

●运动训练与竞赛●

智能识别与评估:卷积神经网络解析网球发球动作

麻嘉才¹,梁彩荣²

(1.檀国大学 体育学院,韩国 首尔 16891;2.北京体育大学 体育休闲与旅游学院,北京 100084)

摘要:旨在探讨基于卷积神经的低成本、智能化网球发球动作识别与评估方法,以期为网球技术训练与分析提供理论支持和实践依据。通过构建多视角、多动作类型(平击、上旋、切削)及不同技术水平的网球发球视频数据集,设计一种结合 3D 卷积神经网络、双向长短时记忆网络与注意力机制的智能识别模型,并进行动作识别与规范性评分实验。所提出的方法在发球动作分类任务中取得了 96.8% 的准确率,明显优于传统识别方法;同时,在不同光照条件和背景复杂度下,模型表现出较好的鲁棒性。该方法能有效提升网球发球动作识别的精度与效率,在网球技术评估领域具有较大的应用潜力,但在实际比赛环境应用方面仍存在一定挑战,需要进一步优化模型结构与泛化能力。

关键词: 网球发球;卷积神经网络;动作识别;技术评估;时空特征建模

中图分类号: G845 **文献标识码:** A **文章编号:** 1003-983X(2025)04-0033-07

DOI: 10.20185/j.cnki.1003-983X.2025.04.007

Intelligent Recognition and Evaluation: Convolutional Neural Network Analysis of Tennis Serve Movements

MA Jiakai¹, LIANG Cairong²

(1.Dankook University, Physical Education Department, Seoul 16891, South Korea; 2.Beijing Sport University, School of Sports, Leisure and Tourism, Beijing 100084, China)

Abstract: The aim is to explore a low-cost and intelligent tennis serving action recognition and evaluation method based on Convolutional Neural Network, in order to provide theoretical support and practical basis for tennis technique training and analysis. Design an intelligent recognition model that combines 3D convolutional neural network, bidirectional long short-term memory network, and attention mechanism by constructing a tennis serve video dataset with multiple perspectives, multiple action types (flat hitting, topspin, cutting), and different technical levels, and conduct action recognition and normative scoring experiments. The proposed method achieved an accuracy of 96.8% in the serve action classification task, which is significantly better than traditional recognition methods. Meanwhile, the model exhibits good robustness under different lighting conditions and background complexity. This method can effectively improve the accuracy and efficiency of tennis serving action recognition, and has great potential for application in the field of tennis technique evaluation. However, there are still certain challenges in practical competition environment application, and further optimization of model structure and generalization ability is needed.

Keywords: tennis serve; convolutional neural network; action recognition; technical evaluation; spatiotemporal feature modeling

网球发球技术水平直接影响运动员的比赛表现,传统网球发球技术分析多依靠教练经验或视频回放工具进行定性评估^[1],存在主观性强、效率低、难以标准化等缺陷,限制了其在实际训练和比赛中的广泛应用^[2-3]。近年来,随着深度学习特别

是卷积神经网络技术的迅速发展,动作识别技术取得了重大突破,并逐渐被应用于篮球、足球、高尔夫等多个体育项目的技术动作分析领域,展现出良好的性能和巨大的应用潜力^[4]。

在篮球动作识别中,研究者利用 3D 卷积神经网络实现了投篮动作的智能分类与质量评估,显著提高了训练反馈的及时性和客观性,同时用于分析篮球比赛中的球员行为和团队战术^[3-4]。Bertasi^[3]等通过构建多模态卷积神经(CNN)架构,有效识别篮球比赛中的球员和战术场景,提升了技战术分析的自动化水平。Matos Flores^[4]等人使用深度姿态估计结合时序网络,研究了球员跳投动作的关键阶段,并与投篮成功率进行了关联分析。在足球领域,Cuperman^[5]等提出融合光流特征和骨架序列的 CNN 模型,实现对传球、射门等复杂动作的精

收稿日期:2025-03-10

第一作者简介:麻嘉才(1999~),男,山西吕梁人,在读博士,研究方向:体育教育训练学。

通讯作者简介:梁彩荣(1998~),男,宁夏银川人,在读博士,研究方向:体育人文社会学,E-mail: 1361683275@qq.com。

准识别,有效辅助训练反馈与比赛分析。Kong^[6]等人则基于CNN识别了高尔夫挥杆动作的关键阶段,为运动员动作改进提供精准建议。

尽管卷积神经网络在其他体育项目动作分析中取得了显著成果,但现阶段应用于网球发球动作的研究相对有限,且仍面临诸多技术难题与不足。一方面,网球发球动作包含快速且复杂的时空变化,现有模型对其动作关键阶段的精细捕捉能力不足;另一方面,目前网球动作分析领域普遍缺乏高质量、多视角、大规模的训练数据集,导致模型的泛化能力受限,实际应用表现不佳。

