

基于改进 NanoDet 的复杂运动场景多人检测算法

刘丛昊¹, 王 军², 谢 非¹, 杨继全¹, 马 磊², 王 琼³

(1. 南京师范大学 电气与自动化工程学院, 江苏 南京 210023)

(2. 南京三万物联网科技有限公司, 江苏 南京 210000)

(3. 南京师范大学 计算机与电子信息学院, 江苏 南京 210023)

[摘要] 在体育赛事场景下, 运动员的行为识别、传球投篮动作次数统计以及 AI 解说等方面都离不开对运动员的人体检测, 这使得复杂场景下对运动员的人体检测速度和精度上均有较高要求。因此, 本文提出一种基于 NanoDet 的复杂运动场景多人检测算法。首先, 该算法使用更平滑的 Mish 函数作为主干网络的激活函数, 改进 ShuffleNetV2 网络, 构建主干网络, 并引入 CBAM 注意力模块, 采用轻量化路径汇集网络进行特征融合, 提高检测准确性; 其次, 使用无锚轻量化检测头 GFLV2 进行回归和分类, 实现复杂运动场景下多人目标检测; 最后, 为了进一步验证提出算法的性能, 将研究算法与目前主流的检测算法进行实验对比, 实验结果表明, 相较于其他算法如 YOLOv3-tiny、YOLOv4-tiny 等本文算法有着更高的检测精度, 比同类型的轻量化检测模型 YOLOv4-tiny 高 14.87%, 此外, 单帧检测时间与 YOLOv4-tiny 的 10.67 ms 相比减少了 31.68%, 由此可见, 本文研究算法在保持检测速度的基础上, 大幅提高了检测精度。

[关键词] 深度学习, 人体检测, 轻量化模型, 注意力机制

[中图分类号] TP391.41, TP18 **[文献标志码]** A **[文章编号]** 1001-4616(2024)02-0140-09

An Improved NanoDet-Based Multi-Human Detection Algorithm for Complex Motion Scenes

Liu Conghao¹, Wang Jun², Xie Fei¹, Yang Jiquan¹, Ma Lei², Wang Qiong³

(1. School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing 210023, China)

(2. Nanjing Sanwan Internet of Things Technology Co., Nanjing 210000, China)

(3. School of Computer and Electronic Information, Nanjing Normal University, Nanjing 210023, China)

Abstract: In sports event scenarios, athlete behaviour recognition, passing and shooting action count and AI commentary are inseparable from human body detection of athletes, which makes high requirements on the speed and accuracy of human body detection of athletes in complex scenarios. Therefore, this paper proposes a NanoDet-based multi-body detection algorithm for complex sports scenes. First, the algorithm uses a smoother Mish function as the activation function of the backbone network, improves the ShuffleNetV2 network, builds the backbone network, and introduces the CBAM attention module, and uses a lightweight path pooling network for feature fusion to improve detection accuracy; next, it uses the anchorless lightweight detection head GFLV2 for regression and classification to achieve multi-body target detection in complex motion scenes. Finally, in order to further verify the performance of the proposed algorithm, the research algorithm is experimentally compared with the current mainstream detection algorithms, the experimental results show that the algorithm in this paper has higher detection accuracy compared to other algorithms such as YOLOv3-tiny and YOLOv4-tiny, which is 14.87% higher than the same type of lightweight detection model YOLOv4-tiny. In addition, the single-frame detection time is reduced by 31.68% compared to the 10.67ms of YOLOv4-tiny, which shows that the investigated method substantially improves the detection accuracy while maintaining the detection speed improvement.

Key words: deep learning, human detection, lightweight model, attention mechanism

收稿日期: 2022-11-04.

基金项目: 国家自然科学基金项目(41974033)、江苏省科技成果转化项目(BA2020004)、江苏省省级工业和信息化产业转型升级专项资金项目(JITC-2000AX0676-71)、江苏省研究生科研与实践创新计划项目。

通讯作者: 谢非, 博士, 副教授, 研究方向: 机器视觉与深度学习. E-mail: xiefei@njnu.edu.cn

近年来,随着深度学习技术和计算机资源的发展,计算机视觉领域取得了长足的进步,目标检测在运动赛事领域有着广泛的应用^[1]. 比如比赛中运动员的行为识别,传球、投篮等动作的次数统计,以及视频分类和 AI 解说等方面都离不开对运动员的检测识别,这使得对运动员人体检测有着速度上和精确度上的较高要求.

