基于自适应关键点热图的遮挡篮球运动员检测算法

任媛¹⁾, 雒江涛^{2)*}, 梁旭鹏³⁾

(重庆邮电大学通信与信息工程学院 重庆 400065)
(重庆邮电大学电子信息与网络工程研究院 重庆 400065)
(重庆邮电大学体育学院 重庆 400065)
(luojt@cqupt.edu.cn)

摘 要: 运动员检测是篮球运动智能化分析的基础,由于篮球视频存在场景复杂、目标运动快速、目标间遮挡严重 的问题,现有目标检测技术不能实现对密集遮挡运动员的精确检测.为此,提出一种基于自适应关键点热图的遮挡 篮球运动员检测算法.首先通过预先构建的全卷积编码-解码网络进行运动员特征提取,利用高斯核函数在特征图上 渲染关键点热图,热图的渲染采用自适应策略,高斯核半径随着目标宽和高的变化而变化,能够加快网络收敛;然 后在热图中提取运动员中心点,回归得到运动员宽高、位置等信息,省去了基于锚框检测中复杂耗时的后处理过程, 更利于在遮挡条件下区分 2 个运动员.在篮球运动数据集 BasketballPlayer 上进行实验的结果表明,在复杂篮球视频 场景下,该算法能有效地解决密集遮挡运动员之间漏检、误检和检测精度不高的问题,处理速度可达到 26 帧/s.

关键词:运动员检测;自适应关键点热图;可变高斯核;密集遮挡中图法分类号:TP391.41 **DOI:**10.3724/SP.J.1089.2021.18707

Algorithm for Detecting Occluded Basketball Players Based on Adaptive Keypoint Heatmap

Ren Yuan¹⁾, Luo Jiangtao^{2)*}, and Liang Xupeng³⁾

¹⁾ (School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065)

²⁾ (Electronic Information and Networking Research Institute, Chongqing University of Posts and Telecommunications, Chongqing 400065)
³⁾ (School of Physical Education, Chongqing University of Posts and Telecommunications, Chongqing 400065)

Abstract: Player detection is the basis of intelligent analysis of basketball events. Due to the complexity of basketball sports, such as fast player movement and serious occlusion of players, existing object detection techniques cannot afford to achieve accurate detection. To address this issue, an object detection algorithm of basketball player is proposed based on an adaptive keypoint heatmap. First, fully-convolutional encoder-decoder networks are built to extract feature of players. Then, the keypoint heatmap is rendered in feature map through an adaptive variable Gaussian kernel radius. The rendering of the keypoint heatmap adopts an adaptive strategy, in which the Gaussian kernel radius is changed with the width and height of the object, so that the network convergence can be accelerated. With the center point of the player extracted from heatmap, the player's size, position and other information are retrieved through regression. Since the player is detected based on the center point, eliminating the complicated and time-consuming post-processing procedure in an-chor-based detections, it is more conducive to distinguish two different players under occlusion conditions. The

收稿日期: 2020-10-10; 修回日期: 2021-04-08. 基金项目: 国家自然科学基金(62003067); 重庆市体育科研重点项目(A2019002). 任媛(1997—), 女,硕士研究生,主要研究方向为计算机视觉; 維江涛(1971—), 男,博士,教授,博士生导师,论文通讯作者,主要 研究方向为移动大数据、新一代网络技术、通信网络测试与优化; 梁旭鹏(1970—), 男,硕士,副教授,主要研究方向为体育产业与经 营管理.

effectiveness of the proposed approach is validated on the BasketballPlayer dataset, and the experimental results show that lots of missed detection, misdetection, and low detection accuracy among densely occluded players are significantly improved at the processing rate of 26 frames per second, compared with existing algorithms.

