计算机技术与发展
Computer Technology and Development
ISSN 1673-629X,CN 61-1450/TP

《计算机技术与发展》网络首发论文

题目：基于深度特征的立定跳远子动作定位方法研究
作者：花延卓，周俊呈，武杰
DOI：10.20165/j.cnki.ISSN1673-629X.2025.0220
收稿日期：2025-06-25
网络首发日期：2025-07-25
引用格式：花延卓，周俊呈，武杰．基于深度特征的立定跳远子动作定位方法研究[J/OL]．计算机技术与发展．https://doi.org/10.20165/j.cnki.ISSN1673-629X.2025.0220

网络首发时间：2025-07-25 11:05:15　网络首发地址：https://link.cnki.net/urlid/61.1450.TP.20250724.1744.001

基于深度特征的立定跳远子动作定位方法研究

花延卓，周俊呈，武 杰

（陕西师范大学 人工智能与计算机学院 西安 710119）

摘　要：立定跳远作为评估中小学生体质健康的重要指标，其子动作定位的精确性直接影响立定跳远动作的细粒度分析效果。然而，由于立定跳远动作的连续性和非匀速性，其相邻子动作视频帧之间具有高度的相似性，且不同跳远者在不同子动作区间内的视频帧数差异较大，这给立定跳远动作的细粒度分析带来了挑战。本研究针对立定跳远视频中子动作定位的需求，深入分析了基于基础深度模型的特征表征方法对该任务的有效性。具体而言，首先基于基础深度模型提取单帧图像的深层视觉特征，其次利用帧与帧之间相似度计算、非极大值抑制和聚类策略，实现了基于基础深度模型特征的无监督立定跳远子动作定位。实验结果表明，利用传统基础深度模型结合无监督动作定位策略可以实现立定跳远子动作的定位。特别地，使用 ResNet50 作为基础深度模型时，取得了 0.9052 的分类准确率和 0.8257 的平均定位精度。这说明，深度特征与无监督动作定位策略的结合能够有效捕捉运动状态的变化点，可为实时运动矫正系统提供有效的动作分析支持。

关键词：　立定跳远；子动作定位；深度学习；非极大值抑制；聚类算法；

中图分类号　TP391

doi:10.20165/j.cnki.ISSN1673-629X.2025.0220

Research on Standing Long Jump Sub-action Localization Method Based on Deep Feature

HUA Yan-zhuo，ZHOU Jun-cheng，WU Jie

(School of Artificial Intelligence and Computing, Shaanxi Normal University, XI'AN, 710119)

Abstract: Standing long jump is a crucial indicator for assessing the physical health of primary and secondary school students. The precision in locating the sub-actions directly impacts the fine-grained analysis of the standing long jump movement. However, due to the continuity and non-uniform speed of the standing long jump, there is a high degree of similarity between adjacent sub-action video frames, and the number of video frames within different sub-action intervals varies significantly among different jumpers. This presents challenges for the fine-grained analysis of standing long jump movements. This study addresses the need for locating sub-actions in standing long jump videos by thoroughly analyzing the effectiveness of feature representation methods based on foundational deep models for this task. Specifically, deep visual features are first extracted from individual video frames using a foundational deep model. Then, by employing similarity calculations between frames, non-maximum suppression, and clustering strategies, unsupervised localization of standing long jump sub-actions is achieved based on the features from the foundational deep model. Experimental results demonstrate that combining traditional foundational deep models with unsupervised action localization strategies can successfully locate standing long jump sub-actions. When using ResNet50 as the foundational deep model, a classification accuracy of 0.9052 and an average localization accuracy of 0.8257 were achieved. This indicates that the integration of deep features with unsupervised action localization strategies can effectively capture the transition points in movement states, providing valuable action analysis support for real-time movement correction systems.

Keywords　standing long jump；sub-action localization；deep learning；non-maximum suppression；clustering algorithm

0 引言

少年强则国强。青少年的强健体魄是民族生命力的重要体现。全面提高体育人才自主培养质量，是落实党中央战略部署、建设体育强国的关键[1]。体育课作为青少年成长过程中的必要科目，对青少

收稿日期：2025-06-25

基金项目：陕西省科技厅项目（No.2023-YBGY-241）；陕西师范大学教育数字化项目（No. JYSZH201307）

作者简介：花延卓（2001-），男，汉族，陕西师范大学人工智能与计算机学院，硕士研究生，研究方向为深度学习。周俊呈（200-），男，汉族，硕士研究生，研究方向为深度学习，证据推理。通讯作者：武杰（1985），男，汉族，博士，副教授，CCF 会员（No. D9020M），研究方向为统计学习、模式识别以及遥感影像处理。

年进行体育锻炼有着重要指导作用。

立定跳远科目作为中小学体育教学和体质测试的重要项目，其成绩直接反映青少年的腿部爆发力与肌肉协调能力，是评估青少年下肢运动能力的有效手段[2]。立定跳远包含预摆、起跳、腾空和落地四个关键技术环节[3]，标准的动作可以有效提升跳远成绩[4]。然而在日常教学环境下，由于立定跳远运动过程时长很短，且不同人在立定跳远不同子动作区间所持续的时间具有较大差异性，教师难以准确把握每一位同学的跳远动作。因此，研究基于视觉分析技术的立定跳远子动作定位有利于帮助教师剖析每一阶段的动作细节[5]，发现不足，制定针对性的训练计划，进而提高青少年的运动表现，并有效避免运动损伤。本研究所提出的基于深度特征的无监督子动作定位方法可为实时运动矫正系统奠定技术基础，有助于教师提升教学的针对性，进而提高效率[6]。