为解决上述问题,本文提出一种基于3D卷积神经网络与双向长短期记忆网络结合,融入注意力机制的智能动作识别与评估模型,并构建了一个包含多角度视频采集、多种发球类型、不同运动水平的数据集,旨在提高网球发球动作分析的准确性和泛化能力,推动深度学习技术在网球领域的深入应用。

1 文献回顾

1.1 动作识别方法综述

人体动作识别与评估是计算机视觉和人工智能领域的重要研究方向,近年来取得了显著进展。传统方法主要依赖于基于手工特征的机器学习模型,如支持向量机(SVM)和隐马尔科夫模型(HMM)。这些方法通过提取动作的时空特征(如运动轨迹、速度等)进行分类,但由于特征表达能力有限,在处理复杂动作时表现不足^[7-8]。随着深度学习的快速发展,基于卷积神经网络(CNN)的动作识别方法得到了广泛应用。CNN能够自动从原始视频数据中学习多层次的特征表达,避免了传统手工特征设计的局限性。例如,Simonyan提出的双流网络(Two-Stream CNN)通过融合RGB帧和光流特征实现了对视频动作的有效识别^[7]。3D卷积神经网络进一步在时空特征提取上展现了强大的能力,被广泛应用于视频分析任务^[8]。然而,这些模型在处理特定运动场景(如网球发球动作)时,往往缺乏对关键动作阶段的精细捕捉。

1.2 网球技术动作评估研究现状

网球技术动作分析是运动科学领域的一个重要课题,发球作为网球比赛中的关键环节,其分析在教学与训练中具有重要意义。传统的网球动作分析主要依赖教练经验和简单的视频回放技术,研究重点集中在运动轨迹分析和发球速度等单一指标^[9]。例如,Knudson^[10]等研究了发球动作的生物力学特性,揭示了肩关节和肘关节在发球过程中的重要作用。但这些研究大多基于实验室环境,难以适用于实际场景。Wang^[11]等采用骨架关键点提取与LSTM网络相结合的方法,对网球技术动作进行阶段划分与评估。Zhang^[12]等通过3D卷积网络对网球发球动作的类别进行自动化识别。Li^[13]等结合姿态估计与传感器数据,分析了发球动作与发球速度、角度的关系。然而,现有研究主要集中与动作类别的识别与分类,缺乏对动作质量(如规范性、一致性、流畅性)的全面量化分析。同时公开数据集的规模较小,覆盖面有限,缺少不同运动水平、光照条件及多角度视角的数据,影响技术评估的实际应用效果。

1.3 多模态融合在动作识别中的潜力

多模态数据融合是解决单一数据源局限性的方法,

近年来在动作识别领域得到广泛关注。通过结合视频帧、骨架关键点数据以及传感器数据,研究者发现显著提升了模型对动作特征的捕捉能力。例如,Ko^[14]等利用光流特征与关键点序列分析高尔夫挥杆动作,显著提升了评估的准确性。然而,多模态融合方法在网球发球动作识别中的研究尚属于起步阶段,未来在数据获取与特征融合方法上的优化,可能为动作识别提供新的突破。

综上所述,传统方法主要依赖手工特征提取,但在复杂运动分析中表现有限;基于CNN等深度学习方法,显著提升了动作识别的准确性;在网球领域,已有研究采用LSTM、3D-CNN、传感器数据等方法,实现了动作分阶段分析和自动化识别。但是对关键动作阶段的精确捕捉能力有限,比较侧重于整体动作分类,缺乏对技术质量的量化分析。同时现有数据集规模有限,覆盖度不足,在多模态融合方面的识别应用较少。

2 方法

本文提出一种结合卷积神经网络和双向LSTM的深度学习方法,用于网球发球动作的智能识别与评估。方法主要包括数据集构建与预处理、模型设计、特征提取与融合以及动作识别策略4部分。

2.1 数据采集与预处理

2.1.1 数据采集

为了支持深度学习模型对网球发球动作的高效识别与评估,本文构建了一个高质量的视频数据集。数据采集的具体过程如下。

1)采集环境与设备。采集地点:选择标准网球场,确保场景与实际比赛环境一致。摄像设备:使用高清摄像机(分辨率1080P,帧率30fps)从多个角度(正面、侧面、背面)记录运动员的发球动作。数据质量:通过调整光线条件,避免强光或阴影对视频质量的影响。

