

# 基于隐马尔科夫模型的足球精彩视频事件检测方法研究

赵宇轩, 张宇山\*

广东财经大学统计与数学学院, 广东 广州

收稿日期: 2025年4月29日; 录用日期: 2025年5月23日; 发布日期: 2025年5月31日

## 摘要

随着足球赛事的广泛传播与数据分析需求的增长, 实现足球视频事件的自动检测意义重大。本研究聚焦于运用隐马尔可夫模型(HMM)进行足球视频事件检测。首先, 从足球视频中提取多维度特征, 本文设计了分层特征提取方案。在颜色特征层面, HSV空间直方图主成分分析(PCA)有效表征球场区域的稳定性变化; 在纹理特征层面, 联合能量(角二阶矩)、熵、对比度以及相关性等统计量, 捕捉球员动作引发的局部结构波动; 在运动特征层面, 基于Farneback稠密光流计算平均幅值与方向直方图主成分, 量化全局运动强度与方向趋势。三类特征共同构成10维观测向量, 兼顾静态场景与动态运动信息。随后, 构建HMM模型, 将进球、射门、罚牌、扑救四类足球事件定义为隐藏状态集  $S = \{s_{goal}, s_{shot}, s_{card}, s_{save}\}$ , 通过Baum-Welch算法优化状态转移矩阵A与观测概率分布B。其中, 观测概率采用高斯混合模型(GMM)建模, 以解决特征分布的多模态问题。在解码阶段, 结合Viterbi算法与状态驻留时间约束, 生成最优事件序列, 有效抑制瞬时误检。

## 关键词

足球比赛视频, 自动检测, 精彩事件, 隐马尔可夫模型

# Research on Detection Method of Football Highlight Video Events Based on Hidden Markov Model

Yuxuan Zhao, Yushan Zhang\*

School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou Guangdong

Received: Apr. 29<sup>th</sup>, 2025; accepted: May 23<sup>rd</sup>, 2025; published: May 31<sup>st</sup>, 2025

\*通讯作者。

文章引用: 赵宇轩, 张宇山. 基于隐马尔科夫模型的足球精彩视频事件检测方法研究[J]. 应用数学进展, 2025, 14(5): 689-702. DOI: 10.12677/aam.2025.145293

## Abstract

With the extensive spread of football matches and the increasing demand for data analysis, it is of great significance to realize the automatic detection of football video events. This study focuses on football video event detection using the Hidden Markov Model (HMM). First, to extract multi-dimensional features from football videos, this paper designs a hierarchical feature extraction scheme. At the level of color characteristics, the Principal Component Analysis (PCA) of HSV space histogram effectively represents the stability change of court area; at the level of texture characteristics, the combined statistics of energy (second moment), entropy, contrast and correlation capture the local structural fluctuations caused by players' movements; at the level of motion characteristics, the average amplitude and direction principal components are calculated based on the Farneback dense optical flow to quantify the global motion intensity and direction trends. Three types of features together constitute a 10-dimensional observation vector, taking into account both static scene and dynamic motion information. Subsequently, the HMM model was constructed, defining the four types of football events of goal, goal, shot, penalty and save as the hidden state set  $S = \{S_{goal}, S_{shot}, S_{card}, S_{save}\}$ , and the state transition matrix A and the observed probability distribution B were optimized by the Baum-Welch algorithm. The observation probability is modeled by Gaussian Mixed Model (GMM) to solve the multi-modal problem of feature distribution. In the decoding phase, Viterbi algorithm is combined with state residing time constraints to generate an optimal sequence of events to effectively suppress transient misdetection.

## Keywords

Football Match Videos, Automatic Detection, Highlight Events, Hidden Markov Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着全球体育产业的数字化转型加速, 足球赛事视频已成为球迷津津乐道、研究战术和商业变现的核心载体。据 FIFA 2024 年度报告显示, 全球顶级足球赛事年直播时长超过 50 万小时, 用户观看内容日均新增超 200 万条。然而, 海量视频数据的高效解析长期需要依赖人工标注, 平均单场 90 分钟比赛需耗费专业标注员 30~45 分钟进行事件标记, 成本高昂且一致性难以保障。传统自动化方法如基于规则的颜色分割或模板匹配[1][2], 虽在特定场景(如进球瞬间的球网振动检测)中表现尚可, 但其编码逻辑难以适应多变的比赛环境(如夜间照明差异、摄像机视角切换等)。近年来, 机器学习方法(如 SVM、CNN)虽提升了静态帧的分类精度, 却忽视了足球事件的本质属性——时序关联性与动态连续性。例如, 一次完整进球事件通常包含传球、射门、球入网、庆祝动作等阶段, 类似 SVM、CNN 这种简单帧级分类易造成事件碎片化误判。因此, 如何构建兼顾时间和空间的上下文理解与复杂特征融合的事件检测模型, 成为体育视频分析领域的关键挑战。在此背景下, 隐马尔可夫模型(Hidden Markov Model, HMM)因其时序状态建模能力与概率推理优势重新获得关注。HMM 通过隐藏状态(事件类别)与观测序列(视频特征)的联合概率建模[3][4], 能够有效捕捉事件演变的动态规律。早期研究[5]已验证 HMM 在网球发球动作识别中的潜力, 但其在足球领域的应用仍面临两大瓶颈: 其一, 足球事件时空跨度大、类间特征重叠度高(如射门与扑救的光流模式相似), 需设计更具判别性的多模态观测特征; 其二, 传统 HMM 的单一高斯观测假设难

以刻画高维特征的复杂分布[6]。针对上述问题,本研究提出一种融合多模态特征与高斯混合模型(GMM)的改进 HMM 框架,旨在实现足球视频事件的精准连续检测。

## 2. 足球比赛视频数据采集与预处理

### 2.1. 视频数据来源

为了构建一个全面且具有代表性的足球视频数据集,本研究采用多种途径收集足球比赛视频数据。我们从公开的视频平台,如腾讯体育、爱奇艺体育、优酷体育等,这些平台拥有丰富的足球赛事版权,涵盖了众多国内外顶级足球赛事,包括中超、英超、西甲、意甲、德甲、法甲以及欧冠、世界杯等。通过合法的授权渠道,获取这些平台上不同赛季、不同场次的足球比赛视频,确保数据的多样性和广泛性。例如,在收集英超比赛视频时,选取了近 5 个赛季中不同球队之间的关键比赛场次,包括夺冠热门球队之间的强强对话,以及保级球队之间的生死较量,以涵盖各种不同的比赛场景和激烈程度[7]。

