基于改进 YOLOv5 与 DeepSORT 的行人多目标跟踪算法研究

赵建光，范晶晶，韩泽山
（河北建筑工程学院，河北 张家口 075000）

摘要：目标跟踪算法在计算机视觉领域一直是研究的难点和热点，但是受行进中人员自身和周边因素的影响跟踪效果一直不佳。文章采用基于检测器的跟踪框架对目标进行跟踪。首先将 YOLOv5s 算法进行改进，为降低模型的计算量引入 GhostNet 轻量化模型，在添加 P-CBAM 注意力机制以增强检测器的有效特征提取；然后跟踪信息使用 DeepSORT 算法实现行人的跟踪。

关键词：多目标跟踪；YOLOv5s；GhostNet；DeepSORT
中图分类号：TP391　　**文献标识码：**A　　**文章编号：**2096-9759(2023)06-0029-04

Research on pedestrian multi-target tracking algorithm based on improved YOLOv5 and DeepSORT

ZHAO Jianguang, FAN Jingjing, HAN Zeshan
（Hebei University of Architecture, Zhangjiakou 075000,China）

Abstract: Target tracking algorithm has always been a difficult and hot topic in the field of computer vision, but the tracking effect is not good because of the influence of people's own and surrounding factors. This paper uses the tracking framework based on detector to track the target. Firstly, the YOLOv5s algorithm was improved, and GhostNet lightweight model was introduced to reduce the calculation amount of the model. P-CBAM attention mechanism was added to enhance the effective feature extraction of the detector. Then the tracking information uses DeepSORT algorithm to track pedestrians.

Key words: multi-target tracking; YOLOv5s; GhostNet; DeepSORT

1 引言

在计算机视觉领域，目标检测是被众多学者较早研究的方向，视频监控、人脸识别、自动驾驶等多个领域并取得长足的进步[1]。但是由于需要检测的目标是行进中的人群，行人受到自身和旁人的影响都会影响检测的准确性，因此对行人的跟踪也成为了计算机视觉领域中的一个难点和热点。

目标跟踪领域分为单目标和多目标跟踪，在现实应用中，单目标跟踪存在众多局限性。所以本文对多目标跟踪进行研究 [2]。传统的多目标目标跟踪方法是以图像为基础对行人进行识别后跟踪，由于缺少帧与帧之间的连续性对目标的跟踪效果不佳，在对视频中的行人进行跟踪时空间特征和时序特征都要考虑[3]，在对行人的跟踪过程中二者起到重要作用。

多目标跟踪算法可分为基于检测的（Tracking by Detection，TBD）和基于初始框的跟踪（Detection Free Tracking，DFT）[4]。DFT 与单目标跟踪的工作原理相似，需要在初始化时人为标记出第一帧的目标，若场景中目标数量偏多容易出现未被标记的目标随后也无法对目标进行跟踪，从而造成结果的不稳定性。因此 TBD 的实用性要优于 DFT，也是目前目标跟踪领域的常用跟踪策略，本文中使用 TBD 对行人进行多目标跟踪。

在目标跟踪中，陆续出现以 Fast-RCNN[5]、Faster-RCNN[6] 等二阶段检测算法和 SSD[7]、YOLO[8] 等一阶段检测算法作为目标跟踪框架的跟踪器。但是与二阶段不同的是一阶段目标检测算法可以直接对目标进行定位和分类，检测速度高于二阶段算法。并且随着神经网络的发展一阶段目标检测算法在精度方面也得到了极大的提升，其中以 YOLOv5 最为突出。

多目标跟踪算法的难点在于数据关联，起始，SORT 算法[9] 的使用较多，使用卡尔曼滤波联合 IOU 构建矩阵之后使用匈牙利算法关联目标轨迹，虽然算法速度快但是跟踪的目标容易出现身份切换。为解决这个问题，DeepSORT 算法[10] 被提出，该算法在 SORT 算法的基础上加入了鉴别网络，大幅度降低了身份切换次数，提高了跟踪精度。