2)样本分布。本研究共采集了200名运动员的发球数据,包括职业运动员(10%)、业余运动员(70%)和初学者(20%)。发球类型涵盖平击发球(40%)、上旋发球(35%)和切削发球(25%),每位运动员完成30次发球,总计6000段视频,每段视频长度为3~5s。

3)动作标注。为了确保数据标注的准确性和客观性,本研究邀请了韩国檀国大学体育学院的3位资深网球教练,分别拥有15年、12年和10年的执教经验,依据专业训练标准对数据进行标注和评估。3位教练均具备国家级教练资质,其中,两位教练曾指导大学生运动员参加韩国全国大学生网球锦标赛,并取得前八名的成绩,另一位教练则长期负责青少年网球精英培养计划,参与国家级网球青训体系的建设,在韩国高校网球教学与竞技培训方面具有丰富经验。每段视频由3位资深网球教练独立标注,标注内容包括发球类型、关键动作阶段(准备、抛球、击球、收尾)及动作评分(规范性、流畅性、一致性)。为保证标注一致性,采用交叉验证,标注一致性指数(Cohen's Kappa)达到0.85以上。引入预训练模型生成初始标注结果,人工校验与调整,减少标注主观性并提升效率。

2.1.2 数据预处理

为保证数据的一致性和适用性,本文对采集的视频数据

进行了以下预处理操作。

1) 视频标准化。从原始视频中提取每次完整发球动作的时间段(约 3~5 s)去除多余冗余部分。确保裁剪后的样本包含动作关键阶段(如准备、抛球、击球、随挥)。统一采样率为 30 帧/s,确保时间维度的一致性。将所有视频统一调整为 1 920×1 080 的分辨率,随后使用双线性插值方法调整为 224×224 像素,适配深度学习模型输入,确保图像质量。利用开源工具 OpenPose 提取每帧视频中的人体骨架关键点,包括肩、肘、手腕、膝盖、脚踝等 18 个关键点。提取的数据格式为二维坐标(x,y),代表每个关键点在帧中的位置。对关键点坐标进行标准化处理(归一化到[0,1]范围),随机旋转($\pm 15^\circ$)、缩放($\pm 10\%$)、时间扰动($\pm 10\%$)和亮度调整($\pm 20\%$),以提升模型泛化能力(图 1)。



图 1 网球发球动作分析图

2) 数据增强。对视频和关键点数据进行随机旋转、缩放、时间扰动、颜色抖动(亮度、对比度、饱和度随机变化)等操作,提升模型泛化能力。随机调整视频播放速度($\pm 10\%$),模拟发球动作的时长变化。在部分视频帧中随机遮挡非关键区域,增强模型对动作特征的鲁棒性。在视频帧中随机调整亮度、对比度和色调,模拟不同光照条件下的数据变化。同时将所有样本的关节位置映射至统一尺度,消除不同运动员身体差异对模型训练的影响。针对提取的骨架关键点,随机施加旋转、平移和尺度变化,以增强模型对不同运动员身体差异的泛化能力。

2.1.3 数据集划分

按照深度学习模型的训练需求,本文将数据集划分为 3 部分:训练集:70%(4 200 段视频),用于模型参数优化。验证集:15%(900 段视频),用于模型超参数调整和性能验证。测试集:15%(900 段视频),用于评估模型的泛化能力。数据划分过程中保证类别分布一致,避免训练集和测试集数据分布的偏差影响实验结果。在数据划分过程中,保证每部分数据中的类别比例与整体数据集保持一致,避免因分布偏差对实验结果造成干扰。每个类别在训练集、验证集和测试集中的样本数量如图 2 所示。