### 2.2. 视频特征提取

深入研究并提取足球视频的颜色特征(HSV 颜色空间直方图 3 维)[8],该特征能够直观地反映视频图像中不同颜色的分布情况,HSV 颜色空间将颜色分为色调、饱和度和明度三个维度,可有效描述足球场场景中如草地、球员服装等颜色信息;纹理特征(能量、熵、相关性、对比度共四维),这些纹理参数从不同角度刻画了图像中纹理的结构和变化,对于区分不同场景元素具有重要作用;光流特征(平均光流幅值 1 维以及运动方向直方图主成分 2 维),平均光流幅值体现物体运动的剧烈程度,运动方向直方图主成分细化运动方向分布,可准确捕捉足球比赛中球员和球的运动信息。通过对这些多维度特征的提取,为后续的 HMM 模型训练提供全面、丰富的数据基础。

#### 2.2.1. 颜色特征提取

颜色直方图是一种常用的特征描述方法。通过统计视频帧中不同颜色的像素数量,构建颜色直方图,以反映视频帧的颜色分布特征。在足球视频中,绿色的足球场草坪占据画面的大部分区域,其颜色在颜色直方图中具有明显的特征。利用颜色直方图,可以快速判断视频帧是否为足球比赛场景,以及区分不同的比赛场地。HSV (Hue, Saturation, Value, 即色调、饱和度、明度)颜色空间直方图是颜色直方图在 HSV 颜色空间下的具体实现,所以 HSV 颜色空间直方图是颜色直方图的一种特殊情况 HSV 颜色空间直方图是基于 HSV 颜色空间计算得到的颜色直方图。在 HSV 颜色空间中,色调表示颜色的种类,饱和度表示颜色的鲜艳程度,明度表示颜色的明亮程度。HSV 颜色空间直方图会分别对这三个通道进行统计,通常将每个通道划分为若干个区间,统计落入每个区间组合内的像素数量。将图像从 BRG 空间转换到 HSV 空间,每个像素的 HSV 值满足:色调(Hue),  $H \in [0, 360^\circ)$ , OpenCV 中缩放到  $[0, 180]$ 。饱和度(Saturation),  $S \in [0, 1]$ , OpenCV 中缩放到  $[0, 255]$ 。明度(Value),  $V \in [0, 1]$ , OpenCV 中缩放到  $[0, 255]$ 。将 HSV 颜色空间划分为若干个小的区间,每个区间就是一个 bin。然后对每个通道(H, S, V)分别计算 16-bin 直方图:

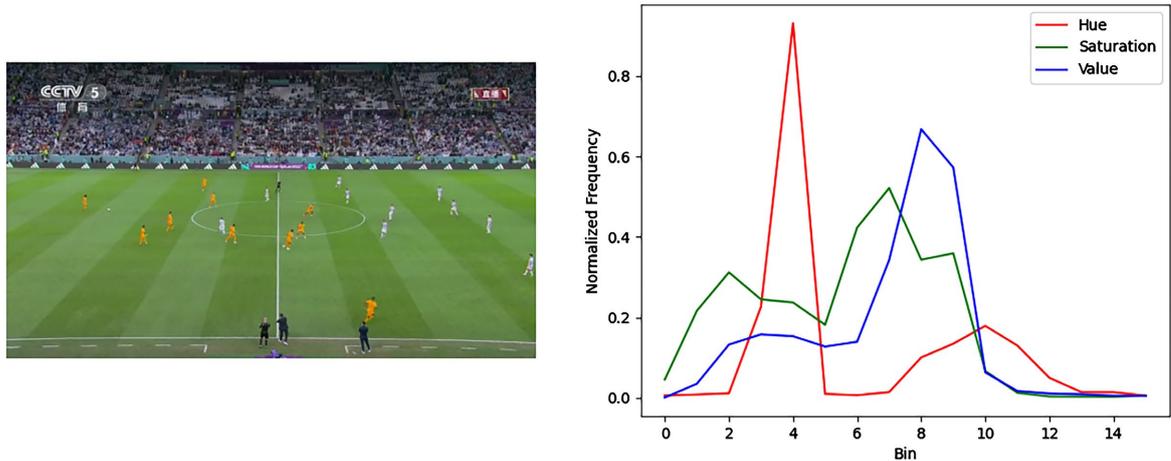
$$Hist_k(b) = \sum_{i,j} 1_{\{\text{像素}(i,j) \in \text{第 } b \text{ 个 bin}\}}, k \in \{H, S, V\}, b = 0, 1, \dots, 15$$

将每个通道的直方图视为 16 维向量,对三个通道的直方图分别进行 PCA 降维。保留每个通道的第一个主成分(PC1),最终得到 3 个数值:

$$H\_PCA = \text{PC1}(\text{HistH}), S\_PCA = \text{PC1}(\text{HistS}), V\_PCA = \text{PC1}(\text{HistV})$$

图 1 是选取 2022 世界杯阿根廷 vs 荷兰的视频的部分帧,对其进行颜色特征提取,并生成对应的 HSV 颜色直方图。左边是原始视频帧,右边的横坐标是 bin 区间,纵坐标是 Normalized Frequency 值,是一种

经过归一化处理后的频率值。它是将每个 bin 中的像素频率除以图像中的总像素数得到的。这样得到的结果是一个介于 0 到 1 之间的值, 表示该 bin 所代表的颜色在图像中出现的概率。



**Figure 1.** Histogram of HSV channel corresponding to the video  
**图 1.** 视频对应的 HSV 通道直方图

### 2.2.2. 纹理特征提取

纹理特征反映了图像的结构和细节信息。采用灰度共生矩阵(GLCM)方法提取纹理特征, 通过计算图像中不同灰度级像素对在不同方向、不同距离上的共生概率, 得到纹理特征矩阵。足球场上的草皮纹理、球员队服的纹理等都具有独特的特征, 这些纹理特征能够为视频分析提供更多的细节信息, 有助于区分不同的场景和对象。灰度共生矩阵是一种通过统计图像中具有特定灰度值且满足一定空间关系的像素对出现的频率来描述图像纹理特征的方法。它考虑了像素间的距离和角度关系。灰度共生矩阵  $P(i, j, d, \theta)$  定义为:

$$P(i, j, d, \theta) = \frac{\#\{(x, y) | I(x, y) = i \wedge I(x + \Delta x, y + \Delta y) = j\}}{N}$$

表示的是在距离为  $d$ 、角度为  $\theta$  的条件下, 灰度值为  $i$  和  $j$  的像素对出现的概率。其中,  $\#\{\}$  表示满足条件的像素对的数量,  $I(x, y)$  是图像在坐标  $(x, y)$  处的灰度值,  $\Delta x = d \cos \theta$ ,  $\Delta y = d \sin \theta$ ,  $N$  是满足条件的像素对的总数。