就目前多目标跟踪算法训练速度慢、检测效果差的问题，提出一种轻量化与注意力机制相结合的模型。使用最为轻便的 YOLOv5s 作为检测器，并引入 GhostNet[11] 减少模型的计算量；为提高模型的检测效果，添加注意力机制 [12]，改进 CBAM，设计并行的 CBAM 添加到检测器部分。增强模型在目标上的有效特征提取。与原模型相比，改进后的模型在计算量和检测效果方面均得到优化。

2 YOLOv5s

YOLOv5 算法是在 YOLOv4[13] 算法的基础之上发展而来，该网络结构主要分为输入、Backbone、Neck 和输出部分。就 YOLOv4 做了较大的改进并去性能得到了极大的提升，YOLOv5 分为 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，以 YOLOv5s 最为轻便。输入端使用 Mosaic[14] 进行数据预处理，还有自适应锚框和自适应图片缩放功能。Backbone 部分使用 Focus 结构对图片进行切片操作之后经过 SPP 结构固定特征图输出的尺寸大小。Neck 部分使用 FPN+PAN 结构进行网络的特征融合，图片先经过自顶向下做上采样的 FPN 随后自底向上做下采样以增强模型对特征图的定位能力。输出部分改进了损失函数，使用 GIoU 作为损失函数并添加了 DIoU 进行预测框的筛选。

收稿日期：2023-03-14
基金项目：河北建筑工程学院硕士研究生创新基金项目《基于 YOLOv5s 算法的吸烟行为检测研究》（XY2023024）。
作者简介：赵建光（1978-），男，河北大名人，研究生，博士，主要研究方向：感知互联与智能计算。
通信作者：韩泽山（1998-），男，河北保定人，研究生，硕士，主要研究方向：目标检测。

3 DeepSORT

DeepSORT 是基于目标检测的多目标跟踪算法。DeepSORT 算法的前身是 SORT 算法，SORT 算法中两个核心算法为卡尔曼滤波和匈牙利算法。卡尔曼滤波算法[15]可以根据 t-1 帧的运动状态预测 t 帧的运动状态，然后将其进行线性加权。匈牙利算法[16]则是解决线性分配问题。

DeepSORT 算法在 SORT 算法的基础上添加鉴别网络提取检测框的外貌特征并且使用级联匹配减少身份切换的次数。在进行多目标跟踪的过程中可能会出现一个行人遮挡其他行人的情况，DeepSORT 算法采用八维状态空间（u，v，r，h，x，y，r，h）定义跟踪场景，分别记录是目标任务的中心点、纵横比、图片高度以及其对应的运动速度。将预测的目标框和当前状态的目标框进行马氏距离的计算，当结果小于指定的阈值则是完成目标的跟踪。如公式（1）所示：

$$d^{(1)}(i,j)=(d_j-y_i)^T S_i^{-1}(d_j-y_i) \tag{1}$$

其中 $d^{(1)}(i,j)$ 表示第 j 个检测目标和第 i 条轨迹之间的运动匹配度，S_i 为使用卡尔曼滤波预测后与当前时刻生成的协方差矩阵，y_i 是轨迹当前时刻的预测量，d_i 则是包含第 j 个目标的位置信息。

设置关联状态后，可以得到示性函数，如公式（2）所示：

$$b_{i,j}^{(1)}=I[d^{(1)}(i,j)]\le t^{(1)} \tag{2}$$

以此类推得到 $d^{(2)}(i,j)$ 和 $b_{i,j}^{(2)}$，最终构成线性加权函数，如公式（3）所示：

$$C_{i,j}=\lambda d^{(1)}(i,j)+(1-d)d^{(2)}(i,j) \tag{3}$$

最后使用匈牙利算法检测当前帧目标是否存在。通过上述的 DeepSORT 算法完成多目标跟踪，在得到当前检测目标后可以准确跟踪到目标下一时刻的位置信息并且解决目标遮挡问题。