Girshick 提出了 R-CNN(Region-based Convolutional Neural Network)算法^[2],这象征 CNN 在目标检测领域应用的开端. R-CNN 在图像中提取 1000~2000 个候选区域,然后 CNN 为每个候选区域提取特征,将特征发送到分类器使检测到的目标得以分类.

Ren 在 2015 年提出了 Faster R-CNN^[3]. 顾名思义,这是 R-CNN 的加速版本. Faster R-CNN 实现了端到端训练,它用 softmax 代替 SVM 进行分类并优化了生成区域建议的方式. 训练速度和检测速度都比 R-CNN 快. 2016 年,Redmon 等^[4]提出 YOLO(You Only Look Once)算法. YOLO 将图像放入多层卷积中以提取特征,并在输出层中获得边界框和分类结果. 现在 YOLO 已经更新到了第五个版本,尽管最后两个版本的作者并不再是 Joseph Redmon,但它依然是一种出色的目标检测器.

Liu 等提出了一种基于单深度神经网络的目标检测方法,并称其为 SSD(Single Shot MultiBox Detector)算法^[5]. SSD 将特征图的边界输出空间离散化为不同比例或不同比例的先验框,它完全忽略了边界框的生成阶段,并将此步骤集成到整个网络中,这种端到端的形式使 SSD 模型的训练更加方便,并且可以集成到检测模块中.

2019 年,FCOS(Fully Convolutional One-Stage Object Detection)算法^[6]被提出. FCOS 不再依赖于锚点框进行目标检测,从而避免了许多与锚点框相关的复杂计算,节省了训练时的内存占用,并可以拓展到图像处理中的其他领域,如语义分割,关键点检测等.

2020 年,Google Brain 团队提出 BiFPN 和联合缩放方法(Compound Scaling)^[7],BiFPN 考虑到不同特征融合对于输出特征的重要性;联合缩放方法综合考虑图像输入分辨率、网络宽度和深度这些因素,权衡了准确率和效率.

CVPR 2022 上,Yang 等提出了焦点和全局知识蒸馏 FGD(Focal and Global Knowledge Distillation for Detectors)算法^[8],焦点蒸馏分离了前景和背景,全局蒸馏重建不同像素之间的关系并将传递给学生,以补偿焦点蒸馏中丢失的全局信息.

2022 年,Gao 等提出一种快速收敛的基于查询的目标检测器 AdaMixer^[9],与传统密集模式的检测器不同,基于查询的检测器通过一组可学习的查询来解码图像特征,并且作者改进了基于查询的解码器过程的适应性.

针对运动赛事场景对运动员人体检测速度和精度上均有较高要求,本文提出了基于改进 NanoDet 的复杂运动场景多人检测算法,设计了基于 ShuffleNetV2 结构的主干网络,引入 CBAM 注意力机制并基于 PANet 进行特征融合,以及基于 GFLV2 的检测头和损失函数,从而解决复杂场景下对运动员人体检测速度和精度不足的问题. 所研究算法的贡献及创新点具体如下:

(1) 针对检测速度不够快,实时性不够高的问题,采用 Mish 函数作为主干网络的激活函数,改进 ShuffleNetV2 进行特征提取,轻量化 Neck 部分,并采用 GFocalV2 无锚轻量化检测头,消除了 FCOS 中的 Center-ness 分支,减少模型参数量,提高了人体检测速度.

(2) 针对复杂场景检测精度较低的问题,引入了 CBAM 注意力机制,分别进行通道注意力和空间注意力运算,再基于轻量化 PANet 进行特征融合,并在检测头里引入分布指导质量预测器,提高了模型的检测精度.

1 基于改进 NanoDet 的复杂运动场景多人检测算法

目标检测器通常由 3 个部分组成,分别为 Backbone、Neck 和 Head. Backbone 常用的算法有 ResNet^[10], VGG^[11], Inception^[12]等,Backbone 通常被作为一个模块用来提取图像中的特征. Neck 则负责融合 Backbone 不同层级之间的特征,常用的有特征金字塔(FPN)^[13]和路径汇聚网络(PANet)^[14]. Head 是目标检测器中最具特点模块,由它输出检测结果,包括图像中目标的所属种类和目标在图像中的位置. 值得注意的是,Head 根据阶段数分为单阶检测器和二阶检测器,根据有无锚点可分为基于锚点和无锚点^[15]. 单阶检测器可以同

