

中图法分类号: TP183 文献标识码: A 文章编号: 1006-8961(2025)01-0148-13

论文引用格式: Zhong L, Ou Q F, Shen M J and Xiong B S. 2025. STNet: a small moving target localization and tracking network for badminton. Journal of Image and Graphics, 30(01):0148-0160(钟亮, 欧巧凤, 沈敏杰, 熊邦书. 2025. STNet: 羽毛球运动小目标定位跟踪网络. 中国图象图形学报, 30(01):0148-0160)[DOI:10.11834/jig.230800]

STNet: 羽毛球运动小目标定位跟踪网络

钟亮, 欧巧凤*, 沈敏杰, 熊邦书

南昌航空大学信息工程学院, 南昌 330063

摘要: 目的 相比于一般视频目标检测跟踪任务, 视频羽毛球的实时定位跟踪主要面临两大难点: 1) 羽毛球属于小目标, 同时伴有严重的运动模糊以及相似目标的干扰, 使用基于矩形框的目标检测跟踪方法准确率低且会带来中心点误差问题; 2) 单帧图像很难准确定位羽毛球目标, 利用视频前后帧的时域特征则可以跟踪到羽毛球目标, 而现有提取时域特征的网络模块结构复杂, 难以满足实时性要求。针对以上问题, 本文使用热力图轮廓检测方法, 提出了羽毛球运动小目标的定位跟踪网络算法(shuttlecock track net, STNet)。方法 网络主体采用“U”型编解码结构; 针对小目标像素信息少的问题, 基于SE(squeeze and excitation)通道注意力与残差结构设计高效特征提取模块(SE channel attention and residual, SECAR), 实现了空域信息的高效提取与传递, 提高了网络的定位性能; 针对目标丢失与相似目标干扰问题, 设计了时序网络(temporal network, TPN)结构用于提取和记忆视频时域特征, 提高了网络跟踪性能。结果 在羽毛球比赛公开数据集TrackNetv2与自制数据集上的实验表明, 本文方法在多个指标上取得了最好的性能表现。相较于现有性能较好的羽毛球定位跟踪方法TrackNetv2, 本文方法在准确率、精确率和F1上分别提高7.5%、15.7%和7.5%, 并且显著降低了参数量, 满足实时处理需求(54帧/s)。结论 本文提出的STNet羽毛球定位跟踪网络, 在面对羽毛球目标外观剧烈变化以及背景干扰严重时, 能够准确定位羽毛球比赛视频帧中可能存在的羽毛球, 实现羽毛球的稳定跟踪, 相比其他羽毛球定位跟踪网络, 具有更优的性能。

关键词: 羽毛球定位跟踪; 小目标; 热力图; 轮廓检测; 编解码网络; 时序网络

STNet: a small moving target localization and tracking network for badminton

Zhong Liang, Ou Qiaofeng*, Shen Minjie, Xiong Bangshu

School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China

Abstract: Objective Badminton has become one of the most popular sports in the world today. Badminton match videos contain a wealth of useful information, and extracting this information can provide data support for precise analysis of sports. Among various types of information, badminton trajectory information is particularly important, as it can extract ball speed, athlete explosiveness, and shuttlecock trajectory data from the positioning information of the shuttlecock. Compared with general video object detection and tracking tasks, real-time localization tracking in badminton videos faces two major challenges: 1) Shuttlecocks are small targets with severe motion blur and interference from similar objects. The accuracy of object detection and tracking methods based on rectangular bounding boxes is low and can cause centroid estimation

收稿日期: 2023-11-20; 修回日期: 2024-07-09; 预印本日期: 2024-07-16

* 通信作者: 欧巧凤 ou.qiaofeng@nchu.edu.cn

基金项目: 国家自然科学基金项目(62365014); 江西省自然科学基金项目(20242BAB202072)

Supported by: National Natural Science Foundation of China (62365014); Natural Science Foundation of Jiangxi Province, China (20242BAB202072)

errors. 2) Locating the shuttlecock target accurately in a single-frame image is difficult, but its position can be tracked using temporal features from adjacent video frames. However, existing network modules for extracting temporal features are complex and struggle to meet real-time requirements. Early research on badminton localization and tracking mainly included threshold-based methods and feature learning-based methods, which performed well in specific scenarios. However, these methods focused only on target information during model construction, making it difficult to utilize background information. As a result, their localization and tracking performance suffer in complex environments. Currently, most deep learning-based object detection and tracking methods are based on rectangular bounding box algorithms. However, shuttlecocks exhibit severe motion blur in single-frame images, making it difficult for the rectangular box algorithm to accurately locate the shuttlecock target's pixel coordinates. Subsequently, a localization method based on heatmap contour detection emerged, which overcame the limitations of rectangular bounding box methods by directly determining the pixel position of the shuttlecock's head center by using segmentation and morphological detection methods. However, current approaches based on heatmap contour detection have weak capabilities in extracting spatiotemporal features for video object localization and tracking, making them inadequate for real-world requirements. To enhance the network's ability to extract temporal and spatial features from video frames, we propose a new badminton localization tracking network algorithm named shuttlecock track net (STNet) based on heatmap contour detection. **Method** The backbone of the STNet network adopts a "U"-shaped encoder-decoder structure. To tackle the issue of limited pixel information in small targets, we design an efficient feature extraction module based on squeeze and excitation (SE) channel attention and residual structure called SE channel attention and residual (SECAR). This module facilitates efficient extraction and transmission of spatial information, improving the network's localization performance. To tackle the challenges of target loss and interference from similar objects, we design a temporal network (TPN) structure for extracting and memorizing temporal features, enhancing the network's tracking performance. The STNet network consists of four main parts: input layer, encoder, decoder, and output layer. The input layer introduces the TPN structure, enabling the network to achieve neural network-style Kalman filtering effects. The encoder and decoder utilize SECAR feature extraction modules, which use multi-level residual bypass channels to improve the network's spatial feature extraction capability and mitigate the problem of information loss for small targets. The output layer is responsible for restoring the output heat map to the predetermined size. After contour detection on the output heat map, pixel coordinate information of the shuttlecock can be obtained. Stable tracking of the shuttlecock is achieved by locating the pixel coordinates of the shuttlecock frame by frame. **Result** Experiments conducted on the open data set TrackNetv2 and a home-brewed data set demonstrated that the STNet network achieved an accuracy of 92.0%, a precision of 96.7%, and an F1 score of 95.0%. Compared with the shuttlecock positioning trajectory method TrackNetv2, our method improved accuracy, precision, and F1 by 7.5%, 15.7%, and 7.5%, respectively, while significantly reducing the parameter count to meet real-time processing requirements (54 frames per second). **Conclusion** This paper proposes a badminton localization tracking algorithm based on SECAR and a temporal encoder-decoder network. The network algorithm employs the SE channel attention mechanism and a residual bypass structure to design the SECAR feature extraction unit, which effectively improves the network's capability to extract spatial features. To extract temporal information from video frames and achieve neural network-style Kalman filtering effects, we introduce the TPN structure. Experimental results demonstrate that even in low-frame-rate badminton match videos with severe motion blur, our proposed network can accurately locate the pixel coordinates of the shuttlecock and achieve stable tracking.

