

基于改进 YOLOv5 的目标智能检测方法 在罚球姿势中的应用

顾玉恒

(上海建桥学院 教育学院, 上海 201306)

摘要: 针对智能体育视域下的罚球智能命中预测问题, 研究提出了基于 YOLOv5 网络改进损失函数和引进注意力机制的目标检测网络完成对比赛视频中罚球球员的准确识别与有效提取; 然后设计了一种基于轨迹优化的识别方法进行罚球球员的姿态检测, 并利用支持向量机分类器进行篮球罚球命中预测; 实验结果显示研究提出的目标检测算法在训练 15 轮时就达到了 85% 的平均精度, 且命中预测算法的进球预测准确性也增加了 5%, 优于实验中的其他算法; 实验结果表明改进损失函数并引入注意力机制的视频罚球球员检测跟踪方法能够有效检测罚球球员所在区域, 并准确将罚球球员的特写视频提取出来; 篮球教练可以根据经验, 结合姿态估计和罚球命中预测数据, 有效评价动作表现并给出指导建议, 既能实现运动姿态矫正, 又能帮助评估运动员水平, 进而建立智能化运动员数据档案。

关键词: 智能体育; 篮球比赛视频; 目标检测; 姿态估计; 罚球命中预测

Application of Target Detection Method Based on Improved YOLOv5 in the Prediction of Different Free Throw Positions

GU Yuheng

(College of Education, Shanghai Jianqiao University, Shanghai 201306, China)

Abstract: Aimed at the problem of intelligent hit prediction of free throws in the context of smart sports, a target detection network based on the improved loss function of YOLOv5 network and the introduction of an attention mechanism is proposed. to complete the accurate identification and effective extraction of free throw players in game videos. Then a recognition method based on trajectory optimization is designed for the pose detection of free throw players, and a support vector machine classifier is used for the basketball free throw hit prediction. The experimental results show that the target detection algorithm proposed in the study achieves an average accuracy of 85% after 15 rounds of training, while the accuracy of the hit prediction algorithm also increased by 5%, which is superior to other algorithms in the experiment. The experimental results show that the improved Loss function and introduction of attention mechanism can effectively detect the area where the free throws are located, and accurately extract the close-up video of the free throws. Basketball coaches can combine posture estimation with the free throw hit prediction data based on experience to effectively evaluate the movement performance and provide guidance suggestions. This can not only achieve posture correction but also help evaluate athlete level, thereby establishing intelligent athlete data archives.

Keywords: intelligent sports; basketball game videos; target detection; attitude estimation; free throw hit prediction

0 引言

计算机视觉中基于视频的目标检测与人体姿态估计一直是研究热门方向, 人体姿态估计的前提就是对人体直观特征的提取, 根据提取出的结果可以进行动作分析、行为检测等^[1-2]。体育赛事和体育训练对运动员的动作规范有较高的要求, 一些错误的动作不仅会影响运动员的训练效果, 甚至可能会引发运动员的伤病, 而人眼观测对运动员的动作评测效率低、准确度差, 且具有一定的风险性。在智能体育视域下, 可以根据提取出的运动员身体参数及运动记

录辅助运动员进行训练, 其效率、准确性和稳定性都要优于人眼观测方法^[3-4]。篮球比赛是热度较高的一项运动, 每年的篮球各大赛事都具有极高的关注度, 这些比赛年龄跨度大, 从小学生到职业队集中了多种体型的运动员^[5]。近些年对于篮球比赛中影响比赛胜负的因素的研究也因此展开, 随着篮球比赛吹罚的尺度越来越严格, 罚球得分在整场比赛得分所占的比重也在逐渐上升, 球队对于球员的罚球命中率也十分的重视。罚球是一项结合全身的动作, 从手部到腿部都要进行配合发力, 一些高水平球队对技术动作的规格化有着显著的相似性。因此, 规范化的罚球姿势

收稿日期: 2023-06-12; 修回日期: 2023-06-16。

作者简介: 顾玉恒(1978-), 男, 大学本科, 讲师。

引用格式: 顾玉恒. 基于改进 YOLOv5 的目标智能检测方法在罚球姿势中的应用[J]. 计算机测量与控制, 2023, 31(12): 290-295.