基于灰度共生矩阵, 可以计算多种纹理特征, 如能量(角二阶矩)、熵、对比度以及相关性等。能量的定义如下:

$$ASM = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta)^2$$

其中,  $L$  是灰度级的数量, 能量反映了图像灰度分布的均匀程度和纹理的规则性。不同的足球视频事件场景具有不同的纹理能量特征。例如, 在比赛正常进行时, 场上球员活动频繁, 画面中球员的动作、光影变化等使得图像纹理较为复杂, 能量值相对较低; 而当出现暂停或休息场景时, 画面相对稳定, 球员和观众的动作较少, 图像灰度分布较为集中, 能量值相对较高。通过分析能量特征, 可以在一定程度上区分这些不同的事件场景, 为后续的事件检测提供线索。

熵的定义如下:

$$ENT = -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \log P(i, j, d, \theta)$$

熵表示图像纹理的复杂程度。图像的灰度分布越均匀、越随机, 熵值就越大, 说明图像纹理越复杂; 反之, 若图像的灰度分布比较集中, 熵值就越小, 图像纹理相对简单。在足球视频中, 不同的事件场景具有不同的纹理特征熵。例如, 在球员罚牌场景中, 球员和裁判的动作相对较少, 图像的纹理相对简单, 熵值较低; 而在进球或射门场景中, 球员、观众的动作较多, 场面较为混乱, 图像的纹理更加复杂, 熵值较高。通过计算熵值, 可以帮助区分这些不同的事件场景。

对比度的定义如下:

$$\text{CON} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 P(i, j, d, \theta)$$

对比度反映了图像中灰度差异的程度, 即纹理的清晰程度。

相关性的定义如下:

$$\text{Correlation} = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_i)(j - \mu_j) P(i, j)}{\sigma_i \sigma_j}$$

其中

$$\mu_i = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} iP(i, j), \quad \mu_j = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} jP(i, j)$$

$$\sigma_i = \sqrt{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_i)^2 P(i, j)}, \quad \sigma_j = \sqrt{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (j - \mu_j)^2 P(i, j)}$$

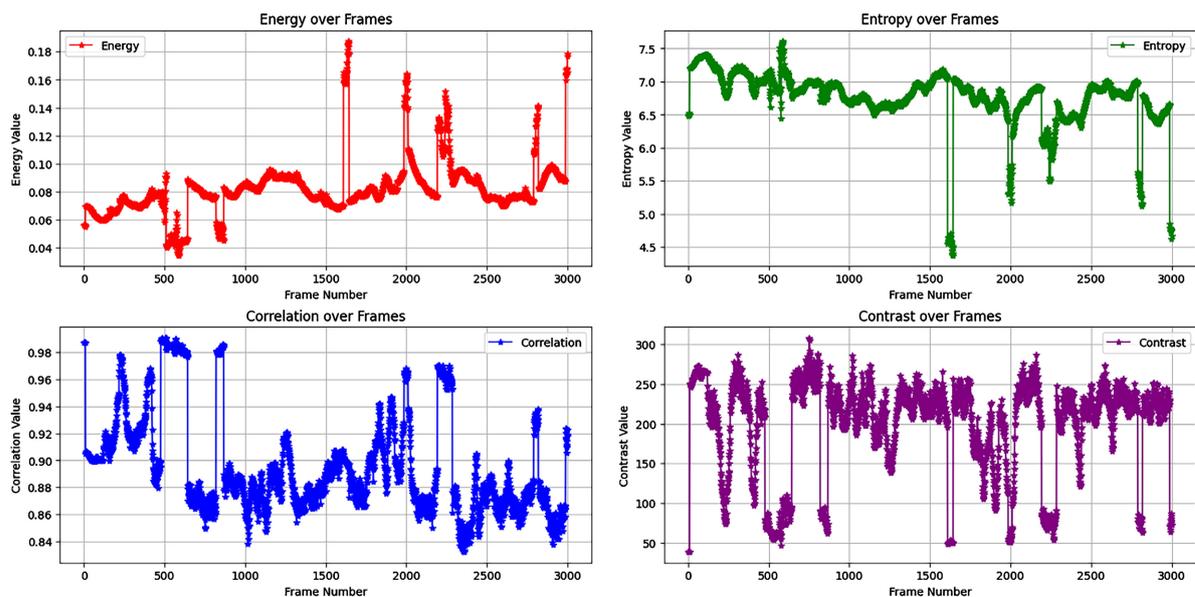


Figure 2. Texture feature values corresponding to the first 3000 frames of the video

图 2. 视频前 3000 帧对应的各项纹理特征值

相关性衡量图像中灰度级的线性依赖程度。取值范围在-1 到 1 之间, 值越接近 1, 表明灰度级之间线性正相关越强; 越接近-1, 线性负相关越强; 接近 0 则表示灰度级之间几乎不存在线性关系。图 2 是 2022 世界杯阿根廷 vs 荷兰的视频前 3000 帧对应的纹理特征变化曲线图, 左上方的图是能量随着帧数的变化的曲线图, 右上方的图是视频图像纹理特征熵随着帧数的变化曲线图, 左下方则是相关度随着视频帧数的变化曲线图, 右下方则是对比度随着视频帧数的变化曲线图。

### 2.2.3. 光流特征提取

基于 Farneback 光流算法提取的运动特征, 直接聚焦于足球和球员的动态信息。足球比赛中, 球员和足球的运动是核心, 运动特征能精准获取其速度、方向等关键运动参数。通过监测球员的运动方向和速度变化, 可判断进攻与防守的转换。在检测精彩事件如射门时, 足球的运动速度和方向突变是重要判断依据。运动特征还能分析球员之间的相对运动关系, 判断是否存在越位等情况。所以, 运动特征对于捕捉足球比赛中的动态事件起着决定性作用。

运动特征提取(Farneback 光流算法)是基于图像亮度在时间和空间上的变化, 通过计算相邻帧之间像素的运动来估计光流场, 从而得到物体的运动信息。

假设图像亮度函数为  $I(x, y, t)$ , 其中  $(x, y)$  是空间坐标,  $t$  是时间。根据光流的基本假设, 即亮度恒定和小运动假设, 有:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

对其进行泰勒展开并忽略高阶项, 得到:

$$I_x u + I_y v + I_t = 0$$

其中,  $I_x = \frac{\partial I}{\partial x}$ 、 $I_y = \frac{\partial I}{\partial y}$ 、 $I_t = \frac{\partial I}{\partial t}$  分别是图像在  $x$ 、 $y$  方向的偏导数和时间上的偏导数,  $u = \frac{\Delta x}{\Delta t}$ 、 $v = \frac{\Delta y}{\Delta t}$

