ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 37,No. 1,Jan. 2022,pp. 108－121
DOI:10. 16337/j. 1004-9037. 2022. 01. 009

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于GhostNet与注意力机制的行人检测跟踪算法

王立辉，杨贤昭，刘惠康，黄晶晶

（武汉科技大学冶金自动化与检测技术教育部工程研究中心，武汉 430081）

摘　要：针对复杂场景下仅依靠传统的目标检测与跟踪算法进行跟踪时准确度低且速度慢的问题，提出一种基于GhostNet与注意力机制结合的行人检测与跟踪算法。首先，将YOLOv3的主干网络替换为GhostNet，保留多尺度预测部分，利用Ghost模块减少深度网络模型参数和计算量，在Ghost模块中融入注意力机制给予重要特征更高的权值。然后，引入目标检测的直接评价指标GIoU来指导回归任务。最后，利用Deep-Sort算法进行跟踪。在公共数据集上实验表明，改进后的模型平均精确度均值（mean Average precision，mAP）达到了92.53%，帧速率是YOLOv3模型的2.5倍；所提算法跟踪准确度优于改进前及其他算法，可以精确有效地跟踪复杂场景下的多目标行人，并具有较强的鲁棒性。

关键词：视频监控；目标检测；行人跟踪；YOLOv3；GhostNet；Deep-Sort跟踪算法

中图分类号：TP391.41　　**文献标志码：**A

Pedestrian Detection and Tracking Algorithm Based on GhostNet and Attention Mechanism

WANG Lihui，YANG Xianzhao，LIU Huikang，HUANG Jingjing

（Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education，Wuhan University of Science and Technology，Wuhan 430081，China）

Abstract：Aiming at the problems of low accuracy and slow speed when only relying on traditional object detection and tracking algorithms in complex scenes，a pedestrian detection and tracking algorithm based on GhostNet and attention mechanism is proposed. First，the backbone network of YOLOv3 is replaced with GhostNet to retain the multi-scale prediction part，the Ghost module is used to reduce the parameters and calculations of the deep network model，and the attention mechanism is integrated into the Ghost module to give higher weight to important features. Then，the direct evaluation index GIoU of object detection is introduced to guide the regression task. Finally，the Deep-Sort algorithm is used for tracking. Experiments on public data sets show that：The mean Average precision（mAP）of the improved model reaches 92.53%，and the frame rate is 2.5 times that of the YOLOv3 model；The tracking accuracy of the proposed algorithm is better than that before the improvement and that of other algorithms；The algorithm can track multi-object pedestrians in complex scenes accurately and effectively，and has strong robustness.

Key words: video surveillance; object detection; pedestrian tracking; YOLOv3; GhostNet; Deep-Sort tracking algorithm

---

基金项目：国家重点研发计划(2017YFC0805100)。
收稿日期：2021-01-23；**修订日期：**2021-05-08

引 言

行人检测与跟踪是计算机视觉领域的研究热点,可应用于交通监测、视频监控、安防等多个领域,具有一定的应用价值和挑战性[1],其实现方式包括检测跟踪和无检测跟踪。由于当前深度神经网络在目标检测领域已经得到了很好的应用,如SSD[2]、Fast RCNN[3]以及YOLO[4-6]等相关算法,其中YOLOv3在速度与性能上优势明显。但随着性能的提升,卷积神经网络结构不断加深,网络参数和计算量也在不断增长,无法达到实时性要求。

为了减小网络参数提升算法实时性,一些轻量化的网络设计应运而生。Xception[7]利用深度卷积运算以更有效地使用模型参数。MobileNet[8]是基于深度可分离卷积的一系列轻量级深度神经网络。MobileNetV2[9]采用反向残差块,而MobileNetV3[10]以更少的浮点数获得更好的性能。ShuffleNet[11]引入了通道混洗操作以改善通道组之间的信息流交换。ShuffleNetV2[12]进一步考虑了紧凑模型设计中的实际速度。尽管这些模型仅用很少的浮点数即可获得出色的性能,但从未充分利用特征图之间的相关性和冗余性。韩凯等[13]提出了一种新型的端侧神经网络架构GhostNet,提供了一个全新的Ghost模块,旨在通过廉价操作生成更多的特征图,以很小的代价生成许多能从原始特征发掘所需信息的幻影特征图。

在跟踪方面,Zhang等[14]提出一种新型检测跟踪方法,即将检测与轨迹联系起来形成长轨迹。Mahmoudi等[15]使用卷积神经网络代替手工标注进行特征提取,并提出一种新的2D在线环境分组方法,具有较高的准确率和实时性。Xiang等[16]设计了一个卷积神经网络来提取针对人的重识别,并使用长期短期记忆网络(Long short-term memory,LSTM)提取运动信息来编码目标的状态。而Deep-Sort多目标跟踪算法[17]则在Sort算法[18]的基础上提取深度表观特征,使跟踪效果有了明显的提升。

本文针对复杂场景下行人跟踪准确度低且速度慢的问题,提出了基于GhostNet与注意力机制的行人检测跟踪算法。实验证明所提算法可以有效精确地跟踪复杂场景下的多目标行人,具有较好的鲁棒性且兼具实时性特点。

1 YOLOv3目标检测算法

YOLOv3是基于回归的目标检测算法,特征提取采用创建的深度残差网络DarkNet53,后采用区域推荐网络中的锚点机制,并将多个尺度融合,结构如图1所示。YOLOv3算法避免了细小物体的漏检问题,故本文基于该算法进行改进。