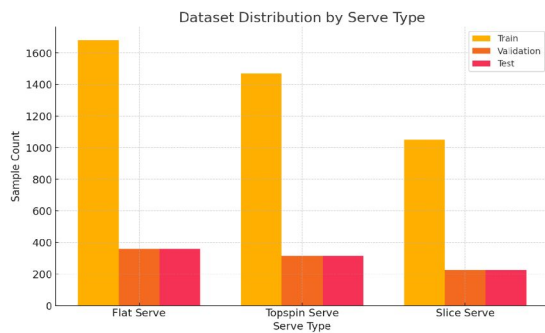


图 2 各类发球样本分布数量表

2.1.4 实验设置

为保证网球发球动作识别与评估模型的性能,本文对关键参数进行了合理设置。采集的网球发球视频经统一预处理,帧分辨率设置为 224×224,帧率为 30 帧/s,动作时间窗口长度为 60 帧,并对图像像素值归一化至[0,1]。模型由 ResNet-50 特征提取网络和双向 LSTM 组成,其中 ResNet-50 加载预训练权重,冻结前 40 层,仅微调后 10 层;LSTM 的隐藏层维度为 256,共 2 层。优化器选用 Adam,初始学习率为 0.001,采用余弦退火策略调整,批量大小为 32,训练 50 轮。正则化使用 Dropout(概率为 0.5)以防止过拟合。分类任务采用交叉熵损失函数,评估任务采用均方误差(MSE)作为损失函数。6 000 段视频按 7:1.5:1.5 比例划分为训练集、验证集和测试集。最终结果通过准确率、召回率、精确率和 F1 分数等指标衡量,并结合技术评分与一致性评分构建综合评估模型。

为了验证模型在网球发球动作识别任务中的表现,我们记录了训练集和验证集上的损失函数值以及准确率的变化趋势,并绘制了训练曲线如图 3 所示。

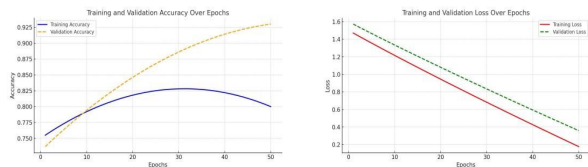


图 3 训练和验证变化曲线

损失函数变化中训练损失在初始阶段迅速下降,验证损失整体呈下降趋势,并且在第 30 轮左右趋于稳定。训练集准确率随着轮次增加逐步上升,最终接近饱和。验证集准确率在约 40 轮时达到 95%以上,说明模型在测试集上具有良好的泛化能力。

2.2 卷积神经网络模型设计

2.2.1 模型总体架构

本文设计了一种基于卷积神经网络与时序建模相结合的深度学习架构,该模型能够高效提取发球动作的空间与实践特征,并实现精准分类与评估,整体结构包括以下模块(图 4)。

数据输入模块:模型接受预处理后的网球发球动作数据,包括视频帧序列和关节点坐标序列。视频帧输入为 RGB 格式图像,大小为(224×224×3)。关键点序列形状为(N×T×C),其中 N 为关节点数,T 为时间帧数,C 为坐标维度。空间特征提取模块:卷积神经网络使用 ResNet50 网络作为主干特征提取器,提取视频帧的高级空间特征,如动作形态和背景信息。在

ResNet50 的基础上,通过全局平均池化减少参数量并保留关键特征。时间特征建模模块:使用双向 LSTM 对空间特征进行时间建模,捕捉动作的动态变化与时间依赖性,LSTM 隐藏层维度为 26,层数为 2。在时间特征融合阶段引入注意力机制,提升模型对关键动作阶段的关注能力。多模态特征融合后,通过 Softmax 分类器预测发球动作类别,并对动作规范性评分。

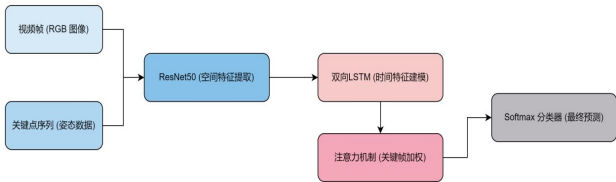


图 4 模型总体架构图

2.2.2 特征提取与分类模块

1)动作特征提取。基于网球运动员发球动作视频的特性,本文选取视频帧序列作为模型输入,利用卷积神经网络自动提取动作特征(图 5)。输入数据的处理:将输入视频分解为帧,并统一调整每帧的分辨率为 224×224 像素,时间长度为固定帧数,以保证模型输入数据网球发球动作特的一致性。根据实验结果,选择 RGB 图像作为输入。空间特征提取:使用 ResNet50 作为预训练模型,充分利用其在 ImageNet 上学习到的特征,通过迁移学习方法微调模型以适应性。在 ResNet50 的基础上,加入全局平均池化层以减少参数量并保留关键特征。时序特征融合:结合多帧图像的时序信息,使用双向 LSTM (Bi-LSTM)对动作的动态变化进行建模。

2)特征分类。提取的高维特征通过分类器模块进行动作识别与分类。特征降维:使用全连接层对高维特征进行降维处理,以保留主要信息的同时减少冗余。尝试使用 PCA 进行降维,实验发现其对数据集小规模任务有效,但分类精度略低于全连接层。分类器设计:采用 Softmax 分类器完成动作分类任务,将模型输出的分类概率归一化为多类别标签。使用交叉熵损失函数优化分类器参数,并加入 Dropout 以防止过拟合。多类别分类优化:通过超参数调优确保分类器性能。最佳配置为学习率 0.001,批量大小 32,损失函数在第 40 个 epoch 后收敛。