对提高罚球命中率有重要作用，对运动员的罚球命中率进行预测和记录反过来又能帮助教练和运动员进行运动姿态矫正。要实现篮球比赛视频中罚球命中的预测，首先需要准确检测出视频中的罚球球员，然后将识别到的特征输入分类器完成命中预测^[6]。文献 [7] 针对驾驶过程中的路况信息提取和目标识别，设计了一种基于 YOLOv3 优化网络的智能系统，该系统通过复制一部分主干网络构建辅助网络，以提升对特征的提取能力，得到了 84.76% 的平均检测准确率^[7]。文献 [8] 针对视频监控，提出了一种基于姿态估计的人体异常行为识别算法，该算法通过提取人体的骨骼关键点坐标提取高级特征，并使用 Softmax 进行行为分类，对实时视频的检测识别帧率达到了 25 帧/秒^[8]。虽然目前对于目标检测和姿态估计的研究均取得了不错的成果，但是目标检测和人体姿态估计结合在体育视频场景中的应用还较少。随着体育赛事尤其是篮球比赛受到的关注越来越多，篮球比赛视频的数量和质量都有很大的提升，使用智能分析辅助体育训练和赛事分析具有很高的可行性和应用意义^[9-10]。因此研究基于 YOLOv5 网络和 OpenPose 算法完成针对篮球视频罚球过程的球员检测与命中预测，目的是辅助教练进行赛事复盘和训练计划制定，运动员也能根据姿态估计和命中率预测的结果分析自己的技能掌握水平，进而实现罚球和整体技能水平的提升。研究期望通过目标检测、姿态估计和命中预测方法为智能体育的发展提供理论依据和应用思路。

1 篮球比赛视频中罚球球员识别与罚球命中概率预测

1.1 引入注意力机制的 YOLOv5 运动员检测方法

在体育领域的视频分析综合功能中，人体目标检测方法可以对球场上的运动员、裁判甚至是观众等展开识别与检测。在进行比赛技术分析的时候，因为只有运动员的表现与比赛的走势有着直接的关系，因此希望尽量忽略裁判和观众的行为，将注意力集中在运动员的行为上。要分析篮球比赛中的罚球情况时，单纯地将球员、观众、裁判等区分开来还远远不够，必须进一步对发球运动员的动作和

行为进行识别与检测。与行人检测相比，对处于运动状态的罚球球员的跟踪更加具有挑战性。研究以区域划分方法为依据，使用回归目标检测算法来对篮球比赛场景下的罚球运动员展开检测，从而实现更加准确、高效的目标对象检测和定位，进而实现对罚球运动员姿态的识别。基于罚球球员检测与姿态估计的篮球罚球命中概率预测方法流程如图 1 所示。

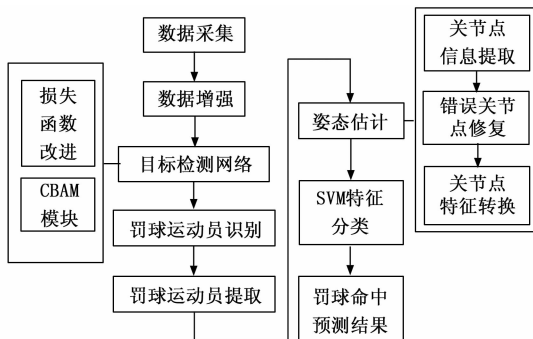


图 1 篮球比赛视频中罚球球员识别与罚球命中概率预测

如图 1 所示，通过线下拍摄和网络收集的方式得到的实验数据集采用基于 YOLOv5 的目标检测网络识别并提取出视频中的罚球运动员，然后对罚球运动员进行姿态估计，最后通过分类器对运动员罚球特征进行分类来输出当前运动员罚球命中与否的结果，实现罚球命中率预测。想要完成运动员罚球命中的预测，首先要对视频中的罚球球员进行准确检测。因此，研究首先针对篮球比赛视频中如何准确检测罚球球员展开讨论。传统的目标检测算法包括候选区域选取、特征提取和训练分类 3 个阶段，但是其存在着许多局限性。而基于深度学习的目标检测方法通过卷积神经网络 (CNN, convolutional neural network) 对深度特征进行提取，对于环境不稳定下的目标识别与定位有更好的效果，尤其是 YOLOv5 在小目标的检测中拥有优越的性能表现，适合用于对篮球比赛视频中的罚球运动员进行检测和识别^[11-12]。YOLOv5 网络由主干网络、中间网络和输出三部分组成，其结构如图 2 所示。