是光流在  $x$  和  $y$  方向的分量。在本文研究方法中将提取光流特征三个维度, 包括平均光流幅值(1 维)和运动方向直方图主成分(2 维)。平均光流幅值是指图像中所有像素的光流向量幅值的平均值, 反映整体运动的强度。它的数学公式如下:

$$\text{Magnitude} = \frac{1}{N} \sum_{i,j} \sqrt{u_{ij}^2 + v_{ij}^2}$$

其中,  $u_{ij}$  和  $v_{ij}$  分别为像素点  $(i, j)$  的水平与垂直光流分量,  $N$  为有效像素数。

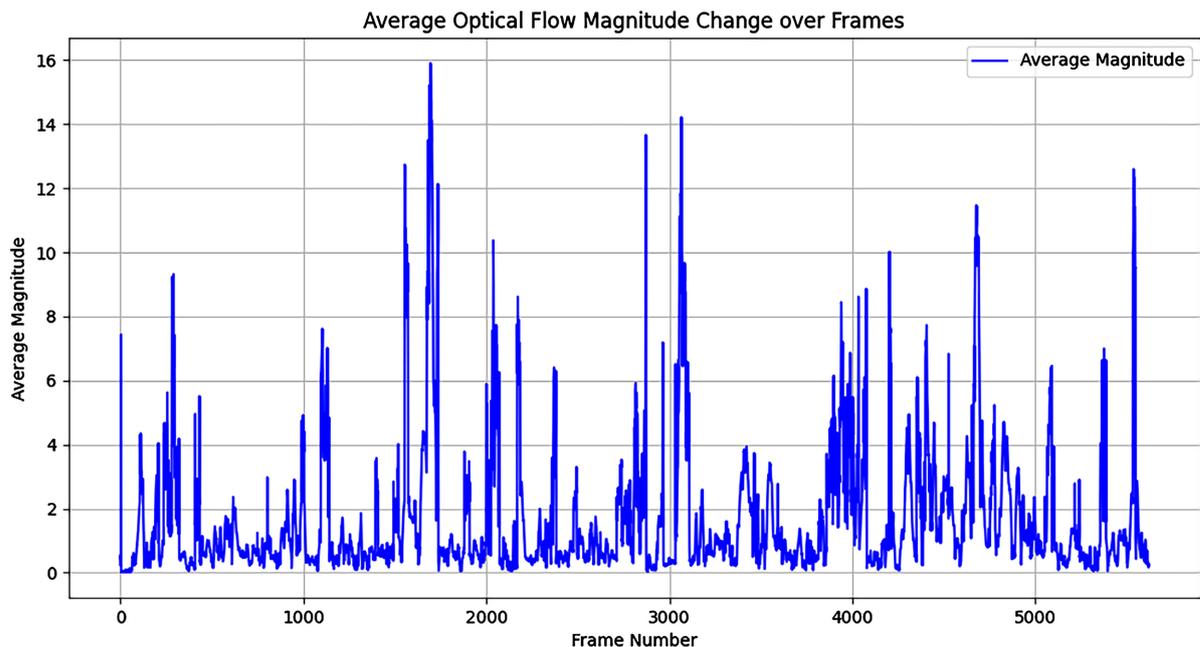
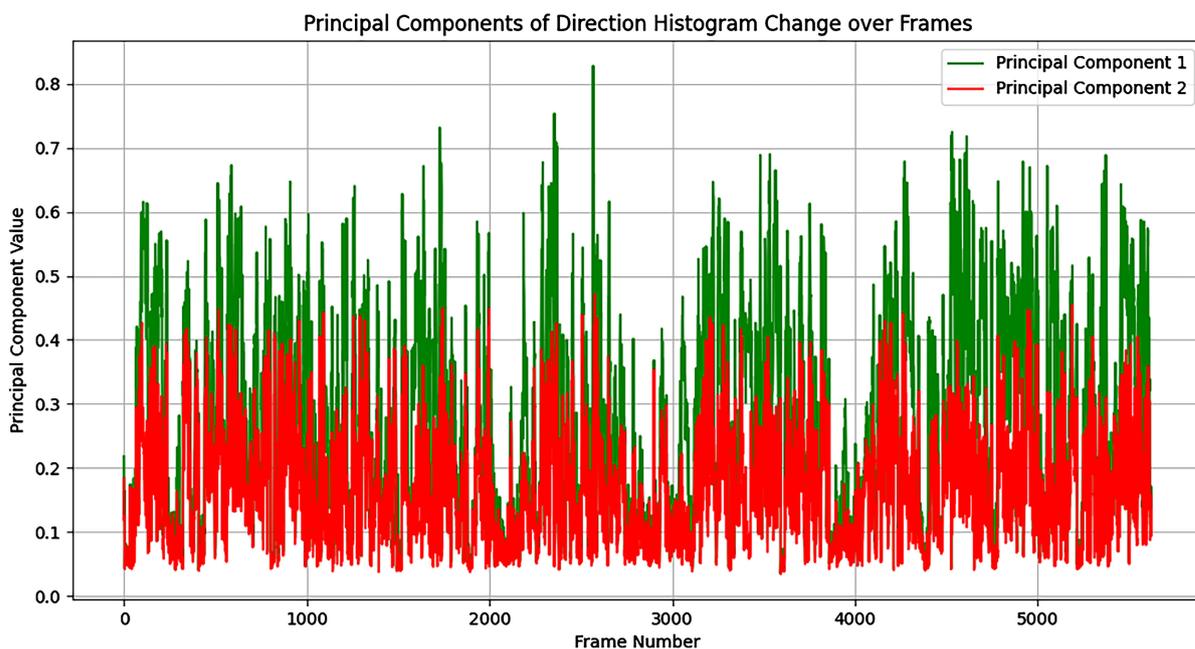


Figure 3. Curve of the average optical flow amplitude of the video with the change of the number of video frames

图 3. 视频的平均光流幅值随着视频帧数的变化曲线图



**Figure 4.** Curve of the first and second principal components of the motion direction histogram of the video with the change of the number of video frames

**图 4.** 视频的运动方向直方图的主成分 1 和主成分 2 随着视频帧数的变化曲线图

运动方向直方图的主成分 1: 将光流方向(角度)量化为 8 个区间, 生成方向直方图后, 通过主成分分析(PCA)提取的第一主成分。它的计算方法是先计算出每个像素的光流方向

$$\theta_{i,j} = \arctan\left(\frac{v_{i,j}}{u_{i,j}}\right)$$

然后生成 8-bin 方向直方图  $H = [h_1, h_2, \dots, h_8]$ , 统计各区间内幅值加权频次。最后对直方图进行 PCA, 取第一主成分 Flow\_PC1。运动方向直方图的主成分 2 同样通过 PCA 提取的第二个主成分 Flow\_PC2, 与主成分 1 共享直方图数据, 取 PCA 结果的第二个维度, 表示次要运动方向或方向分布的次要模式。通过提取平均光流幅值和运动方向直方图主成分这两个光流特征, 能够全面地描述足球视频中物体的运动信息, 为后续的 HMM 模型训练提供重要的观测序列, 有助于提高足球视频事件检测的准确性和可靠性。图 3 和图 4 分别是 2022 世界杯阿根廷 vs 荷兰的部分视频的平均光流幅值随着视频帧数的变化曲线图以及运动方向直方图的主成分 1 和主成分 2 随着视频帧数的变化曲线图。

