设计保证了在加深网络的情况下不丢失特征信息。试验结果表明，与其他优秀的视频分析模型相比，TA3D 在推理速度和准确率方面均达到了最佳效果。

参考文献：

[1] 黄建平. 种猪肢蹄病的常见原因分析[J]. 猪业科学, 2020, 37(5): 106-107.

[2] JØRGENSEN B. Influence of floor type and stocking density on leg weakness, osteochondrosis and claw disorders in slaughter pigs[J]. Animal Science, 2003, 77(3): 439-449.

[3] 刘瑞玲. 猪群肢蹄病发病症状、病因及防治措施[J]. 国外畜牧学 (猪与禽), 2011, 31(1): 88-90.

[4] 王怀中, 张印, 孙海涛. 猪肢蹄病的预防[J]. 养猪, 2010 (2): 39-40.

[5] 王超, 魏宏逵, 彭健. 广西公猪站公猪淘汰原因及在群猪肢蹄病发病规律调查研究[C]//中国畜牧兽医学会动物营养学分会. 第七届中国饲料营养学术研讨会论文集. 郑州: 中国农业大学出版社, 2014: 1.

[6] 杨亮, 王辉, 陈睿鹏, 等. 智能养猪工厂的研究进展与展望[J]. 华南农业大学学报, 2023, 44(1): 13-23.

[7] 刘波, 朱伟兴, 杨建军, 等. 基于深度图像和生猪骨架端点分析的生猪步频特征提取[J]. 农业工程学报, 2014, 30(10): 131-137.

[8] 朱家骥, 朱伟兴. 基于星状骨架模型的猪步态分析[J]. 江苏农业科学, 2015(12): 453-457.
[9] 李前, 初梦苑, 康熙, 等. 基于计算机视觉的奶牛跛行识别技术研究进展[J]. 农业工程学报, 2022, 38(15): 159-169.
[10] ZHAO K, BEWLEY J M, HE D, et al. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique[J]. Computers and Electronics in Agriculture, 2018, 148: 226-236.
[11] 康熙, 李树东, 张旭东, 等. 基于热红外视频的奶牛跛行运动特征提取与检测[J]. 农业工程学报, 2021, 37(23): 169-178.
[12] JIANG B, SONG H, WANG H, et al. Dairy cow lameness detection using a back curvature feature[J]. Computers and Electronics in Agriculture, 2022, 194: 106729.
[13] POURSABERI A, BAHR C, PLUK A, et al. Online lameness detection in dairy cattle using Body Movement Pattern (BMP)[C]//IEEE. 2011 11th International Conference on Intelligent Systems Design and Applications. Cordoba, Spain: IEEE, 2011: 732-736.
[14] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks [C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 4489-4497.
[15] YANG K, QIAO P, LI D, et al. Exploring temporal preservation networks for precise temporal action localization[C]//AAAI. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, California, USA: AAAI, 2018: 7477-7484.
[16] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 7794-7803.
[17] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Springer. Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 3-19.
[18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
[19] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 6848-6856.
[20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
[21] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
[22] CARREIRA J, ZISSERMAN A. Quo vadis, action recognition? a new model and the kinetics dataset[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6299-6308.
[23] LIU Z, NING J, CAO Y, et al. Video swin transformer [C]//IEEE/CVF. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 3202-3211.
[24] FEICHTENHOFER C, FAN H, MALIK J, et al. Slowfast networks for video recognition[C]//IEEE/CVF. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 6202-6211.
[25] LIU Z, WANG L, WU W, et al. TAM: Temporal adaptive module for video recognition[C]//IEEE/CVF. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 13688-13698.
[26] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//IEEE/CVF. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 10012-10022.

【责任编辑 庄 延】