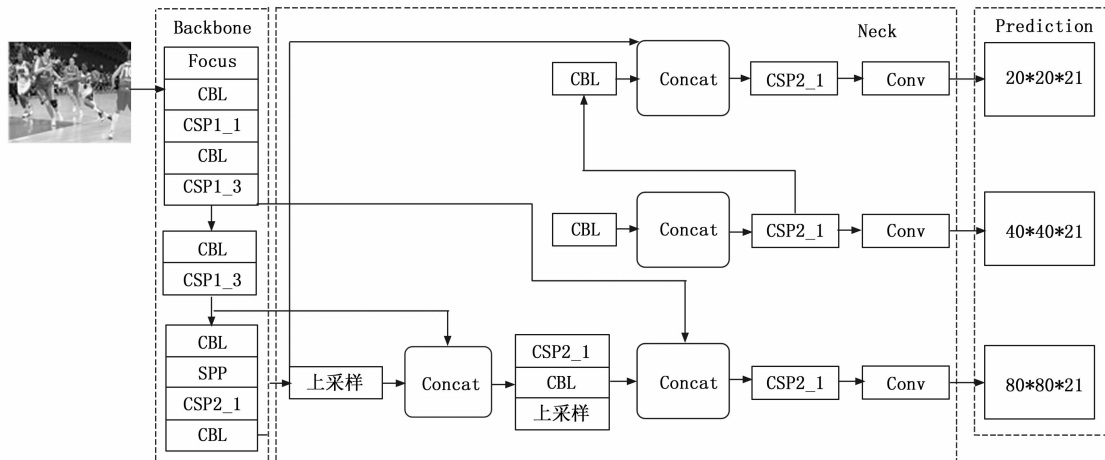


图 2 YOLOv5 检测网络的结构图

如图 2 所示, YOLOv5 的主干网络通过 CNN 生成四层特征图, 并通过多次卷积从输入图像中进行多尺度的特征映射提取。YOLOv5 的中间网络将主干网络提取的特征图进行融合以获得更多的上下文信息。中间网络提取的新特征映射导入到输出部分实现对目标的检测与识别分类。在 YOLOv5 的输入端部分增加了自适应锚框计算和自适应图片缩放, 这样在每次训练的时候通过重新计算锚框来匹配不同的训练集以获得最合适的锚框值。自适应图片缩放使用缩放系数和添加黑色像素块来完成, 通过压缩图像宽度减少上下黑边的宽度, 改进了黑边造成的信息冗余和检测速度下降。Backbone 是网络的主干, 其中 Focus 模块的作用是进行图片的切片和连接, 使得下采样时可以更好地提取特征^[13]。Neck 连接了 Backbone 和输出层, 使得特征提取更加充分, 为网络输出做准备, 输出层将此前提取到的特征或处理特征的工作转换成项目期望的结果进行输出。为了进一步优化网络模型的目标检测精度, 研究引入卷积注意力模块 (CBAM, convolutional block attention module) 加强对检测目标的关注, 用于解决环境复杂导致的精度下降问题。CBAM 包括输入模块、通道注意力模块、空间注意力模块和输出模块^[14-15]。特征图经过两个注意力模块生成数个含权重的注意力图, 通过权重与输入进行乘积运算输出更加突出的特征图。CBAM 的机制与结构如图 3 所示。

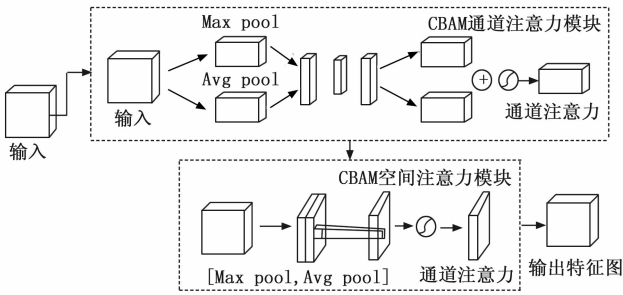


图 3 CBAM 的运行机制与组织结构图

如图 3 所示, 特征图经过通道注意力机制在空间维度上进行最大值池化和平均值池化, 得到的两个只有通道维度的向量进行连接和相加后通过 Sigmoid 函数得到通道注意力向量。然后经过空间注意力机制在通道维度上进行最大值池化和平均值池化, 得到的结果基于通道进行连接和卷积后降维为 1 个通道, 仍旧通过 Sigmoid 函数生成空间注意力向量。传统的 YOLOv5 算法采取引入了惩罚项的广义交并比 (GIoU, generalized intersection over union) 损失函数对候选框置信度误差计算进行优化, 解决了 IoU 损失函数在预测框不相交时产生梯度的问题。但是 GIoU 为了达到更好的检测效果, 会通过增大检测框来减小空白面积的值, 导致网络的收敛速度下降, 于是考虑预测框与目标框中心点位置的距离 IoU (DIoU, distance-IoU) 应运而生^[16-17]。由于研究的检测对象是篮球视频中的罚球球员, 预测框的