Key words: shuttlecock localization and tracking; small target; heatmap; contour detection; encoder-decoder network; temporal network

0 引言

数字化、智能化电子信息技术在现代体育运动中广泛应用,使用先进电子信息技术对运动进行精

密化分析正在成为运动员技术分析与能力提升的重要方法(Liu等,2023)。例如,通过对羽毛球运动员的各项指标进行数字化提取,可以建立运动员成长履历、身体机能和战术素养等方面的具体评价指标,对羽毛球运动员训练和竞赛具有很好的指导作

用。目前主要有穿戴式测量方法(Jacob等,2016; Raina等,2017)与视觉测量方法(陶树和王美丽,2022)。

非接触式视觉测量方法通过相机采集羽毛球比赛视频,进而对所需信息进行处理与分析,如运动员动作、羽毛球轨迹和场地边界等信息。其中,羽毛球定位跟踪是实现击球速度测试、运动员击球爆发力等运动信息数字化分析的关键技术和难点之一。本文主要研究羽毛球定位跟踪方法轨迹信息的提取。

羽毛球大型赛事使用的“鹰眼系统”,可以实现高精度的羽毛球三维轨迹重建,进而判定羽毛球落地是否出界(Liu等,2022)。“鹰眼系统”的硬件成像设备昂贵、帧速高,部署光照环境要求严格,目前常用于人工回放视频进行辅助裁判。与此不同,本文方法不仅适用于理想条件下的羽毛球比赛视频,而且适用于通用低帧率网络相机采集的低质量羽毛球比赛视频,实现比赛场地全景视频内羽毛球目标的实时定位跟踪。方法可应用于羽毛球比赛实时智能裁判系统、运动员数据分析以及电视转播等专业场景,具有重要的应用价值。

传统的羽毛球定位与跟踪方法主要包括基于阈值的方法和基于特征学习的方法。基于阈值的方法依靠羽毛球视频帧间差分信息,配合手动设计的多层滤波器检测跟踪羽毛球目标(Shishido等,2014);基于特征学习的方法使用Adaboost算法(Chen等,2019)以及哈尔特征(Haar-like features)、有向梯度直方图(histogram of oriented gradient, HOG)、局部二值模式(local binary patterns, LBP)等算法提取羽毛球特征来检测飞行的羽毛球。这些方法通过提取目标特征构建表观模型,在图像中搜索与模型最匹配的区域作为跟踪结果。由于模型构建时只关注目标信息,很难利用背景信息,并且容易受到非目标对象(如运动员)的影响,在面对复杂环境时,定位跟踪性能与泛化性较差。

近年来,基于深度学习的羽毛球定位跟踪算法性能提升到了新的高度。主要包括基于矩形框的羽毛球检测跟踪方法和基于热力图轮廓检测的羽毛球定位跟踪方法。

基于矩形框的目标检测网络,根据预测框生成方法可分为一阶段和二阶段目标检测网络(刘颖等,2020;曹家乐等,2022)。其中,一阶段网络以

YOLO(you only look once)(Redmon和Farhadi,2018;Bochkovskiy等,2020;Wang等,2023)系列改进网络为代表,二阶段目标检测网络以R-CNN(region-based convolutional neural network)(Girshick等,2014)、Fast R-CNN(Girshick,2015)、Faster R-CNN(Ren等,2017)为代表。由于一阶段网络实时性较好,比二阶段网络更适合实现羽毛球实时检测。

Cao等人(2021)基于Tiny YOLOv2的架构,提出了M-YOLOv2和YOLOBR两种新的变体。由于羽毛球属于小目标检测对象,为了更加适应小目标检测任务,M-YOLOv2对YOLOv2损失函数进行了改进,忽略分类损失,突出了边界框的宽度和高度损失。为了更好地保留羽毛球小目标对象的语义信息,YOLOBR相比YOLOv2减少了Maxpool层的数量,并且根据因式分解思想使用 1×1 和 3×3 卷积核替换2个连续的 3×3 卷积核。为了测试网络性能,作者在多个背景环境下近距离采集羽毛球图像,制作目标检测数据集。与其他检测方法相比,M-YOLOv2和YOLOBR在检测精度与速度方面表现出了更好的性能。然而,由于数据集背景并不是羽毛球完整场地,并且羽毛球目标在图像中尺度变化较小,难以应对全场景比赛视频中羽毛球目标的检测。

Zhang等人(2022)基于YOLOv4-Tiny提出了基于YOLO-BTM的羽毛球检测方法。与YOLOv4-Tiny相比,YOLO-BTM对骨干网络CSPDarkNet53-tiny进行了改进。通过使用TreLU(triangular rectified linear unit)激活函数和使用步长为2、卷积核为 2×2 的卷积核替换平均池化操作,减少了主干网络的参数,提高了推理速度。同时,作者在特征金字塔网络(feature pyramid network, FPN)中引入3个通道注意力模块,提高了小目标检测的性能。

Goh等人(2023)使用YOLOv5网络对包括羽毛球、运动员、球拍等目标进行检测。为了克服单帧图像内羽毛球目标数量的失衡,作者制作了静态情况下多羽毛球目标的图像数据集。由于缺乏动态条件下羽毛球目标的数据,该方法无法准确检测运动羽毛球目标。

以上算法在检测无拖影模糊、对比清晰的羽毛球目标时效果良好,而对于拖尾模糊严重的目标会产生漏检。此外,为了给出羽毛球是否出界判定结果,需要确定羽毛球球头中心点坐标,如果将矩形框中心点或者角点坐标作为羽毛球像素坐标,将会带

来较大的随机误差。

基于矩形框的目标检测方法由于预测框大小以及比例的限制,这些算法对低速、像素区域占比较大的目标进行检测与跟踪效果很好,但是并不适用于小目标且形变严重的羽毛球目标的定位与跟踪。为此,针对球类小目标的定位跟踪,出现了基于热力图轮廓检测的方法。

TrackNet(track network)网络(Huang等,2019)是最早将热力图轮廓检测方法应用于球类目标定位跟踪的算法,网络通过提取每一帧网球中心点位置信息,实现网球目标的跟踪。网络由于多次使用下采样和上采样复原操作,造成小目标特征信息的丢失,导致目标定位准确率较低。

Komorowski等人(2019)使用全卷积操作,设计了基于DeepBall的球类目标检测网络。该网络可以处理任何尺寸的图像,并生成表征球位置的置信度图。网络使用Hypercolumn结构,将来自不同层级级别的卷积网络的特征图进行合并,共同送入分类输出层。由于DeepBall网络无法引入任何时序信息,在对视频图像目标进行检测跟踪时效果较差。

Van Zandycke和De Vleeschouwer(2019)将球检测问题表述为球分割问题,设计了BallSeg网络。在篮球检测任务中,为了利用运动信息,BallSeg网络除了输入单帧RGB图像,还包括相邻帧三通道差分图像。实验结果表明,对于视频球类目标检测跟踪任务,融合多帧信息对网络性能具有显著帮助。

随后出现的TrackNetv2网络(Sun等,2020),使用U-Net网络的跳跃连接结构,有效防止了解码层信息的丢失,显著提高了羽毛球定位跟踪性能。连同该论文开源的TrackNetv2羽毛球数据集是目前羽毛球定位跟踪任务使用最广泛的公共数据集,TrackNetv2网络成为羽毛球定位跟踪任务使用范围最广的基线网络。TrackNetv2网络虽然避免了上采样复原操作产生的信息丢失问题,但是由于多次使用3层多通道连续卷积的简单结构,网络参数量大,实时性较差。尽管网络使用了3个连续帧滑窗输入和3个连续帧热力图输出的结构来提升帧间信息提取能力和运算效率,但是存在无法利用不同滑窗间的时序信息的问题,如果使用重叠的滑窗,则会存在滑窗重叠帧重复推理的问题。