4 注意力机制

4.1 CBAM 注意力机制

CBAM 是 2018 年提出的一种轻量注意力模块，如图 1 是 CBAM 注意力机制，该注意力机制分为通道和空间两个注意力模块，通过通道维度和空间维度的结合进行特征的提取。

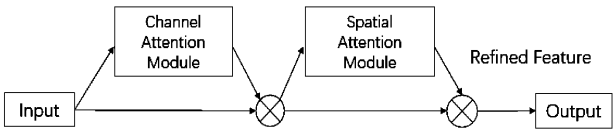

图 1 CBAM 注意力机制

输入的特征图先进入通道注意力进行池化、卷积、归一化后经过 Sigmoid 激活函数输出，重新定义权重。再进入空间注意力模块进行 GAP 和 GMP 操作，之后进入一个多层感知机得到注意力权重，然后通过 Sigomid 函数与原通道权重相乘重新标定权重。重新标定如公式（4）所示，输入特征图 F，W_0 和 W_1 分别为通道注意力和空间注意力权重。

$$\begin{aligned} M_C(F)&=\sigma(MLP(AvgPool(F))+MLP(MaxPool(F)))\\ &=\sigma(W_1(W_0(F_{\max}^c))) \end{aligned} \tag{4}$$

随后经过空间注意力对特征图进行标定，如公式（5）所示，f 为特征图经过 7×7 卷积核的卷积操作。

$$\begin{aligned} M_S(F)&=\sigma(f^{7\times7}([AvgPool(F);MaxPool(F)]))\\ &=\sigma(f^{7\times7}([F_{avg}^s;F_{\max}^s])) \end{aligned} \tag{5}$$

4.2 P-CBAM 注意力机制

如图 1 所示，CBAM 首先生成通道注意力特征图，用作空间注意力的输入，最终将两模块混合作为 CBAM 的输出。CBAM 中两模块串行工作，虽然可以提高模型的性能，但是缺乏灵活性。在实际应用中，不同卷积神经网络有不同侧重点，有些层更加注重通道的表达，有些层注重空间的表达，在这种情况下还引入空间注意力或通道注意力反而会造成不好的效果，产生过拟合现象。为了解决这一问题，本文将 CBAM 的串行模块改为并行模块，赋予两模块相同的优先级，采用加权的方式充分利用通道和空间维度提取的特征图信息，将并行的 CBAM 注意力机制称为 P-CBAM。

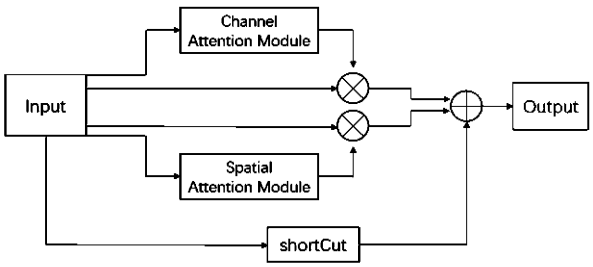

图 2 P-CBAM 注意力机制

P-CBAM 注意力机制如图 2 所示，基于 CBAM 中的公式（4）与公式（5），P-CBAM 的输出公式如（6）、（7）、（8）所示。

$$F_C=M_C(F)\otimes F \tag{6}$$

$$F_S=M_S(F)\otimes F \tag{7}$$

$$F_{Out}=\begin{cases}\dfrac{w_1\cdot F_C+w_2\cdot F_S}{w_1+w_2}(shortCut=False)\\ \dfrac{w_1\cdot F_C+w_2\cdot F_S+w_3\cdot F}{w_1+w_2+w_3}(shortCut=Ture)\end{cases} \tag{8}$$

5 模型轻量化

由于内存和计算资源有限，在 CVPR2020 的一篇论文提出 GhostNet，通过廉价的操作获取更多的特征图，很多卷积神经网络为了获取高准确率，通常的做法就是增加卷积层，增加参数，以增加模型的复杂度提升模型精度。而 GhostConv 通过常规的卷积、Ghost 特征图生成和特征图拼接三个步骤实现计算量的压缩。