长宽比变化波动比较小, 因此研究在 DIoU 的基础上考虑候选框的长宽比因素, 通过改进后的完整 IoU (CIoU, complete-IoU) 损失函数进行网络训练。CIoU 损失函数 L_{CIoU} 的定义如式 (1) 所示:

$$\begin{cases} L_{CIoU} = 1 - IoU(B_P, B_T) + \frac{\rho^2(B_{P_c}, B_{T_c})}{l^2} + \alpha v \\ v = \frac{4}{\pi^2} \left(\arctan \frac{w_{B_P}}{h_{B_P}} - \arctan \frac{w_{B_T}}{h_{B_T}} \right) \end{cases} \quad (1)$$

式 (1) 中, α 为权重函数, v 用来度量目标框 B_T 与检测框 B_P 之间的相似性, 作用通过控制检测框的大小使其在较短的时间内接近目标框的大小; w_{B_P} 和 h_{B_P} 为预测框的宽和高, w_{B_T} 和 h_{B_T} 为目标框的宽和高; $\rho(B_{P_c}, B_{T_c})$ 为检测框中心点 B_{P_c} 和目标框中心点 B_{T_c} 的欧式距离; l 是同时包围预测框和目标框的最小区域的对角线长度。

1.2 基于运动员姿态估计的罚球命中率预测模型

罚球命中的结果主要根据运动员罚球时的人体姿态进行预测, 首先要求对罚球队员进行目标检测, 研究已经基于 YOLOv5 建立了目标运动员检测算法。因此接下来基于目标检测算法对识别提取后的罚球队员进行人体姿态估计, 以完成对运动员罚球的命中率预测。人体姿态估计是进一步对人体行为进行分析处理的关键问题, 也是当前领域研究的热点之一。对于不同角度不同尺度的人体姿态识别, 由于图片或视频中人与其他事物之间存在遮挡容易对识别效果产生干扰, 并且光线、衣着以及环境变化也会对姿态估计的效果产生影响^[18]。由于静态图片只能获取到固定的肢体信息, 而视频分析能够根据前后时间段信息获取到更多的肢体信息从而进行修正, 因此基于视频的人体姿态估计相较于基于静态的人体姿态估计更有优势。研究在 OpenPose 算法的基础上, 通过基于轨迹优化的识别方法对人体姿态检测方法在视频中提取出的结果进行初步筛选, 然后根据结果进行肢体特征角度转换和特征融合, 最后通过支持向量机 (SVM, support vector machine) 分类器完成命中预测。OpenPose 算法是一种多目标二维姿态检测算法, 基于连续的运动视频帧提取运动员的关节点特征信息, 并以此对运动员的状态进行分析^[19-20]。OpenPose 网络的结构如图 4 所示。

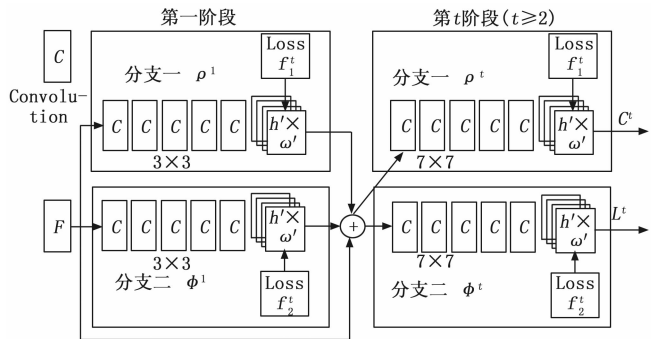


图 4 OpenPose 网络的结构图

OpenPose 的主体网络使用 vggNet 为框架，其中的两个分支分别用于骨骼关节检测和关节连接匹配度向量场预测。网络的第一阶段从分支 ρ 和 φ 输出一组关键点预测置信图 $C = (C_1, C_2, \dots, C_j)$ 和一组关节连接匹配度向量场 $L = (L_1, L_2, \dots, L_m)$ 。两个分支通过迭代操作不断提高预测的精度。分支将前一个阶段的分支网络输出 C^{t-1} 、 L^{t-1} 和输入图像的特征映射 F 作为输入进行迭代运算，输出 C^t 和 L^t 的计算如式 (2) 所示：

$$\begin{cases} C^t = \rho^t(F, C^{t-1}, L^{t-1}), \forall t \geq 2 \\ L^t = \varphi^t(F, C^{t-1}, L^{t-1}), \forall t \geq 2 \end{cases} \quad (2)$$

为了网络结构可以迭代地预测各分支的结果，在上下分支网络的每个阶段结束时设置了 L2 损失函数。第 t 个阶段的两个分支网络的损失函数 f_c^t 和 f_l^t 如式 (3) 所示：

$$\begin{cases} f_c^t = \sum_j \sum_p W(p) \cdot \|C_j^*(p) - C_j^*(p)\|_2^2 \\ f_l^t = \sum_{m=1}^M \sum_p W(p) \cdot \|L_m^t(p) - L_m^*(p)\|_2^2 \end{cases} \quad (3)$$

