

控制与决策

Control and Decision

基于MobileNet的多目标跟踪深度学习算法

薛俊韬, 马若寒, 胡超芳

引用本文:

薛俊韬, 马若寒, 胡超芳. 基于MobileNet的多目标跟踪深度学习算法[J]. *控制与决策*, 2021, 36(8): 1991–1996.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1424>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度时序特征迁移的轴承剩余寿命预测方法

Remaining useful life prediction of bearing based on deep temporal feature transfer

控制与决策. 2021, 36(7): 1699–1706 <https://doi.org/10.13195/j.kzyjc.2019.1809>

输入约束不确定系统的点对点迭代学习控制与优化

Point-to-point iterative learning control and optimization for uncertain systems with constrained input

控制与决策. 2021, 36(6): 1435–1441 <https://doi.org/10.13195/j.kzyjc.2019.0908>

基于条件对抗生成孪生网络的目标跟踪

Conditional generative adversarial siamese networks for object tracking

控制与决策. 2021, 36(5): 1110–1118 <https://doi.org/10.13195/j.kzyjc.2019.1215>

尺度自适应的多特征融合相关滤波目标跟踪算法

Scale adaptation and multi-feature fusion correlation filtering object tracking algorithm

控制与决策. 2021, 36(2): 429–435 <https://doi.org/10.13195/j.kzyjc.2019.0445>

基于稀疏度阶数优化的杂波密度估计算法

A clutter density estimation algorithm by optimized sparsity order

控制与决策. 2020, 35(12): 2923–2930 <https://doi.org/10.13195/j.kzyjc.2019.0429>

基于MobileNet的多目标跟踪深度学习算法

薛俊韬¹, 马若寒¹, 胡超芳^{1,2†}

(1. 天津大学 电气自动化与信息工程学院, 天津 300072;
2. 天津大学 微光机电系统技术教育部重点实验室, 天津 300072)

摘要: 针对深度学习算法在多目标跟踪中的实时性问题, 提出一种基于 MobileNet 的多目标跟踪算法. 借助于 MobileNet 深度可分离卷积能够对深度网络模型进行压缩的原理, 将 YOLOv3 主干网络替换为 MobileNet. 通过将标准卷积分解为深度卷积和逐点卷积, 保留多尺度预测部分, 以有效减少参数量. 对于检测得到的边框信息, 利用 Deep-SORT 算法进行跟踪. 实验结果表明, 所提出方法在跟踪效果基本不变的情况下可提升处理速度近 50%.

关键词: 深度学习; 多目标跟踪; 目标检测; YOLOv3; deep-SORT; MobileNet

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1424

开放科学(资源服务)标识码(OSID):



引用格式: 薛俊韬, 马若寒, 胡超芳. 基于 MobileNet 的多目标跟踪深度学习算法[J]. 控制与决策, 2021, 36(8): 1991-1996.

Deep learning algorithm based on MobileNet for multi-target tracking

XUE Jun-tao¹, MA Ruo-han¹, HU Chao-fang^{1,2†}

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; 2. Key Laboratory of Micro Opto-electro Mechanical System Technology of Ministry of Education, Tianjin University, Tianjin 300072, China)

Abstract: For the real-time problem of deep learning algorithm in multi-target tracking, a multi-target tracking algorithm based on MobileNet is proposed. Using the principle that deepwise separable convolution of MobileNet can compress deep network models, the YOLOv3 backbone network is replaced by MobileNet, but multiscale predictions are remained, which reduces effectively the amount of parameters by factorizing a standard convolution into a depthwise convolution and a pointwise convolution. For the detected bounding box information, the tracking task is performed using the Deep-SORT algorithm. The experimental results show that the processing speed is improved by nearly 50% under the condition of maintaining the tracking result.