2 本文算法

本文采用改进的YOLOv3目标检测算法对行人目标进行检测,将YOLOv3的主干网络替换为GhostNet,保留多尺度预测部分,减少深度网络模型参数和计算量以加快检测速度;加入SE(Squeeze-and-excitation)注意力机制给予重要特征更高的权值以提高检测跟踪的准确度,引入目标检测的直接评价指标GIoU来指导回归任务;最后将基于GhostNet的目标检测算法与Deep-Sort相结合,进行行人检测与跟踪。

2.1 GhostNet算法

GhostNet结合了线性运算和普通卷积,由已生成的普通卷积特征图进行线性变换得到相似特征图,从而产生高维卷积效果,减少了模型参数和计算量。结合了卷积运算和线性运算的模块称为Ghost模块,如图2所示。图中,Y表示通过卷积生成的固有特征图,Y'表示通过线性运算生成的冗余特征图。

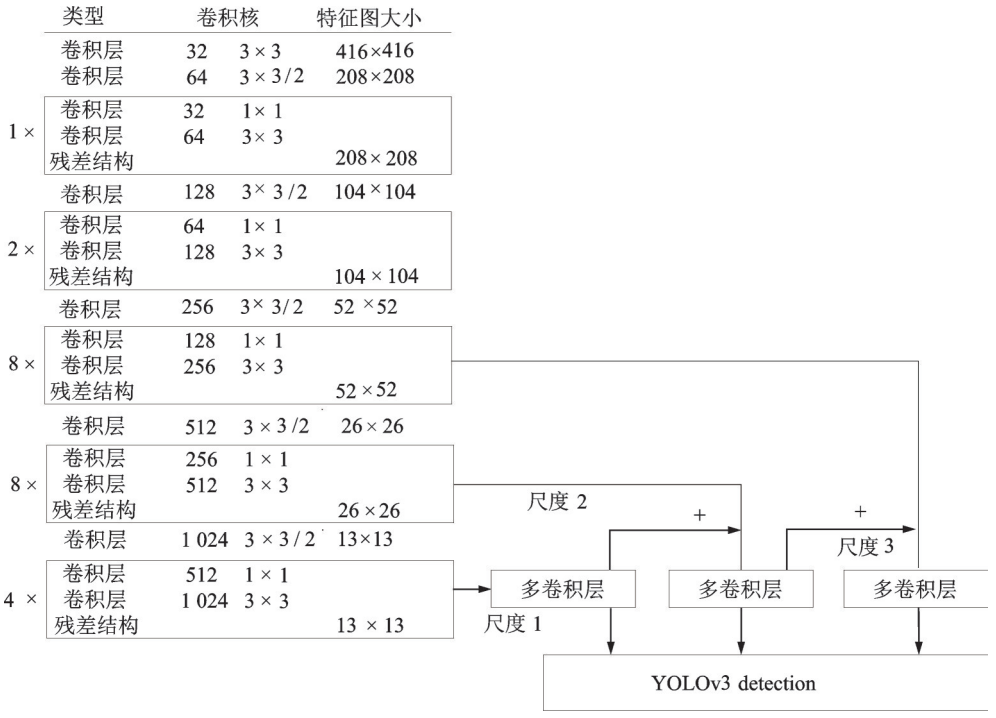


图 1 YOLOv3多尺度预测部分结构图

Fig.1 Structure of multi-scale prediction of YOLOv3

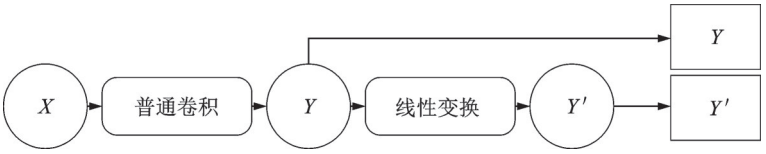


图 2 Ghost模块原理图

Fig.2 Schematic diagram of Ghost module

对于任意卷积层生成n个特征图的操作可以表示为

$$\boldsymbol{Y}_0 = \boldsymbol{X} * \boldsymbol{f} + b \tag{1}$$

式中:输入数据$\boldsymbol{X} \in \mathbf{R}^{c \times h \times w}$;$\boldsymbol{f} \in \mathbf{R}^{c \times k \times k \times n}$为该层的卷积核;*表示卷积操作;$b$为偏置项。此时得到的特征图$\boldsymbol{Y}_0 \in \mathbf{R}^{h' \times w' \times n}$。这个卷积过程需要的浮点数为$n \cdot h' \cdot w' \cdot c \cdot k \cdot k$。原输出的特征为某些内在特征且通常数量都很少,可以通过一个普通卷积操作生成,即

$$\boldsymbol{Y} = \boldsymbol{X} * \boldsymbol{f}' \tag{2}$$

式中:$\boldsymbol{Y} \in \mathbf{R}^{h' \times w' \times m}$为普通卷积输出;$\boldsymbol{f}' \in \mathbf{R}^{c \times k \times k \times m}$为使用的卷积核;由于$m \leqslant n$,将偏置项简化。

现在需要得到n维的特征图,对得到只有m维的固有特征图进行一系列简单线性变换为

$$y_{ij} = \Phi_{ij}(y_i') \qquad \forall i = 1, \cdots, m; j = 1, \cdots, s \tag{3}$$

式中:y_i'为固有特征图中的第i个特征图;$\Phi_{i,j}$为第i个特征图进行的第j个线性变换的线性变换函数。最后,增加1个恒等映射$\Phi_{i,s}$,将固有特征图叠加到经线性变换得到的特征图上,以保留固有特征图。