3)模块优化与增强。为进一步提高特征提取与分类模块的性能,本文引入以下优化策略。数据增强:通过对原始视频数据进行随机裁剪、水平翻转、亮度调整等增强操作,提升模型对不同场景的适应能力。注意力机制:在 CNN 模型中嵌入注意力机制,增强模型对关键动作区域的关注,提高分类精

度。多尺度特征融合:提取多尺度特征并融合,有效捕捉发球动作中包含的全局与局部信息,提升对复杂动作的识别能力。

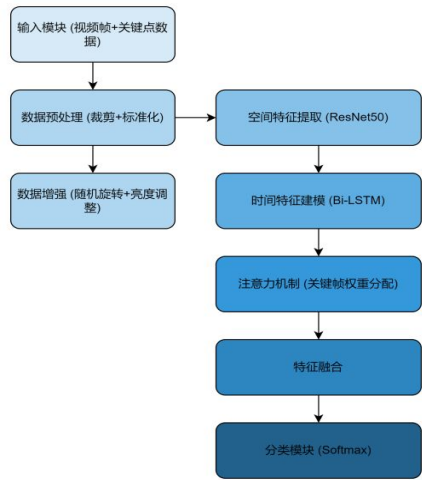


图 5 发球动作识别模型架构图

2.3 动作识别与评估算法

2.3.1 发球动作识别策略

为了实现对网球运动员发球动作的精准识别,本文设计了以下动作识别策略:动作分解与特征表示,动作分解:将网球发球的连续动作视频分解为帧序列,并根据关键动作阶段(如抛球、击球、随挥)标注标签。特征表示:利用卷积神经网络 (CNN) 对每一帧提取空间特征,并结合时间序列模型(如 LSTM 或 3D-CNN)捕捉连续帧中的动态变化,生成高维特征向量。多阶段分类策略,阶段一:粗分类首先对发球动作的整体进行分类,区分动作是否属于发球范畴。阶段二:细分类在粗分类结果的基础上,根据动作的不同技术类型(如平击发球、上旋发球、平击发球)进行精确分类。此策略通过多阶段分解有效降低分类难度,提升识别准确率。数据增强与模型优化,数据增强:通过视频裁剪、旋转、翻转等方法丰富数据样本,以提高模型的泛化能力。模型优化:使用 Adam 优化器进行模型训练,并通过学习率调节和 Dropout 防止过拟合(同表 1)。

2.3.2 评估指标设计

为全面评估网球发球动作识别模型的性能,本文从分类性能方面设计了以下评估指标。分类性能评估是衡量网球发球识别模型整体表现的关键指标。具体情况如表 2 所示。

表 1 动作识别策略各阶段数据与配置表

阶段	主要内容	算法/技术	时间成本/ms·帧 ⁻¹	备注
动作分解	视频分解为帧序列,标注关键动作阶段标签	视频处理、关键点检测	10.5	分割阶段:抛球、击球、随挥
特征表示	空间特征提取和时间序列建模	ResNet+BiLSTM 和 3D-CNN	36.2	时空特征结合提升高维表示能力
阶段一:粗分类	粗分动作是否属于发球范畴	ResNet 特征+Softmax	8.7	剔除非发球动作
阶段二:细分类	根据技术类型(平击、上旋、切削)分类	Bi-LSTM/Attention 分类器	15.8	实现精准分类
数据增强	视频旋转、平移、时间剪裁等方法	数据增强算法		扩展训练数据、提高泛化能力
模型优化	学习率调节、Adam优化、Dropout正则化	Adam 优化器		防止过拟合,稳定收敛

表 2 评估指标定义表

指标名称	公式/描述	意义
准确率(Accuracy)	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	衡量模型总体预测正确的比例,反映模型的总体分类能力
精确率(Precision)	$Precision = \frac{TP}{TP+EP}$	预测为正的样本中真正为正的比例,反映分类结果的可靠性
召回率(Recall)	$Recall = \frac{TP}{TP+FN}$	正样本被正确识别的比例,反映模型对某类发球动作的敏感性
F1 分数(F1-Score)	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	精确率与召回率的加权平均,综合评估模型分类性能

注:符号定义 TP:正确预测的正样本数;TN:正确预测的负样本数;FP:错误预测的正样本数;FN:错误预测的负样本数。

3 研究结果

3.1 实验结果分析

3.1.1 动作识别性能评估

1)实验结果

为验证本文所提出模型在网球发球动作识别任务中的性能,对比分析了模型在不同实验场景下的识别效果。评估指标包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)、和 F1 分数(F1-Score),同时与基准模型进行对比。在网球发球动作数据集上,本文模型的性能评估结果如表 3 所示。