时做到对图像中的目标进行定位和分类,但是其参数学习难度大,因此一阶检测器运行速度一般较高,但是准确度可能略差;相反,二阶检测器具有较高准确度,但由于它对目标的定位和分类是分开进行的,因此运行速度相较于单阶检测器慢。

1.1 总体框架

本文算法在轻量目标检测网络 NanoDet 的基础上进行改进,在保持检测速度较快的前提下,进一步提高了检测平均精度,以满足实际场景的需求。本文模型构建基于无锚的一阶目标检测算法,如图 1 所示,

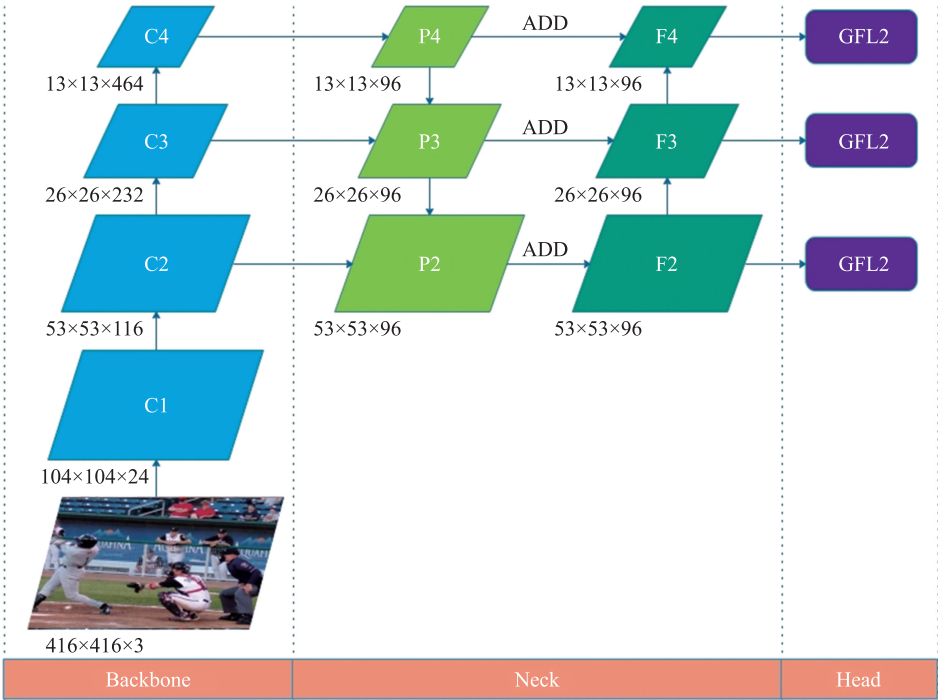


图 1 本文方法结构框架
Fig. 1 The framework of the proposed method

主要由以下 3 部分构成:Backbone 部分使用 Mish 作为激活函数,采用 ShuffleNetV2 1.0x^[16] 删除最后卷积层,抽取 8,16 和 32 倍下采样特征作为下层输入;Neck 部分由 CBAM^[17] 和轻量化 PANet 构成,进行特征融合;输出的特征图在 Head 部分,采用 GFLV2^[18] 进行分类和回归。

1.2 基于 ShuffleNetV2 的 Backbone

首先,主干网络上,本文 Backbone 基于 ShuffleNetV2 进行改进,并选用 Mish 作为 ShuffleNetV2 的激活函数,使网络得到更好的准确性和泛化性。

ShuffleNetV2 作为轻量级 Backbone,其中包括卷积块,最大池化块,深度卷积块和 ShuffleNetV2 块。本文主干网络采用 ShuffleNetV2 1.0x,并删除最后卷积层,抽取 8,16 和 32 倍下采样特征作为下层输入,如图 2 所示,其共有四个阶段。