假设Ghost模块包含1个固有特征图和$m \cdot (s-1) = \dfrac{n}{s} \cdot (s-1)$个线性变换操作,每个操作核大小

为$d\times d$,Ghost模块升级普通卷积的理论加速比为

$$r_s=\frac{n\cdot h'\cdot w'\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot h'\cdot w'\cdot c\cdot k\cdot k+(s-1)\cdot\frac{n}{s}\cdot h'\cdot w'\cdot d\cdot d}=\frac{c\cdot k\cdot k}{\frac{1}{s}\cdot c\cdot k\cdot k+\frac{s-1}{s}\cdot d\cdot d}\approx\frac{s\cdot c}{s+c-1}\approx s \tag{4}$$

式中$d\times d$与$k\times k$幅度相似,$s\ll c$,则理论的参数压缩比为

$$r_c=\frac{n\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot c\cdot k\cdot k+(s-1)\cdot\frac{n}{s}\cdot d\cdot d}\approx\frac{s\cdot c}{s+c-1}\approx s \tag{5}$$

用Ghost模块升级普通卷积的理论参数压缩比大约等于理论加速比,线性操作每个通道的计算成本远小于普通的卷积。设计Ghost瓶颈层用来存储Ghost模块,如图3所示。

Ghost瓶颈层类似于ResNet[19]中的基本残差块,其中集成了多个卷积层和shortcut。Ghost瓶颈层主要由2个堆叠的Ghost模块组成:第1个Ghost模块用作扩展层来增加通道数;第2个Ghost模块减少通道数,使其与shortcut一致,然后连接这2个Ghost模块的输入和输出。步长为2的Ghost瓶颈层插入了深度可分离卷积层,减小特征几何变化的影响,降低了参数规模。

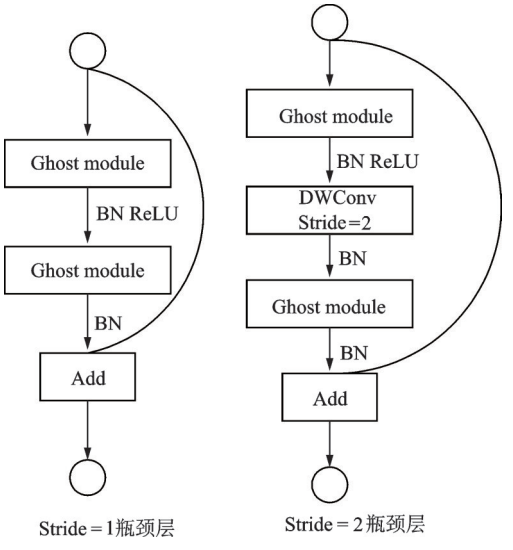


图3 Ghost瓶颈层

Fig.3 Ghost bottleneck layer

2.2 SE注意力模块

Hu等[20]提出了通道注意力模块,通过对不同通道赋予不同的权重来获取每个特征通道的重要程度。图4为加入到GhostNet中的SE注意力模块。图中:F_{tr}表示传统卷积操作;X表示F_{tr}的输入;U为F_{tr}的输出;C表示图像通道数;W表示图像宽度;H表示图像高度;C'、H'、W'为卷积操作F_{tr}之前的图像通道数、图像高度和宽度。注意力模块是U后面的结构:首先对U进行一次全局平均池化,对应图中的$F_{sq}(\ )$;然后将输出得到的$1\times1\times C$数据经过两级全连接,对应图中的$F_{ex}(\ )$;最后用sigmoid归一化化至0~1范围,将这个值作为重要因子乘到U的C个通道上,作为下一级的输入数据。通过注意力机制,给予重要的行人目标特征更多的关注,从而让提取的特征指向性更强,特征利用更充分。

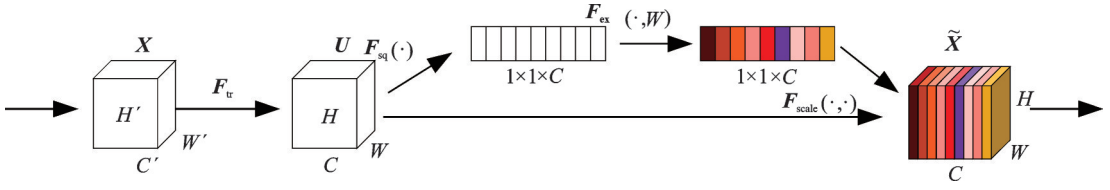


图4 SE注意力模块原理图

Fig.4 Schematic diagram of SE attention module

2.3 YOLOv3-GhostNet-SE网络结构

为了简化模型大小,本文对YOLOv3的特征提取层重新设计了轻量级网络单元。将GhostNet网络中14×14、28×28和56×56三种大小不同的特征图与YOLOv3多尺度特征进行拼接,加入SE注意力模块,构成最终目标检测模型。替换后的网络结构如图5所示。