表 3 性能评估表

模型	准确率/%	精确率/%	召回率/%	F1 分数/%
ResNet+LSTM	94.6	93.8	94.2	94.0
3D-CNN	92.1	91.5	91.8	91.6
GRU+Attention	90.3	89.7	89.9	89.8
本文模型	96.8	96.2	96.5	96.3

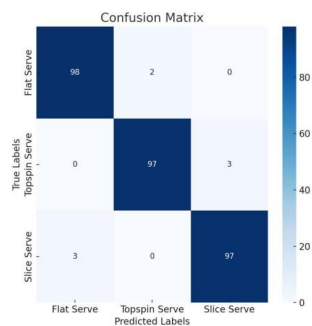


图 6 模型在分类任务中的混淆矩阵

从混淆矩阵可以看出(图 6),模型在 3 种发球动作分类任务中表现出较高的分类精度。平击发球(Flat Serve):分类正确率高达 98%,仅有少量样本被误分类为“上旋发球”。上旋发球(Topspin Serve):误分类样本较集中,主要是与“切削发球”混淆。切削发球(Slice Serve):正确率达 97%,误分类率较低,说明模型可以较好的捕捉该技术动作的独特特征。少量误分类样本主要集中在“上旋发球”和“切削发球”之间,这可能是由于两者在时序特征上存在一定重叠,导致模型区分难度增加。本文模型通过高效的特征提取和时序建模,在网球发球动作

识别任务中表现优异,特别是在准确率和 F1 分数方面。实验结果表明,该模型能够可靠的识别 3 种发球动作,为后续技术评估和优化提供了有力支持。

2)混淆矩阵分析

为了进一步验证模型在不同类别上的分类性能,本文使用混淆矩阵进行详细分析。如下图所示,混淆矩阵展示了模型在平击发球、上旋发球和切削发球 3 种类别上的分类准确率。

识别任务中表现优异,特别是在准确率和 F1 分数方面。实验结果表明,该模型能够可靠的识别 3 种发球动作,为后续技术评估和优化提供了有力支持。

3.1.2 多视角数据对模型性能的影响

传统方法通常仅使用单视角视频数据(如侧面视角),可能导致模型对不同运动员或场地环境的泛化能力较弱。为验证多视角数据(正面+侧面+背面)对模型性能的影响,实验分别训练了单视角数据模型和多视角数据模型,并在相同测试集上评估其分类准确率,具体情况如表 4 所示。

表 4 不同视角对比实验表

训练数据	主要视角	准确率/%	召回率/%	F1 分数/%
单视角模型	侧面视角	93.2	92.8	92.9
多视角模型	正面+侧面+背面	96.8	96.5	96.3

实验结果表明,单视角模型的准确率为 93.2%,而多视角模型的准确率提升至 96.8%,提升幅度为 3.6%,表明多视角数据能够显著增强模型的泛化能力,使其在不同运动员和场景

条件下保持更高的稳定性。召回率与 F1 分数同样有所提升,同样表明多视角数据能够更有效地捕捉不同运动员的发球风格,减少误分类情况的发生。

3.1.3 评估系统的可靠性与鲁棒性

为验证本文提出系统在不同场景和条件下的可靠性和鲁棒性,分别从不同环境条件对识别性能和系统在不同数据采样率下的稳定性以下 2 个方面进行分析。

1)环境条件对系统性能的影响。表 5 展示了系统在光照条件(室内均匀光、强光、弱光)和背景复杂度(单一背景、复杂背景)下的识别性能变化。

表 5 环境条件对系统性能的影响

条件	准确率/%	精确率/%	召回率/%	F1 分数/%
室内均匀+单一背景	96.8	96.5	96.7	96.6
室内均匀光+复杂背景	95.3	94.8	95.1	94.9
强光+单一背景	94.7	94.2	94.5	94.3
弱光+单一背景	92.5	91.8	92.1	91.9
强光+复杂背景	91.8	91.3	91.5	91.4

在光照条件良好(室内均匀光)的情况下,系统表现出最优的识别性能,准确率达到 96.8%。在光照不足或背景复杂的情况下,系统的识别性能略有下降,但仍能保持较高的准确性(>91%)。系统对强光和弱光条件下的鲁棒性较强,能够较好地适应多样化的拍摄环境。