目前研究中，动作定位旨在从未剪辑视频中定位出不同动作类别的起止时间[7][8]。通常，这些动作类别之间往往存在较大的差异性。而子动作定位则聚焦于在一段连续的动作视频中，分割出属于同一动作类别中不同阶段子动作开始和结束时刻的边界。由于子动作是同一动作的不同阶段，相邻子动作间具有高度的视觉相似性，这使得精确定位各个子动作面临显著挑战。相比而言，现有动作定位方法所研究的动作类别间具有较为明显的差异性，而子动作定位则需要捕捉更为细微的特征变化，才能实现精确的动作边界划分。目前，作为动作定位中更细粒度的研究问题，连续动作中子动作的定位问题尚未得到充分且深入地探讨。

本研究利用基础深度模型结合无监督动作定位方法，初步探索并形成了立定跳远子动作定位的解决方案。本研究方法采用基于两阶段的设计方案：首先，利用深度学习方法提取单个视频帧的动作特征，接着使用无监督动作定位方法估计子动作边界的位置信息。基于所定位的子动作，教师可在体育教学中进行更加准确、有效的动作指导，有助于预防由于不良动作习惯所导致的运动损伤。

1 相关工作

子动作定位相较于仅需识别动作类别的动作定位任务更为精细，不仅要求将连续动作分解为若干子阶段，还需精准识别子动作阶段之间的转换时刻（即子动作边界）。因此，子动作定位是对动作定位问题的进一步深入。动作定位作为计算机视觉的重要研究方向，近年来在视频内容识别[9]、智能监控[10]、体育运动检测[11]等领域展现出广泛应用价值。现有的动作定位方法主要可分为弱监督学习、基于目标检测和多阶段处理三类。然而，这些方法在立定跳远子动作定位任务中均面临一定局限性。

弱监督学习方法因动作速度快、跨度大及人工标注时序动作的困难而受到关注。例如，ActionBytes 方法[12]通过将视频分割为短时片段，利用聚类修剪视频中的动作单元并生成伪标签，实现未修剪长视频的动作定位；UntrimmedNets[13]结合分类与选择模块实现端到端动作识别与定位；文献[14]则通过稀疏时序池化和时序类激活图，利用视频级标签定位动作时间区间。这些方法减轻了标注负担，但主要针对不同动作类别的分类问题。在立定跳远子动作定位中，所有子动作均隶属同一宏观动作类别，弱监督模型因缺乏细粒度监督信号，难以捕捉子动作间的细微差异及边界，导致定位性能受限。

基于目标检测的方法通过动作边界编码，将任务迁移至目标检测框架，采用基于锚点、无锚点或查询等机制进行定位。例如，G-TAD[15]通过图卷积网络构建视频时序图，并利用多级语义上下文融合将问题转化为子图定位问题；ReAct[16]则通过引入关系注意力机制增强查询建模并预测定位质量；AFO-TAD[17]通过无锚点预测动作类别与边界，动态调整接收场以适应不同动作长度的定位。然而，这些方法受初始参考对象质量影响，且立定跳远子动作持续时间短且分布不均（如起跳准备阶段长，蹬地发力阶段短），通用锚点或时间窗口难以捕捉短时动作，且冗余检测往往导致效率降低。

多阶段处理方法通过特征提取、候选提名和边界细化等模块化流程解决定位问题。例如，PSDF[18]利用 C3D 网络生成时空特征，通过滑动窗口提名并优化边界；TAL-Net[19]基于 I3D 网络改进 Faster R-CNN，结合多尺度锚点进行提名，并对提名的动作边界进行优化；SSN[20]通过 TSN 网络提取特征与时间金字塔结合生成多样化提名，并通过分类与边界回归提升精度。在动作定位中，此类方法展现了较好的灵活性与高精度，但其核心模块（包括候选提名和边界细化等）依赖监督信息。在立定跳远子动作定位中，各子动作帧数差异显著，基于监督信息的模型难以有效训练与泛化，导致其在本研究问题中的适用性受限。

综上可见，现有方法在处理立定跳远子动作定

位时存在局限性：弱监督学习方法[12][13][14]难以捕捉高度相似的子动作的细微差异；基于目标检测的方法[15][16][17]难以适应子动作的短暂性和不均匀分布；多阶段处理方法[18][19][20]则对于监督信号具有较大依赖性。为此，本研究针对立定跳远子动作定位任务，借鉴模块化设计理念，采用基于深度特征结合无监督动作定位的方法，规避了定位时对监督信息的依赖。首先，利用基础深度模型提取表征动作变化的帧级特征；随后，通过帧间相似度计算结合非极大值抑制生成候选边界；最后，通过聚类策略细化候选边界，生成最终的边界序列。该方法减少了对标注数据的依赖，且具有良好的实时性。

2 本文的方法

为了实现对立定跳远中子动作的精确定位，本研究采用基础深度模型结合无监督动作定位方法。对于输入的一段完整的立定跳远视频，首先经过卷积神经网络提取单帧图像的空间特征，并将输出的概率分布向量作为该帧的特征向量，然后利用余弦相似度计算相邻两帧间的特征相似性获得相似度曲线，接着采用非极大值抑制[21]方法提取候选边界，最后采用聚类算法[22]对候选边界进行细化，进而完成对跳远动作中子动作的定位任务。基于ResNet50 模型与无监督动作定位方法的子动作定位框架如图 1（a）所示。

2.1 基础深度模型

在目前的基础深度模型中，ResNet 网络[23]通过残差连接有效地缓解训练过程中存在的梯度弥散或梯度爆炸问题。该模型通过不断地堆叠残差块，有效地扩展了深度卷积神经网络模型的表征能力。ResNet 网络结构如图 1（b）所示。其中，残差块是 ResNet 网络的基本构建块，它由主分支和跳跃分支两个分支组成。残差块的结构如图 1（c）所示。