Key words: player detection; adaptive keypoint heatmap; variable Gaussian kernel; dense occlusion

随着计算机视觉技术的发展,视频分析技术 已经成为提高运动训练水平和体育比赛成绩的重 要工具,也是克服我国篮球运动训练中仅凭教练 人工重复指导或依赖昂贵的辅助训练传感设备等 传统训练方法的重要手段.基于篮球运动视频对 篮球运动员的精准识别是训练效果评估、比赛技战 术分析的基础.区别于普通行人的检测与识别,篮 球视频场景下的运动员检测与识别主要存在以下 难点.

(1)场景复杂.在篮球运动场景下,除了参加 比赛的运动员外,还有裁判、观众、教练、替补运 动员以及其他后勤人员,为运动员检测带来极大 的干扰.

(2) 目标运动快速.不同于传统视频监控中的 行人检测与追踪,篮球比赛中运动员运动快、变化 情况剧烈,为运动员定位和识别带来难度.

(3)目标间相互遮挡严重.篮球运动为多人协 作运动,在篮球视频中,由于运动员身着统一队 服,外观上更为相似.在运动员们发生激烈碰撞、 交叉遮挡时的识别是极具挑战的任务.

针对运动员检测问题,国内外学者展开了大 量的研究. Lu 等^[1]研究认为, 使用可变形零件模型 (deformable part model, DPM)检测运动员, 会存在 部分遮挡目标丢失问题,因此使用颜色直方图结 合逻辑回归分类器, 通过2个球队队服颜色特征进 一步对运动员检测和分类. Chen 等^[2]同样使用 DPM 进行形状检测, 根据球衣颜色特征, 利用支 持向量机(support vector machine, SVM)区分运动 员与裁判. Santhosh 等^[3]利用方向梯度直方图人工 提取特征,结合 SVM 得到球场内所有人的边界框, 但还需要人工提取球衣颜色特征来区分运动员、裁 判员和观众. Nost 等^[4]提出背景分离算法, 将背景 和运动员分离开,再利用 Suzuki 等^[5]提出的算法 识别出不同运动员. Tsai 等^[6]提出一种非监督算法, 根据运动员边界框内外特征响应的差异评估每个 特征的目标区分能力, 然后动态地生成所有特征 的权重,将其组合进行目标分类. Li 等^[7]使用光流 法形成运动矢量,用运动矢量记录位置信息,该信息指示 2 个连续帧之间像素的源坐标和目标坐标,由于目标由一组运动矢量表示,因此无需预先定义任何目标形状模板,根据运动矢量方向即可确定运动员位置信息.还有利用低级手工提取特征的最大稳定极值区域^[8]、尺度不变特征变换^[9]等算法进行运动员检测,但由于需要人工进行过程繁复的特征提取,均具有一定局限性.Yoon等^[10]采用卷积神经网络(convolutional neural network, CNN)和实时区域建议算法,直接检测每帧中是否存在运动员目标,但检测精度不高,在遮挡条件下存在漏检、误检问题.

随着深度神经网络和高性能计算硬件的迅速 发展,将 CNN 用于特征提取的效果明显优于传统 机器学习算法.目前已有的目标检测算法主要分 为3类.

 (1) 基于区域建议的目标检测算法.如区域卷 积神经网络(region CNN, R-CNN)^[11]和快速区域卷 积神经网络(Faster R-CNN)^[12]等.

(2) 基于目标回归的目标检测算法.如 YOLO(you only look once)^[13]和单阶段目标检测模型^[14]等.

(3) 基于无锚框的目标检测算法. 基于区域建 议和目标回归的目标检测算法往往依赖于锚框, 而基于锚框的算法常存在正负样本不均衡、超参难 调、匹配耗时严重的问题. 在篮球视频场景下对运 动员检测与识别的精确度和实时性提出了更高的 要求,基于无锚框的检测框架由于其简单、快速、 高效的优点引起广泛的关注. CornerNet^[15]通过检 测目标边界框的一对角点检测出目标,提出角合 并以更好地定位目标边界框的角点,将基于关键 点的无锚框检测带入了公众视野. ExtremeNet^[16]将 关键点定义为极值点,根据几何结构对关键点进 行分组,通过对每个目标类预测4个多峰值的热图 寻找极值点. FCOS^[17]将检测与其他使用全卷积网 络的任务统一起来,减少超参数的设计,可拓展性 强,达到了单阶段检测的顶尖性能. CenterNet^[18]通 过高斯核将关键点映射到特征图上,然而高斯核 半径为一确定值,使渲染的热图为圆形,不随目标 形状变化而变化,真值的边界框外也存在非零值, 导致负样本以一定权重倾向于正样本,模糊网络 学习,在训练的过程中不容易收敛.