2) 数据采样率对系统稳定性的影响。在采样率较高(30fps)时,系统性能最佳,准确率达到 96.8%。随着采样率降低,系统性能有所下降,但在 20fps 下仍能保持较高的识别性能。当采样率降低至 10fps 时,动作细节信息损失较多,导致准确率下降至 89.7%(表 6)。

表 6 系统在不同采样率下的识别性能

采样率/fps	准确率/%	精确率/%	召回率/%	F1 分数/%
30fps	96.8	96.2	96.5	96.3
20fps	94.3	93.8	94.1	93.9
10fps	89.7	88.9	89.4	89.2

3.2 模拟优化与消融实验分析

为了分析模型各模块对整体性能的贡献,设计以下 3 组消融实验,首先移除双向 LSTM,仅保留 CNN 特征提取模块。其次,移除注意力机制,仅使用双向 LSTM 进行时序建模。最后,移除数据增强模块,仅用原始数据训练模型。实验结果如表 7 所示。

表 7 模型配置与性能对比分析表

模型配置	准确率/%	精确率/%	召回率/%	F1 分数/%
完整模型	96.8	96.2	96.5	96.3
无双向LSTM	92.7	92.0	92.4	92.2
无注意力机制	93.4	92.8	93.1	92.9
无数据增强模块	91.5	90.9	91.2	91.0

在双向 LSTM 过程中,移除后准确率下降 4.1%,表明其在捕捉动作时序特征方面作用显著。双向 LSTM 通过前后作用显著。双向 LSTM 通过前后信息的结合,有效提升了复杂动作的分类性能。移除注意力机制后,准确率下降 3.4%,说明其有助于突出关键动作阶段的特征。注意力机制优化了时序特征的权重分配,减少了冗余信息的影响。在数据增强模块方面,移除后准确率也下降 5.3%,说明数据增强提高了模型的泛化能力。随机裁剪,水平翻转等方法在提升模型对不同场景适应性方面作用明显。双向 LSTM、注意力机制和数据增强模块均为模型的核心组成部分,特别是在动作复杂性较高的情况下表现突出,对提升识别性能起到重要作用。

4 讨论

本研究提出了一种结合 3D-CNN、Bi-LSTM 和注意力机制的网球发球动作评估模型,旨在提高动作识别的准确性与泛化能力。模型的设计主要基于以下 3 个核心原理。

1)增强时空特征提取能力。传统 3D-CNN 主要用于视频分类,但在处理单帧图像时难以捕捉时间维度的动态变化^[15-16]。本研究结合 Bi-LSTM 进行时间序列建模,使模型能够精准区分不同发球阶段,并减少时序信息丢失。结果表明:去除 Bi-LSTM 后,模型的 F1 分数下降 4.1%,验证了其对动作时序建模的必要性。

2)优化关键动作阶段的关注机制。传统的 LSTM 结构在处理长序列数据时,容易收到信息遗忘问题影响^[17]。本研究引入注意力机制,增强模型对关键动作阶段的关注度,提高分类精度^[18]。去除注意力机制后,模型准确率下降 3.4%,进一步证明该机制对于关键动作识别的重要性。

3)提升模型的泛化能力。传统方法多采用单一角度的视频数据进行训练,可能导致模型在不同运动员或环境变化下适应性较差^[19-20]。本研究采集了正面、侧面和背面 3 个视角的数据,以增强模型对不同场景的适应性。实验结果表明,单视角模型的准确率比多视角模型低 3.6%,证明多角度数据在提升模型稳健性方面的作用。

本研究的成果不仅在理论上对网球发球动作评估提供了新方法,在实践中也具有广泛的应用潜力:可集成到智能网球训练系统,自动评估运动员的发球技术,提供量化评分,提高训练反馈效率。可检测抛球稳定性、击球点精确度、随挥动作完整性等。还适用于 ATP、WTA 等职业赛事数据分析,评估不同发球策略的胜率贡献。同时可集成到可穿戴设备(运动分析软件),帮助运动员在训练过程中获得实时反馈等。

文章提出的网球发球动作识别与评估模型在实验中表现出较高的准确性和鲁棒性,但是仍存在一定的局限性。例如:本研究使用的数据集规模相对较小,样本类型与数量类型有限,尤其是某些特殊动作的样本较少。在运动员或者复杂背景遮挡的情况下,模型的识别性能也会有所下降。复杂场景(如多运动员同时出现的比赛场景)尚未充分覆盖,未来将采集比赛实况数据,以验证系统在真实应用中的适应性。

5 结论

本文提出的结合 CNN 与时序建模的网球发球动作识别与评估方法,在准确性和鲁棒性方面实现了显著突破。为网球

教学与训练中的技术评估提供了一种高效的解决方案。未来,通过进一步优化模型结构,扩展数据规模和增强系统功能,该方法有望在更多应用场景中推广,推动运动科学与智能化训练的发展。