Vgg16[24]通过堆叠多层相同的小卷积核（3×3）来增加网络深度，进而提高图像特征的提取能力，是深度模型研究中的重要基础模型之一。

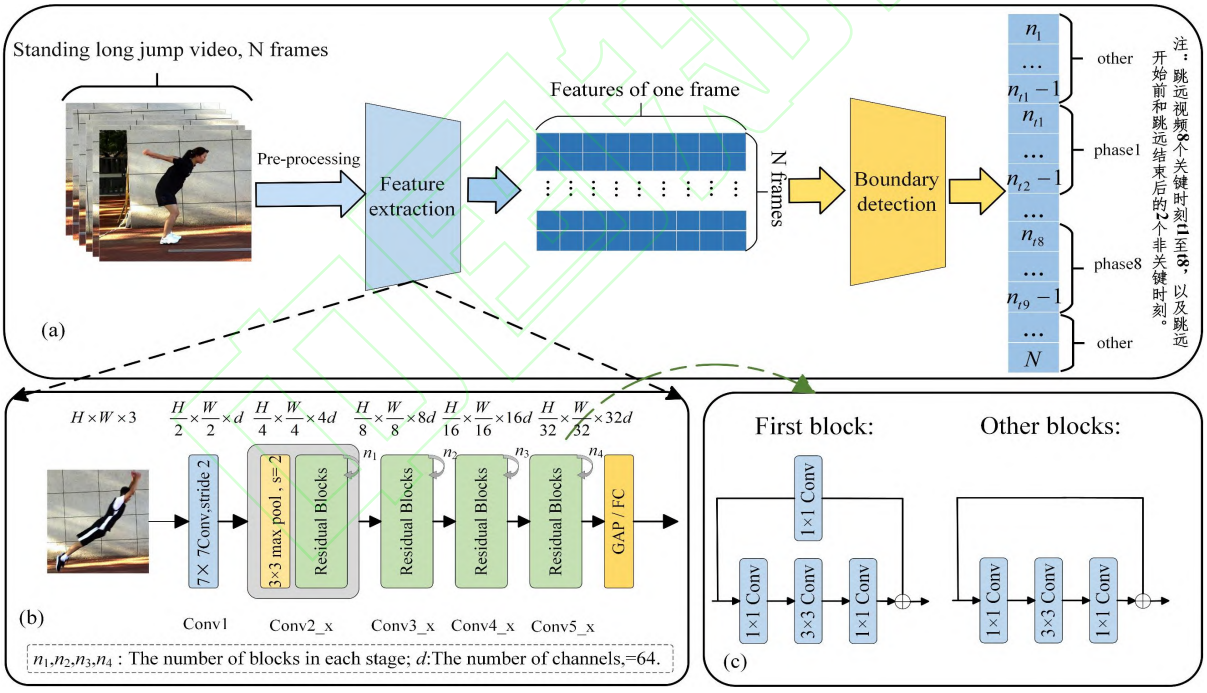

图 1 基于 ResNet50 模型和无监督动作定位方法的子动作定位框架

相比于上述两种模型，SqueezeNet[25]则采用轻量化设计，利用 squeeze 和 expand 模块在减少网络参数量的同时，具有较高的准确率，有效平衡了模型大小与计算成本之间的关系，适用于计算资源受限的环境。MobileNetV3[26]则结合了神经网络架构搜索（NAS）技术和先进的模型压缩策略，不仅在模型大小和计算效率上得到了显著的提升，而且在保证模型推理速度的同时，具有较好的模型性能。

因此，本研究选择上述 5 个经典且具代表性的模型，用以验证所提框架在解决立定跳远子动作精确定位问题上的有效性。

需要说明的是，在基础深度模型训练过程中，本文基于不同子动作类别，采用加权交叉熵损失进行模型参数学习。具体而言，本文首先按照立定跳远视频中帧图像的前后顺序，依次将单帧图像输入

到预训练的模型中，并将输出的类别概率向量作为该帧图像的特征向量；接着，通过依次拼接所有帧的特征向量生成动作定位模块的输入特征矩阵 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_{N-1},\boldsymbol{x}_N\}$。其中 $\boldsymbol{x}_i$ 表示第 i 帧图像的特征向量。特征矩阵 $\boldsymbol{X}$ 的行数表示视频的帧数，列数表示每一帧图像的特征向量的维度，与输出的类别数相关。

2.2 动作定位方法

基于所计算的特征矩阵 $\boldsymbol{X}$，本研究通过动作定位（Action localization，AL）模块将其分为 K 个类别。具体流程为：首先对帧图像进行平滑处理；然后计算平滑后的帧间余弦相似度，生成相似度曲线；接着采用非极大值抑制方法获取初始候选边界；最后通过聚类操作细化候选边界，生成最终立定跳远子动作边界序列，完成定位任务。

2.2.1 帧间相似性计算

在理想情况下，同一子动作内部的帧特征应具有高度相似性，而不同子动作之间的帧特征具有高度不相似性。因此，在时间维度上，通过计算相邻两帧的余弦相似度，并比较它们之间的相似性，即可实现子动作候选边界的定位。余弦相似度计算公式如（1）所示。

$$\boldsymbol{S}_t=\frac{\boldsymbol{x}_t\cdot\boldsymbol{x}_{t+1}}{\|\boldsymbol{x}_t\|\cdot\|\boldsymbol{x}_{t+1}\|} \tag{1}$$