式 (3) 中， $C_j^*(p)$ 是第 p 帧的真实标签的身体部位位置信度图， L_m^* 是真实标签的部分亲和力向量场， W 是一个在缺少图像位置标签等情况下用于避免错误惩罚的进制掩码。复杂环境下的人体姿态估计需要解决的一个关键问题就是背景或遮挡被误检为关节点，必须对这些错误识别的关节点需要修复操作，否则会影响球员检测的效果。通过计算相邻帧的姿态帧间距离 $d(K_1, K_2)$ 对视频中的球员进行判断追踪，其计算公式如式 (4) 所示：

$$d(K_1, K_2) = \sum_i \frac{y_i}{x_i} \quad (4)$$

式 (4) 中，围绕相邻帧姿态 k_1 和 k_2 的第 i 个身体部位 k_1^i 与 k_2^i 分别提取边界框 B_1^i 和 B_2^i ， x_i 为从 B_1^i 中提取出的特征点个数， y_i 为从 B_2^i 中提取出的特征点个数。当身体部位位置信度低于设置的阈值 th_b 时，以此将该身体部分关节点对应的姿态边界框放大一、二、三倍，并且分别用 tr_1 、 tr_2 、 tr_3 倍尺度采样后重新进行姿态识别，若仍低于设置阈值，则计算前一帧与当前姿态的相似度 $Sc_{g,n}^i$ ，计算方式如式 (5) 所示：

$$Sc_{p,q}^i = \alpha * \sum_i \frac{n_i}{m_i} + (1 + \alpha) * \|H_p - H_q\| \quad (5)$$

式 (5) 中， m_i 为第 p 帧姿态中第 i 个关节点边界框的特征点个数， n_i 是第 q 帧中第 i 个关节点边界框与之相匹配的特征点个数。相似度高于阈值 th_b 时，该帧的候选关节点为前序帧的身体部位关节点，相似度低于阈值 th_b 时，该帧的身体部位关节点信息将被清除。罚球姿势的分析首先是对球员的身体姿态进行识别，然后以人体关节点的相对位置信息为分类器输入特征进行分类，这就需要采集球员的身体形态数据，消除身体形态差异导致的同一球员出现不同的特征值。因此研究用角度特征代替关节点的相对位置，对提取到的关键点展开联系，并将关键点之间的联系作为特征。完成一个完整的投篮动作的关键部位是下蹲时

的膝盖弯曲程度和投篮手抬起时手肘的弯曲程度，即关键特征为大腿与小腿的夹角、大臂与小臂的夹角、大臂与躯体的夹角。在关键特征的分析中将 3 个关节点表示为两个向量，以右手手臂为例，定义右手腕的坐标为 $R_w(x_0, y_0)$ ，右手肘的坐标为 $R_e(x_1, y_1)$ ，右肩膀关节点的坐标为 $R_s(x_2, y_2)$ ，则右手手臂的夹角如式 (6) 所示：

$$\beta = \cos^{-1} \left(\frac{I_1 \cdot I_2}{\|I_1\| \|I_2\|} \right) \quad (6)$$

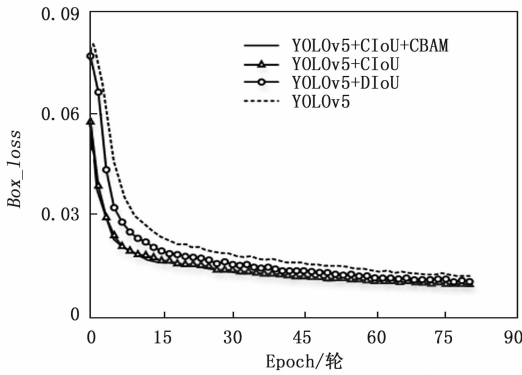
式 (6) 中， I_1 与 I_2 分别为右小臂与右大臂的向量，且 $I_1 = (x_0 - x_1, y_0 - y_1)$ ， $I_2 = (x_2 - x_1, y_2 - y_1)$ 。人体姿态识别出球员的关节点绝对位置坐标后，对连接两个肢体的 3 个坐标点进行角度转换，然后将角度特征送入 SVM 分类器进行预测。分类学习的基本思想是基于样本集在样本空间中寻找一个能够将两个种类划分的超平面，分类任务就是寻找用于分类学习的最优超平面。SVM 分类器具有数学理论的支撑，同时摆脱了统计方法的依赖，简化了分类问题，但是其训练时间较长，不适用于样本数量较大的任务训练。