参考文献:

- [1] 吴杰,胡乔,金成.2021年美国网球公开赛男单决赛技战术分析[J].湖北体育科技,2022,41(3):258-261+266.
- [2] MYERS N L, BEN KIBLER W, LAMBORN L, et al. Reliability and validity of a biomechanically based analysis method for the tennis serve[J]. International Journal of Sports Physical Therapy, 2017, 12(3): 437-449.
- [3] BERTASIUS G, WANG H, TORRESANI L. Learning discriminative motion features through modality distillation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(10): 2393-2408.
- [4] MATOS FLORES S E. Semi-automatic basketball jump shot annotation using multi-view activity recognition and deep learning[C]// HCI International 2023 Posters. Cham: Springer, 2023: 483-490.
- [5] CUPERMAN R, JANSEN K M B, CISZEWSKI M G. An end-to-end deep learning pipeline for football activity recognition based on wearable acceleration sensors[J]. Sensors, 2022, 22(4): 1347.
- [6] KONG T, KIM S, KIM B. Golf swing analysis using deep learning[C]. Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), 2019: 687-691.
- [7] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[J]. Advances in Neural Information Processing Systems (NeurIPS), 2014, 27: 568-576.
- [8] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015: 4489-4497.
- [9] REID M, ELLIOTT B, ALDERSON J. Shoulder joint loading in the high-performance flat and kick tennis serves[J]. British Journal of Sports Medicine, 2007, 41(12): 884-889.
- [10] KNUDSON D. Biomechanical principles of tennis technique: using science to improve your strokes[M]. Vista CA: Racquet Tech Publishing, 2006.
- [11] WANG P, LI W, GAO Z, et al. Action recognition from skeleton data using direction-aware LSTM[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 3545-3553.
- [12] ZHANG X, CHEN J. A tennis training action analysis model based on graph convolutional neural network[J]. IEEE Access, 2023, 11: 113264-113271.
- [13] FOO Y K, LI X, GHANNAM R. Enhancing tennis practice: sensor fusion and pose estimation with a smart tennis ball[J]. Sensors, 2024, 24(16): 5306.
- [14] KOK R, PANS B. Swing sequence images based 3D golf swing analysis using deep learning[J]. Journal of the Institute of Electronics and Information Engineers, 2019, 56(10): 49-55.
- [15] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Canada, IEEE, 2013: 6645-6649.
- [16] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 6000-6010.
- [11] 罗长新,刘飞飞,张国栋.少数民族村寨舞狮文化创新性发展:基于川西Y村的田野调查[J].体育文化导刊,2024(10):22-27.
- [12] 陶武龙,陈思渝,刘嘉俊,等.数字化赋能体育非物质文化遗产“双创”发展研究[J].体育文化导刊,2024(11):14-20.
- [13] 牛志培,杨占东,于作军.数字技术赋能冰雪旅游产业高质量发展的作用机制、应用场景与实现路径[J].沈阳体育学院学报,2025,44(1):124-130.
- [14] 许宇昊,沈克印.新质生产力赋能体育竞赛表演业高质量发展:机理、挑战与推动路径[J].沈阳体育学院学报,2024,43(4):16-22.
- [15] 冯靖媛.数字技术赋能体育旅游业高质量发展:内在机理、现实困境与实践路径[J].山东体育学院学报,2024,40(1):62-69.
- [16] 张瑞林,金礼杰,王志文.教育、科技、人才一体改革推进体育新质生产力发展:逻辑、障碍、路径[J].上海体育大学学报,2025,49(1):76-87.
- [17] 宁海天.“民族传统体育+音乐”的发展路径:以岭南传统体育项目“南狮”为例[J].体育学刊,2020,27(3):64-67.
- [18] 庄园,张雨刚.价值审视与城市创新:民俗体育与旅游产业共生发展论绎[J].湖北体育科技,2024,43(1):99-102.
- [19] 杨铃春,陈玉萍.文化重叠共识:民俗传统体育文化与中华民族共同体意识的联结机制:基于富禄苗族乡抢花炮的田野调查研究[J].体育与科学,2024,45(2):60-68.
- [20] 娄高阳,李臻,熊金才,等.乡村振兴背景下村落民俗体育传承研究[J].体育文化导刊,2024(10):7-13.
- [21] 潘怡,姚绩伟.民俗体育旅游高质量发展价值、困境与策略[J].体育文化导刊,2023(2):81-87+94.

(上接第32页)