Keywords: deep learning; multi-target tracking; object detection; YOLOv3; deep-SORT; MobileNet

0 引言

多目标跟踪是计算机视觉领域的研究热点, 可应用于交通监测、安防等多个领域, 具有一定的应用价值和挑战性^[1]. 检测方式可以分为检测跟踪和无检测跟踪两类, 前者需要检测目标后再进行跟踪; 后者需要在第 1 帧手动初始化目标, 然后进行跟踪. 在目标跟踪中, 涌现出许多具有良好性能的算法, 如 SSD^[2]、R-CNN^[3-4] 以及 YOLO 系列^[5-7], 其中 YOLOv3^[7] 算法在检测跟踪中体现出较强的优势.

由于深度学习的发展, 卷积神经网络模型逐渐替代了传统手工设计的特征, 提供了一种端到端的处理

方法, 精度也大幅提高. 但 CNN 模型在不断提高精度的同时, 其网络深度和尺寸也在成倍增长, 需要 GPU 来进行加速, 使得基于深度学习的跟踪算法无法直接应用于移动设备, 导致难以符合实时性要求. 因此降低算法复杂度、提高实时性、简化和加速模型便成为亟待解决的问题. 文献[8-10]使用剪枝方法对神经网络进行网络压缩. 文献[11]提出从零开始训练低秩约束卷积神经网络模型的方法, 不仅速度得到提升, 而且在一些情况下模型性能也有所提高. 目前, 深度网络模型压缩方法分为两个方向: 一是对已经训练好的深度网络模型进行压缩得到小型化模型; 二

收稿日期: 2019-10-10; 修回日期: 2020-03-13.

基金项目: 国家自然科学基金项目(61773279); 天津市科技计划项目(19YFHBQY00040); 微光机电系统技术教育部重点实验室(天津大学)开放基金项目(MOMST2016-4).

责任编辑: 陈家伟.

†通讯作者. E-mail: cfhu@tju.edu.cn.

是直接设计小型化模型进行训练,如 SqueezeNet^[12]、ShuffleNet^[13]、MobileNet^[14]等。

在多目标跟踪方面,Zhang等^[15]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^[16]使用卷积神经网络代替手工标注进行特征提取,以改善算法精度,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^[17]设计了一个卷积神经网络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;此外,还设计了基于递归神经网络的贝叶斯过滤模块,并将LSTM网络的隐藏状态作为输入,执行递归预测和更新目标状态。而Deep-SORT^[18]多目标跟踪算法则在SORT^[19]算法的基础上,提取深度表观特征,使跟踪效果有了明显的提升。

本文针对算法的实时性问题,结合深度学习目标检测、深度网络模型压缩以及多目标跟踪算法,提出基于MobileNet的多目标跟踪算法,在保证精度的前提下,有效改善深度网络模型庞大以及计算复杂的问题,提高了算法的执行速度。

1 YOLOv3目标检测算法

YOLOv3算法的基本思想是:将输入图像分割为 $S \times S$ 个单元格,每个单元格用于检测中心点落在该网格内的目标,并预测 B 个边界框和置信度。边界框用 (x, y, w, h, c) 5个参数表达,其中 (x, y) 为目标中心相对于单元格左上角的相对坐标,而 w 和 h 则分别是目标与整张图像的宽和高之比。单元格还预测 C 个类别的各类概率值。因此,每个单元格共预测 $B \times (5 + C)$ 个值。YOLOv3借鉴ResNet的理念,采用YOLOv2的Darknet-19,并创建一个新的特征提取网络Darknet-53^[20]。YOLOv3采用类似FPN的上采样和特征融合机制,使用不同的特征尺度进行预测。考虑到其对小目标检测的良好效果,本文采用该算法进行目标检测。

2 基于MobileNet的多目标跟踪算法

利用MobileNet将标准卷积分解为深度卷积和逐点卷积以减少参数数量的特点,替换YOLOv3目标检测算法的主干网络框架,并保留多尺度预测,形成基于MobileNet的目标检测算法,再进一步结合Deep-SORT算法进行多目标跟踪。