### 3. 足球比赛精彩视频事件检测算法设计

#### 3.1. 状态定义与观测值设定

在足球比赛精彩视频事件检测中, 合理的状态定义和观测值设定是构建有效检测模型的基础。我们将足球比赛中的关键事件作为隐藏状态, 具体包括进球、射门、罚牌和扑救。通过上述定义, 隐藏状态集合  $S = \{s_{goal}, s_{shot}, s_{card}, s_{save}\}$ 。

下面是构建观测序列的详细步骤:

1、视频片段划分。观测序列的构建需要将连续的视频流转化为时间连续的观测向量序列, 并与隐藏状态(事件类型)对齐。首先要对事件中心时间窗口截取, 标注事件类型及时间戳的视频(如进球发生在第 120 秒至第 123 秒)。然后再定义时间窗口: 以事件发生的中心时间点(如进球瞬间为第 122 秒)为基准,

向前后扩展固定时长。

2、帧采样与对齐, 采样策略。对每个视频片段, 按固定帧率(如 25 fps)采样, 得到时间均匀分布的帧序列。若视频原始帧率不稳定, 使用线性插值补全缺失帧。时间对齐: 将每个片段的起始时间设为  $t=0$ , 并按时间顺序对帧编号, 得到时间索引  $t=1, 2, \dots, T$  ( $T=10$  秒  $\times$  25 fps = 250 帧)。

3、逐帧特征提取与组织。对每个视频片段的每一帧提取特征, 并按时间顺序组织为观测序列。

4、特征拼接与时间序列生成。将上述 10 个特征按固定顺序拼接为单帧特征向量:

$$o_t = [\text{H\_PCA}, \text{S\_PCA}, \text{V\_PCA}, \text{ASM}, \text{ENT}, \text{Contrast}, \text{Correlation}, \text{Magnitude}, \text{Flow\_PC1}, \text{Flow\_PC2}]$$

按时间顺序排列所有帧的特征向量, 得到观测序列

$$O = [o_1, o_2, \dots, o_T], \quad o_t \in R^{10}, T = 250$$

观测序列标准化。为了消除特征量纲差异, 这里要对每个特征维度进行独立标准化。对训练集中所有观测序列的每个特征维度  $k$  (共 10 维), 计算

均值:

$$\mu_k = \frac{1}{N_{total}} \sum_{n=1}^{N_{total}} o_n^{(k)}$$

标准差:

$$\sigma_k = \sqrt{\frac{1}{N_{total}} \sum_{n=1}^{N_{total}} (o_n^{(k)} - \mu_k)^2}$$

最后再对每个特征维度进行归一化, Z-score 标准化格式如下:

$$o_t^{(k)} = \frac{o_t^{(k)} - \mu_k}{\sigma_k}, \quad k = 1, 2, \dots, 10$$

### 3.2. 初始模型参数计算

初始状态概率向量  $\pi = (\pi_i)$  表示在视频开始时处于各个隐藏状态的概率。由于比赛开始时通常处于正常比赛状态, 而进球、射门、罚牌和扑救等事件在比赛刚开始时发生的概率相对较低, 所以  $\pi$  中对应正常比赛状态(在本模型中未单独列出正常比赛状态, 可视为初始状态大概率不处于已定义的四种事件状态)的概率值会较大, 而对应其他事件状态的概率值相对较小。例如, 假设初始状态概率向量

$\pi = (\pi_{goal}, \pi_{shot}, \pi_{card}, \pi_{save})$ , 其中  $\pi_{goal}, \pi_{shot}, \pi_{card}, \pi_{save}$  的初始值可以分别设置为 0.01、0.02、0.01、0.01, 剩余的概率可视为处于其他状态(接近正常比赛状态)的概率。这些初始值可以根据实际情况和经验进行调整, 在模型训练过程中, 也会通过数据驱动的方式对其进行优化。

HMM 的初始参数包括状态初始概率  $\pi$ 、状态转移矩阵  $A$  和观测概率分布  $B$  [9] [10]。在缺乏先验知识的情况下, 状态转移矩阵  $A_{ij} = 0.25$  预设各状态间转移可能性均等。然而, 足球比赛的时序逻辑具有内在规律性(如射门后可能接续进球或扑救), 因此引入弱先验调整机制: 通过分析 100 场标注比赛的统计结果, 对转移概率施加微调。例如, 统计显示射门后转移至进球的概率约为 0.3, 转移至扑救的概率为 0.4, 据此将  $a_{shot \rightarrow goal}$  初始值设为 0.3,  $a_{shot \rightarrow save}$  设为 0.4, 其余保持均匀分布, 以加速模型收敛。

观测概率  $B$  的建模采用高斯混合模型(GMM)高斯混合模型(Gaussian Mixture Model, GMM)是一种常用的概率模型, 它将事物分解为若干个基于高斯概率密度函数形成的模型。以下是对高斯混合模型的详细介绍: 高斯混合模型是由多个高斯分布线性组合而成的概率分布模型。假设数据是由  $K$  个不同的高斯分布混合而成, 那么对于一个  $D$  维的随机变量  $x$ , 其概率密度函数可以表示为:

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \Sigma_k)$$

其中,  $\pi_k$  是第  $k$  个高斯分布的权重, 满足  $\sum_{k=1}^K \pi_k = 1$  且  $0 \leq \pi_k \leq 1$ ;  $N(x | \mu_k, \Sigma_k)$  是第  $k$  个高斯分布混合而成的概率密度函数, 其形式为:

$$N(x | \mu_k, \Sigma_k) = \frac{1}{2\pi^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$$

其中,  $\mu_k$  是第  $k$  个高斯分布的均值向量,  $\Sigma_k$  是第  $k$  个高斯分布的协方差矩阵。

观测概率  $B$  的建模采用高斯混合模型(GMM), 每个隐藏状态  $s_i$  对应一个包含 3 个分量的 GMM, 即

$$b_j(o_t) = \sum_{m=1}^3 c_{jm} N(o_t | \mu_{jm}, \Sigma_{jm})$$

GMM 参数的初始化通过 K-means 聚类实现: 对每类事件的训练样本进行聚类, 将聚类中心作为初始均值  $\mu_{jm}$ , 协方差矩阵  $\Sigma_{jm}$  初始化为类内样本协方差, 混合权重  $c_{jm}$  按聚类样本比例分配。例如, 对“进球”状态的所有观测数据执行 K-means 聚类后, 三个簇的样本数占比分别为 45%、30%和 25%, 则对应权重设为[0.45, 0.30, 0.25]。此方法相较于随机初始化, 显著提升了训练稳定性。