本文针对运动员之间遮挡严重的复杂篮球视 频场景,提出一种基于自适应关键点热图的遮挡 篮球运动员识别算法,通过自适应可变高斯核半 径渲染关键点热图,提取到运动员目标的中心点; 然后通过运动员目标宽和高的信息,回归得到运 动员目标宽高、位置等信息.本文基于无锚框的检 测能够简化有锚框检测中复杂耗时的后处理过程, 自适应的关键热图算法能够优化固定高斯核半径 渲染热图中多关键点重合的问题,在复杂篮球视 频场景下能有效地解决密集遮挡运动员之间漏检、 误检和检测精度不高的问题.

1 本文算法

本文从密集遮挡、快速变化的复杂篮球视频场 景出发,提出图1所示基于自适应关键点热图的运 动员检测算法.对于一个运动员目标,首先根据自 适应关键点热图提取其中心点,然后根据中心点 回归出其宽高和位置信息.



图 1 自适应关键点热图识别运动员目标

1.1 特征提取与输出网络

本文算法使用 3 个不同的全卷积编码-解码网 络作为骨干网络进行特征提取,分别为深度特征 融合网络(deep layer aggregation, DLA)^[19]、高阶沙 漏网络 Hourglass^[20]和上卷积残差网络 ResNet^[21]. 虽然采用了 3 个不同的特征提取网络,但网络输出 预测部分都是由热图、宽高和回归 3 个子网络构成,如图 2 所示.其中,热图子网络输出目标的高斯分 布,表示目标的中心点;宽高子网络输出目标的宽 和高;回归子网络输出目标中心整数化的偏置误差.

(1) 改进的 DLA-34^[19]. 使用可变形卷积和上 采样层组合进行特征提取,提升空间分辨率. DLA 是带多级跳跃连接的图像分类网络,可变形卷积 跳跃连接低层和输出层,与固定的 3×3 卷积相比,可变形卷积可以使网络的感受野变得更加精确. 上采样的过程采用转置卷积,能够更好地还原图像的语义信息和位置信息. DLA-34 特征提取网络的总体结构如图 3 所示.



图 3 特征提取网络结构

(2) Hourglass^[20]. 堆叠的 Hourglass 网络通过 2 个连续的 Hourglass 模块对输入图像进行 4 倍下采 样,每个 Hourglass 模块由对称的 5 层上卷积网络 和下卷积网络组成,且带有跳跃连接.该网络结构 较大,故通常会生成最好的关键点估计.

(3) ResNet^[21]. Xiao 等^[22]对标准的 ResNet 增加 3 个上卷积网络来提高输出图像的分辨率,将 3 个 上卷积网络的输出通道分别修改为 256, 128 和 64, 以节省计算量.

假设有输入图像 $I \in \mathbb{R}^{W \times H \times 3}$,其中,W和H 分别为图像的宽和高.预测时得到的关键点热图 为 $\hat{Y} \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}$,其中,R为输出步长,C对应 检测类别数. 对于标签图中的某一类别*C*,利用 其左上角坐标(x_1, y_1)和右下角坐标(x_2, y_2)计算出 中心点坐标 $p = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$,对 p进行下采 样,得到低分辨率的中心点坐标 $\tilde{p} = \left[\frac{p}{R}\right]$. 然后用 高斯核函数将中心点渲染到特征图上,即

$$Y_{xyc} = \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right)$$
(1)

其中, $\sigma_x = \frac{\mu w}{6}$; $\sigma_y = \frac{\mu h}{6}$; μ 为缩放因子, 用于 调节高斯核函数渲染范围. 渲染后的特征图上每 个点为 $Y \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}$, Y=1的点即为目标中心点.