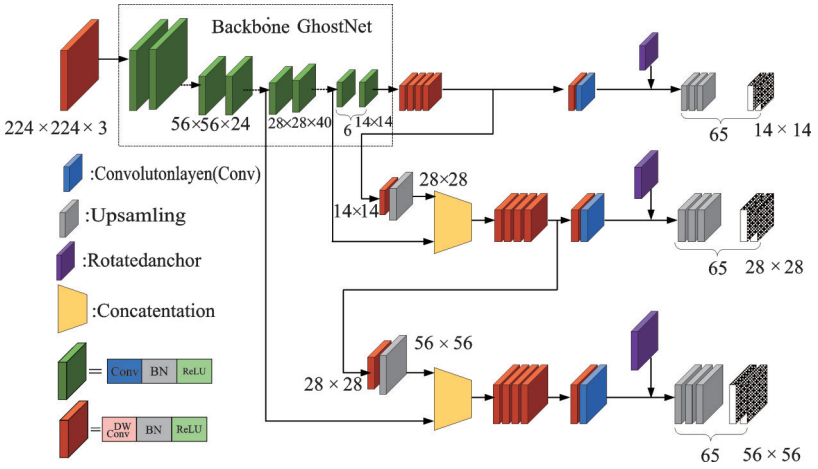


图5 YOLOv3-GhostNet-SE网络结构图

Fig.5 Network structure of YOLOv3-GhostNet-SE

2.4 模型参数细节

模型的重要参数信息如表1所示。

将YOLOv3原有的Darknet53特征提取网络替换为结构更简单、运算复杂度更低的GhostNet网络,替换后可实现卷积降维,从而对整体神经网络架构进行加速,采用多个尺度融合的方式来进行预测,在保存全面特征信息的同时减少了计算量。通过YOLOv3-GhostNet-SE对行人目标进行检测,利用有限的计算资源达到高精度效果。

表1 模型参数信息表

Table 1 Model parameter information table

输入	操作方法	步长	输出	SE注意力模块
$224^2\times3$	二维卷积3×3	2	16	
$112^2\times16$	Ghost瓶颈层1	1	16	
$112^2\times16$	Ghost瓶颈层2	2	24	
$56^2\times24$	Ghost瓶颈层1	1	24	
$56^2\times24$	Ghost瓶颈层2	2	40	
$28^2\times40$	Ghost瓶颈层1	1	40	
$28^2\times40$	Ghost瓶颈层2	2	80	
$14^2\times80$	Ghost瓶颈层1	1	80	
$14^2\times80$	Ghost瓶颈层1	1	80	
$14^2\times80$	Ghost瓶颈层1	1	80	
$14^2\times80$	Ghost瓶颈层1	1	112	1
$14^2\times112$	Ghost瓶颈层1	1	112	1
$14^2\times112$	Ghost瓶颈层2	2	160	1
$7^2\times160$	Ghost瓶颈层1	1	160	
$7^2\times160$	Ghost瓶颈层1	1	160	1
$7^2\times160$	Ghost瓶颈层1	1	160	
$7^2\times160$	Ghost瓶颈层1	1	160	1
$7^2\times160$	二维卷积1×1	1	960	
$7^2\times960$	池化层7×7			
$1^2\times960$	二维卷积1×1	1	1 280	
$1^2\times1\,280$	全连接层		1 000	

2.5 Deep-Sort目标跟踪算法

Deep-Sort算法[17]在Sort[18]中并入了外观度量信息能够更好地解决遮挡问题。故本文将基于GhostNet与注意力机制的目标检测算法与Deep-Sort相结合,进行行人跟踪。算法结合过程分为以下几个阶段:

(1)目标检测

利用改进后的检测模型对输入视频进行检测,得到目标边框及特征信息。

(2)轨迹处理和状态估计

轨迹处理采用与Sort算法中相同方式,在具有8维状态空间$(u,v,\gamma,h,\boldsymbol{u},\boldsymbol{v},\boldsymbol{\gamma},\boldsymbol{h})$中实施估计。

其中(u,v)代表边界框的中心位置,γ代表纵横比,h为高度。使用卡尔曼滤波器来对运动位置实施估计,即预测结果为$(\boldsymbol{u},\boldsymbol{v},\boldsymbol{\gamma},\boldsymbol{h})$。

(3)数据关联

运动信息关联:第j个检测结果和第i条轨迹之间的运动匹配度计算公式为

$$d^{(1)}(i,j)=(\boldsymbol{d}_j-\boldsymbol{y}_i)^{\mathrm{T}}\boldsymbol{S}_i^{-1}(\boldsymbol{d}_j-\boldsymbol{y}_i) \tag{6}$$

式中:$\boldsymbol{S}_i$为卡尔曼滤波器当前时刻观测空间的协方差矩阵;$\boldsymbol{y}_i$表示当前时刻的预测观测量;$\boldsymbol{d}_j$为第j个检测的状态(u,v,γ,h)。通过逆卡方分布的0.95分位点作为阈值$t^{(1)}$,指标函数定义为

$$b_{i,j}{}^{(1)}=\prod[d^{(1)}(i,j)\leqslant t^{(1)}] \tag{7}$$

外观信息的关联:通过每个边界框d_j,得到$\left\|\boldsymbol{r}_i\right\|=1$的外观状态$r_i$。此外,为每个轨迹$k$建立外观描述符图库$R_K=\left\{r^{(i)}\right\}_{K=1}^{L_K}$,以便存储近100帧已经关联成功的特征向量。其次,计算外观信息中第i个轨迹与第j个检测之间的最小余弦距离,表达式为

$$d^{(2)}(i,j)=\min\left\{1-\boldsymbol{r}_J^{\mathrm{T}}r_k^{(i)}\in\mathbf{R}_i\right\} \tag{8}$$

$$b_{i,j}^{(2)}=\prod[d^{(2)}(i,j)\leqslant t^{(2)}] \tag{9}$$

若距离小于指定阈值时,如式(9),即外观匹配关联成功,将加权总和联结2个指标,表达式为

$$c_{i,j}=\lambda d^{(1)}(i,j)+(1-\lambda)d^{(2)}(i,j) \tag{10}$$

(4)级联匹配

当物体被长时间遮挡时,会出现目标状态预测准确度降低的情景。减小任何检测的标准偏差与投影轨迹均值之间的距离,可能出现轨迹碎片增加与不稳定的轨迹。因此,Deep-Sort引入了级联匹配,出现更加频繁的目标会被分配更大的权重,算法伪代码如下:

输入:行人跟踪$T=\{1,\cdots,N\}$,检测输出边框$D=\{1,\cdots,M\}$,遮挡最大帧数$B_{\max}$

计算外观组合匹配度$\boldsymbol{A}=[\boldsymbol{A}_{i,j}]$

设置目标与行人边框阈值$\boldsymbol{C}=[\boldsymbol{C}_{i,j}]$

for $n\in\{1,\cdots,B_{\max}\}$ do

依据被遮挡的帧数跟踪目标,$T_n\leftarrow\{i\in T\mid B_i=n\}$

$[x_{i,j}]\leftarrow$min_cost_matching(C,T_n,O)

$P\leftarrow P\cup\{(i,j)\mid b_{i,j}\cdot x_{i,j}>0\}$

$O\leftarrow O\setminus\{j\mid\sum b_{i,j}\cdot x_{i,j}>0\}$

End for

Return P,O

3 实验验证及分析

3.1 实验环境

本文算法使用pytorch框架实现,在PASCAL VOC2007和VOC2012数据集上进行训练。训练环境为Inter(R)Core i7-8750HCPU,运行内存16 GB,GPU为NVIDIA GeForce GTX 2080Ti,操作系统ubuntu 16.04。

3.2 损失函数

原始YOLOv3中使用均方误差(Mean square error,MSE)作为损失函数来进行检测框的回归。

MSE对目标的尺度十分敏感,在YOLOv3中通过对目标的长宽开根号的方式降低尺度对回归精度的干扰,但并未彻底解决这个问题。预测框与真实框之间的交并比(Intersection-over-union,IoU)可以反映预测检测框与真实检测框的检测效果,具备尺度上的健壮性,更能体现回归框的质量,其定义为

$$\mathrm{IoU}=\frac{|A\cap B|}{|A\cup B|} \tag{11}$$

式中:A为目标的预测框;B为目标的真实框。

采用IoU作为损失函数时,遇到轴对齐的二维边界框不相交情况,依据IoU计算公式,此时IoU为零,无法进行模型训练。因此Rezatofighi等[21]提出了一种既能维持IoU尺度不变性,还能弥补IoU无法衡量无重叠框之间的距离缺点的评价指标GIoU,其定义为

$$\mathrm{GIoU}=\mathrm{IoU}-\frac{|C\backslash(A\cup B)|}{|C|} \tag{12}$$

式中C为预测框和真实框的最小框面积。

由式(12)可知,GIou引入了包含A、B两个形状的C,所以当A、B不重合时,依然可以进行边界框回归优化,因此本文采用GIou作为损失函数进行回归。

3.3 模型训练

模型参数设置如表2所示。训练阶段采用动量项为0.9的异步随机梯度下降,采用小批量随机梯度下降进行优化,1个样本设置为64张图片,每训练完1个样本进行1次参数更新。总训练周期为180周期。网络学习率衡量网络学习训练样本的速率。权值衰减正则项在训练中防止过拟合。

表2 训练参数设置

Table 2 Training parameter setting

参数名称	参数值
样本数量	64
组数	180
学习率	0.001
动量项	0.9
衰减正则项	0.000 5

3.4 目标检测算法对比实验

对目标的种类检测正确并且检测框中心坐标和检测框维度在一定范围内的检测结果记为正样本(True positive,TP),目标类别识别错误或者检测框不在设定的阀值之内记为负样本(False positive,FP),被错误地划分为负样本的个数记为FN(False negatives),被错误地划分为正样本的个数记为TN(True negatives),具体定义如表3所示。

表3 检测指标定义

Table 3 Definition of test indexes

预测	相关,正类	无关,负类
被检测到	TP	FP
未被检测到	FN	TN

查全率(召回率)R即被检测到目标个数与样本集中所有目标的比值,表达式为

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{13}$$

查准率(精确率)P表示目标检测过程中正确检测的目标与所检测目标个数的比值,表达式为

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{14}$$

F_1为精确率与召回率的调和平均,表达式为

$$F_1=\frac{2PR}{P+R} \tag{15}$$

为了证明目标检测模型的有效性,本文进行了充分的对比实验。改进前后的YOLOv3在VOC2007和VOC2012数据集上的精确率-召回率对比曲线如图6所示,F_1指标对比如图7所示。从图6,7可以看

出,改进后YOLOv3的精确率-召回率曲线与坐标轴之间的面积更大,效果更好,同时F_1指标也优于改进前YOLOv3。表4为目标检测算法对比结果。由表4可看出,加入Ghost模块和SE注意力机制的检测模型mAP提高了19.84%,帧速率达到YOLOv3算法的2.5倍。结果证明了Ghost模块能够减小特征几何变化的影响,同时减少深度网络模型参数和计算量,提取到更丰富、更有效的特征信息。

表4 目标检测算法对比结果

Table 4 Comparison results of target detection algorithms

算法	参数量/MB	mAP/%	速率/(帧•s^{-1})
YOLOv3	237	72.69	20.67
本文算法	95	92.53	51.32

以上实验结果表明,Ghost模块能够依靠线性变换得到丰富的目标特征的多通道特征图,步长为2的Ghost瓶颈层插入了深度可分离卷积层,减小特征几何变化的影响,提高了准确率和召回率,同时减少了深度网络模型参数和计算量;加入SE注意力机制给予重要的行人目标特征更多的关注,减少背景特征的影响,从而提高了精确率。综上证明,本文所提的目标检测算法能够有效提升特征图的利用效率,减少参数,提升检测结果。