然而，在现实场景中，受遮挡、阴影或者光照变化等因素影响（如图 2 所示），视频帧中会包含大量复杂的干扰因素。这使得同一子动作内的帧特征在时间维度上呈现差异性。

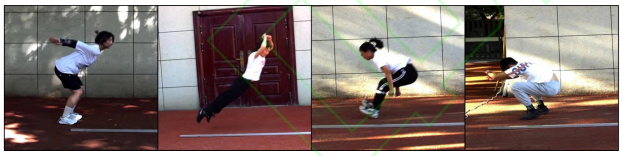

图 2 图像背景干扰

为了缓解噪声对于子动作定位的干扰并保持边界间的区分度。本研究采用了 Bartlett 窗口对时间维度进行加权平滑处理。即在局部窗口内应用三角形权重，赋予中心帧最大权重，并向窗口两侧线性递减（如窗口大小为 5 时，权重为 [1,2,3,2,1]）。此方法可有效抑制了遮挡、光照等噪音，同时避免过度平滑导致子动作边界特征模糊，保留动作转换点的区分度。具体计算如公式（2）所示。

$$\boldsymbol{g}_t=\frac{\sum_{i=t-k}^{t+k}w_i\boldsymbol{x}_i}{\sum_{i=t-k}^{t+k}w_i} \tag{2}$$

其中，$\boldsymbol{g}_t$ 表示第 t 帧的平滑后的特征向量；$\boldsymbol{x}_i$ 表示第 i 帧的原始特征向量；w_i 表示第 i 帧的权重值；i 表示当前处理的帧的时间索引；k 表示窗口半径。

对于输入的立定跳远帧图像，每一帧的特征向量都先经过平滑滤波处理之后再计算相邻帧间的相似度，并采用 $\boldsymbol{g}_t$ 和 $\boldsymbol{g}_{t+1}$ 间的相似度来代替原来帧间相似度。

2.2.2 动作边界检测

理想情况下，考虑到相邻子动作之间的差异性，真实动作边界的位置应与相似度较低的变化点位置一致。为了有效地提取子动作的边界，本研究首先沿时间维度排序相邻帧的相似度，然后采用基于局部邻域的非极大值抑制方法，从中筛选出局部极小值点作为子动作边界的初始候选边界。

2.2.3 聚类细化

假设 $B=\{b_1,b_2,b_3,...,b_M\}$ 表示通过非极大值抑制操作得到候选子动作边界集合。其中，$b_M=N$（N 表示视频帧数），M 表示候选边界数。

在聚类操作中，首先计算相邻候选子动作边界间原始视频帧特征向量的平均向量，即原始视频帧向量 $\boldsymbol{X}$ 被合并为 $\boldsymbol{X}'=\{\hat{\boldsymbol{x}}_1,\hat{\boldsymbol{x}}_2,...,\hat{\boldsymbol{x}}_{M+1}\}$，其中 $\hat{\boldsymbol{x}}_m$ 表示合并后的第 m 个的特征向量；其次，利用公式（1）计算 $\boldsymbol{X}'$ 中任意相邻两特征向量间的相似度，并合并相似度最高的两个相邻特征向量；接着，用该对相邻特征向量的平均向量替换被合并的特征向量；最后，通过重复上述过程，直到动作类别的数量减少到 K 为止，得到长度为 K 的预测边界序列 $B_R=\{b_1,b_2,..,b_K\}$。综上，本文所采用的动作定位方法的具体流程如算法 1 所示。

算法 1 无监督的动作定位算法

输入:（$\boldsymbol{X}$, K）

输出：B

1、初始化平滑后的帧特征：$\boldsymbol{G}\leftarrow\boldsymbol{X}$

2、计算帧与帧之间相似度：$\boldsymbol{S}\leftarrow\boldsymbol{G}$

3、用非极大值抑制策略初始化候选边界: $B\leftarrow\boldsymbol{S}_t$

4、初始化视频段特征：$\boldsymbol{X}'\leftarrow B$

$While(length(\boldsymbol{X}')>K)do$

5、计算任意相邻视频段特征向量间的相似度矩阵：$S\leftarrow\boldsymbol{X}'$

6、比较每对相邻特征向量的相似度，计算相似度最大的一对视频段所对应原始视频帧特征的平均特征向量；

7、以该平均特征向量替代被合并视频段特征向量，并以特征向量所对应视频帧位置进行排序，更新 $\boldsymbol{X}'$；

end while

3 实验验证

3.1 立定跳远数据集

从体育运动学的角度看，立定跳远包括预摆、起跳、腾空、落地四个子动作阶段[3]。这一分阶段框架为理解跳跃的整体流程提供了体育运动理论基础，并且每一阶段的规范性都会显著影响最终成绩。

然而，对于以指导分析跳远动作为目的的子动作分类任务而言，传统四阶段划分具有明显的粗粒度性，难以捕捉跳远动作的关键细节。其局限在于将功能不同的“微观动作”合并，例如：“起跳”阶段模糊了重心下降到最低点与随后的爆发性伸展，而这两个过程恰恰是爆发力的关键；“腾空”阶段也未区分主动收腿与主动伸腿这两个功能不同的动作。

为了实现立定跳远动作的精细化分析，经与体育专家深入探讨，本研究选择采用更为细致的分析框架，着眼于识别出整个动作流中 9 个具有明确生物力学意义的“关键时刻”（如图 3 所示），包括：手腕后摆至最大、脚后跟离地、脚尖离地、充分展体、大腿垂直于地面、膝盖贴近胸前、脚后跟着地、全脚掌着地，直至髋关节达到最低点。