在TrackNetv2网络的基础上,MonoTrack网络(Liu和Wang,2022)通过使用更多的残差连接以及

引入Dice损失,缓解了羽毛球有效信息丢失和目标像素占比失衡问题,但是仍然存在时序信息无法充分利用的不足。

为了充分利用原图像数据的高分辨率语义信息,Tarashima等人(2023)设计了基于SBDT(spatially balanced deep transmission network)网络的球类目标检测方法。利用与HRNet(high-resolution network)网络类似的并行多分辨率分支结构,SBDT网络引入了HRMs(high-resolution modules)高分辨率特征提取模块;为了增加网络对球位置的敏感度,降低蒙版类别不平衡的影响,网络对蒙版标签的目标区域外轮廓进行软编码,并使用焦点损失函数。SBDT网络通过反复的多尺度信息交换,增加了图像数据的空间语义信息利用效率,但仍然无法有效引入视频图像的时序信息。

针对矩形检测框深度网络存在定位随机误差以及现有方法对时域信息利用率低的问题,本文设计了STNet(shuttlecock track net)网络算法,主要创新包括以下两个方面:1)为了抑制小目标羽毛球细节特征信息的逐层流失问题,使用改进的结合SE(squeeze and excitation)通道注意力机制与残差结构的SECAR(SE channel attention and residual)特征提取模块,提高了网络的空域特征提取能力。2)为了避免跟踪目标的丢失与相似目标干扰问题,设计了TPN(temporal network)结构,实现了网络在时序上的连续卷积,提高了羽毛球运动目标跟踪能力。

STNet网络参数量小,对羽毛球的定位准确率、精确率和召回率高,且鲁棒性好,实现了多场景下羽毛球目标的准确定位与稳定跟踪。

1 方法

本文提出的羽毛球定位跟踪网络STNet,如图1所示。接下来分5个部分对网络设计、训练和推理进行详细描述。

1.1 网络主体结构

STNet网络主体结构采用改进的“U”型结构,输入3帧羽毛球比赛视频图像,输出3帧热力图,通过热力图轮廓检测获取羽毛球球头中心的像素坐标,逐帧获取定位信息,从而实现羽毛球的稳定跟踪。对“U”型网络结构进行以下两个改进:

1)设计了SECAR特征提取模块替代传统“U”型

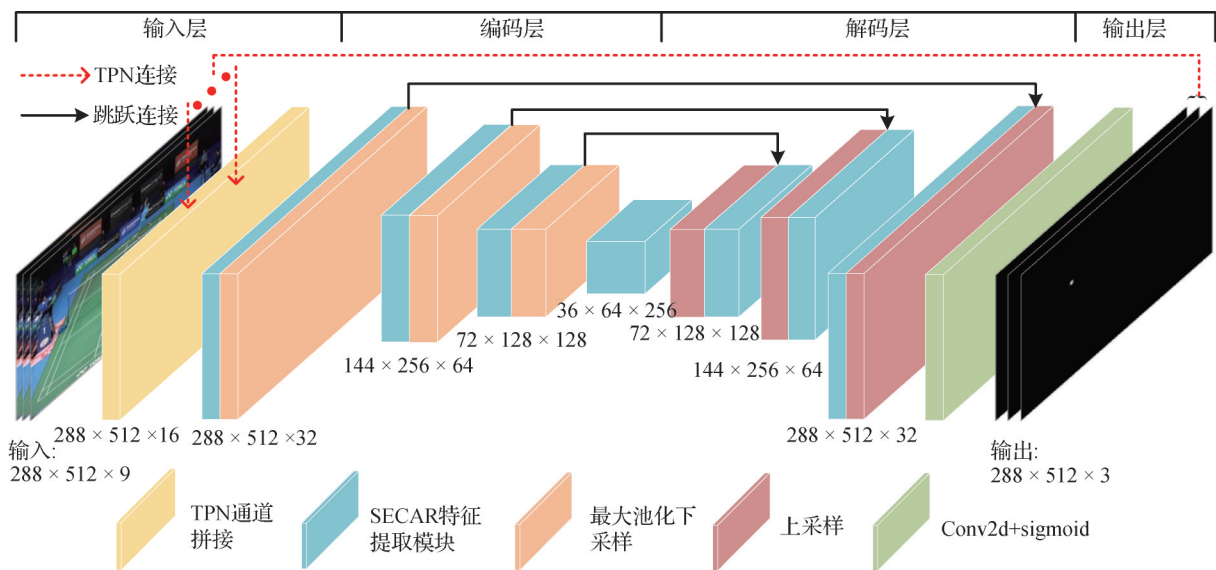


图 1 STNet 网络架构

Fig. 1 STNet network architecture

网络编解码层中单一的卷积单元,使得网络更加关注羽毛球目标的前景信息,并且多级残差结构实现了信息的有效传递,用以提高网络空间特征提取能力。

网络编码层使用了3层SECAR特征提取模块与3层下采样模块,用以提取输入视频帧特征信息,并逐步压缩编码和提取深层信息。网络解码层由3层上采样与3层SECAR特征提取模块构成,解码层负责将压缩特征信息逐步复原为网络输入图像尺寸。

由于卷积层越深越容易引入无用的噪声信息,解码层需要借助跳跃连接来获取编码层更详尽的特征信息,因此将编码层上采样输出与解码层特征提取模块的输出在通道维度上拼接,共同送入下一层解码层特征提取模块。

2)网络输入层引入了TPN结构,根据历史帧信息推断当前帧羽毛球位置信息,实现神经网络类型的卡尔曼滤波效果。TPN结构提高了时域特征提取能力,避免了多帧输入多帧输出带来的重复推理与时序信息切割问题。

输出层主要负责调整网络输出通道数,并将特征图经过sigmoid激活函数运算,得到归一化热力图。其中,输出热力图的每个像素值表征属于羽毛球区域的概率值。基于热力图方法可以有效去除干扰信息,目前已应用于许多目标定位任务,作为联系分割任务与目标定位任务的桥梁。

1.2 羽毛球小目标特征提取SECAR模块

在羽毛球比赛视频帧中,相机的摆放位置通常是不变的,因此图像的核心信息在于运动羽毛球所在的占比较小前景区域。对于小目标的定位跟踪,如果使用基础卷积进行简单堆砌,在经过多层卷积处理后,特征图不断变小,会造成小目标信息的丢失。此时,如果添加残差结构,网络借助残差旁路将低层次的特征与高层次的特征结合起来,能够提高信息的传递效率。此外,多级残差旁路的引入可以减少过拟合,加速训练收敛,避免深层网络出现梯度消失或梯度爆炸等问题,从而降低训练难度。

除了使用残差结构抑制细节信息丢失和混淆的问题外,本文还通过SE通道注意力增强神经网络对小目标的感知能力。

通过引入SE通道注意力模块,网络可以自动学习重要的通道特征,使得网络更加关注小目标所在的通道,从而提升小目标的定位能力。如图2所示,本文网络关注于图像中的运动信息,在添加SE通道注意力后,网络可以提高对运动羽毛球目标的敏感度,提高网络性能。