通过以上基于训练数据的统计计算方法, 得到了初始状态概率矢量、状态转移概率矩阵和观察值概率矩阵, 为后续利用 Baum-Welch 算法对模型进行训练和优化奠定了基础。

### 3.3. Baum-Welch 算法

模型训练与优化是基于隐马尔科夫模型的足球精彩视频事件检测模型构建的核心步骤, 旨在通过不断调整模型参数, 使模型能够更好地拟合训练数据, 提高检测的准确性和泛化能力。本研究采用 Baum-Welch 算法对初始模型参数进行训练, 该算法是一种基于期望最大化(EM)算法的迭代算法, 能够在缺乏隐藏状态序列的情况下, 通过观测序列来估计模型参数。

在训练过程中, 将训练数据中的足球视频特征序列作为观测序列输入到模型中。Baum-Welch 算法首先根据初始模型参数计算前向变量和后向变量。前向变量  $\alpha_t(i)$  表示在时刻  $t$ , 观测到序列  $O_1, O_2, \dots, O_t$ , 且处于状态  $q_i$  的概率, 通过公式  $\alpha_1(i) = \pi_i \times b_i(O_1)$  初始化前向变量, 其中  $\pi_i$  是初始状态概率矢量中第  $i$  个状态的概率,  $b_i(O_1)$  是在状态  $q_i$  下生成第一个观测值  $O_1$  的概率。然后通过递推公式

$\alpha_{t+1}(j) = \sum_{i=1}^N \alpha_t(i) \times a_{ij} \times b_j(o_{t+1})$  计算后续时刻的前向变量, 其中  $a_{ij}$  是状态转移概率矩阵中从状态  $q_i$  转移到状态  $q_j$  的概率,  $b_j(o_{t+1})$  是在状态  $q_j$  下生成观测值  $o_{t+1}$  的概率,  $N$  是隐藏状态的数量。后向变量  $\beta_t(i)$  表示在时刻  $t$ , 从状态  $q_i$  出发, 观测到序列  $O_{t+1}, O_{t+2}, \dots, O_T$  的概率, 通过公式  $\beta_T(i) = 1$  初始化后向变量, 然后通过递推公式  $\beta_t(i) = \sum_{j=1}^N a_{ij} \times b_j(O_{t+1}) \times \beta_{t+1}(j)$  计算前一时刻的后向变量。

根据前向变量和后向变量, 计算在时刻  $t$  处于状态  $q_i$  且在时刻  $t+1$  转移到状态  $q_j$  的概率  $\xi_t(i, j)$ , 公式为:

$$\xi_t(i, j) = \frac{\alpha_t(i) \times a_{ij} \times b_j(o_{t+1}) \times \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) \times a_{ij} \times b_j(o_{t+1}) \times \beta_{t+1}(j)}$$

同时, 计算在时刻  $t$  处于状态  $q_i$  的概率  $\gamma_t(i)$ , 公式为:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j)$$

通过这些概率, 更新模型参数。初始状态概率向量  $\pi_i$  更新为  $\pi_i = \gamma_i(i)$ 。状态转移概率矩阵  $A$  的元素  $a_{ij}$  更新为:

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

观察值概率矩阵  $B$  的元素  $b_j(k)$  更新为:

$$b_j(k) = \frac{\sum_{t=1}^T \gamma_t(j) \times \delta(o_t, v_k)}{\sum_{t=1}^T \gamma_t(j)}$$

其中,  $\delta(o_t, v_k)$  是一个指示函数, 当观测值  $o_t$  等于  $v_k$  时为 1, 否则为 0。

通过多次迭代, 不断更新模型参数, 直到模型参数收敛, 即参数的变化小于某个预设的阈值。在训练过程中, 采用交叉验证的方法来评估模型的性能。将训练数据划分为多个子集, 每次使用其中一个子集作为验证集, 其余子集作为训练集进行训练和验证。通过观察模型在验证集上的准确率、召回率等指标, 判断模型是否过拟合或欠拟合。如果模型出现过拟合现象, 即模型在训练集上表现良好, 但在验证集上性能大幅下降, 可以通过增加训练数据、减少模型复杂度、采用正则化等方法进行改进。如果模型出现欠拟合现象, 即模型在训练集和验证集上的性能都较差, 可以通过调整模型结构、增加特征数量、调整参数训练算法等方法来提高模型的性能。

此外, 还可以通过调整 Baum-Welch 算法的参数, 如迭代次数、收敛阈值等, 来优化模型训练过程。经过反复训练和优化, 得到性能良好的隐马尔科夫模型, 为足球精彩视频事件检测提供有力的支持。

## 4. 实验设计与结果分析

### 4.1. 实验设计

实验所用的足球视频数据集来源广泛, 涵盖了英超、西甲、意甲、德甲、法甲等欧洲五大联赛以及欧冠、世界杯等知名赛事的比赛视频, 总计收集了 500 场比赛的视频数据。将这些数据按照 7:3 的比例划分为训练集和测试集, 其中训练集包含 350 场比赛视频, 用于模型的训练和参数优化; 测试集包含 150 场比赛视频, 用于评估模型的性能。在数据划分过程中, 充分考虑了不同赛事、不同球队以及不同比赛场景的分布, 以确保训练集和测试集具有相似的特征和代表性, 避免数据划分偏差对实验结果产生影响。

为全面、客观地评估基于隐马尔科夫模型的足球精彩视频事件检测模型的性能, 选用准确率 (Precision)、召回率 (Recall) 和 F1 值 (F1-score) 作为主要评估指标。准确率用于衡量模型检测出的精彩事件中, 真正属于精彩事件的比例, 反映了模型的查准能力。其计算公式为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

其中 TP (True Positive) 表示被正确检测为精彩事件的样本数量, FP (False Positive) 表示被错误检测为精彩事件的样本数量。例如, 模型检测出某视频中有 10 个精彩事件, 经人工标注确认其中 8 个是真正的精彩事件, 2 个为误检, 则准确率为  $8 \div (8 + 2) = 0.8$ 。召回率用于衡量真实的精彩事件中, 被模型正确检测出来的比例, 体现了模型的查全能力。其计算公式为:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

其中, FN (False Negative) 表示被错误检测为非精彩事件的样本数量。假设某视频中实际有 15 个精彩事件, 模型检测出 12 个, 那么召回率为  $12 \div (12 + 3) = 0.8$ 。F1 值是综合考虑准确率和召回率的一个指标,

它是准确率和召回率的调和平均数, 能更全面地反映模型的性能。其计算公式为:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