高斯核函数的标准差随着目标的宽和高的变 化而变化,能够自适应目标的尺寸进行热图渲染, 与 CenterNet^[18]中将高斯核半径设置为一固定常数 相比,本文算法更为合理,2种算法渲染的热图对 比如图 4 所示. 前者由于高斯核半径为常量, 渲染 出的热图为圆形,在目标边界框外有权重非零点 存在,网络会将这些点倾向于正样本进行学习,但 通常定义边界框外的点为严格负样本[17],这种方 式会造成正负样本模糊, 使网络难以收敛. 本文提 出的自适应关键点热图的优点是, 高斯核半径自 动随着目标的宽和高变化而变化, 使最终渲染出 的热图形状为椭圆, 正样本严格存在于目标边界 框内, 能够极大地减少模糊样本; 特别是在篮球比 赛场景下,运动员之间遮挡严重,当热图为圆形时 可能会发生2个运动员目标中心点重叠的情况,导 致漏检、误检.



图 4 2 种算法渲染热图对比

1.2 损失函数

总体损失函数 *L*_{det} 由 3 部分组成, 分别是中心 点位置损失 *L_k*、物体宽高损失 *L*_{size} 和中心点偏移 损失 *L*_{off}.总体损失为三者的加权和, 即

$$L_{\text{det}} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}}$$
(2)

其中, A_{size} 和 A_{off} 分别为物体宽高损失和中心点偏移损失所占权重.

中心点位置损失由改进的 Focal Loss^[23]衡量,即

$$L_{k} = \frac{-1}{N} \sum_{xyc} \begin{cases} \left(1 - \hat{Y}_{xyc}\right)^{\alpha} \ln\left(\hat{Y}_{xyc}\right), & Y_{xyc} = 1\\ \left(1 - Y_{xyc}\right)^{\beta} \left(\hat{Y}_{xyc}\right)^{\alpha} \times \ln\left(1 - \hat{Y}_{xyc}\right), & \ddagger \text{th} \end{cases}$$
(3)

其中, $\alpha \ \pi \ \beta$ 为 Focal Loss 的超参数; *N* 为图像 *I* 的关键点数量.

Focal Loss 解决了正负样本比例失衡的问题, 弱化了实际中心点周围其他负样本的损失比重.

假设 $(x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)})$ 为目标 k,所属类别为 c_k ,它的中心点为

$$p_k = \left(\frac{x_1^{(k)} + x_2^{(k)}}{2}, \frac{y_1^{(k)} + y_2^{(k)}}{2}\right)$$
(4)

使用关键点热图 \hat{Y} 预测所有中心点,然后对 每个目标 k 的尺寸进行回归,最终回归到 $s_k = (x_2^{(k)} - x_1^{(k)}, y_2^{(k)} - y_1^{(k)})$,这个值是在训练前计算出 来进行了下采样之后的长宽值.为了降低回归的 难度,本文使用 $\hat{S} \in \mathbb{R}^{\frac{W'}{R} \times \frac{H}{R} \times 2}$ 作为预测值,使用 L_1 损失函数衡量物体宽高损失,即

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^{N} \left| \hat{S} p_k - s_k \right|$$
(5)

上文中对图像进行了R=4的下采样,且由于 特征图重新映射到原始图像上时会带来精度误差, 因此,对于每个中心点,本文采用一个局部偏置 $\hat{O} \in \mathbb{R}^{\frac{W}{R} \times \frac{H}{R} \times 2}$ 去补偿它.所有类c的中心点共享同 一个偏置值预测,这个偏置值也是使用L_损失函

一个偏置值预测,这个偏置值也是使用 L 损失函 数来训练,即

$$L_{\text{off}} = \frac{1}{N} \sum_{p} \left| \hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right|$$
(6)