本文结合SE通道注意力机制和残差结构,设计了一种高效的羽毛球特征提取模块SECAR,如图3(c)所示。首先,使用1层基础卷积(如图3(a)所示)进行特征提取;然后,使用SE通道注意力模块(如图3(b)所示)对卷积输出进行通道标定;接着,再经过可选参数的M层基础卷积进行特征信息的进一步



(a) 添加SE前

(b) 添加SE后

图2 添加SE前后热力图与原图叠加可视化结果

Fig. 2 Visualization results of overlaying the thermal map with the original map before and after adding SE
((a) before adding SE; (b) after adding SE)

提取;最后,将首层卷积输出特征信息通过残差旁路与末尾的卷积输出相加,构成 SECAR 特征提取模块,实现羽毛球小目标信息的有效传递。

1.3 时序网络结构

时序网络在不同领域有着不同定义,在本文中,时序指在视频时间线上的序列帧,视频图像相比图像图像具有更丰富的时域特征。在视觉领域,更丰富的运动信息和更明显的动态上下文对小目标的定位跟踪具有显著帮助。

羽毛球在视频中的轨迹符合运动学规律,羽毛球在当前帧中的定位与前帧具有强关联性,即当前

帧中羽毛球的定位同时取决于过去帧羽毛球的定位,因此,融合历史帧,提取其中的时域特征信息,能够有效提高羽毛球的定位跟踪性能。

目前,针对一维时间序列数据,网络模型主要通过 TCN (temporal convolutional network) 结构和一维卷积层结构提取时域特征信息 (Lea 等, 2017); 二维时间序列数据主要是使用三维卷积操作或者 RCNN (recurrent convolutional neural network) 结构,实现时域特征的提取 (McLaughlin 等, 2016; Lai 等, 2015), 但是三维卷积以及 RCNN 结构的运算量较大,难以满足实时性要求。

2020年提出的 TrackNetv2 网络采用了多帧输入多帧输出结构,利用了相邻历史帧信息,实现了较好的定位跟踪性能,但由于使用 3 帧输入的结构,视频帧被手动分割,破坏了时序信息的连续性。在面对羽毛球目标静止或者遮挡时,TrackNetv2 网络容易产生目标丢失或者定位错误的现象。

针对上述问题,本文设计了一种 TPN 结构对基于热力图轮廓检测的目标定位方法进行改进。将图 1 中输出层输出的热力图作为短时记忆信息进行时序上的暂存与拼接,能够实现时域特征的融合利用。

设在时序 k 上获取羽毛球比赛视频帧 $P_{rgb}(k)$, $P_{rgb}(k)$ 在 STNet 网络上推理输出得到热力图 $H(k)$,

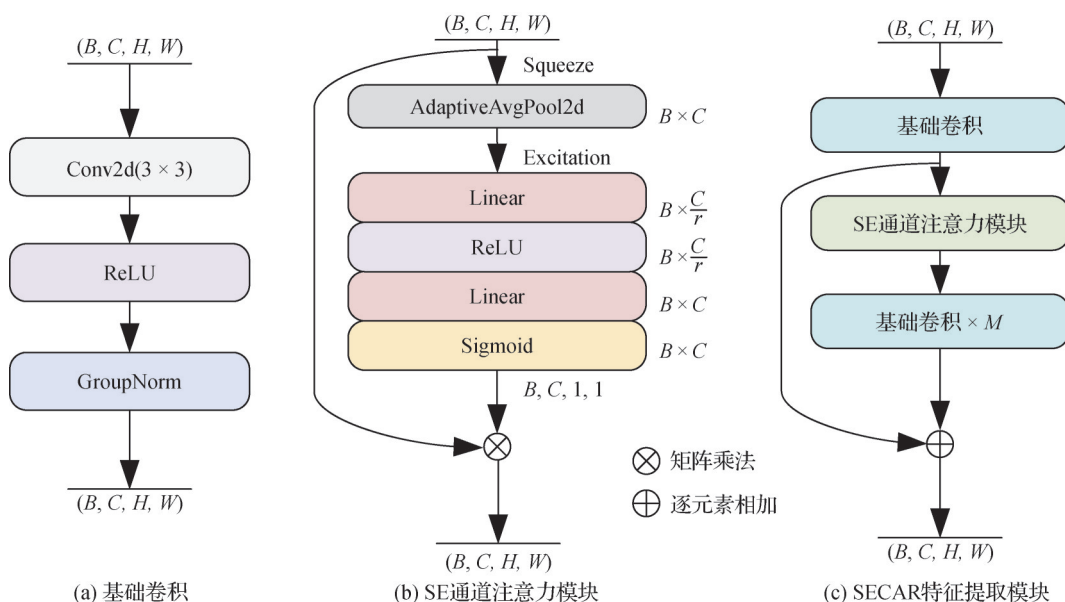


图3 SECAR 特征提取模块结构

Fig. 3 Structure of SECAR feature extraction module

((a) basic convolution; (b) SE channel attention module; (c) SECAR feature extraction module)

时序信息帧长度为 l 。使用时序网络结构,则可以得到STNet网络输出与输入的关系,具体为

$$H(k-2:k) = \text{STNet}(P_{\text{rgb}}(k-2:k) + H(k-2-l:k-3)) \quad (1)$$

式中,STNet网络使用了3帧输入、3帧输出的结构;“ $k-2:k$ ”表示第 $k-2$ 帧到第 k 帧;“+”表示图像在通道维度上进行拼接。

当视频图像帧滑动采样窗口步长为1,TPN时域特征拼接实现方法如图4所示。首先,滑动窗口顺序取样3帧视频帧;然后与暂存的7帧历史输出热力图在通道上进行拼接,最后将拼接好的16通道特征图前向推理得到与输入3帧视频帧对应的热力图,并将第一位置的热力图进行缓存。当缓存的历史输出热力图不足7帧时,使用零矩阵进行填充。