2 基于视频的罚球球员检测与罚球命中概率预测结果分析

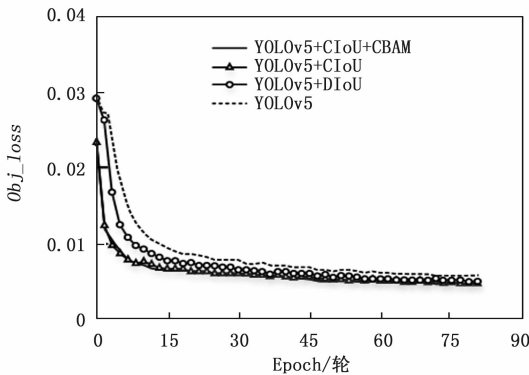
2.1 基于 YOLOv5 的运动员检测跟踪算法性能分析

篮球比赛罚球运动员的目标检测实验要求数据集应包括多场地、多尺度和多角度的比赛场景，因此研究在线下拍摄了大量罚球训练的视频，同时在网络上收集了篮球比赛现场的视频。然后对视频进行预处理，通过关键帧提取得到罚球的视频片段，并将视频分为两部分，分别用于罚球运动员识别和罚球命中预测的训练中。用于目标识别的数据集需要人工标注图片中特定对象的类型和位置信息。罚球命中预测中使用标签标记视频中的罚球队员和对应的罚球结果，以此作为训练的数据集。为了增加训练样本的丰富性和模型的泛化性能，采用数据增强的方法处理训练数据集。采集和处理后得到的实验数据集中训练集与测试集的图像数量为 8:2。罚球运动员检测中存在 3 种目标类型：罚球球员、普通球员、观众，因此需要通过修改分类器减少网络参数的个数，这样才能实现较小计算量下的高效率高精度检测，具体操作是对 YOLOv5 分类器进行修改并且设置仅识别罚球球员。目标框与预测框匹配的损失和目标检测的损失曲线如图 5 所示。

图 5 (a) 中， Box_loss 为目标框与预测框匹配的损失值，图 5 (b) 中， Obj_loss 为目标检测的损失值。由于目标检测中修改分类器仅识别罚球球员，分类损失为 0，因此不进行分类损失的数值比较。如图 5 (a) 所示，4 种算法损失逐渐收敛至稳定时的训练轮数分别为 60 轮、45 轮、30 轮、30 轮，而研究提出的识别检测算法的初始损失值在 0.06 以下，是 4 种方法中最低的，说明研究提出的目标检测算法 YOLOv5+CioU+CBAM 的效果是最好的，能够用



(a) 目标框与预测框的匹配损失曲线



(b) 目标检测的损失曲线

图 5 目标框与预测框的匹配损失曲线和目标检测的损失曲线

于命中预测的罚球球员跟踪检测。4 种算法的平均识别精度结果如图 6 所示。

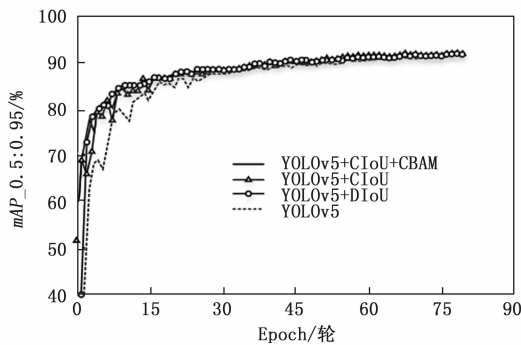


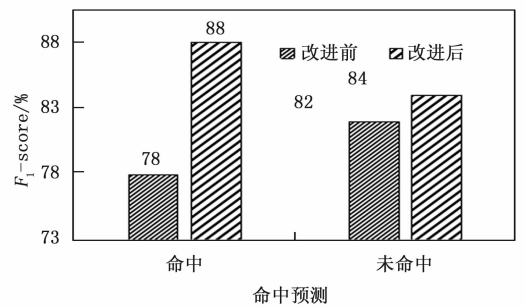
图 6 不同目标检测算法的平均精度曲线图

图 6 中，改进后的模型与原始的 YOLOv5 模型在相同配置下分别训练 80 轮，模型在训练 60 轮后逐渐趋于稳定，平均精度稳定在 90% 左右。采用几种不同损失函数的模型训练效果均较为良好，没有出现拟合与欠拟合的现象。原始的 YOLOv5 网络在 30 轮时达到了 85% 的平均精度，而研究提出的 YOLOv5 + CIoU + CBAM 算法在 15 轮时的平均精度就达到 85%，表现出更优秀的运行效率和收敛精度。目标检测模型的有效性检验结果验证了模型改进的可行性与优越性，能够满足后续姿态估计和罚球命中预测的