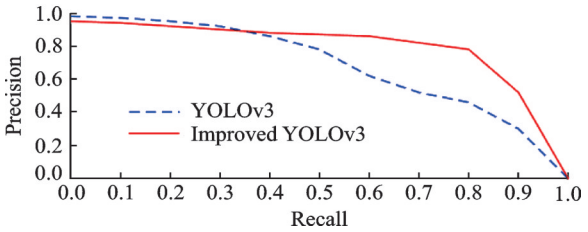


图6 改进前后YOLOv3的精确率-召回率曲线

Fig.6 Precision-Recall curves of YOLOv3 and improved YOLOv3

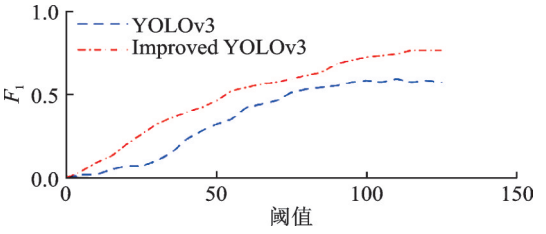


图7 改进前后YOLOv3的F_1曲线

Fig.7 F_1 curves of YOLOv3 and improved YOLOv3

3.5 行人跟踪算法对比试验

3.5.1 数据集对比实验

为了验证本文算法的准确度,分别对DNS[22]、LSSCDL[23]、Joint Learning[24]、Gate S-CNN[25]、SC-FPD[26]和YOLOv3-DeepSort[17]等经典算法在Market1501、CUHK03、PRID、VIPeR和CUHK01五个行人公共数据集进行对比实验。

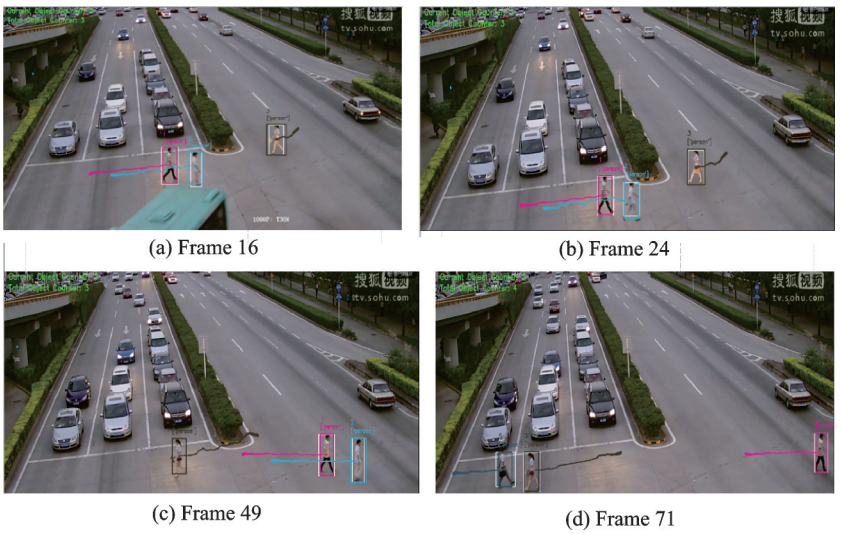
CUHK03数据集含有1 467张行人的14 096张图像,通过10个摄像头拍摄了5个场景,分别包括标准的200个训练和测试切片,100个被随机选择的身份作为测试数据,100个用来评估,其他的1 267张图片作为训练样本。Market1501数据集包含了32 668张检测图像,采集自清华大学里的6个不同的摄像头。PRID2011数据集由2个静态监视摄像机记录获取,摄像机A和B分别有385和749个行人,其中200个行人在2个视角中都会出现。

本实验使用rank-n指标来评估行人跟踪算法的性能,表示在相似度最高的前n张图中有正确结果的概率,最后的结果中对多个查询的rank-n取平均值。评估指标准确度rank-1结果如表5所示。实验结果表明,本文算法rank-1与未改进的YOLOv3-DeepSort相比提高了7.5,与其他行人检测跟踪算法相比效果提升明显。同时,图8中给出了Market1501数据集上的跟踪结果,ID为3的行人目标在与相对行来的两人发生遮挡后仍能被准确跟踪,且ID不变。为了验证本文算法的鲁棒性,给出了在CUHK03数据集上的测试效果,跟踪效果良好,如图9所示。由以上分析可以看出,所提算法可以有效地跟踪多目标行人。

表5 5个数据集上评估指标准确度rank-1结果对比

Table 5 Comparison of rank-1 results of evaluation index standard accuracy on five datasets

算法	Market1501	CUHK03	PRID	VIPeR	CUHK01
DNS[22]	61.0	54.7	40.9	51.2	69.1
LSSCDL[23]	53.4	51.2	38.4	52.7	64.8
Joint Learning[24]	48.6	52.2	36.3	35.8	65.2
Gate S-CNN[25]	65.9	61.8	39.7	37.8	66.1
SC-FPD[26]	66.5	65.9	40.2	51.2	69.3
YOLOv3-Deep-Sort[17]	68.2	70.1	41.9	49.9	68.0
本文算法	75.7	70.6	42.5	53.4	72.2



(a) Frame 16 (b) Frame 24

(c) Frame 49 (d) Frame 71

图8 Market1501数据集行人目标跟踪结果

Fig.8 Pedestrian target tracking results on Market151 dataset

(a) Frame 45 (b) Frame 68

图9 CUHK03数据集行人目标跟踪结果

Fig.9 Pedestrian target tracking results on CUHK03 dataset

3.5.2 GhostNet与SE注意力模块对行人跟踪的影响

为了进一步分析GhostNet与SE注意力模块在行人跟踪任务中的影响,本文分别对缺少某一模块

作用下与所提算法整体作用下的效果进行了对比。实验在TUD-standemitte数据集上进行。TUD-standemitte视频序列有179帧,分辨率为640像素×480像素。GhostNet能够依靠线性变换得到丰富行人目标特征的多通道特征图,步长为2的Ghost瓶颈层插入了深度可分离卷积层,可以减小特征几何变化的影响,以提高跟踪的准确率和召回率,结果如表6所示。加入SE注意力机制,可以给予重要的行人目标特征更多的关注,减少视频背景特征的影响,从而来提高跟踪准确率和精确率,但是由于对关注的局部特征给予了过高的权值,会导致对尺度变化较大和淡出视野的行人跟踪效果不佳,导致召回率的降低,结果如表7所示。