（1）在常规卷积操作中，计算量为 w'×h'×c×k×k×n。经过常规的卷积操作，得到 m 个 intrinsic feature maps，这一过程计算量为 $Y'=X*f'=h'\times w'\times n\times c\times k\times k$。其中 $Y'\in R^{h'\times w'\times m}$，$X\in R^{c\times h\times w}$，$f'\in R^{c\times k\times k\times m}$，卷积核大小为 k×k。

（2）将 Y' 每个经过常规卷积得到的特征图进行 cheap linear operations（∅操作），在这一过程中每个线性操作的卷积核大小均为 d×d，每个特征图生成 s 张特征图，共生成 m×s 张特征图，如公式（9）所示，最终可以得到 Ghost 特征图 y_{ij}。

$$y_{ij}=\phi_{i,j}(y_i),\forall i=1,...,m,j=1,...,s \tag{9}$$

（3）将步骤（1）得到的 intrinsic feature maps 和步骤（2）得到的 Ghost 特征图通过 identity 连接在一起得到最后的输出。

普通卷积和 GhostConv 的计算量之比如公式（10）所示，从结果可以看出普通卷积近似为 GhostConv 的 s 倍。

$$\begin{aligned} r&=\frac{n\cdot h'\cdot w'\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot h'\cdot w'\cdot c\cdot k\cdot k+(s-1)\cdot\frac{n}{s}\cdot h'\cdot w'\cdot d\cdot d}\\ &=\frac{c\cdot k\cdot k}{\frac{1}{s}\cdot c\cdot k\cdot k+\frac{s-1}{s}\cdot d\cdot d}\approx\frac{s\cdot c}{s+c-1}\approx s \end{aligned} \tag{10}$$

6 实验过程及结果

6.1 数据集介绍及实验配置

本次实验数据集方面分为两个部分，第一部分是检测器权重的训练，采用 WiderPerson dataset 数据集，该数据集包含 13382 张行人的图片，训练轮次为 200 轮，根据改进模型训练出预训练权重。第二部分训练出跟踪部分的权重，使用 DeepSORT 算法在 MOT16 数据集中训练 200 轮次。进行实验的配置如表 1 所示。

表 1 实验配置

实验数据	配置
操作系统	Windows 10
CPU	Intel(R) Core(TM) i5-7300HQ CPU @ 2.50GHZ
RAM	8GB
GPU	GeForce GTX 1050
显存	4GB
轮次	200

6.2 实验结果

对检测器进行轻量化处理以降低模型训练时的计算量，为降低算法模型在内存空间和计算资源的占用，我将原模型的 CSP 模块换成 GhostBottleneckCSP，将普通卷积层换为 Ghost 卷积层。从表 2 中反映出，在维持算法准度率的基础上，使用修改后的 YOLOv5s 模型在运算量上大幅度的降低，在“每秒执行浮点运算”的这一指标中，使用 GhostNet 轻量化的模型 GFlops 仅为原模型的 37%，参数量仅为原模型的 40%。

表 2 改进模型与不同模型的性能比较

算法	Precision	Parameters	GFlops
YOLOv3-tiny	74.7%	8669002	12.9
YOLOv4-tiny	77.4%	6399178	21.8
YOLOv5s	81.5%	7056607	16.4
YOLOv5s+GhostCSP	80.4%	4019095	8.4
GhostCSP+GhostConv	80.2%	2806775	6.1

检测器部分使用 YOLOv5s 模型对目标进行检测，对目标添加注意力机制以增强模型的检测效果，起初添加 CBAM、SE 等注意力机制，但是检测效果并未达到实验预期。如图 3 所示，(a) 是在使用 YOLOv5s 模型作为检测器时模型的跟踪效果，模型将橱窗中的模特也当做行人进行的跟踪。(b) 为对检测器添加 CBAM 注意力机制，模型的检测效果并未得到优化反而将橱窗中更多的模特当做需要跟踪的行人，错检现象更加严重。(c) 为对 YOLOv5s 添加 P-CBAM 注意力机制后的检测效果，从图中可以看出添加 P-CBMA 注意力机制后模型的错检现象得到了解决，模型并未对错误的目标进行跟踪，检测效果得到了优化。