要求。

2.2 基于人体姿态估计的篮球罚球命中预测模型性能分析

研究提出的识别检测较为稳定，能够准确、完整地将运动员罚球过程中的人体姿态抽取出来，并且将抽取出的运动员姿态以视频帧的形式保存，为罚球命中预测提供了前提条件。基于人体姿态估计的篮球罚球命中预测模型的性能分析实验基于目标检测实验中提取到的关键帧进行，首先构建 OpenPose 网络进行罚球运动员的关节特征信息提取，然后将罚球过程关键帧的肢体夹角作为分类器的特征输入。改进前后的算法准确性测试结果与 ROC 曲线如图 7 所示。



(a) 改进前后的算法准确性测试结果

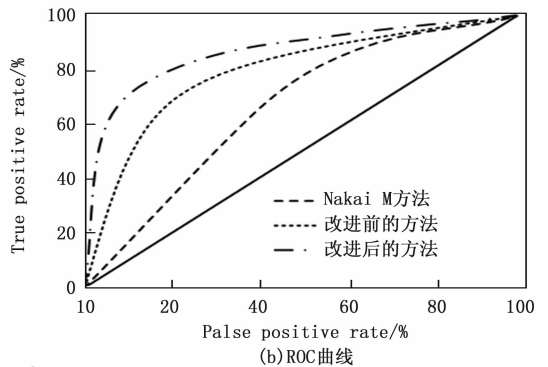


图 7 改进前后的算法准确性测试结果与 ROC 曲线图

研究使用 Nakai M 方法与改进前的姿态估计方法作为对比算法。Nakai M 方法使用 OpenPos 骨骼识别数据建立姿态分析模型，该模型采用逻辑回归来预测篮球罚球的命中概率。图 7 (a) 中，改进后算法的罚球命中预测准确率较改进前增加了 5%，证明运动员体型的变化会对部位关节的相对位置产生影响，仅使用相对位置作为特征输入会影响预测的效果。从图 7 (b) 的 ROC 曲线对比结果可以看出，研究改进后的预测模型明显能够得到更好的预测效果。不同特征对罚球命中预测结果的影响如图 8 所示。

图 8 是通过观察不同肢体夹角特征对罚球命中预测结果的影响度，来研究投篮过程中的关键特征。具体措施是通过在原实验的基础上依次将特定肢体特征从实验中去除后，再进行投篮命中预测。根据不同特征对实验结果的影响度来分析判断各个关键特征对罚球结果的作用程度，以

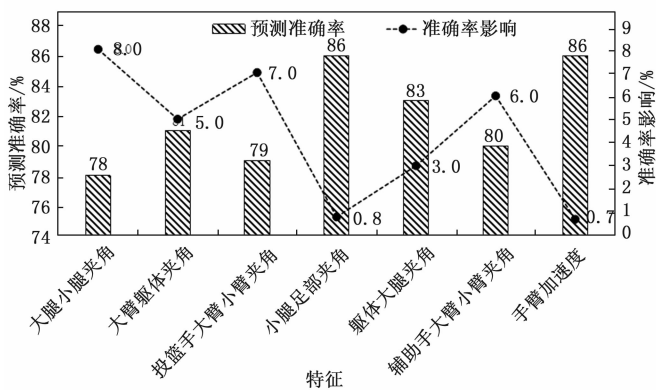


图 8 不同特征对罚球命中预测结果的影响

此作为改进投篮动作的指导理论。如图 8 所示，大腿与小腿的夹角、投篮手大臂与小臂的夹角和辅助手大臂与小臂的夹角对预测准确率的影响均在 6% 以上，去除这些特征后，对罚球命中的预测率均下降至 80% 及以下，结果表明这些夹角特征对投篮命中有重要影响，因此要着重训练有关特征夹角的规范度，这样能够较为迅速地提高罚球命中水平。小腿足部夹角和手臂加速度对预测准确率的影响程度不足 1%，对实验结果的影响较小。具体分析关键夹角发现，当运动员下蹲至夹角等于 120° 时为最佳进球角度，而进球动作初始的大臂与小臂的最佳夹角为 80° ，出手后的大臂躯干夹角则在 130° 左右，在罚球训练时应重视这 3 个特征角度的规范性。