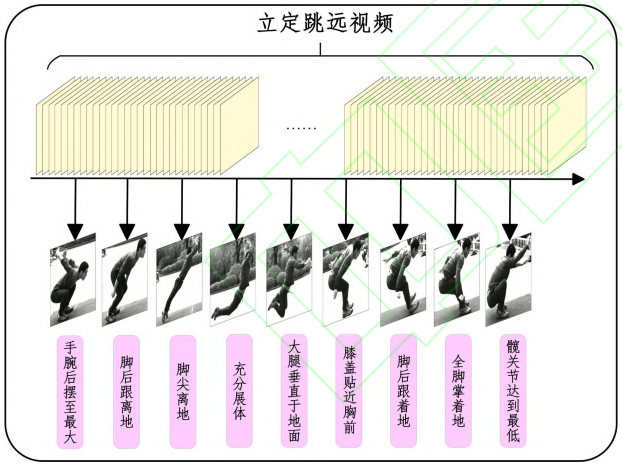

图 3 跳远 9 个关键时刻点

这种基于关键时刻的划分将立定跳远动作分为 10 个片段。其中，本文将动作的起始和结束部分（即视频开始到手腕后摆至最大、髋关节达到最低点到视频结束）归为“other”类，其余构成了 8 个核心子动作阶段（phase 1 至 phase 8）。因此，本文将一个立定跳远视频内的所有帧图像划分为 9 个类别。

表 1 数据集详情

Label	Numbers
other	11565
phase1	2018
phase2	2228
phase3	368
phase4	771
phase5	1267
phase6	872
phase7	367
phase8	1439

本研究使用 100 帧/秒的高速摄像机在自然场景中采集立定跳远运动数据集。该数据集包含 31 名女生 62 次、23 名男生 43 次跳远视频，共计 20895 张图像。依据子动作分析框架将其分为 9 类，具体数据分布如表 1 所示。本文以跳远者为单位进行划分，训练集 16716 张图像，测试集 4179 张图像，确保两者中不存在同一人员的数据。

众所周知，立定跳远子动作分析存在复杂因素：第一，各个阶段持续时间存在差异，导致各类别样本数不平衡。第二，不同个体的身体素质不同，导致同一子动作阶段的帧数也不相同。为了了解这些复杂性，本研究对每个人各类别的平均帧数以及帧数的离散程度进行分析。各类别的平均帧数和标准差误差棒如图 4 所示。结果表明，不同跳远者在同一阶段上的帧数不同，同一跳远者在不同阶段上的帧数也不同。各阶段的平均帧数及标准差反映了立定跳远动作的复杂性和个体差异。

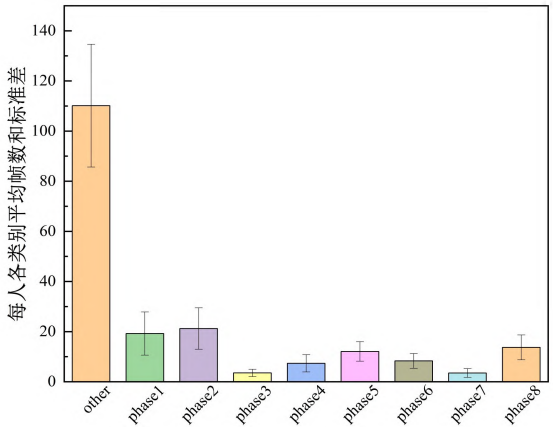

图 4 每人各类别的平均帧数和标准差

3.2 评估指标

由于立定跳远数据集各个类别（阶段）的重要性相同，本研究采用分类准确率(*Accuracy*)和宏平均精确率（ $macro\_P$ ）、宏平均召回率（ $macro\_R$ ）和宏平均 F1-score ($macro\_F1$)作为评价指标。具体计算如公式（3）至（6）所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{3}$$

$$macro\_P = \frac{1}{9}\sum_{i=1}^{9} Precision_i \quad (4)$$

$$macro\_R = \frac{1}{9}\sum_{i=1}^{9} Recall_i \quad (5)$$

$$macro\_F1 = \frac{1}{9}\sum_{i=1}^{9} F1-score_i \quad (6)$$

其中，TP（真正例）表示实际为正且预测为正的样本数；TN（真负例）表示实际为负且预测为负的样本数；FP（假真例）表示实际为负但预测为正的样本数；FN（假负例）表示实际为正但预测为负的样本数。$Precision_i$ 表示第 i 类的精确率，$Recall_i$ 表示第 i 类的召回率，$F1-score_i$ 表示第 i 类的 F1-score。

本研究评估了无监督动作定位方法在立定跳远子动作定位中的性能，采用人工标注的真实边界作为基准，并选择边缘检测分析中常用的测度 R 值（如公式（7）所示）进行结果评价。

$$R = \frac{1}{m}\sum_{k=1}^{m} \frac{1}{1+\alpha d^2(k)} \quad (7)$$

其中，m 表示预测边界和真实边界的个数，$d(k)$ 表示各个子动作的预测边界与真实边界之间的几何距离。分析可见，平均距离越小，说明视频中预测边界越接近真实边界，即 R 值越大。

需要说明的是，真实边界的引入仅用于评估方法的有效性，其本身并未参与模型的训练或边界的预测过程。

3.3 实验细节

本研究选用 ResNet50、ResNet101、Vgg16、SqueezeNet 和 MoblieNetV3 五个基础深度模型进行特征提取。所有基础深度模型均基于 PyTroch 深度学习框架进行仿真实现，且模型训练均在 NVIDIA GeForce GTX 3090 GPUs 上进行。

在训练阶段，首先将输入图像大小统一调整为 256×256 尺寸，然后依次进行左右翻转、上下翻转、随机裁剪至 224×224 尺寸（squeezeNet 输入尺寸裁剪为 227×227）以及颜色抖动（亮度 0.6、对比度 0.7、饱和度 0.5、色调 0.1）等预处理操作以增强数据多样性。