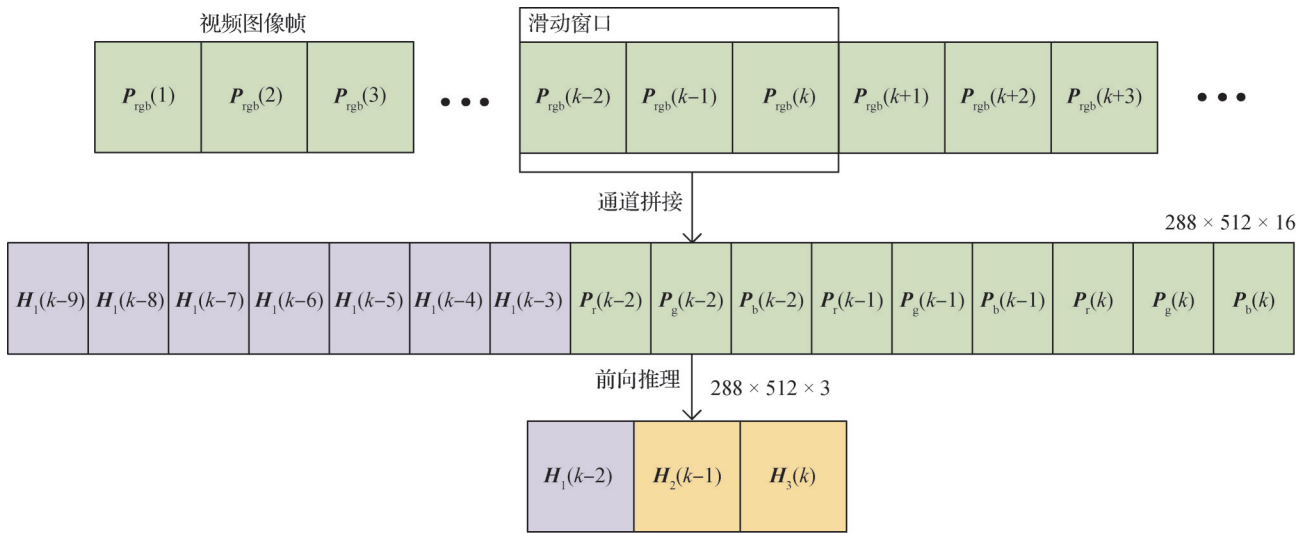


图4 TPN时域特征拼接示意图

Fig. 4 Schematic diagram of TPN time-domain feature stitching

将输出热力图转换成灰度图,通过伪色彩转换拼接在原图上,得到采用TPN结构改进网络前后羽毛球小目标定位结果,如图5所示。从图5(d)中可以观察到在羽毛球落地停止运动后,未加入TPN结构的网络容易丢失以及错误定位羽毛球目标,而TPN结构的引入使得帧间信息得以连续,不仅避免了目标丢失问题,而且能够克服人物运动等带来的

相似背景干扰。

在网络中,该结构起到了关联羽毛球前后帧坐标信息的作用,实现了神经网络类型的卡尔曼滤波效果。同时,由于TPN结构的引入需要设置视频帧数据为顺序采样训练,归一化方法设置为组归一化,因此网络训练时数据加载批数可以设置为1,最终网络不仅减小了训练所需的硬件要求,而且有效提

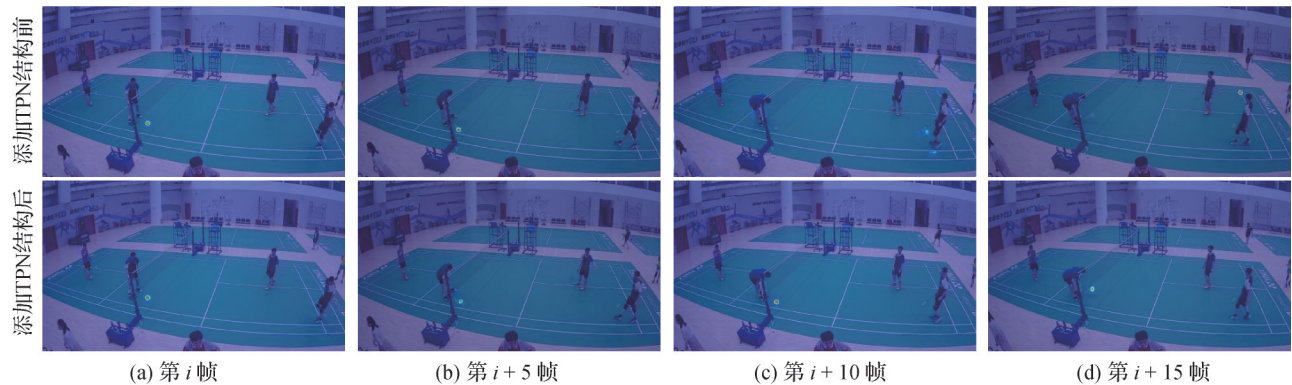


图5 原图与热力图叠加效果比较

Fig. 5 Comparison of original images with heatmap overlay effects((a) frame i ; (b) frame $i+5$; (c) frame $i+10$; (d) frame $i+15$)

升了网络模型的定位跟踪性能。

1.4 热力图轮廓检测

使用点坐标信息的羽毛球定位跟踪任务不需要羽毛球的分割区域信息,而是需要提取羽毛球球头中心的像素坐标。从热力图中计算羽毛球球头中心像素坐标,包括二值分割、轮廓像素提取和最小外接正矩形计算3个步骤。

1)由于归一化热力图每个像素值对应属于羽毛球目标的概率值,因此以概率0.5作为阈值对热力图进行二值化。

2)对二值分割得到的各个候选闭合区域进行外轮廓提取。

3)计算各个候选闭合轮廓的最小外接正矩形,如图6(c)所示;最后,以候选最小外接正矩形中最大面积矩形的中心点坐标作为羽毛球球头中心位置,如图6(d)所示。

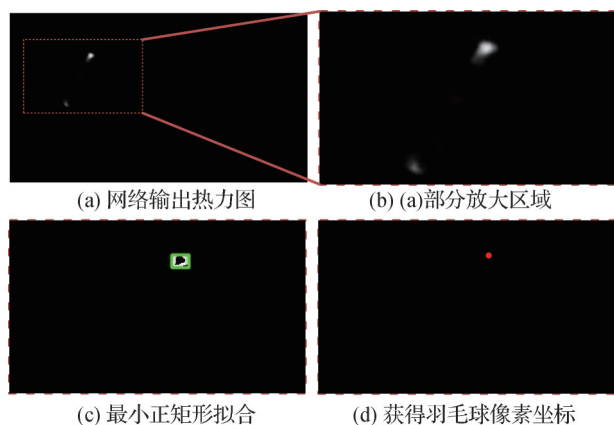


图6 热力图轮廓检测效果示意图

Fig. 6 Schematic diagram of thermal diagram contour detection effect ((a) thermal diagram output from the network; (b) enlarged area in (a); (c) minimum rectangle fit; (d) obtain badminton pixel coordinates)

1.5 网络训练与推理

训练时需要利用视频帧时序信息,因此设置单个视频内的图像帧为滑窗连续采样。设置TPN结构滑窗尺寸为3,以步长1移动,时序信息帧长度为7。学习率(learning rate)设置为等差下降方式,第 n 轮学习率 $lr(n)$ 的计算为

$$lr(n) = \alpha - (\alpha - \varepsilon) \times \frac{n}{epochs} \quad (2)$$

式中, α 为初始学习率, ε 为小于1E-6的微小常量, $epochs$ 为训练周期总数。

为了缓解羽毛球像素占比失衡问题,本文采用加权二值交叉熵(weighted binary cross-entropy, WBCE)损失进行模型训练,最终的损失函数可以表示为

$$L_{WB}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^N \left[(1-w)^2 y_i \log(\hat{y}_i) + w^2 (1-y_i) \log(1-\hat{y}_i) \right] \quad (3)$$

式中, N 为像素索引, $\hat{\mathbf{y}}$ 为模型预测值, $\mathbf{y}$ 为标签真实值, w 为加权系数,本文设置 $w = y_i$ 。通过加权策略,损失向预测值与标签值相差大的像素倾斜,有助于模型收敛。训练时保留验证集准确率最高的一组模型权重用于网络推理。