2.1 MobileNet算法

MobileNet是Google为移动端和嵌入式设备提出的高效模型。使用深度可分离卷积构建深度神经

网络,并通过两个超参数,从通道数量和特征图大小两个方面减少计算量。分解过程如图1所示。

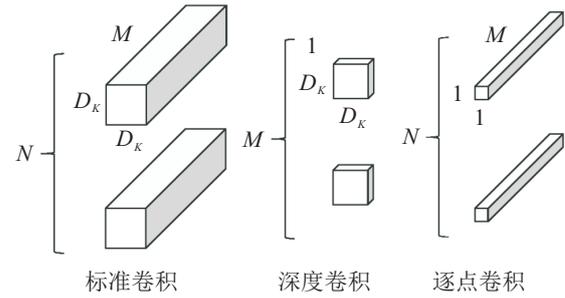


图1 深度可分离卷积分解过程

假设输入的特征映射 F 尺寸为 (D_F, D_F, M) ,采用的标准卷积 K 为 (D_K, D_K, M, N) ,则输出的映射 G 尺寸为 (D_F, D_F, N) 。 M 为输入通道数, N 为输出通道数,对应的计算量为

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F. \quad (1)$$

将标准卷积分解为深度卷积和逐点卷积,深度卷积起滤波作用,尺寸大小为 $(D_K, D_K, 1, M)$,输出特征映射尺寸为 (D_F, D_F, M) ;逐点卷积用于通道转换,尺寸大小为 $(1, 1, M, N)$,输出映射尺寸为 (D_F, D_F, N) 。两者对应的计算量为

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F. \quad (2)$$

相比于标准卷积,深度可分离卷积计算量减少,即

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}. \quad (3)$$

此外,MobileNet还引入了宽度乘子 α 和分辨率乘子 ρ 两个超参数,分别用于改变通道数和特征图分辨率,便于控制模型大小,降低参数量。

2.2 基于MobileNet的目标检测模型

MobileNet网络基于深度可分离卷积,除全连接层外,所有层后跟有BatchNorm和ReLU,最后使用softmax进行分类。MobileNet将深度卷积和逐点卷积计为单独的层,共有28层。

针对实时性问题,本文结合卷积神经网络模型压缩方法,选用MobileNet与YOLOv3检测模型相结合,用前者代替后者的网络框架,同时保留YOLOv3的多尺度预测,最终得到轻量级的检测模型。

YOLOv3中包括 13×13 、 26×26 和 52×52 三个尺度,在MobileNet网络中找到对应 $13 \times 13 \times 1024$ 、 $26 \times 26 \times 512$ 和 $52 \times 52 \times 256$ 的部分,使用YOLOv3多尺度预测的方法进行融合,替换后的网络结构如图2所示。

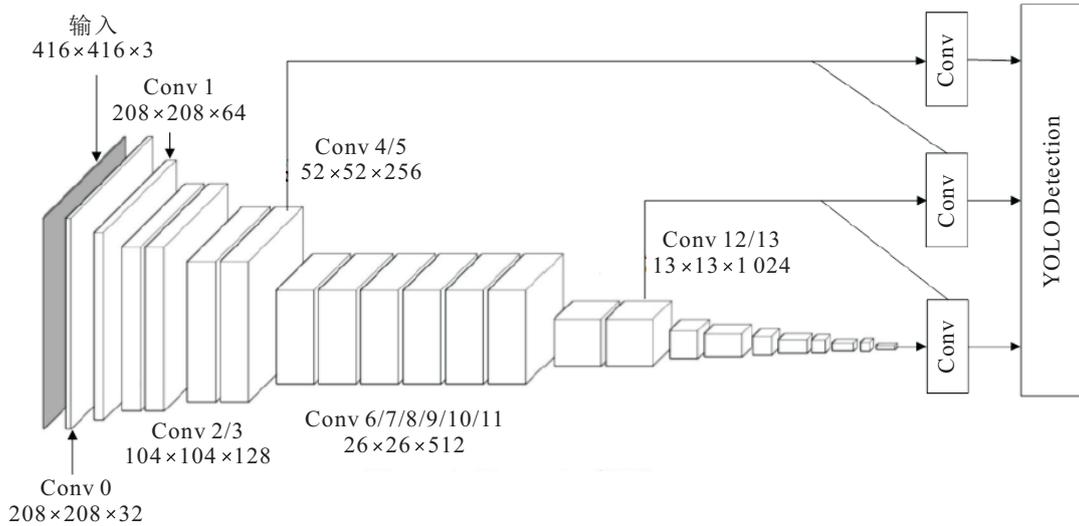


图2 YOLOv3-MobileNet框架

2.3 多目标跟踪

由于Deep-SORT算法加入了深度表观特征,并具有较高的准确度和实时性,本文将基于MobileNet的目标检测算法与Deep-SORT相结合进行多目标跟踪,主要分为以下部分:

1) 目标检测. 首先对输入的视频进行目标检测,得到目标的边框及特征信息,并根据置信度和非极大值抑制进行边框过滤.

2) 轨迹处理和状态估计. 运动状态估计中使用8个参数($u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h}$)进行运动状态的描述,其中(u, v)为边框的中心坐标, γ 为长宽比, h 为高度,这4个参数来源于目标检测部分,其余4个参数表示对应图像坐标系中的速度信息.使用卡尔曼滤波器对运动状态进行预测.

3) 对跟踪器参数和特征集进行更新,判断有无目标消失或者有无新目标出现. 对每个目标,记录其上次检测结果和跟踪结果匹配后的帧数 a_k ,只要检测结果与跟踪结果正确关联,就将该参数置为0. 如果 a_k 超过了设置的最大阈值 A_{max} ,则结束对该目标的跟踪.

4) 检测结果与跟踪预测结果匹配. 区分已确认和未确认状态的跟踪器,对已确认状态的跟踪器进行匹配指派. 其中指派问题使用了匈牙利算法,并同时考虑运动信息的关联和目标外观信息的关联.

运动信息关联: 使用卡尔曼滤波器预测状态和新测量之间的马氏距离,以此表达运动信息,有

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i). \quad (4)$$

式(4)表示第 j 个检测结果与第 i 条轨迹之间的运动匹配度. 其中: S_i 为卡尔曼滤波器当前时刻观测空间的协方差矩阵, y_i 为当前时刻的预测观测测量, d_j 为第

个检测的状态(u, v, γ, h). 马氏距离通过测量远离平均轨道位置的标准偏差考虑状态估计的不确定性,通过逆卡方分布的0.95分位点作为阈值 $t^{(1)}$,指标函数定义如下:

$$b_{i,j}^{(1)} = 1[d^{(1)}(i, j) \leq t^{(1)}]. \quad (5)$$

目标外观信息关联: 由于相机运动会使马氏距离度量方法失效,引入第2种关联方法,对每个跟踪目标构建一个库,存储每个跟踪目标成功关联的最近100帧特征向量,则第 i 个跟踪器与当前帧第 j 个检测

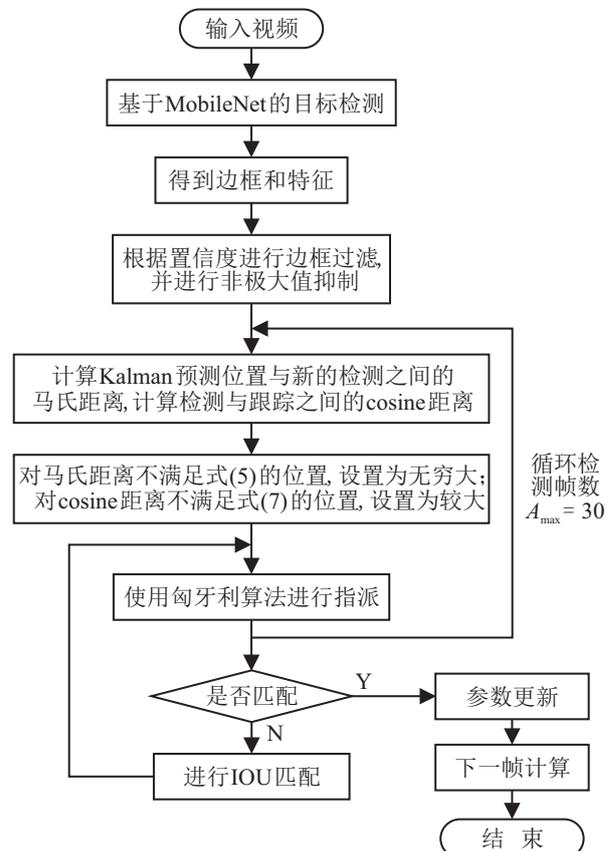


图3 多目标跟踪算法流程

结果之间的表观匹配度为

$$d^{(2)}(i, j) = \min\{1 - r_i^T r_k^{(i)} \| r_k^{(i)} \leq \mathfrak{R}_i\}. \quad (6)$$