在测试阶段，直接将输入图像统一调整至适合模型输入的尺寸进行处理。

在训练设置中，所有基础深度模型均以在 ImageNet 上预训练的参数作为初始化进行学习，初始学习率设置为 0.0001，训练周期为 500。训练过程中采用加权交叉熵损失进行优化。其中，各类别的权重参数由公式（8）计算得到，优化器选用随机梯度下降（SGD）方法，动量设置为 0.9，权重衰减设置为 5e-4。

$$w_i = \frac{\sum_{i=1}^{9} c_i}{c_i} \quad (8)$$

其中 c_i 表示训练样本中各类别的样本数。

3.4 对比实验

3.4.1 模型性能比较

（a）性能指标分析

为评估不同模型对于立定跳远帧图像的特征表征性能，本文针对各基础深度模型的分类性能，分别利用公式（3）至（6）来计算分类结果的准确率（$Accuracy$）和宏平均精确率（$macro\_P$）、宏平均召回率（$macro\_R$）和宏平均 F1-score（$macro\_F1$）。具体数值结果如表 2 所示。

表 2 各基础模型性能比较

Model	*Accuracy*	*macro_P*	*macro_R*	*macro_F1*
ResNet50	**0.9052**	**0.8822**	0.8449	**0.8583**
ResNet101	0.8983	0.8479	**0.8550**	0.8481
Vgg16	0.8993	0.8578	0.8465	0.8465
SqueezeNet	0.8813	0.8372	0.8401	0.8373
MobileNetV3	0.8720	0.8146	0.8222	0.8141

从表 2 中可见，ResNet50 综合表现最优。其在准确率（$Accuracy$）、宏平均精确率（$macro\_P$）和宏平均F1-score（$macro\_F1$）上均领先其他模型。这说明 ResNet50 能够较为有效地对数据集样本进行分类，且具有较好的泛化能力。但 $macro\_R$ 略低于 ResNet101，这可能是由于 ResNet101 模型的更深层次网络学习了更复杂，更抽象的特征表示。

相比之下，ResNet101 尽管参数量更大，但 $macro\_P$ 显著下降。这可能是由于模型复杂度增加导致模型辨识能力下降。Vgg16 模型在性能上接近 ResNet101 模型，但分类准确率稍逊于 ResNet50。轻量化的 SqueezeNet 与 MobileNetV3 在所有评估指标上均显著落后，这可能与模型本身的轻量化设计策略有关。

（b）混淆矩阵分析

为了深入分析模型对于立定跳远各个子动作的分类性能，图 5 到 9 展示了不同基础深度模型分类结果的混淆矩阵。从混淆矩阵分析可知，所有模型的误分类样本主要集中于相邻子动作之间。特别的，other 类别易与 phase1 和 phase8 类别混淆，这源于立定跳远相邻子动作间存在高度相似性且在时间维度上该几类子动作之间具有邻近特性。