3 结束语

随着深度学习与姿态识别在近年来的研究中取得了越来越多的研究成果，如何应用在体育赛事中成为了研究热点。研究以篮球比赛视频中的罚球片段为对象，通过数据增强和添加标记进行数据处理，对基于 YOLOv5 的目标检测网络进行了损失函数改进和注意力机制引进，以此完成罚球球员的有效提取。在准确检测目标运动员的基础上，进行基于罚球姿态估计的命中预测方法使用角度特征替换位置信息进行改进，提取转换后的特征使用 SVM 分类器进行预测。实验结果表明研究提出的目标检测算法的初始损失值在 0.06 以下，且训练 15 轮时的平均精度就达到了 85%，其检测性能能够用于罚球命中预测。改进后的命中预测算法的进球预测准确性较改进前增加了 5%，且其 ROC 曲线的表现优于实验中的其他算法，验证了使用肢体夹角作为特征输入的必要性 and 基于姿态估计的命中预测方法的有效性与优越性。不过研究仅考虑了室内比赛的场景，后续应针对室外或特殊场景进行扩展与优化。

参考文献：

- [1] 冯杰, 郑建立. 基于卷积与 Transformer 的人体姿态估计方法对比研究 [J]. 软件工程, 2023, 26 (3): 18-24.
- [2] WANG J, SUN K, CHENG T, et al. Deep high-resolution rep-

resentation learning for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43 (10): 3349-3364.

- [3] 贺义兴, 于少勇. 信息化时代智能体育及其产业研究 [J]. 当代体育科技, 2021, 11 (9): 177-180.
- [4] 宗立波, 宋一凡, 王熠明, 等. 体育视频分析中姿态估计进展的综述 [J]. 小型微型计算机系统, 2020, 41 (8): 1751-1757.
- [5] 张禹, 魏一凡, 陶靖婷, 等. 篮球运动员的多目标追踪与轨迹预测 [J]. 中国运动医学杂志, 2021, 40 (10): 800-809.
- [6] 马骁, 闫育东. 基于多尺度时空特征的篮球场景中人体姿态估计 [J]. 中南民族大学学报 (自然科学版), 2023, 42 (1): 95-102.
- [7] XU Q, LIN R, YUE H, et al. Research on small target detection in driving scenarios based on improved yolo network [J]. IEEE Access, 2020, 8: 27574-27583.
- [8] 李建更, 谢海征. 基于姿态估计的人体异常行为识别算法 [J]. 北京工业大学学报, 2022, 48 (7): 710-720.
- [9] LU S, WANG B, WANG H, et al. A real-time object detection algorithm for video [J]. Computers & Electrical Engineering, 2019, 77: 398-408.
- [10] 张红民, 庄旭, 郑敬添, 等. 优化 YOLO 网络的人体异常行为检测方法 [J]. 计算机工程与应用, 2023, 59 (7): 242-249.
- [11] 奉志强, 谢志军, 包正伟, 等. 基于改进 YOLOv5 的无人机实时密集小目标检测算法 [J]. 航空学报, 2023, 44 (7): 251-265.
- [12] 李丽, 张荣芬, 刘宇红, 等. 基于多尺度注意力机制的高分辨率网络人体姿态估计 [J]. 计算机应用研究, 2022, 39 (11): 3487-3491.
- [13] TANG Y, ZOU W, HUA Y, et al. Video salient object detection via spatiotemporal attention neural networks [J]. Neurocomputing, 2020, 377: 27-37.
- [14] 周升儒, 陈志刚, 邓伊琴. 基于 PoseC3D 的网球动作识别及评价方法 [J]. 计算机工程与科学, 2023, 45 (1): 95-103.
- [15] 蔡兴泉, 霍宇晴, 李发建, 等. 面向太极拳学习的人体姿态估计及相似度计算 [J]. 图学学报, 2022, 43 (4): 695-706.
- [16] 贺愉婷, 车进, 吴金蔓. 基于 YOLOv5 和重识别的行人多目标跟踪方法 [J]. 液晶与显示, 2022, 37 (7): 880-890.
- [17] 杨文涛, 张维光. 基于改进 YOLOv5m 的弱小目标识别方法 [J]. 计算机测量与控制, 2022, 30 (12): 218-223, 231.
- [18] 王芙蓉, 吴静静. 基于关键姿态的快递场景人-物交互行为识别方法 [J]. 计算机测量与控制, 2022, 30 (6): 182-189.
- [19] 张富凯, 贺天成. 结合轻量 Openpose 和注意力引导图卷积的动作识别 [J]. 计算机工程与应用, 2022, 58 (18): 180-187.
- [20] 杨君, 张素君, 张创豪, 等. 基于 OpenPose 的人体动作识别对比研究 [J]. 传感器与微系统, 2021, 40 (1): 5-8.