指标函数表示为

$$b_{i,j}^{(2)} = 1[d^{(2)}(i, j) \leq t^{(2)}]. \quad (7)$$

采用上述两种度量的线性加权作为最终度量,有

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j). \quad (8)$$

只有当 $c_{i,j}$ 位于两种度量阈值的交集内时,才认为实现了正确的关联. 当指派完成后,分类出未匹配的检测和跟踪器.

5) 对未确认状态的跟踪器、未匹配的跟踪器和未匹配的检测进行 IOU 匹配,再次使用匈牙利算法进行指派.

6) 对于匹配的跟踪器进行参数更新,删除再次未匹配的跟踪器,未匹配的检测初始化为新目标.

算法整体流程如图3所示.

3 实验结果

本文算法使用 keras 实现,并利用 PASCAL VOC 2007 和 VOC 2012 数据集进行训练. 作为视觉分类检测的基准测试, PASCAL VOC 数据集包含 20 类物体,如人、动物、交通工具、家具等,每张图片均有标注. 本文使用 VOC 2007 trainval 数据集和 VOC 2012 trainval 数据集,共 16 551 张图片进行训练,训练环境为 Google Colab 云平台. 所使用的操作系统为 Ubuntu 18.04.1, GPU 为 Tesla K80. 其他测试均在 Windows 8.1 下进行, CPU 为 Intel(R) Core(TM) i5-5200U, 无 GPU.

在目标检测评价中,常用的评价指标为 mAP (mean average precision),表示对目标类的 AP 值取平均,本文采用 mAP、单张图片检测时间及模型大小等指标对目标检测模型进行评估. 对于多目标跟踪算法,由于很难使用单一评分评估多目标跟踪性能,采用 MOT Challenge 评价标准.

3.1 目标检测算法对比实验

表 1 为 YOLOv3 目标检测算法、基于 MobileNet 的目标检测算法和其他检测算法实验对比结果.

表 1 目标检测算法性能对比结果

算 法	mAP/%	模型大小/MB	单张检测时间/s
YOLOv3	75.71	237	6.46
SSD	61.76	201	4.10
Fater R-CNN	73.2	109	—
YOLOv3-MobileNet(本文)	63.28	93	3.14

本文选取比较流行的单阶段检测算法和两阶段检测算法进行对比实验. 由表 1 可见,基于 MobileNet 的目标检测算法模型是 YOLOv3 算法模型的 1/3,并且相比 SSD 和 Fater R-CNN,模型最小;单张图片检测时间相比 YOLOv3 模型提高了 2 倍;在精度方面,本文检测算法相比 YOLOv3 模型下降约 12.4%,但相比 SSD 提高 1.5%. 总体而言,虽然本文检测算法在检测精度上有所降低,但可以满足检测跟踪的实际需求,并且在模型大小和单张检测时间上具有优势. 精度降低的原因是由于 YOLOv3 的网络结构中使用了残差网络,并且网络层数更深,而 MobileNet 网络结构仅使用深度可分离卷积,没有使用残差结构,在网络深度方面也比 YOLOv3 更浅,因此造成本文检测算法在 mAP 值上与 YOLOv3 差距较大. 虽然本文检测算法在一定程度上牺牲了精度,但实现了模型的小型化和速度上的提升.