1.3 推理阶段

在推断预测阶段对图像进行下采样,在下采 样的特征图中预测中心点.首先提取热图上的所 有峰值点,峰值点定义为8邻域内的最大值,利用 3×3的最大池化选取该热图上100个峰值点;然后 设定一个阈值,从这100个峰值点中选取大于或等 于该阈值的峰值点作为最终确定的关键点.假设 $\hat{\mathcal{P}} = \{(\hat{x}_i, \hat{y}_i)\}_{i=1}^n 为检测到的关键点,每个关键点的$ $位置表示为整型坐标(<math>x_i, y_i$),使用 $\hat{Y}_{x_iy_ic}$ 表示当前 关键点的置信度,推理出目标边界框的公式为

$$B = \left(\hat{x}_{i} + \delta \hat{x}_{i} - \hat{w}_{i}/2, \hat{y}_{i} + \delta \hat{y}_{i} - \hat{h}_{i}/2 \hat{x}_{i} + \delta \hat{x}_{i} + \hat{w}_{i}/2, \hat{y}_{i} + \delta \hat{y}_{i} + \hat{h}_{i}/2\right)$$
(7)

其中, $(\delta \hat{x}_i, \delta \hat{y}_i) = \hat{O}_{(\hat{x}_i, \hat{y}_i)}$ 为当前关键点对应原始图 像的偏置点; $(\hat{w}_i, \hat{h}_i) = \hat{S}_{(\hat{x}, \hat{y}_i)}$ 为预测出来当前关键 点对应目标的宽和高. 所有的输出结果都直接从 特征图上获得,无需非极大值抑制等繁复的后处 理步骤.

实验及结果分析 2

2.1 数据集及评价指标

本文构建 BasketballPlayer 数据集对算法效果 进行训练和评估、该数据集是在大量篮球比赛、训 练视频场景中以不同角度拍摄录制的俯视视频, 且在时间维度上进行了规范化,将视频分帧进行 标注. 如图 5 所示, 在标注运动员整体边框的基础 上,将运动员整体划分为多部分并分别标注,以降 低运动员之间交叉遮挡带来的处理难度. 最终构 建的数据集特征表达包括目标分类、目标局部和整 体位置、所属帧号、目标身份信息等要素.标注的 数据集有 5141 张图片, 通过翻转、旋转、比例缩 放、裁剪和添加高斯噪声等数据增强方式、最终增 强后的数据集共有图片 7958 张, 随机划分数据集, 其中, 训练集 5049 张, 验证集 1552 张, 测试集 1357张.



目标局部位置标注降低运动员间交叉遮挡 图 5

本文采用平均精确度(average precision, AP)、 训练时间、处理速度作为评价指标,对检测器性能 进行评估. AP 表示为

$$AP = \frac{\sum P}{N}$$
(8)

其中,

$$P = \frac{N_{\text{true}}}{M} \tag{9}$$

N_{tme} 为成功检测的运动员目标个数; M 为检测的 运动员目标总数; N 为图片总数. 实验中 AP 为 AP@{50.0%~95.0%},即在锚框与真实框的交并比 (intersection over union, IoU)的值为 50.0%~95.0%, 以 0.05 步长取 10 个 IoU 阈值下计算的 AP.

2.2 实验设置

实验在 Tesla V100 GPU 显卡环境下进行, 显 存为 32GB. 实验平台的操作系统为 Ubuntu16.04, 采用 PyTorch 框架, Python 版本为 3.6.2, CUDA 10.0、CUDNN 7.6. 实验中、模型输入的图像大小 为 1920×1080, 将图像中的每个像素点的值除以 255 实现归一化,设置批次大小为 16,初始化学习 率为1.25×10⁻⁵,每100个epoch按10%的速率衰减, 总训练周期为 270 个 epoch, 使用 Adam 优化器^[24] 进行优化.实验中设置关键点选取阈值大小为 0.75, 高斯核函数缩放因子 μ=1.15, 损失权重 $\lambda_{\text{size}} = 0.1$, $\lambda_{\text{off}} = 1$, Focal Loss 超参数 $\alpha = 2$, $\beta = 4$.