第一个阶段由一个内核为 3,步幅为 2 和填充为 1 的卷积块再加一个内核为 3,步幅为 2 和填充为 1 的最大池化块构成。

之后三个阶段均由 ShuffleNetV2 块构成。每个 ShuffleNetV2 块均具有两个分支,其中左侧分支结构受

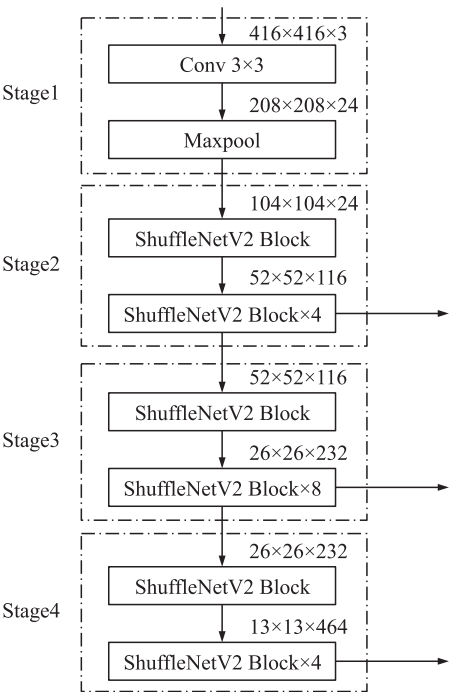


图 2 ShuffleNetV2 结构框架
Fig. 2 The structure framework of ShuffleNetV2

步幅设定影响,图 3、图 4 分别为步幅为 2 和 1 时 ShuffleNetV2 块的结构. 在后三个阶段中,每个阶段要先进行一个步幅为 2 的 ShuffleNetV2 块,如图 3,使得输出尺寸长宽变为输入的一半,再进行一个步幅为 1 的 ShuffleNetV2 块,如图 4,使通道数保持不变传入下一阶段并同时输出特征图.

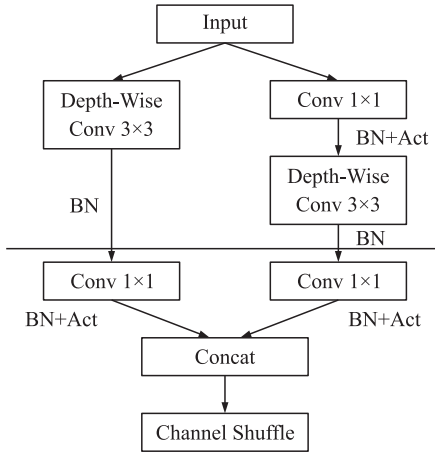


图 3 步幅为 2 时 ShuffleNetV2 块的结构框架

Fig. 3 The structure framework of ShuffleNetV2 in step 2

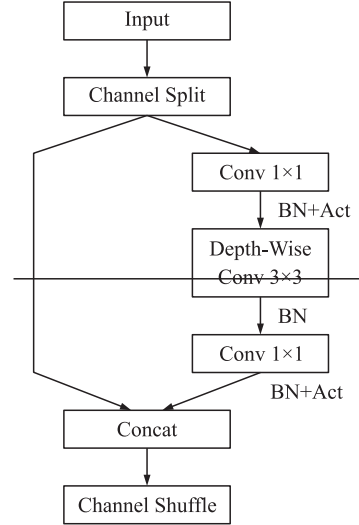


图 4 步幅为 1 时 ShuffleNetV2 块的结构框架

Fig. 4 The structure framework of ShuffleNetV2 in step 1

在主干网络中,采用 Mish 代替了 LeakyRelu 激活函数. 与 Relu 和 LeakyRelu 相比,使用 Mish 作为激活函数可以确保每个点的平滑性,从而使神经网络获得更好的梯度下降效果,更高的精度和泛化性;Mish 是非单调函数,有助于保持较小的负值并稳定网络梯度流. Mish 激活函数公式为(1),LeakyRelu 激活函数公式为(2),其中 x 代表输入, y 代表输出.

$$y = x * \tanh(\ln(1 + e^x)) \quad (1)$$

$$y = \max(0, x) + 0.1 * \min(0, x) \quad (2)$$

1.3 CBAM 注意力机制和轻量化路径汇聚网络

在上一阶段中,Backbone 输出了三个特征图,尺寸分别为 $52 * 52 * 116$, $26 * 26 * 232$ 和 $13 * 13 * 464$. 对这些特征图进行 $1 * 1$ 卷积运算,将它们通道数调整为 96. 然后将特征图输入 CBAM,输入特征先经过通道注意力模块,然后再经过空间注意力模块,以增强对重要信息的关注. CBAM 结构框架如图 5 所示.