表6 GhostNet对算法的影响

Table 6 Influence of GhostNet on the algorithm %

算法	MOTA	MOTP	Recall	Precision
无GhostNet	60.2	61.4	67.5	82.0
包含GhostNet	73.4	65.7	79.1	99.8

表7 SE注意力模块对算法的影响

Table 7 Influence of SE attention module on the algorithm %

算法	MOTA	MOTP	Recall	Precision
无SE注意力模块	59.6	54.3	83.5	79.6
包含SE注意力模块	73.4	65.7	79.1	99.8

根据表6和表7结果可知,当本文所提算法中缺少GhostNet时,跟踪准确率为60.2%,降低了约17%。加入了GhostNet后,召回率达到了79.1%,说明GhostNet减小了行人特征几何变化的影响,增强了模型的特征表达能力。添加SE注意力模块后,跟踪准确率和精确率分别达到了73.4%及65.7%,说明SE注意力机制可以让提取的目标特征指向性更强,特征利用更充分。召回率降低了5.27%,这是由于对关注的局部行人特征给予了过高的权值,且TUD-standemitte视频序列中目标遮挡严重。综合以上结果可以看出,无论是忽略GhostNet还是SE注意力机制,都会导致跟踪的总体性能下降,而本文所提算法获得的跟踪效果最佳,验证了本文所提模型在行人跟踪任务中的有效性。

3.5.3 复杂场景的实验对比

S2视频序列共795帧,分辨率为768像素×576像素。TUD-standemitte视频序列有179帧,分辨率为640像素×480像素。视频中行人运动轨迹复杂且遮挡严重,很大程度上增加了目标检测跟踪难度。采用如下评价指标评价跟踪算法的性能:

(1)多目标跟踪准确率MOTA表达式为

$$\mathrm{MOTA}=1-\frac{\sum_t(\mathrm{FP}_t+\mathrm{FN}_t+\mathrm{IDS}_t)}{\sum_t \mathrm{GT}_t} \tag{16}$$

式中:IDS为跟踪过程中目标身份转换数;GT为人工标注目标的数量。

(2)多目标跟踪精确率MOTP表达式为

$$\mathrm{MOTP}=\frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \tag{17}$$

式中：d为检测目标i和给它分配的GT之间在所有帧中的平均度量距离；c为在当前帧匹配成功的数目。跟踪结果与对应目标的平均偏差，值越大表明跟踪的轨迹越接近目标实际的运动轨迹。

在公共数据集S2上测试结果如表8所示，可以看出本文算法在准确率与精确率上都超过了Sort与YOLOv3-DeepSort算法。部分测试效果图如图10所示，其中图10(a)为YOLOv3-DeepSort算法效果图，图左上角为该帧在视频中出现的帧序号，分别为第288、293、301、304帧。从图中很明显看到原YOLOv3-DeepSort算法有部分目标出现漏检。改进后的算法相应帧跟踪结果如图10(b)所示，漏检目标能被连续跟踪，修复了轨迹断裂问题，例如图中第18号目标在检测过程中出现漏检情况，导致轨迹不连续，改进后的跟踪算法保证了目标轨迹的连续性。

表8 S2序列跟踪结果对比

Table 8 Comparison of tracking results on S2 sequence %

算法	MOTA	MOTP	Recall	Precision
RMOT[27]	67.4	70.9	72.6	80.3
SORT[18]	80.3	64.2	97.1	94.4
MDP[28]	87.9	64.5	98.6	90.8
YOLOv3-DeepSort[17]	88.4	56.8	91.1	97.9
本文算法	90.2	66.7	91.7	98.5

(a) Tracking results by YOLOv3-DeepSort algoritnm

(b) Tracking results by the proposed algorithm

图10 S2数据集跟踪结果对比

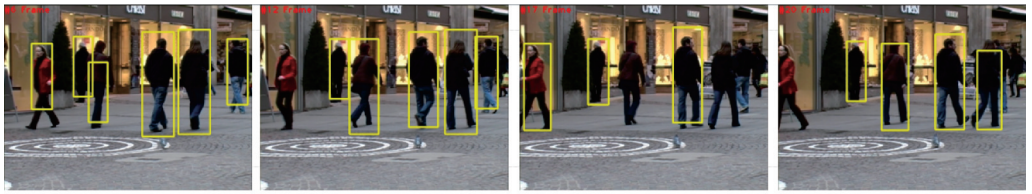
Fig.10 Comparison of tracking results on S2 data set

在TUD-standemitte上测试结果如表9所示。本文算法MOTA最佳，Recall相对较低是由于视频中目标尺度相差较大且目标遮挡严重。原YOLOv3-DeepSort跟踪效果如图11(a)所示，分别为视频帧第6、12、17、20帧。在跟踪过程中，部分尺度较小、遮挡严重的目标难以被检测出。本文算法跟踪后对应视频帧检测跟踪结果如图11(b)所示，可以看出遮挡后的目标在后续帧又被跟踪成功，部分尺度较小和淡出视野的行人也能持续跟踪，具有较强的鲁棒性。