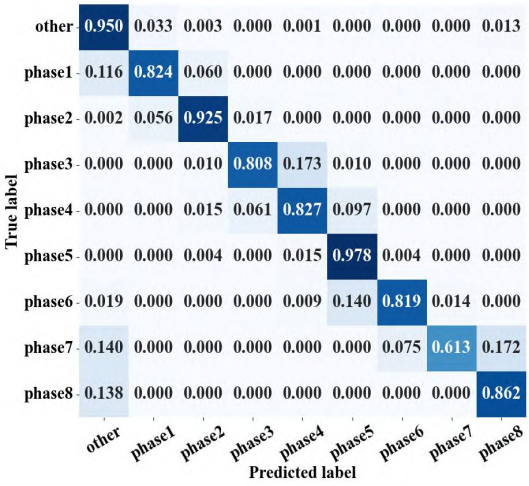

图 5 ResNet50 混淆矩阵

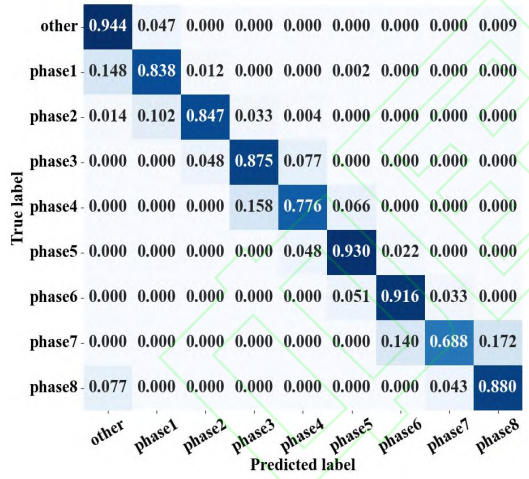

图 6 ResNet101 混淆矩阵

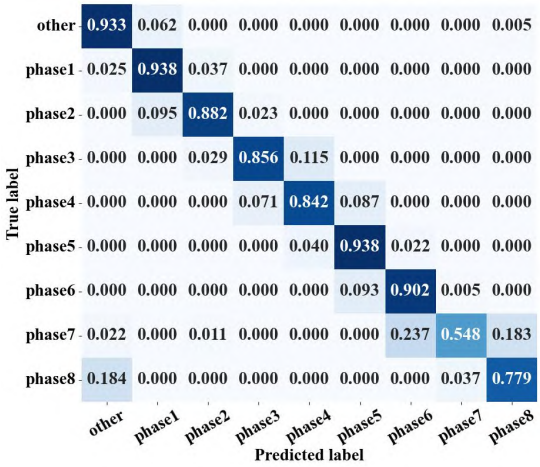

图 7 Vgg16 混淆矩阵

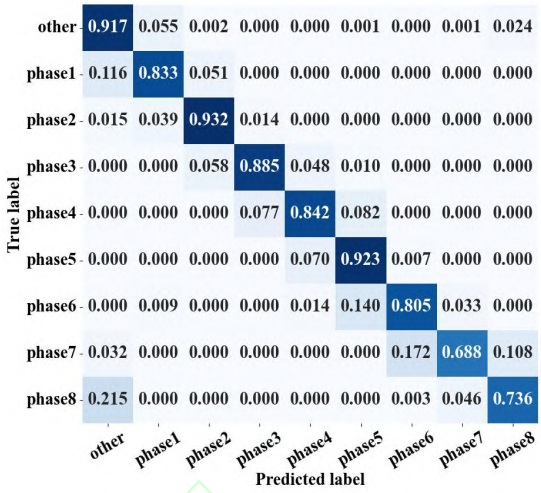

图 8 SqueezeNet 混淆矩阵

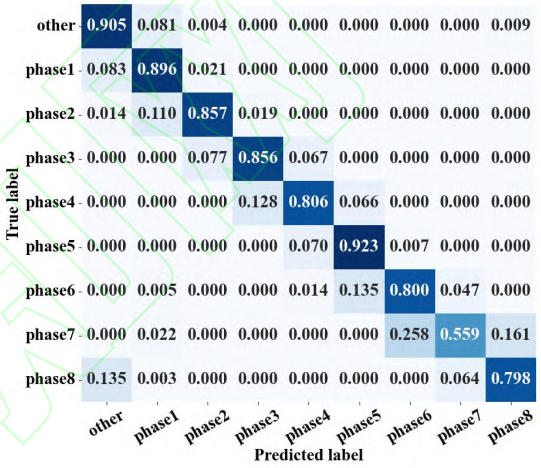

图 9 MobileNetV3 混淆矩阵

对于不同类别而言，模型在类别 phase4（充分展体）和 phase7（全脚掌着地）的分类性能普遍较差。其具体原因在于：1、phase4 和 phase7 的样本数量相对较少，导致模型难以充分学习此类别的特征。2、不同实验对象在 phase4 和 phase7 阶段的姿态差异显著（例如展体幅度、着地姿势），这增加了模型捕捉类别内部共性特征的难度。

在所采用的基础深度模型中，ResNet50 和 ResNet101 模型对各个类别的分类精确率较为均衡（如图 5、6 所示），除 phase7 和 phase4 外的所有子动作类别上均取得了 0.8 以上的高精确率，体现出较强的鲁棒性。相比之下，vgg16 模型在 phase7 这一小样本类别的分类精确率显著偏低。从图 7 中可以看出，vgg16 倾向于将 phase7 的样本误分类为其他类别（如 phase6 和 phase8），反映出其在小样本场景下的局限性。从图 8 和图 9 的混淆矩阵可以看出，SqueezeNet 与 MobileNetV3 模型在多个类别上的错误率普遍较高。特别是在 phase6 和 other 两个类别上，其误分类率均高于其他模型。这表明两模型在处理相似特征的区分能力较弱。

3.4.2 子动作定位性能比较

为了评估深度特征在立定跳远子动作定位中的性能，本研究将 3.2 节中的无监督动作定位方法与基础深度模型相结合。基于 107 个未修剪视频的真实边界，采用公式（7）来计算 *R* 值。表 3 统计了各个模型对应的所有视频 *R* 值的均值(AVG_*R*)与方差(VAR_*R*)。

表 3 各模型 R 值的平均值与方差

Model	AVG_*R*	VAR_*R*
ResNet50+AL	0.8257	0.0100
ResNet101+AL	0.7893	0.0202
Vgg16+AL	0.7717	0.0181
SqueezeNet+AL	0.7440	0.0216
MobileNet_V3+AL	0.7654	0.0213

注：AVG_*R* 衡量子动作定位的平均准确度，VAR_*R* 反映模型的稳定性（方差越低，泛化能力越强）。

实验结果表明 ResNet50+AL 组合在立定跳远子动作定位结果上取得最高 AVG_*R*（0.8257）和最小 VAR_*R* （0.01），验证了该模型特征提取的有效性，进而有利于立定跳远的子动作定位。相比之下，ResNet101 的性能可能由于参数较大导致模型无法收敛到全局最优，从而造成 AVG_*R* 下降 4.4%，VAR_*R* 翻倍。其他模型，例如 Vgg16 所获得的 AVG_*R* 更低（相对于 ResNet50 下降 6.54%），但其方差（VAR_*R*）较 ResNet101 略低一些。SequeezeNet 和 MobileNetV3 两者因轻量化网络结构使其特征提取能力不足，子动作定位精度显著降低（相对于 ResNet50，AVG_*R* 分别降低 9.89%和 7.3%），凸显了其在复杂任务中的局限性。本研究认为 ResNet50+AL 的子动作定位方法在邻近动作具有较高相似性的子动作分析场景中具有显著优势，且泛化能力较好。

4 结束语

国家将立定跳远作为衡量中小学生体质测试的标准之一，而立定跳远子动作的精确定位对于帮助学生查看动作的规范性、预防受伤至关重要。本文针对于立定跳远的子动作定位问题，使用基础深度模型提取深度特征并结合无监督动作定位方法对其进行了探索。该方法通过模型提取单帧图像的深度特征，结合帧间相似度计算、非极大值抑制及聚类策略，实现了子动作边界的精确划分。实验结果表明，ResNet50+AL 方法在分类准确率（0.9052）和子动作定位精度（0.8257）上均显著优于其他模型，验证了其在高精度动作分析与时序定位任务中的有效性与鲁棒性。