网络部署推理时可设置采样滑窗步长为3,此时网络推理一次可直接对3帧连续图像进行羽毛球定位,与步长为1相比,此时视频帧羽毛球实际定位速度是网络推理速度的3倍。

2 实验

2.1 数据集

本文实验采用了公开数据集TrackNetv2和自制数据集。

TrackNetv2数据集公开于2020年,包含职业和业余羽毛球比赛视频,以及对应视频帧的羽毛球像素坐标和目标可见性标志。如图7所示,TrackNetv2数据集的羽毛球比赛视频大部分来自联赛转播镜头,视频采用固定相机,从不同高度俯视拍摄完整的羽毛球场地,视频分辨率为1280×720像素,包含不同颜色背景的球场。由于使用了均匀光源,视频背景干扰较小,成像效果较好。受透视投影效果影响,在视频帧中,羽毛球目标的尺寸大致在10±3像素,拖影最长可达40像素。经过对源数据集的整理,得到78558帧训练数据、12656帧验证数据。在训练集中含有羽毛球目标的帧数占比为87.1%,验证集中含有羽毛球目标的帧数占比为87.5%。

本文自制数据集采集了多个赛场共58870帧羽毛球比赛视频,其中10470帧图像用于验证集,训练集中含有羽毛球目标的帧数占比为86.3%,验证集中含有羽毛球目标的帧数占比为82.4%。由于采集的视频内存在间隔较大的未出现羽毛球目标的无用视频段,本文制作的数据集首先经过了切片处理,将其中较长的无用片段进行切除,使用该方法处理得

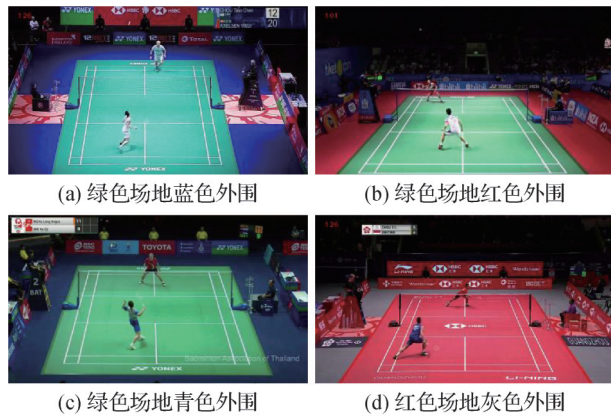


图7 TrackNetv2数据集图像示例

Fig. 7 Examples of images within the TrackNetv2 dataset
((a) green court with blue perimeter; (b) green court with red perimeter; (c) green court with cyan perimeter; (d) red court with gray perimeter)

到共728个不同长度的视频。数据集采集使用了家用监控摄像头,刷新率为25帧/s。图8展示了其中4个场地的视频帧效果,在图8(a)(b)中,灯光的反射干扰严重,同时在图8(a)(b)(c)中存在来自相邻场的多目标与相似目标干扰,在图8(d)中则是自然光干扰严重。图8(a)(b)(d)视频分辨率为 $2\,304 \times 1\,296$ 像素,图像帧中羽毛球像素范围在 30 ± 20 像素,最长拖影在120像素。图8(c)视频分辨率为 $3\,840 \times 2\,160$ 像素,图像帧中羽毛球像素范围在 47 ± 22 像素,拖影长度最长可达320像素。

与公共数据集相比,自制数据集视频采集帧率更低、光照等背景更加复杂、目标拖尾现象更加严



图8 自制数据集图像示例

Fig. 8 Examples of self-made dataset images
((a) light reflection and adjacent field interference; (b) adjacent field interference; (c) multi target interference; (d) natural light reflection interference)

重、定位跟踪难度更大。

与原始帧图像对应的标签为与图像尺寸相同的二值矩阵,其中“255”表示羽毛球目标区域圆形掩膜,“0”表示背景像素,二值矩阵点 (x, y) 处的像素值 $F(x, y)$ 可表述为

$$F(x, y) = \begin{cases} 0 & (x - x_0)^2 + (y - y_0)^2 > r^2 \\ 255 & (x - x_0)^2 + (y - y_0)^2 \leq r^2 \end{cases} \quad (4)$$

式中, (x_0, y_0) 为羽毛球球头中心像素坐标, r 表示圆形掩膜半径,实验时取 r 为图像短边长度的2.8%。由于羽毛球目标像素占比过小,网络训练收敛困难且不利于网络模型输出后处理中的轮廓检测,掩膜区域将大于羽毛球所占实际区域。

2.2 实验设置

本文使用PyTorch11.3深度学习框架训练与验证网络模型。实验使用的硬件环境主要包括Intel i5-12600KF CPU、Nvidia 3060 GPU、16 GB DDR4内存。

实验时设置初始学习率 α 为 $1E-3$ 、 ε 为 $1E-8$,训练周期总数epochs为30。网络权重迭代使用自适应矩估计(adaptive moment estimation, Adam)优化器。将输入图像帧高宽尺寸设为 288×512 像素,多帧输入的数据取样滑动窗口长度设为3,并设置训练数据集顺序采样的滑动窗口步距为1,验证数据集顺序采样的滑动窗口步距为3。在每个epoch内以视频为对象对训练数据集进行随机排列,并在单个视频内顺序采样。TPN结构缓存长度设为7帧,编码层与解码层SECAR特征提取模块内含基础卷积个数依次为 $(0, 0, 1, 2, 1, 0, 0)$ 。设置羽毛球真实像素坐标与网络预测坐标的允许误差为4 px,训练时使用20%概率值对3帧图像同步进行高斯滤波、中值滤波和指数范围为 $(0.95, 1.05)$ 的亮度调整,用以增加训练图像的多样性。

在本文中,任务的核心是在每一帧图像中定位羽毛球的像素坐标,但是预测结果与真实标签并不一定含有羽毛球目标。基于已有研究,本文主要采用准确率(accuracy, Acc)、精确率(precision, Prec)、召回率(recall, Rec)、F1、参数量(Params)和网络每秒推理帧数(frames per second, FPS)作为羽毛球定位网络算法的评估指标。

2.3 消融实验

为了验证本文所提网络对于羽毛球定位跟踪的

准确性和效率,设计了SECAR模块和TPN结构消融实验。以U-Net网络作为基准网络,依次使用SE通道注意力机制、Res残差结构、SECAR特征提取模块、TPN结构对基准网络进行改进。实验使用TrackNetv2数据集,保持训练和验证的超参数一致。实验中,SE通道注意力机制、Res残差结构和SECAR特征提取模块的加入将替代基准网络中的普通卷积模块。设计的TPN结构主要作用于网络输入输出,将TPN结构缓存的时间序列长度设为7帧。

实验结果如表1所示,相比基准网络,SE通道注意力、Res残差结构、SECAR特征提取模块以及TPN结构的加入均能使得Acc和F1两个有效性验证指标有所提升。其中,Res残差结构和TPN结构的引入对网络性能的影响最为明显。网络改进后,虽然参数量有所下降,但是因为使用了更复杂的结构,导致推理速度的略微下降,不过仍然满足羽毛球定位跟踪的实时性需求。

表1 消融实验结果

Table 1 Results of ablation experiments

结构	Acc/%	Prec/%	F1/%	Params/M	FPS /(帧/s)
基准网络	82.9	95.4	82.6	2.5	90
SE改进	83.2	95.1	84.2	2.5	76
Res改进	85.1	95.5	88.9	2.5	80
SECAR改进	88.2	93.4	91.4	2.5	73
TPN改进	89.1	92.2	94.5	2.5	59
STNet(步距=1)	91.7	94.9	94.9	2.5	54
STNet(步距=3)	92.0	96.7	95.0	2.5	54×3