3.2 多目标跟踪算法对比实验

在 MOT 基准数据库提供的序列集上评估跟踪性能,以 MOT Challenge 标准对多目标跟踪算法进行评价,具体评价指标如表 2 所示,各指标含义如下:

MOTA (\uparrow): 多目标跟踪准确度;

MOTP (\uparrow): 多目标跟踪精度;

MT (\uparrow): 目标跟踪轨迹占真实长度 80% 以上的轨迹数目;

表 2 评价指标对比

算 法	MOTA \uparrow	MOTP \uparrow	MT \uparrow	ML \downarrow	IDs \downarrow	FP \downarrow	FN \downarrow	FPS \uparrow
TC_ODAL ^[21]	15.1	70.5	3.2	55.8	637	13 171	38 538	—
RMOT ^[22]	18.6	69.6	5.3	53.3	684	12 473	36 835	—
SORT ^[19]	22.0	76.7	16.2	41.4	1 375	19 298	65 446	—
MDP ^[23]	30.1	71.6	10.4	41.3	690	8 789	33 479	—
YOLOv3-Deep-SORT	30.7	78.0	8.7	53.6	363	5 871	70 256	0.15
MOTDT ^[24]	36.3	75.9	12.6	42.1	192	1 915	30 072	—
YOLOv3-MobileNet-Deep-SORT(本文)	23.5	77.8	9.5	52.2	437	14 396	69 595	0.55

ML (↓): 目标跟踪丢失轨迹占真实长度至多 20% 的轨迹数目;

IDs (↓): 目标 ID 发生变化的次数;

FP (↓): 误报总数;

FN (↓): 未命中目标的总数;

FPS (↑): 帧频率.

对于带有 (↑) 的评价指标, 分数越高表示效果越好; 对于带有 (↓) 的评价指标, 结果相反.

由表 2 可见: 本文算法的 MOTP 值与 YOLOv3-Deep-SORT 算法相差不大, 但均高于表中其他算法; 被跟踪的轨迹比例 (MT) 相比 YOLOv3-Deep-SORT 算法有所上升, 跟丢轨迹占比 (ML) 有所下降, 均优于 TC_ODAL 和 RMOT 算法; ID 变换次数低于表中前 4 个算法 (即 TC_ODAL、RMOT、SORT、MDP), 但相比 YOLOv3-Deep-SORT 算法有所上升; 帧频率提高为该算法的 3 倍, 检测跟踪时间提高显著. 本文算法虽然在跟踪准确度、ID 变换次数和误报总数上不及 YOLOv3-Deep-SORT 算法, 但其他跟踪指标均优于该算法, 并且本文算法相比于 TC_ODAL、RMOT 和 SORT 等算法优势明显. 为进一步表现本文算法的优越性, 以 ID 标号为 3 的目标为例, 给出如图 4 所示的部分跟踪结果, 且跟踪框上带有数字 ID 标识. 由图 4 可见, 该算法在第 200 帧、260 帧、320 帧均能连续跟踪, 跟踪效果良好.



(a) 第 200 帧



(b) 第 260 帧



(c) 第 320 帧

图 4 跟踪结果

3.3 无人机人流监控实验

以某城市交通路口人流监控为例, 将本文算法应用于无人机, 利用其高空独特视角, 对十字路口行人进行统计, 实时监测人流密度以及人流密度变化情况, 以配合地面警力巡查, 对各类警情及交通情况进行预警, 辅助警力调度, 实现对城市环境下的安全监控. 此类监管系统的关键是要具有良好的实时性, 因此降低网络参数量, 提高实时性显得十分重要.

本文以无人机数据集进行实验, 选取第 76、133 和 182 帧, 如图 5 所示. 由图 5 可见, 本文算法可有效检测和跟踪大部分行人, 且实时性与之之前检测跟踪实验结果具有相同数量级, 但对于相互遮挡和全身信息不全的行人检测跟踪效果相对较差.



(a) 第 76 帧



(b) 第 133 帧



(c) 第 182 帧

图 5 无人机数据集跟踪结果

4 结论

本文在 YOLOv3 和 Deep-SORT 算法的基础上, 针对实时性问题, 使用轻量级网络 MobileNet 代替原有网络结构对检测模型进行压缩. 实验结果表明, 尽管改进算法由于更浅的网络结构和未使用残差网络等因素导致精度下降, 但实时性提高显著, 实现了精度和速度间的折衷, 达到了快速有效跟踪的目的.