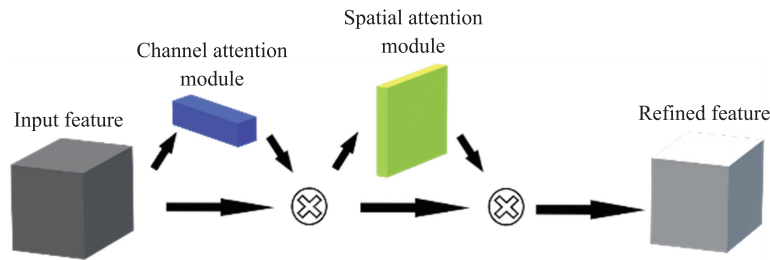


图 5 注意力机制 CBAM 结构框架

Fig. 5 The structure framework of attention mechanism CBAM

CBAM 是一种将通道注意力和空间注意力结合在一起的注意力模块. 与 SENet^[19]相比,SENet 只关注在通道上取得更好的效果,而 CBAM 不仅包含通道注意力模块,还包含空间注意力模块. 在通道注意力模块中,输入特征图将采用全局最大池化和全局平均池化,然后经过多层感知器. 多层感知器对输出特征图进行基于元素的加法运算,然后执行 Sigmoid 函数(3)激活以生成最终的通道注意力特征图.

$$Sigmoid(x) = \frac{1}{1 + e^x} \quad (3)$$

将通道注意力特征图和输入特征图相乘,以生成空间注意力模块所需的输入特征图. 空间关注模块

的输入特征图先进行基于通道的全局最大池和全局平均池,然后对这两个结果在通道上进行 concat 操作进行融合. 然后,在卷积运算后,将通道减小为 1,执行 Sigmoid 激活以生成空间注意力特征图. 最后,将空间关注模块的特征图和输入的特征图相乘,以获得最终的特征图,并将其输入到轻量化路径汇聚网络中.

路径汇聚网络是基于特征金字塔构建的,特征金字塔被包含在路径汇聚网络中. 特征金字塔是将功能图的自上而下的进行融合,并通过上采样和低级特征融合高级特征. 而路径汇聚网络利用底层特征的精确定位信息,增强了自下而上的路径. 本文所采用的轻量化路径汇聚网络作为图 1 中的 Neck 部分,与普通路径汇聚网络有所不同,它使用加法运算来替代 concat 运算,并且仅使用必要的双线性插值进行上采样和下采样,而无需进行任何卷积和非线性运算,以此来保持它的轻量,使本文算法的参数量更少.

1.4 GFLV2 检测头和损失函数

在原方法中检测头是基于 GFocalV1^[18] 改进的,在进行分类任务时,有 Classification 和 Center-ness 两个分支,本文算法检测头部分采用 GFocalV2 无锚轻量化检测头,消除了 Center-ness 分支,并在回归分支构建了分布指导质量预测器(Distribution-Guided Quality Predictor, DGQP)来指导对检测到的对象进行分类.

与原检测头相比本文引入了 DGQP,它可以将学习到的分布特性去指导分类,从而提高检测质量,可以实现现在提高检测准确性的同时,保持原有的检测速度且不会对训练和测试产生影响,GFLV2 如图 6 所示.

GFLV2 具有两个分支,分别为回归分支和分类分支. 回归分支对输入的特征图执行卷积运算,并输出边界框的预测. 分类分支也对输入特征图执行卷积和 Sigmoid 激活操作,其输出需要乘以 DGQP 输出的质量评估分数,然后分类分支输出对象类别的预测. DGQP 结构框架如图 7 所示.

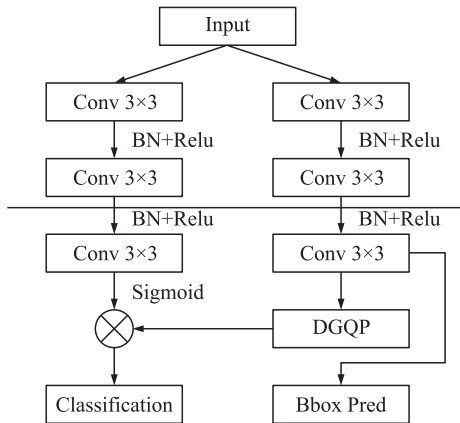


图 6 基于 GFLV2 检测头结构框架

Fig. 6 The structure framework of detection head based on GFLV2

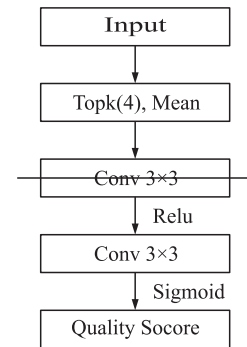


图 7 基于 DGQP 结构框架

Fig. 7 The structure framework of DGQP

DGQP 从回归分支中选出输出的前四个最大值,并计算他们的平均值,然后对它们和他们的平均执行 concat 操作,融合到一张特征图. 通过对该特征图进行卷积和 Sigmoid 操作来输出质量评估得分,质量评估得分可以指导分类分支以生成分类预测.