为了充分利用已有数据集,网络训练数据集滑动采样窗口长度设置为3,步距设置为1,此时除去单个视频开头和结尾外的绝大部分图像帧将会出现在输入窗口的3个顺序位置。对于验证集,当滑动取样窗口长度设置为3、步距设置为3时,每帧图像将只被推理一次,并且单个视频末尾将会有对3求余的数据帧被丢弃,此时网络推理帧数将会是网络推理次数的3倍。在TrackNetv2验证集的不同步距实验结果如表2所示。

2.4 与现有羽毛球定位跟踪方法对比

为了验证本文方法性能的优越性,将所提网络算法STNet与目前常用的羽毛球定位跟踪算法进

表2 STNet在验证集不同步距的实验结果

Table 2 Experimental results of STNet on the validation set with different strides

步距	推理帧数	TP	TN	FP1	FP2	FN
1	12 598×3	29 131	5 513	546	1 023	1 581
3	12 627	9 604	2 008	135	190	690

行横向比较,对比结果如表3所示。由于大部分网络源代码并未公开,并且复现无法保证评价细节一致,因此,本文采用了可靠性较高的“Papers with Code”公开数据。实验表明,本文提出的STNet方法在准确率、精密度和F1指标上均达到最优结果,实验结果显示了本文所提网络算法具有一定的优越性。

表3 不同网络的对比实验结果

Table 3 Comparison results of different networks

方法	Acc/%	Prec/%	F1/%	Params/M
DeepBall	38.6	60.0	52.4	0.1
DeepBall-Large	36.8	59.5	50.6	1.0
BallSeg	72.2	68.4	79.9	12.7
ResTrackNetV2	84.0	82.2	89.4	1.2
TrackNetv2(基线)	85.6	83.6	90.5	11.3
MonoTrack	85.9	84.9	90.9	2.9
WASB	87.0	88.5	91.6	1.5
STNet(本文)	92.0	96.7	95.0	2.46

注:加粗字体表示各列最优结果。

2.5 网络超参数设置

为了提高网络性能,本文方法使用了多个细节改进方法,包括多帧输入、TPN时序网络结构等,各个参数实验细节如下。

2.5.1 滑动窗口长度选择

在TrackNetv2与MonoTrack中,多帧输入的滑动取样窗口长度为3。为了验证多帧输入多帧输出结构的可靠性,在保持其他参数不变的情况下,本文对不同滑动窗口长度进行实验,结果如表4所示。可以观察到,当窗口长度在2~5时网络能够达到最佳性能,并且随着滑动窗口长度的增加,网络对于目标的召回率,即查全性能将会不断提高,但精确度,即查准率将会出现下降。出于网络综合性能考虑,本文将滑动窗口长度设为3。

表4 滑动窗口长度影响实验结果

Table 4 Experimental results with sliding windows of different lengths				
滑动窗口长度	Acc	Prec	Rec	F1
1	84.7	92.0	88.4	90.1
2	91.0	95.9	93.0	94.4
3	92.0	96.7	93.3	95.0
4	90.4	93.4	94.7	94.1
5	91.0	93.7	95.2	94.4
6	90.4	93.2	95.0	94.0
7	89.7	91.8	95.6	93.6

注:加粗字体表示各列最优结果。

2.5.2 TPN缓存帧长度选择

本文设计了TPN时序网络结构,用以引入时序信息。为了验证TPN缓存帧长度对网络性能的影响,设置了对比实验。保持滑动窗口长度一致,网络训练与验证参数一致,得到表5所示结果,当缓存帧长度为7时本文网络达到最佳性能。

表5 不同TPN长度影响实验结果

Table 5 Experimental results of the impact of different TPN lengths				
TNP长度	Acc	Prec	Rec	F1
3	88.4	92.4	93.2	92.8
4	90.7	93.9	94.7	94.3
5	86.6	89.8	93.4	91.8
6	91.6	93.8	94.8	94.3
7	92.0	96.7	93.3	95.0
8	89.6	92.4	94.9	93.6
9	90.4	93.4	94.6	94.0

注:加粗字体表示各列最优结果。

2.5.3 定位误差允许值

在TrackNetv2等方法内,网络预测定位与真实定位允许误差值设为4像素。在输出热力图分辨率高宽设为288×512像素时,设置不同的误差允许值,使用本文训练好的STNet网络对验证集进行测试,可得到表6结果。当定位误差允许值大于4像素时,在客观评价指标上并未取得较多增长。

表6 不同允许误差实验结果

Table 6 Experimental results for different allowable errors				
允许误差值/像素	Acc	Prec	Rec	F1
6	92.6	97.5	93.3	95.4
5	92.5	97.4	93.4	95.4
4	92.0	96.7	93.3	95.0
3	88.7	92.5	93.0	92.8
2	71.7	71.0	91.1	79.7
1	45.0	36.9	84.1	51.3

2.6 自制数据集实验

2.3节消融实验和2.5节对比实验使用的TrackNetv2数据集,主要是采用专业录像设备录制的正式羽毛球比赛视频,成像灯光环境较好。为了验证本文方法在普通羽毛球训练场地、光照和拍摄设备等条件下的适应性,本文利用通用网络摄像头采集和制作了一个羽毛球比赛视频数据集进行实验。部分网络实验结果如表7所示。

表7 自制数据集实验结果

Table 7 Experimental results of self-made dataset			
方法	Acc	Prec	F1
TrackNetv2	57.6	84.0	68.1
MonoTrack	65.7	90.4	75.5
STNet(本文)	91.0	94.0	94.7

注:加粗字体表示各列最优结果。

从表7可以看出,在面对复杂环境下羽毛球的定位跟踪任务时,本文方法具有更好的性能表现。实验发现,STNet网络在面对拖影严重和停止运动的羽毛球目标时,具有显著的性能优势,部分验证集的定位结果如图9所示。由图9可以观察到,网络在多种场景下能够有效克服相似目标、多目标和光照带来的干扰以及拖尾现象,实现了在低帧率成像设备下对羽毛球目标的准确定位与稳定跟踪。

3 结论

面向智能裁判和羽毛球运动数字化分析应用,



图9 部分图像帧与热力图定位结果

Fig. 9 Partial image frames and thermal map positioning results ((a) light reflection interference; (b) natural light interference; (c) multi target interference; (d) target deformation interference)

本文研究了羽毛球定位和跟踪关键技术,提出了一种基于热力图轮廓检测的羽毛球定位跟踪网络算法STNet。STNet网络主要进行了两方面的创新设计。首先,通过结合SE通道注意力和残差结构,设计了SECAR特征提取模块,提升了网络对图像的空域特征提取能力。多级残差旁路有效缓解了深层网络的有效信息丢失问题。其次,在使用3帧输入3帧输出的网络结构基础上,设计了TPN结构,实现了深度神经网络类型的卡尔曼滤波算法,有效克服了目标丢失以及相似目标干扰问题。

在TrackNetv2公共数据集与自制多场景数据集的实验表明,相较于现有羽毛球定位跟踪优秀方法,所提出的STNet网络具有更高的准确率、精确率和F1,并且参数量显著下降,满足实时性要求。本文方法可为构建羽毛球比赛智能裁判、羽毛球比赛战术指导以及运动员能力分析等系统提供技术支撑,促进运动数字化发展。