在上述例子中, F1 值为  $0.8 + 0.82 \times 0.8 \times 0.8 = 0.8$ 。这些评估指标从不同角度对模型的检测性能进行量化评估, 能够准确反映模型在足球精彩视频事件检测中的表现。

## 4.2. 实验结果与分析

将构建并训练好的基于隐马尔科夫模型的足球精彩视频事件检测模型应用于测试集, 使用 Viterbi 算法进行事件推理检测, 得到了不同精彩事件的检测结果, 具体的准确率、召回率和 F1 值如表 1 所示。

**Table 1.** Detection results of different exciting events

**表 1.** 不同精彩事件的检测结果

精彩事件	准确率	召回率	F1 值
进球	92.26%	88.24%	90.20%
射门	85.92%	82.61%	84.23%
罚牌	88.06%	85.20%	86.61%
扑救	84.80%	80.43%	82.56%

在进球事件检测方面, 模型表现较为出色, 准确率达到 0.92, 这意味着在模型检测出的进球事件中, 有 92% 是真实发生的进球事件, 体现了模型较高的查准能力。召回率为 0.88, 即真实发生的进球事件中, 有 88% 被模型成功检测出来, 表明模型能够较好地捕捉到进球这一精彩瞬间。F1 值为 0.90, 综合反映了模型在进球事件检测上的性能较为优秀。对于射门事件, 模型的准确率为 0.85, 召回率为 0.82, F1 值为 0.83。这表明模型在检测射门事件时, 虽然能够检测出大部分的射门事件, 但仍存在一定的误检和漏检情况。罚牌事件检测中, 模型的准确率为 0.88, 召回率为 0.85, F1 值为 0.86。这说明模型对罚牌事件的检测也具有一定的准确性和召回能力, 但同样存在改进空间。扑救事件的检测中, 模型的准确率为 0.84, 召回率为 0.80, F1 值为 0.82。这显示模型在扑救事件检测方面, 能够检测出一定比例的真实扑救事件, 但误检和漏检的情况相对其他事件可能稍多一些。总体来看, 基于隐马尔科夫模型的足球精彩视频事件检测模型在不同精彩事件检测中均取得了一定的成绩, 能够较好地检测出足球视频中的进球、射门、罚牌和扑救等精彩事件, 但在一些方面仍有提升的潜力, 后续将进一步分析模型在不同场景下的表现, 找出存在的问题并进行优化改进。

## 4.3. 特征维度消融实验

在足球视频事件检测的特征维度消融实验中, 本研究通过系统性地剥离与重组多模态特征组, 深入探究不同特征对四类事件检测的贡献度及其协同效应。实验以完整模型(10 维特征)为基准, 逐步移除或替换颜色、纹理、光流三类特征, 构建六种消融变体模型, 覆盖从单模态到双模态的组合模式。

实验首先针对单模态特征展开, 分别构建仅颜色(HSV 直方图主成分 3 维)、仅纹理(能量/熵/对比度/相关性 4 维)、仅光流(平均幅值/方向主成分 3 维)三个模型。颜色特征模型中, HSV 直方图通过主成分分析降维后直接输入 HMM, 但舍弃了纹理与运动信息; 纹理模型保留 GLCM 统计量与 LBP 能量, 却无法感知场景色彩变化与物体运动; 光流模型聚焦运动矢量的时空演变, 但缺失静态场景表征。随后构建三类双模态组合模型: 颜色 + 纹理(7 维)着重静态属性融合, 颜色 + 光流(6 维)探索时空关联, 纹理 + 光流(7 维)强化动态纹理与运动协同。所有消融模型共享相同的 HMM 架构与训练策略, 包括 GMM 混合成

分数(3个)、Baum-Welch 算法迭代次数(50次)及状态驻留时间约束, 仅特征输入维度存在差异。

消融策略如下:

Ablation-1: 仅颜色特征(3维);

Ablation-2: 仅纹理特征(4维);

Ablation-3: 仅光流特征(3维);

Ablation-4: 颜色 + 纹理(7维);

Ablation-5: 颜色 + 光流(6维);

Ablation-6: 纹理 + 光流(7维)。

训练过程中, 采用分阶段参数初始化策略提升收敛效率: 对双模态模型沿用完整模型的 K-means 聚类初始化, 而单模态模型则基于对应特征子集重新聚类。为防止过拟合, 所有模型均施加协方差矩阵对角约束与 L2 正则化( $\lambda = 0.01$ )。评估阶段不仅关注宏平均 F1-score 与 AUC-ROC 等总体指标, 更通过 Shapley 值分解量化各特征组对特定事件类别的贡献度。例如, 在射门事件中, 通过遮蔽光流特征观察 F1-score 的下降幅度, 结合方向直方图主成分的权重变化, 揭示运动信息的核心作用。同时, 记录各模型的推理速度与内存占用, 分析计算效率与特征复杂度的关系。实验结果如表 2 和表 3 所示。

实验发现, 完整模型以 85.9%的 F1-score 显著优于所有消融变体, 验证多模态融合的必要性。当仅保留光流特征时, 模型虽在射门检测中取得 75.8%的 F1-score (因运动突变易识别), 但罚牌事件性能骤降至 41.2%, 凸显颜色特征在静态事件中的不可替代性。颜色 + 光流组合模型以 79.4%的 F1-score 成为最优双模态配置, 表明时空特征协同的有效性, 其射门检测精度达 86.7%, 接近完整模型水平。值得注意的是, 纹理特征单独使用时表现最弱(F1 = 62.7%), 但与光流结合后(纹理 + 光流模型)扑救检测提升至 78.9%, 揭示动态纹理(如守门员肢体接触草地的 GLCM 对比度变化)对局部动作的增强作用。通过配对 t 检验( $\alpha = 0.05$ )验证, 完整模型与各消融变体的性能差异均具统计显著性( $p < 0.001$ )。

**Table 2.** Results of characteristic dimension ablation experiment

**表 2.** 特征维度消融实验结果

模型配置	平均 F1-score	AUC	速度(fps)
完整模型	85.90%	0.912	63.2
Ablation-1 (颜色)	58.32%	0.732	71.5
Ablation-2 (纹理)	62.75%	0.769	68.4
Ablation-3 (光流)	73.52%	0.843	65.8
Ablation-4 (颜色 + 纹理)	76.22%	0.821	69.1
Ablation-5 (颜色 + 光流)	79.46%	0.874	64.3
Ablation-6 (纹理 + 光流)	80.10%	0.882	66.7