GFLV2 的损失函数包括 Distribution Focal Loss (DFL), Quality Focal Loss (QFL) 和 GIoU Loss (GIL).

DFL 迫使检测器迅速将注意力集中在标签 y 附近的值上,它显式增大最接近标签 y 两侧 y_i 和 y_{i+1} 的概率($y_i \leq y \leq y_{i+1}$). DFL 定义如公式(4).

$$DFL = -((y_{i+1} - y) \log(P(y_i)) + (y - y_i) \log(P(y_{i+1}))) \quad (4)$$

QFL 可以平衡负样本和正样本,它还可以支持从 0 到 1 的连续标签. $y=0$ 表示分类分数为 0 的负样本,而 $y_i < y \leq y_{i+1}$ 代表正样本分类分数为 y . 采用带有 sigmoid 运算的多重二分类,并将输出标记为 σ . DFL 定义如公式(5).

$$QFL = -|y - \sigma|^2((1 - y) \log(1 - \sigma) + y \log(\sigma)) \quad (5)$$

交并比(IoU)是边界框和地面实况框的交集和并集的比率. GIoU^[20] 与 IoU 相似,可以看到一个距离, GIL 定义如公式(6).

$$GIL = 1 - GIoU \quad (6)$$

GIoU 的值总是小于等于 IoU, IoU 的范围为 $0 \leq IoU \leq 1$, GIoU 范围则为 $-1 \leq GIoU \leq 1$. 当预测的边界框

和真实框完全重合时, $GIoU=IoU=1$. 我们可以找到一个包含边界框和地面实况框的最小形状, 将形状的面积标记为 C , 将边界框和地面实况框的并集标记为 U . $GIoU$ 可以计算公式如(7). 并根据公式(6), 即可得到 GIL.

$$GIoU=IoU-\frac{C-U}{C} \quad (7)$$

综上所述, 本文方法使用具有 Mish 激活函数的 ShuffleNetV2 的 Backbone, 使用 CBAM 增强对重要信息的关注, 并使用轻量化路径汇聚网络作为 Neck, 采用无锚轻量化检测头 GFLV2 作为 Head, 实现了在检测速度保证的同时大幅提高了检测精度.

2 实验与分析

为了验证本文算法的有效性, 采用 CPU: Intel Core i9-9900K 3.6GHz, GPU: Nvidia Titan-RTX 24GB, 内存: 64G RAM, 操作系统: Ubuntu20.04 作为实验环境进行实验.

本实验模型首先在 COCO 数据集上进行预训练, COCO 数据集有 80 个类别, 这里我们只需要训练 person 的类别即可, 再用各类比赛视频进行抽帧得到的图片以及网络收集到的比赛图片, 使用 Labelimg 对这些图片进行标注出人体, 制成共 5 996 张图片的自制数据集. 再将此数据集其随机划分为 5 397 张作为训练集, 599 张作为测试集, 进行实验.

由于模型输入图片的大小越大, 检测器可以提取的特征就越多. 将输入图像的分辨率设置为 $416 * 416$, 替代它原有的配置 $320 * 320$, 以从图像中获取更多信息并提高准确性. 经过特征提取和特征融合, 特征图的最终输出大小分别为 $13 * 13 * 96$, $26 * 26 * 96$ 和 $53 * 53 * 96$, 分别减小为 $1/32$, $1/16$ 和 $1/8$.