表9 TUD-standemitte序列跟踪结果对比

Table 9 Comparison of tracking results on TUD-standemitte sequence %

算法	MOTA	MOTP	Recall	Precision
RMOT[27]	53.5	56.5	74.7	78.7
SORT[18]	60.2	72.4	75.1	82.0
MDP[28]	69.7	53.4	75.6	94.3
YOLOv3-DeepSort[17]	72.4	52.6	83.7	98.6
本文算法	73.4	65.7	79.1	99.8



(a) Tracking results by YOLOv3-DeepSort algoritnm

(b) Tracking results by the proposed algorithm

图11 TUD-standemitte数据集跟踪结果对比

Fig.11 Comparison of tracking results on TUD-standemitte data set

4 结束语

本文针对复杂场景下行人跟踪准确度低且速度慢的问题,提出了基于GhostNet与注意力机制的行人检测跟踪算法。将YOLOv3的主干网络替换为GhostNet,减少了深度网络模型参数和计算量,提高了检测速度。加入SE注意力机制给予不同的特征不同的权值,引入了目标检测的直接评价指标GIoU来指导回归任务,提高了检测跟踪的准确度。将基于GhostNet的目标检测算法与Deep-Sort相结合进行行人检测与跟踪。实验结果表明,改进算法可以区分大规模目标行人,能够有效地处理复杂场景下行人跟踪时的遮挡问题,提高了跟踪的准确度,并具有较强的鲁棒性。

参考文献:

[1] 韦皓瀚, 曹国, 尚岩峰,等. 一种改进聚合通道特征的行人检测方法[J]. 数据采集与处理, 2018, 33(3): 521-529.
WEI Haohan, CAO Guo, SHANG Yanfeng, et al. Improved pedestrian detection method of modified aggregate channel feature[J]. Journal of Data Acquisition and Processing, 2018, 33(3): 521-529.

[2] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 21-37.

[3] REN Shaoqing, HE Kaiming, 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

IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE Computer Society, 2016: 779-788.

[5] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE Computer Society, 2017: 6517-6525.

[6] REDMON J, FARHADI A. YOLOv3: An incremental improvement [EB/OL]. (2018-04-08) [2021-01-12]. .https://arxiv.org/abs/1804.02767.

[7] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE Computer Society, 2017: 1251-1258.

[8] HOWARD A G, ZHU Menglong, CHEN Bo, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2021-01-12]. http://arXiv/abs/1704.04861.

[9] SANDLER M, HOWARD A G, ZHU Menglong, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Computer Society, 2018: 4510-4520.

[10] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE Computer Society, 2019: 1314-1324.

[11] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Computer Society, 2018: 6848-6856.

[12] MA Ningning, ZHANG Xiangyu, ZHENG Haitao, et al. Shufflenetv2: Practical guidelines for efficient CNN architecture design[C]//Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018: 122-138.

[13] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: More features from cheap operations[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE Computer Society, 2020: 1580-1589.

[14] ZHANG Shun, WANG Jinjun, WANG Zelun, et al. Multi-target tracking by learning local-to-global trajectory models[J]. Pattern Recognition, 2015, 48(2): 580-590.

[15] MAHMOUDI N, AHADI S M, RAHMATI M, et al. Multi-target tracking using CNN-based features: CNNMTT[J]. Multimedia Tools and Applications, 2019, 78(6): 7077-7096.

[16] XIANG Jun, ZHANG Guoshuai, HOU Jianhua. Online multi-object tracking based on feature representation and Bayesian filtering within a deep learning architecture[J]. IEEE Access, 2019, 7: 27923-27935.

[17] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//Proceedings of IEEE International Conference on Image Processing. Piscataway, NJ, USA: IEEE Press, 2017: 3645-3649.

[18] BEWLEY A, GE Zongyuan, OTT L, et al. Simple online and realtime tracking[C]//Proceedings of IEEE International Conference on Image Processing. Piscataway, NJ, USA: IEEE Press, 2016: 3464-3468.

[19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE Computer Society, 2016: 770-778.

[20] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(8): 2011-2023.

[21] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE Computer Society, 2019: 658-666.

[22] ZHANG L, XIANG T, GONG S G. Learning a discriminative null space for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 1239-1248.

[23] ZHANG Y, LI B H, LU H C, et al. Sample-specific SVM learning for person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE Computer Society, 2016: 1278-1287.

[24] HAN C, YE J, ZHONG Y, et al. RE-ID driven localization refinement for person search[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE Computer Society, 2019: 9814-9823.

[25] TAKIKAWA T, ACUNA D, JAMPANI V, et al. Gated-SCNN: Gated shape CNNS for semantic segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE Computer Society, 2019: 5229-5238.

[26] COSTEA A D, NEDEVSCHI S. Semantic channels for fast pedestrian Detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE Computer Society, 2016: 2360-2368.

[27] JU H Y, YANG M H, LIM J, et al. Bayesian multi-object tracking using motion context from multiple objects[C]//Proceedings of Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE Computer Society, 2015: 33-40.

[28] XIANG Y, ALAHI A, SAVARESE S. Learning to track: Online multi-object tracking by decision making[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE Computer Society, 2015: 4705-4713.

作者简介:

王立辉(1993-),男,硕士研究生,研究方向:深度学习、目标检测与跟踪,E-mail:437405253@qq.com。

杨贤昭(1978-),**通信作者**,男,副教授,研究方向:智能控制、信号处理等,E-mail: yangxianzhao@wust.edu.cn。

刘惠康(1963-),男,教授,研究方向:自动控制、新型电气传动、故障诊断技术等。

黄晶晶(1996-),男,硕士研究生,研究方向:智能控制、机器视觉。

(编辑:张黄群)