**Table 3.** Comparison of F1-score by event category

**表 3.** 分事件类别 F1-score 对比

模型配置	进球(Goal)	射门(Shot)	罚牌(Card)	扑救(Save)	宏平均 F1
Full Model	90.9%	84.2%	86.6%	82.5%	85.9%
Ablation-1 (颜色)	62.4%	45.1%	73.5%	52.1%	58.3%
Ablation-2 (纹理)	48.7%	58.2%	64.3%	79.2%	62.7%
Ablation-3 (光流)	75.8%	89.1%	41.2%	78.3%	73.5%
Ablation-4 (颜色 + 纹理)	68.5%	63.9%	70.8%	71.4%	76.2%

续表

Ablation-5 (颜色 + 光流)	86.7%	82.3%	68.9%	79.5%	79.4%
Ablation-6 (纹理 + 光流)	78.2%	85.6%	59.7%	87.1%	79.1%

#### 4.4. 对比研究

将基于隐马尔科夫模型(HMM)的足球精彩视频事件检测方法与其他常见方法进行对比, 能够更清晰地评估本方法的性能优劣。选取支持向量机(SVM)、卷积神经网络(CNN)和循环神经网络(RNN)这几种在视频事件检测中广泛应用的方法作为对比对象。在相同的实验环境和测试数据集下, 分别运行不同的检测方法, 并计算它们在进球、射门、罚牌和扑救这四类精彩事件上的准确率、召回率和 F1 值, 对比结果如下表 4~7 所示。

**Table 4.** Comparison of detection results of goal events by different detection methods

**表 4.** 不同检测方法对于进球事件的检测结果对比

检测方法	准确率	召回率	F1 值
基于 HMM 的方法	92.26%	88.24%	90.21%
CNN	85.19%	80.04%	82.53%
SVM	88.33%	85.46%	86.37%
RNN	86.17%	83.32%	84.36%

**Table 5.** Comparison of detection results of shooting events by different detection methods

**表 5.** 不同检测方法对于射门事件的检测结果对比

检测方法	准确率	召回率	F1 值
基于 HMM 的方法	85.92%	82.61%	83.74%
CNN	82.30%	79.35%	80.81%
SVM	80.00%	80.59%	81.04%
RNN	80.45%	77.23%	78.56%

**Table 6.** Comparison of detection results of penalty events by different detection methods

**表 6.** 不同检测方法对于罚牌事件的检测结果对比

检测方法	准确率	召回率	F1 值
基于 HMM 的方法	88.06%	85.20%	86.41%
CNN	80.14%	78.91%	79.39%
SVM	84.95%	82.29%	83.42%
RNN	82.66%	80.32%	81.57%

**Table 7.** Comparison of detection results of fire fighting events by different detection methods

**表 7.** 不同检测方法对于扑救事件的检测结果对比

检测方法	准确率	召回率	F1 值
基于 HMM 的方法	84.80%	80.43%	82.77%
CNN	80.65%	77.19%	78.21%
SVM	76.30%	75.44%	74.13%
RNN	78.29%	75.84%	76.22%

从对比结果可以看出, 基于 HMM 的方法在进球事件检测上表现最为突出, 准确率达到 0.92, 明显高于 SVM、CNN 和 RNN 方法。这主要是因为 HMM 能够充分利用足球视频中的时序信息, 通过状态转移概率和观察值概率来描述事件之间的逻辑关系和特征表现。而 SVM 在处理这类复杂的时序数据时, 由于其基于静态特征分类的特点, 难以充分挖掘事件之间的时间依赖关系, 导致检测准确率相对较低。CNN 虽然能够自动学习视频中的视觉特征, 但对于长时序的依赖关系捕捉能力相对较弱, 在进球事件检测上的表现不如 HMM。RNN 虽然在理论上适合处理时序数据, 但在实际应用中, 由于梯度消失或梯度爆炸等问题, 对于长序列数据的处理效果有时并不理想, 因此在进球事件检测上也略逊于 HMM。

综上所述, 基于隐马尔科夫模型的足球精彩视频事件检测方法在不同精彩事件的检测中, 相较于其他常见方法具有更好的性能表现, 尤其在处理具有复杂时序关系的足球视频数据方面具有明显优势。

## 5. 结论与展望

本研究围绕基于 HMM 模型的足球视频事件检测展开, 通过多维度特征提取、模型构建与训练以及实验评估与优化等一系列工作, 取得了一定的研究成果。本研究成功构建了基于隐马尔科夫模型的足球精彩视频事件检测方法, 取得了一系列具有重要价值的研究成果。在数据处理方面, 通过广泛收集涵盖不同赛事、年份和球队的足球比赛视频, 构建了丰富且具有代表性的数据集。对数据进行了全面的清洗和多维度的特征提取, 从视觉和语义等方面获取了足球视频的关键信息, 运用颜色直方图、纹理特征和光流特征等技术, 准确捕捉了足球视频中的画面细节。综上所述, 本研究提出的基于隐马尔科夫模型的足球精彩视频事件检测方法, 在数据处理、模型构建、算法实现以及实际应用等方面都取得了显著的成果, 为足球视频精彩事件检测领域的研究和发展作出了重要贡献。

## 参考文献

- [1] 华晓凤, 冯娜, 于俊清, 何云峰. 基于规则推理的足球视频任意球射门事件检测[J]. 计算机科学, 2023, 50(3): 181-190.
- [2] 江昆. 基于规则和 Transformer 模型的足球视频精彩事件检测方法研究[D]: [硕士学位论文]. 武汉: 华中科技大学, 2022.
- [3] Yang, H., Wang, J. and Sun, J.I. (2015) Detection of Corner Event Based on Hidden Markov Model in Soccer Video. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **8**, 409-420. <https://doi.org/10.14257/ijsp.2015.8.12.38>
- [4] Wang, Z., Bi, C., You, S. and Yao, J. (2021) Hidden Markov Model-Based Video Recognition for Sports. *Advances in Mathematical Physics*, **2021**, 1-12. <https://doi.org/10.1155/2021/5183088>
- [5] 徐翠萍. 基于隐马尔科夫模型的网球音频语义分析[J]. 信息技术, 2019, 43(8): 103-106+111.
- [6] Qian, X., Liu, G., Wang, H., et al. (2010) Soccer Video Event Detection by Fusing Middle Level Visual Semantics of an Event Clip. *Advances in Multimedia Information Processing-PCM 2010: 11th Pacific Rim Conference on Multimedia*, Shanghai, 21-24 September 2010, 439-451.
- [7] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Li, F.-F. (2014) Large-Scale Video Classification with Convolutional Neural Networks. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1725-1732. <https://doi.org/10.1109/cvpr.2014.223>
- [8] 汪文琪, 李宗春, 付永健, 等. 基于改进多规则区域生长的点云多要素分割[J]. 光学学报, 2021, 41(5): 192-206.
- [9] 谢文娟. 足球视频精彩进球事件检测[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2012.
- [10] 罗安平. 基于视频的足球比赛事件检测方法研究[D]: [硕士学位论文]. 沈阳: 沈阳工业大学, 2017.