图 3 跟踪效果图

6 结语

由于 YOLOv5s 模型的轻便和准确率高的特性，使用 YOLOv5s 与 DeepSORT 算法相结合分别作为目标跟踪框架的检测器和跟踪器。为进一步降低模型在训练时的计算量，将 GhostNet 与 YOLOv5s 结合，设计 GhostCSP+GhostConv。随后为解决 YOLOv5s 在进行目标跟踪时出现错检的现象为检测器添加注意力机制，将 CBAM 注意力机制设计为并行的 P-CBAM 在解决错检的问题上起到很好的效果，提高了模型的检测效果。

参考文献：

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton. ImageNet Classification with DeepConvolutional Neural Networks[J]. Advances in neural information processing systems, 2012,25: 1097-1105.

[2] 韩瑞泽,冯伟,郭青,胡清华.视频单目标跟踪研究进展综述[J].计算机学报,2022,45(09):1877-1907.

[3] 王子越,陈贤富.一种基于机器视觉的机场场面多目标跟踪算法[J].电子技术,2022,51(12):16-18.

[4] 罗茜,赵睿,庄慧珊,罗宏刚.YOLOv5 与 Deep-SORT 联合优化的无人机多目标跟踪算法[J].信号处理,2022,38(12):2628-2638.DOI:10.16798/j.issn.1003-0530.2022.12.017.

[5] 梁俊欢,董峦,孙宗玖,马海燕,艾尼玩·艾买尔,阿仁,阿斯娅·曼力克,郑逢令.基于改进 Faster-RCNN 模型的无人机影像白喉乌头物种的检测[J].新疆农业大学学报,2022,45(04):323-329.

[6] 贺艺斌,田圣哲,兰贵龙.基于改进 Faster-RCNN 算法的行人检测[J].汽车实用技术,2022,47(05):34-37.DOI:10.16638/j.cnki.1671-7988.2022.005.008.

[7] 韦正璐,王家晨,刘庆华.基于改进 SSD 算法的路面破损检测 [J]. 电子设计工程, 2023,31 (03): 63-68. DOI: 10.14022/j.issn1674-6236.2023.03.013.