2.3 结果与分析

计算 Basketballplayer 数据集在 3 种不同特征 提取网络上的表现,并与未使用自适应调整高斯 核半径的方式进行比较,结果如表1所示.可以看 出,相比于传统关键点热图的渲染算法,使用自适 应关键点热图能够减少正负样本模糊问题,加速 网络收敛,在不同骨干网络上都能减少接近一半 的模型训练时间.同时,在3种不同的骨干网络上 都能提高一定的精确度:在 Hourglass-104 骨干网 络上能达到最高 58.2%的精确度, 比未使用自适应 调整高斯核半径的方式提高了 3.6%的精确度; 在 ResNet-18 骨干网络上提高了 2.5%的精确度; 在 DLA-34 骨干网络上提高了 2.9%的精确度.

表1 自适应关键点热图算法评估结果

骨干网络	使用自适应	AP/%	训练时间/h
PasNat ^[22]		44.3	96
Resident	\checkmark	46.8	50
DLA ^[19]		50.8	192
	\checkmark	53.7	94
Hourglass ^[20]		54.6	240
	\checkmark	58.2	178

在 BasketballPlayer 数据集上将本文算法与已 有目标检测算法进行对比,表2所示为不同算法的 对比结果. ResNeXt^[25]将 ResNet 块中输入数据的通 道划分到不同组,每组的计算过程相对独立,将所 有组的计算结果进行空间聚合,作为最终的输出. 可以看出, ResNeXt 可以在不增加参数量的情况下

进一步提高 ResNet 的特征提取能力, 表现出更好 的网络性能. 特征金字塔(feature pyramid network, FPN)^[26]使用具有横向连接的自上而下的体系结构, 可以根据单比例输入构建网络内要素金字塔.具 有 FPN 的基础网络可以根据 FPN 的规模从不同级 别的功能金字塔中提取感兴趣区域特征. 使用 ResNet-101-FPN 作为骨干网络进行特征提取, 可 以在准确性和速度上取得具有竞争力的结果. 无论 是基于 DLA-34 还是 Hourglass-104 的骨干网络, 本 文算法均能取得高于其他算法的精确度.在 DLA-34 上本文算法最高处理速度为 26 帧/s, 远远 快于 YOLOv3^[27], RetinaNet^[23]等基于锚框的检测算 法; Hourglass-104 由于网络结构更大, 处理速度稍 慢于 DLA-34, 然而与 CornerNet^[15]和 ExtremeNet^[16] 等同样利用 Hourglass-104 作为骨干网络的基于无 锚框的目标检测算法相比, 在处理速度上领先接 近2倍,有明显优势.在ResNet-101-FPN骨干网络 上, 本文算法较 Faster R-CNN^[12]提高了 5.7%的精 确度,还领先了近2倍的处理速度;较FCOS^[17]提高 了1.5%的精确度,同时在处理速度上领先了4.5帧/s, 主要原因是 FCOS 具有耗时的非极大值抑制后处 理过程.图6所示为本文算法在篮球视频场景下篮

算法	骨干网络	处理速度/(帧·s ⁻¹)	AP/%
YOLOv3 ^[27]	DarkNet-53	18.0	44.3
RetinaNet ^[23]	ResNeXt-101-FPN	5.7	51.2
CornerNet ^[15]	Hourglass-104	4.2	50.8
ExtremeNet ^[16]	Hourglass-104	3.6	50.1
Faster R-CNN ^[12]	ResNet-101-FPN	7.8	50.7
FCOS ^[17]	ResNet-101-FPN	9.7	54.9
本文	ResNet-101-FPN	14.2	56.4
本文	DLA-34	26.0	53.7
本文	Hourglass-104	10.3	58.2



图 6 密集遮挡条件下的运动员检测

球运动员的检测结果,可以看出,在密集遮挡的 条件下,本文算法能够实现对每个运动员的精准 识别.