为推动体育教学改革并加速技术成果的落地应用，子动作定位方法可作为实时动作分析系统的技术基础，助力于体育教学的有效开展。未来，该技术可能集成到便携设备中，确保在课堂或训练场等多种场景下的实时性。通过精确定位学生跳远子动作，有利于教师对学生动作进行可视化分析反馈，给学生呈现子动作与标准子动作的偏差。同时，基于子动作定位的动作分析系统将有利于老师分阶段查看学生动作的规范性，为学生提供个性化的改进建议，提高立定跳远教学的效率。

本研究的创新性在于其基于深度特征的无监督子动作定位策略，通过在聚类阶段利用帧间相似性细化候选边界，克服了子动作定位依赖精细标注的局限，提升了方法的普适性与可扩展性。除了立定跳远，该方法还可应用于三级跳远、跳高、投篮等复杂运动项目的子动作分析。可见，本研究具有在多种场景的运动教学中的适用性。

然而，本研究还存在以下问题：立定跳远子动作类别（阶段）的划分依赖人工标注，易产生主观偏差，影响模型的客观性；在基础模型训练阶段，仅利用了单帧图像的深度特征，未充分考虑子动作边界邻近帧的时序信息，忽视了视频动作在时间维度上的动态变化特征。此外，该方法虽在单人场景下表现良好，但在多人场景中存在局限性。由于当前设计主要针对单一对象的动作分析，当视频中出现多个运动主体时，可能难以准确区分和定位各个个体的子动作，这限制了其在复杂教学或比赛环境中的应用潜力。

未来研究方向：第一，探索轻量化网络架构，结合注意力机制提升特征提取能力，在降低计算复杂度的同时增强时序建模的精度。第二，引入子动作边界邻近帧的时序特征，增强模型鲁棒性与性能。第三，针对多人场景的挑战，未来可引入多目标跟踪技术，增强该方法在复杂环境下的适应性，进而拓宽其应用范围。

参考文献

[1] 黄汉升.全面提高体育人才自主培养质量,加快建设体育强国[J].武汉体育学院学报,2023,57(01):5-13.

[2] 季浏,尹小俭,吴慧攀,等.“体教融合”背景下我国儿童青少年体质健康评价标准的探索性研究[J].体育科学,2021,41(03):42-54.

[3] 王艺霏.基于表面肌电的青少年立定跳远动作不

同部位肌力对远度的影响[D].鲁东大学,2023.
[4] 巨鑫.起跳姿势控制训练对初中男生立定跳远成绩的影响研究[D].首都体育学院,2022.
[5] 梁万宗.立定跳远的动作技术分析及教学探究[J].河南教育(基教版),2025,(06):84-85.
[6] 李晟,宋可儿,欧阳柏强,等.基于人体姿态识别的立定跳远动作智能评估系统[J].人工智能,2022,(02):75-87.
[7] XIA H, ZHAN Y. A survey on temporal action localization[J]. IEEE Access, 2020, 8: 70477-70487.
[8] XIA K, WANG L, SHEN Y, et al. Exploring action centers for temporal action localization[J]. IEEE Transactions on Multimedia,2023,25:9425-9436.
[9] WANG Y, ZHANG H, YUE Y, et al .Uni-AdaFocus: Spatial-Temporal Dynamic Computation for Video Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025,47(3),1782-1799.
[10] LI Z, XU C, WEI X, et al. Online action detection in streaming videos with time buffers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 9669-9683.
[11] KAHATAPITIYA K, RYOO M S. Coarse-fine networks for temporal activity detection in videos[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 8385-8394.
[12] JAIN M, GHODRATI A, SNOEK C G M. Actionbytes: Learning from trimmed videos to localize actions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE,2020: 1171-1180.
[13] WANG L, XIONG Y, LIN D, et al. Untrimmednets for weakly supervised action recognition and detection[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 4325-4334.
[14] NGUYEN, PHUC X et al. Weakly Supervised Action Localization by Sparse Temporal Pooling Network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6752-6761.
[15] XU M, ZHAO C, et al. G-TAD: Sub-graph localization for temporal action detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 10156-10165.
[16] SHI D, ZHONG Y, CAO Q, et al. React: Temporal action detection with relational queries[C]//European conference on computer vision. Tel Aviv, Israel: Springer Nature Switzerland, 2022: 105-121.
[17] TANG Y, NIU C, DONG M, et al. AFO-TAD: Anchor-free one-stage detector for temporal action detection[J]. arXiv preprint arXiv:1910.08250, 2019. https://arxiv.org/pdf/1910.08250.
[18] SHOU Z, WANG D, CHANG S F. Temporal action localization with pyramid of score distribution features[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 3093-3102.
[19] CHAO Y W, VIJAYANARASIMHAN S, SEYBOLD B, et al. Rethinking the Faster R-CNN architecture for temporal action localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:1130-1139.
[20] ZHAO Y, XIONG Y, WANG L, et al. Temporal action detection with structured segment networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2914-2923.
[21] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression[C]//18th international conference on pattern recognition. Hong Kong, China: IEEE, 2006, 3: 850-855.
[22] BINDRA K, MISHRA A. A detailed study of clustering algorithms[C]//2017 6th international conference on reliability, infocom technologies

and optimization (trends and future directions)(ICRITO). Noida, India: IEEE, 2017: 371-376.

[23] He K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.

[24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014. https://arxiv.org/pdf/1409.1556.

[25] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016. https://arxiv.org/abs/1602.07360.

[26] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE,2019: 1314-1324.