（下转第 36 页）

dose computed tomographic screening [J]. N Engl J Med, 2011, 306(5):395-332.
[2] International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF, et al. Survival of patients with stage I lung cancer detected on CT screening[J]. N Engl J Med, 2006, 355(17):1763-1771.
[3] Siegel, R., D. Naishadham, and A. Jemal, Caner statistics, 2013.CA Cancer J Clin,2013,63(1):p.11-30.
[4] Natbnal Lung screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening [J]. N Engl J Med, 2011, 365(5):395-409.
[5] Li, D., A. Djulovic, and J. Xu, A Study of k NN using ICU Multivariate Time Series Data [C], in The 9th International Conference on Data Mining (DMIN'13). 2013, CSREA: Las Vegas. 211~217.
[6] Thongkam J, Xu G, Zhang Y. Ada Boost algorithm with random forests for predicting breast cancer survivability [C]// Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE, 2008: 3062~3069.
[7] Cheng C W, Chanani N, Maher K, et al. icu ARM-II: improving the reliability of personalized risk prediction in pediatric intensive care units [C]//Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. ACM, 2014: 211-219.
[8] Nakhaei, Fardis, Irannajad, et al. Application and comparison of RNN, RBFNN and MNLR approaches on prediction of flotation column performance[J]. International Journal of Mining Science and Technology, 2015, 25(6):983-990.
[9] Dastider A G ,Sadik F , Fattah S A . An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound[J]. Computers in Biology and Medicine, 2021, 132(2):104296.
[10] Islam M Z , Islam M M , Asraf A . A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images[J].Informatics in Medicine Unlocked, 2020, 20:100412.
[11] 张宇冲.基于深度学习预测 CT 图像中肺结节的病理亚型[D].中国医科大学,2022.
[12] Ketu S, Mishra P K. India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability[J]. Soft Computing, 2022, 26(2): 645-664.
[13] Qiang Y, Zhang X, Ji G, et al. Automated Lung Nodule Segmentation Using an Active Contour Model Based on PET/CT Images [J]. Journal of Computational and Theoretical Nanoscience, 2015, 12(8): 1972-1976.
[14] Kasun L L C, Zhou H, Huang G B, et al. Representational learning with ELMs for big data[J]. IEEE Intelligent Systems, 2013, 28(6): 31-34.
[15] 赵鑫,强彦,葛磊.基于双模态深度自编码的孤立性肺结节诊断方法[J].计算机科学,2017,44(08):312-317.
[16] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997, 9(8):1735-1780.
[17]]Polat, K., and Gunes, S., Priciples component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lungcancer. Expert Syst. Appl. 34:214-221,2008.
[18] Suzuki K, Armato III S G, Li F, et al. Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography[J].Medical Physics,2003, 30(7): 1602-1617.
[19] Keshani M, Azimifar Z, Tajeripour F, Boostani R. Lung nodule segmentation and recognition using SVM classifier and active contour modeling: a complete intelligent system. Comput Biol Med. 2013;43:287-300.
[20] Jing Z, Bin L, Lianfang T. Lung nodule classification combining rule-based and SVM. In: Edited by Li K,Proceedings of the IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications: 23-26 September 2010; Changsha, China. Piscataway,NJ: IEEE Computer Society;2010. p.1033-36
[21] Dolejsi, M., Kybic, J., Polovincak, M.,Tuma, S.,2009.The Lung TIME: annotatedlung nodule dataset and nodule detection framework, vol. 7260, SPIE, 72601U, doi: http://dx.doi.org/10.1117/12.811645, URL <http://link.aip.org/link/PSI/7260/72601U/1>.

（上接第 31 页）

[8] 王彤,李琦.基于残差网络与特征融合的改进 YOLO 目标检测算法研究[J].河北工业大学学报,2023,52(01):41-49.DOI:10.14081/j.cnki.hgdxb.2023.01.006.
[9] 陈锋军,朱学岩,周文静,郑一力,顾梦梦,赵燕东.利用无人机航拍视频结合 YOLOv3 模型和 SORT 算法统计云杉数量[J].农业工程学报,2021,37(20):81-89.
[10] 罗茜,赵睿,庄慧珊,罗宏刚.YOLOv5 与 Deep-SORT 联合优化的无人机多目标跟踪算法[J].信号处理,2022,38(12):2628-2638.DOI:10.16798/j.issn.1003-0530.2022.12.017.
[11] 石博雅,董学峰.融合 GhostNet 的 YOLOv4 轻量化网络设计与实现 [J/OL]. 小型微型计算机系统: 1-9 [2023-03-13].DOI:10.20009/j.cnki.21-1106/TP.2021-0516.
[12] 曾凯,李响,陈宏君,文继锋.引入注意力机制的改进型 YOLOv5 网络研究[J].软件工程,2023,26(01):55-58+54.
[13] 徐翔,蔡茂国,唐剑兰.基于改进的 YOLOv4 的目标识别研究[J].信息技术,2022,46(12):107-111+117.DOI:10.13274/j.cnki.hdzj.2022.12.019.
[14] 陈翠琴,范亚臣,王林.基于改进 Mosaic 数据增强和特征融合的Logo检测[J].计算机测量与控制,2022,30(10):188-194+201.DOI:10.16526/j.cnki.11-4762/tp.2022.10.029.
[15] 沈皓,常军.基于改进卡尔曼滤波的结构损伤识别方法研究[J].苏州科技大学学报(工程技术版),2022,35(04):14-19.
[16] 邬春学,贺欣欣.基于改进匈牙利算法对多人人体关键点匹配的研究[J].信息技术与网络安全,2022,41(05):45-50+58.DOI:10.19358/j.issn.2096-5133.2022.05.007