将训练好的模型在测试集进行推理计算, 并输出计算结果, 图 8 为不同复杂场景下的多人检测效果

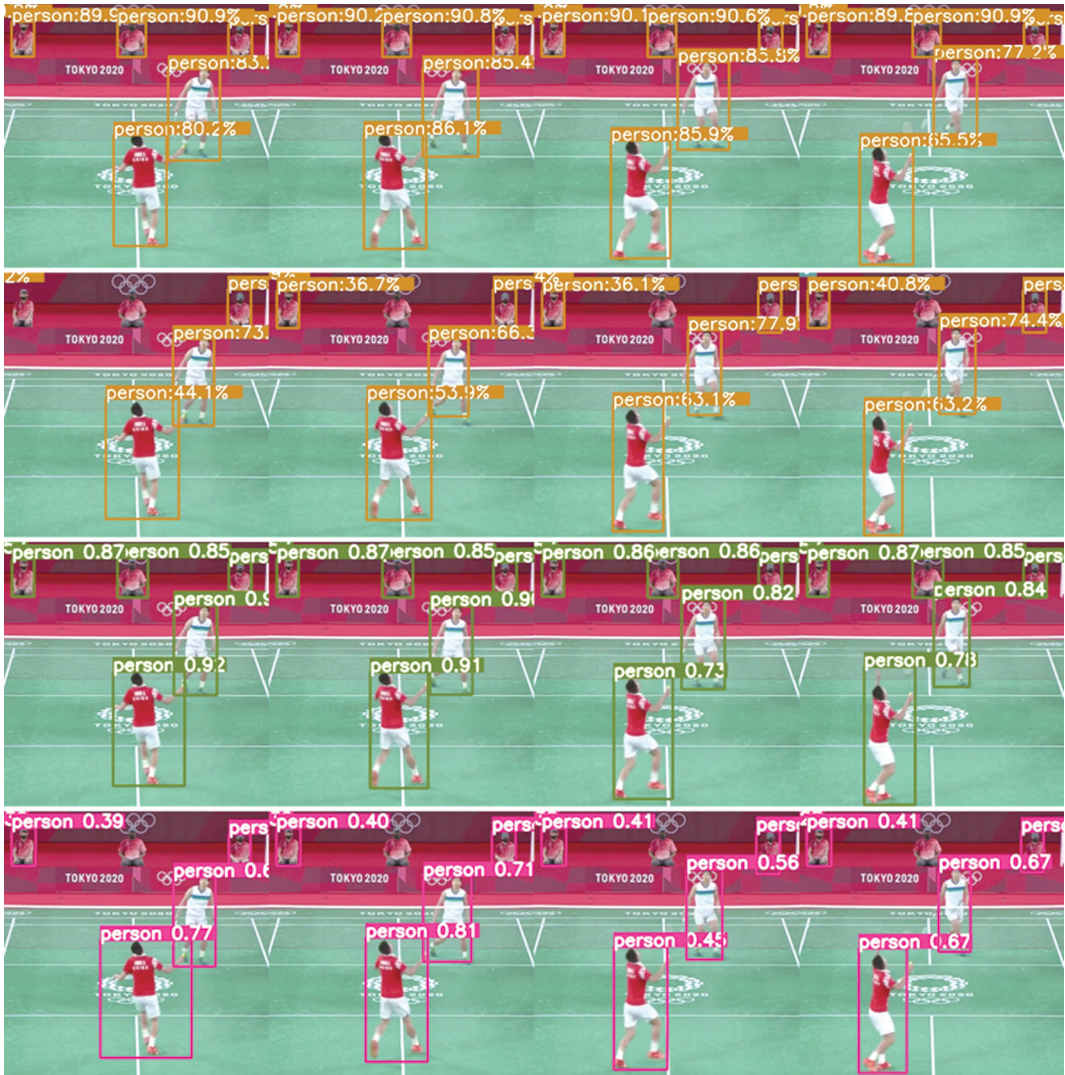


图 8 不同复杂场景下的多人检测效果图

Fig. 8 The multi-body detection results in different complex scenes

图. 可以看出在各复杂运动场景, 本文算法均可以准确检测出多人目标, 甚至部分遮挡情况下也可以准确检测出人体目标.

除此之外, 还将训练好的算法对连续的图片进行推理计算. 本文选用羽毛球比赛视频, 对其抽帧并进行推理计算, 效果如图 9.



算法从上向下依次为: 本文算法、NanoDet、Yolov3、Yolov3-tiny
图 9 四种算法对同一组连续图片的检测效果对比图

Fig. 9 The comparison of detection results with the same continuous pictures using four different algorithms

从图 9 中可以看出本文算法与 NanoDet^[21] 和 Yolov3-tiny 相比, 检测出了所有人体, 没有出现漏检的情况; 与同样没有漏检的 Yolov3 相比, 检测效果略好.

为了更好的评估本文算法, 设置目标检测交并比 IoU 阈值为 0.50, 即当检测框与真值框重叠区域超过两者并集区域的 50% 时, 认为该框检测正确, 并采用 IoU 为 0.50 时的平均精确率 AP50, 模型检测时间 Times, 每秒传输帧数 FPS, 模型参数量 Params 以及模型大小 Model Size 作为算法衡量指标.

表 1 为不同方法对人体目标检测性能对比, 在上述评价指标中, IoU 阈值为 0.5 时本文方法平均精度在 93.15%, 比其他方法准确率高; 每秒传输帧数为 137.2, 比多数方法高; 模型参数量和模型大小也比多数方法要小, 总体优于其他方法.