目前,STNet网络主要针对球类细小单目标的定位跟踪,在今后的工作中,将进一步研究如何将本文方法应用于多目标定位跟踪任务,提高本文方法

的应用价值与应用范围。

参考文献(References)

- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. Yolov4: optimal speed and accuracy of object detection [EB/OL]. [2023-11-20]. <https://arxiv.org/pdf/2004.10934.pdf>
- Cao J L, Li Y L, Sun H Q, Xie J, Huang K Q and Pang Y W. 2022. A survey on deep learning based visual object detection. *Journal of Image and Graphics*, 27(6): 1697-1722 (曹家乐, 李亚利, 孙汉卿, 谢今, 黄凯奇, 庞彦伟. 2022. 基于深度学习的视觉目标检测技术综述. *中国图象图形学报*, 27(6): 1697-1722) [DOI: 10.11834/jig.220069]
- Cao Z G, Liao T B, Song W, Chen Z H and Li C S. 2021. Detecting the shuttlecock for a badminton robot: a YOLO based approach. *Expert Systems with Applications*, 164: #113833 [DOI: 10.1016/j.eswa.2020.113833]
- Chen W, Liao T B, Li Z H, Lin H Z, Xue H, Zhang L, Guo J and Cao Z G. 2019. Using FTOC to track shuttlecock for the badminton robot. *Neurocomputing*, 334: 182-196 [DOI: 10.1016/j.neucom.2019.01.023]
- Girshick R. 2015. Fast R-CNN//*Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]
- Goh G L, Goh G D, Pan J W, Teng P S P and Kong P W. 2023. Automated service height fault detection using computer vision and machine learning for badminton matches. *Sensors*, 23(24): #9759 [DOI: 10.3390/s23249759]
- Huang Y C, Liao I N, Chen C H, Ik T U and Peng W C. 2019. Track-Net: a deep learning network for tracking high-speed and tiny objects in sports applications//*Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Taipei, China: IEEE: 1-8 [DOI: 10.1109/AVSS.2019.8909871]
- Jacob A, Wan Zakaria W N and Md Tomari M R B. 2016. Implementation of IMU sensor for elbow movement measurement of Badminton players//*Proceedings of the 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*. Ipoh, Malaysia: IEEE: 1-6 [DOI: 10.1109/ROMA.2016.7847813]
- Komorowski J, Kurzejamski G and Sarwas G. 2019. DeepBall: deep neural-network ball detector [EB/OL]. [2023-11-20]. <https://arxiv.org/pdf/1902.07304.pdf>
- Lai S W, Xu L H, Liu K and Zhao J. 2015. Recurrent convolutional neu-

- ral networks for text classification//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA: AAAI: 2267-2273 [DOI: 10.1609/aaai.v29i1.9513]
- Lea C, Flynn M D, Vidal R, Reiter A and Hager G D. 2017. Temporal convolutional networks for action segmentation and detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 1003-1012 [DOI: 10.1109/CVPR.2017.113]
- Liu C X, Wang W, Liu H R and Wang J. 2022. Application of hawk-eye technology to sports events//Proceedings of the 2nd International Conference on Information Technology and Contemporary Sports (TCS). Guangzhou, China: IEEE: 1-5 [DOI: 10.1109/TCS56119.2022.9918811]
- Liu J B, Huang G G, Hyppä J, Li J, Gong X D and Jiang X F. 2023. A survey on location and motion tracking technologies, methodologies and applications in precision sports. Expert Systems with Applications, 229: #120492 [DOI: 10.1016/j.eswa.2023.120492]
- Liu P and Wang J H. 2022. MonoTrack: shuttle trajectory reconstruction from monocular badminton video//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, USA: IEEE: 3512-3521 [DOI: 10.1109/CVPRW56347.2022.00395]
- Liu Y, Liu H Y, Fan J L, Gong Y C, Li Y H, Wang F P and Lu J. 2020. A Survey of research and application of small object detection based on deep learning. Acta Electronica Sinica, 48(3): 590-601 (刘颖, 刘红燕, 范九伦, 公衍超, 李莹华, 王富平, 卢津. 2020. 基于深度学习的小目标检测研究与应用综述. 电子学报, 48(3): 590-601) [DOI: 10.3969/j.issn.0372-2112.2020.03.024]
- McLaughlin N, del Rincon J M and Miller P. 2016. Recurrent convolutional network for video-based person re-identification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 1325-1334 [DOI: 10.1109/CVPR.2016.148]
- Raina A, Lakshmi T G and Murthy S. 2017. CoMBaT: wearable technology based training system for novice badminton players//Proceedings of the 17th IEEE International Conference on Advanced Learning Technologies (ICALT). Timisoara, Romania: IEEE: 153-157 [DOI: 10.1109/ICALT.2017.96]
- Redmon J and Farhadi A. 2018. Yolov3: an incremental improvement [EB/OL]. [2023-11-20]. <https://arxiv.org/pdf/1804.02767.pdf>
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Shishido H, Kitahara I, Kameda Y and Ohta Y. 2014. A trajectory estimation method for badminton shuttlecock utilizing motion blur//Proceeding of the 6th Pacific-Rim Symposium Image and Video Technology. Guanajuato, Mexico: Springer: 325-336 [DOI: 10.1007/978-3-642-53842-1_28]
- Sun N E, Lin Y C, Chuang S P, Hsu T H, Yu D R, Chung H Y and Ik T U. 2020. TrackNetV2: efficient shuttlecock tracking network//Proceedings of 2020 International Conference on Pervasive Artificial Intelligence (ICPAI). Taipei, China: IEEE: 86-91 [DOI: 10.1109/ICPAI51961.2020.00023]
- Tao S and Wang M L. 2022. Stroke recognition in badminton videos based on pose estimation and temporal segment networks analysis. Journal of Image and Graphics, 27(11): 3280-3291 (陶树, 王美丽. 2022. 结合姿态估计和时序分段网络分析的羽毛球视频动作识别. 中国图象图形学报, 27(11): 3280-3291) [DOI: 10.11834/jig.210407]
- Tarashima S, Haq M A, Wang Y S and Tagawa N. 2023. Widely applicable strong baseline for sports ball detection and tracking [EB/OL]. [2023-11-20]. <https://arxiv.org/pdf/2311.05237.pdf>
- Van Zandycke G and De Vleeschouwer C. 2019. Real-time CNN-based segmentation architecture for ball detection in a single view setup//Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports. Nice, France: ACM: 51-58 [DOI: 10.1145/3347318.3355517]
- Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]
- Zhang Y M, Chen C X and Hu R L. 2022. YOLO-BTM: a novel shuttlecock detection method for embedded badminton robots//Proceedings of 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE). Wuhan, China: IEEE: 1-6 [DOI: 10.1109/ICARCE55724.2022.10046579]

作者简介

钟亮,男,硕士研究生,主要研究方向为视频目标检测跟踪和行为识别。E-mail: jinrannuo@icloud.com

欧巧凤,通信作者,女,副教授,主要研究方向为图像处理、机器视觉和人工智能。E-mail: ou.qiaofeng@nchu.edu.cn

沈敏杰,男,硕士研究生,主要研究方向为图像处理和人体行为识别。E-mail: 3487302550@qq.com

熊邦书,男,教授,主要研究方向为图像处理、计算机视觉和直升机故障诊断技术。E-mail: xiongbs@126.com