3 结 语

本文提出一种基于自适应关键点热图的遮挡 篮球运动员识别算法,通过最大池化获得8邻域内 的峰值点,省去了基于锚框检测中复杂耗时的后 处理过程,更利于在遮挡条件下区分开2个不同运 动员目标.热图的渲染采用自适应策略,高斯核半 径随着目标的宽和高的变化而变化;将目标边界 框外的样本严格定义为负样本,减少正负样本模 糊问题,在遮挡的条件下预测的边界框更加紧凑; 计算的坐标范围由整个特征图缩小为目标边界框 内,能够加快网络收敛,同时提高检测器精确度. 利用 BasketballPlayer 数据集在不同主流目标检测 算法上的实验结果表明,本文算法在复杂篮球视 频场景下能有效地解决密集遮挡运动员之间漏检、 误检、检测精度不高的问题.

参考文献(References):

- Lu W L, Ting J A, Little J J, *et al.* Learning to track and identify players from broadcast sports videos[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(7): 1704-1716
- [2] Chen L H, Su C W, Hsiao H A. Player trajectory reconstruction for tactical analysis[J]. Multimedia Tools and Applications, 2018, 77(23): 30475-30486
- [3] Santhosh P K, Kaarthick B. An automated player detection and tracking in basketball game[J]. Computers, Materials & Continua, 2019, 58(3): 625-639
- [4] Nost C L, Lefevre F, Bombardier V, et al. Automatic video editing: original tracking method applied to basketball players in video sequences[C] //Proceedings of International Conference on Image and Signal. Heidelberg: Springer, 2018: 117-126
- [5] Suzuki S, Abe K. Topological structural analysis of digitized binary images by border following[J]. Computer Vision, Graphics, and Image Processing, 1985, 30(1): 32-46
- [6] Tsai T Y, Lin Y Y, Liao H Y M, et al. Precise player segmentation in team sports videos using contrast-aware co-segmentation [C] //Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society Press, 2016: 1826-1830
- [7] Li W S, Powers D. Multiple object tracking using motion vectors from compressed video[C] //Proceedings of the International Conference on Digital Image Computing: Techniques and Applications. Los Alamitos: IEEE Computer Society Press, 2017: 1-5

- [8] Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image & Vision Computing, 2004, 22(10): 761-767
- [9] Lowe D. Distinctive image features from scale-invariant key-points[J]. International Journal of Computer Vision, 2004, 20: 91-110
- [10] Yoon Y, Hwang H, Choi Y J, et al. Analyzing basketball movements and pass relationships using realtime object tracking techniques based on deep learning[J]. IEEE Access, 2019, 7: 56564-56576
- [11] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 580-587
- [12] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149
- [13] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 779-788
- [14] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2016: 21-37
- [15] Law H, Deng J. CornerNet: detecting objects as paired keypoints[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 734-750
- [16] Zhou X Y, Zhou J C, Krähenbühl P. Bottom-Up object detection by grouping extreme and center points[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 850-859
- [17] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional

one-stage object detection[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9626-9635

- [18] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points[OL]. [2020-10-10]. https://arxiv.org/abs/1904.07850v1
- [19] Yu F, Wang D Q, Shelhamer E, et al. Deep layer aggregation[C] //Proceedings of the IEEE Conference on Computer Vision and Pat-tern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 2403-2412
- [20] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2016: 483-499
- [21] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [22] Xiao B, Wu H P, Wei Y C. Simple baselines for human pose estimation and tracking[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 472-487
- [23] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 318-327
- [24] Kingma D P, Ba J. Adam: a method for stochastic optimization[OL]. [2020-10-10]. https://arxiv.org/abs/1412.6980v8
- [25] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5987-5995
- [26] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[OL]. [2020-10-10]. https://arxiv.org/abs/ 1612.03144
- [27] Redmon J, Farhadi A. YOLOv3: an incremental improvement[OL]. [2020-10-10]. https://arxiv.org/abs/1804.02767