具体来看, 在平均精度上, 本文方法比改进前的 NanoDet^[21] 的 AP50 提高了 16.06%, 比同类型的轻量化检测模型 Yolov4-tiny 高 14.87%; 在检测速度上, 本文方法平均检测一张图片只需要 7.29 ms, 与改进前相比虽然时间增加了 0.84 ms, 但是换得了检测精度的大幅提高, 并且与同类型的轻量化检测模型 Yolov4-tiny 10.67 ms 的检测时间相比, 检测所用的时间减少了 31.68%; 在模型参数量和模型大小上, 由于加入了

CBAM 等模块,本文方法参数量和模型大小比 NanoDet 略有增加,但与轻量化检测模型 Yolov4-tiny 相比,模型参数量和模型大小只有其约 1/6。总体来看,本文方法与 Yolov3、Yolov4、Faster R-CNN 等算法相比,在检测速度上有巨大优势,同时在检测精度上更高和在模型大小上更小;本文方法与 Yolov3-tiny、Yolov4-tiny、NanoDet 算法相比,在检测精度上有质的提升,并且不输于对比方法的检测速度,并保持了轻量级的模型参数量和模型大小。

表 1 不同方法对人体目标检测性能对比

Table 1 The performance comparison of human body detection using different

Method	AP50	Times(ms)	FPS	Params	Model Size
Faster R-CNN	75.68	407.62	2.5	>100 M	>1 GB
SSD	83.24	70.17	14.3	>100 M	275 MB
Yolov3	88.74	28.65	34.9	61.9 M	118 M
Yolov3-tiny	62.65	13.12	76.2	8.9 M	34 M
Yolov4	92.71	136.44	7.3	65 M	246 M
Yolov4-tiny	78.28	10.67	93.7	6.1 M	23 M
NanoDet	77.09	6.45	155.0	0.95 M	4 M
Ours	93.15	7.29	137.2	1.1 M	5 M

综上所述,本文方法在复杂场景下可以准确检测出多人目标,对连续序列图像的处理效果优于其他对比方法,并且与其他对比方法相比本文方法模型小且参数量少,使其在保证检测速度的前提下进一步提高了检测精度,有效改善了复杂场景下对运动员人体检测速度和精度不足的问题。

3 结论

本文基于 NanoDet 改进的运动赛事场景人体检测算法,采用 CFocalV2 无锚轻量化检测头,并引入 CBAM 注意力机制,在牺牲少量检测时间的基础上大幅提高了检测精度,相较于其他主流算法,本文算法在保证更低检测时间的同时,其检测精度有着极大优势;同时本文算法对于不同比赛场景、有一定遮挡情况以及多目标检测下均能实现很好的检测效果。该算法能够适应各种实时性高的检测应用任务,并为未来运动员的动作统计、运动识别以及 AI 解说等方面打下基础。

[参考文献]

- [1] EINFALT M, LIENHART R. Decoupling video and human motion: towards practical event detection in athlete recordings[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Virtual, 2020: 892–893.
- [2] 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. Columbus, Ohio, USA, 2014: 580–587.
- [3] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137–1149.
- [4] 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. Las Vegas, USA, 2016: 779–788.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]//European Conference on Computer Vision. Springer, Cham, Amsterdam, Netherlands, 2016: 21–37.
- [6] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South), 2019: 9627–9636.
- [7] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, 2020: 10781–10790.
- [8] YANG Z, LI Z, JIANG X, et al. Focal and global knowledge distillation for detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA, 2022: 4643–4652.
- [9] GAO Z, WANG L, HAN B, et al. AdaMixer: A Fast-Converging Query-Based Object Detector [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA, 2022: 5364–5373.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on

- Computer Vision and Pattern Recognition. Las, Vegas, USA, 2016;770–778.
- [11] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv Preprint arXiv:1409.1556, 2014.
- [12] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Thirty-first AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017;4278–4284.
- [13] LIN T Y, DOLLÁ R P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San, Juan, USA, 2017;2117–2125.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018;8759–8768.
- [15] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020;9759–9768.
- [16] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018;6848–6856.
- [17] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich, Germany, 2018;3–19.
- [18] LI X, WANG W, WU L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. Advances in neural information processing systems, 2020, 33:21002–21012.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018;7132–7141.
- [20] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019;658–666.
- [21] 刘方坚, 李媛. 基于视觉显著性的 SAR 遥感图像 NanoDet 舰船检测方法[J]. 雷达学报, 2021, 10(6):885–894.

[责任编辑:陆炳新]