参考文献 (References)

[1] Luo W, Xing J, Milan A, et al. Multiple object tracking: A literature review[J]. 2014, arXiv: 1409.07618.

[2] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multi

- box detector[C]. Proceedings of European Conference on Computer Vision (ECCV). New York: Springer Press, 2016: 21-37.
- [3] Girshick R. Fast R-CNN[C]. Proceedings of International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society Press, 2015: 1440-1448.
- [4] 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.
- [5] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of Computer Vision and Pattern Recognition (CVPR). Washington DC: IEEE Computer Society Press, 2015: 779-788.
- [6] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]. Proceedings of Computer Vision and Pattern Recognition (CVPR). Washington DC: IEEE Computer Society Press, 2017: 6517-6525.
- [7] Redmon J, Farhadi A. YOLOv3: An incremental improvement[C]. Proceedings of Computer Vision and Pattern Recognition (CVPR). Los Alamitos: IEEE Computer Society Press, 2018: 1-6.
- [8] Lei Wu, Han Yue, Peng Chen, et al. A novel dynamic network pruning via smooth initialization and its potential applications in machine learning based security solutions[J]. IEEE Access, 2019, 7: 91667-91678.
- [9] 王生生, 王萌, 王光耀. 基于深度神经网络剪枝的两阶段遥感图像目标检测[J]. 东北大学学报: 自然科学版, 2019, 40(2): 174-179.
(Wang S S, Wang M, Wang G Y. Deep neural network pruning based two-stage remote sensing image object detection[J]. Journal of Northeastern University: Natural Science, 2019, 40(2): 174-179.)
- [10] Adam Polyak, Lior Wolf. Channel-level acceleration of deep face representations[J]. IEEE Access, 2015, 3: 2163-2175.
- [11] Zhang Xiangyu, Zou Jianhua, He Kaiming, et al. Accelerating very deep convolutional networks for classification and detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 1943-1955.
- [12] Iandola F N, Moskewicz M W, Ashraf K, et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 1mb model size[J]. 2016, arXiv: 1602.07360.
- [13] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[J]. 2017, arXiv: 1707.01083.
- [14] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. 2017, arXiv: 1704.04861.
- [15] 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.
- [16] Mahmoudi N, Ahadi S M, Rahmati M. Multi-target tracking using CNN-based features: CNNMTT[J]. Multimedia Tools and Applications, 2019, 78(6): 7077-7096.
- [17] 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.
- [18] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]. Proceedings of International Conference on Image Processing (ICIP). Piscataway: IEEE, 2017: 3645-3649.
- [19] Bewley A, Zongyuan G, Ramos F, et al. Simple online and realtime tracking[C]. Proceedings of International Conference on Image Processing (ICIP). Piscataway: IEEE, 2016: 3464-3468.
- [20] Wang Hai, Yu Yijie, Cai Yingfeng, et al. A comparative study of state-of-the-art deep learning algorithms for vehicle detection[J]. IEEE Intelligent Transportation Systems Magazine, 2019, 11(2): 82-95.
- [21] Bae S H, Yoon K J. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning[C]. Proceedings of Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE, 2014: 1218-1225.
- [22] Yoon J H, Yang M H, Lim J, et al. Bayesian multi-object tracking using motion context from multiple objects[C]. Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2015: 33-40.
- [23] Xiang Y, Alahi A, Savarese S. Learning to track: Online multi-object tracking by decision making[C]. Proceedings of International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 4705-4713.
- [24] Chen Long, Ai Haizhou, Zhuang Zijie, et al. Real-time multiple people tracking with deeply learned candidate selection and person re-identification[J]. 2018, arXiv: 1809.04427.

作者简介

薛俊韬(1973—), 女, 副教授, 从事图像处理、语音识别、智能交通、检测技术与智能装置等研究, E-mail: xuejt@tju.edu.cn;

马若寒(1996—), 女, 硕士生, 从事模式识别与智能系统、目标检测与跟踪等研究, E-mail: maruohan96@163.com;

胡超芳(1973—), 男, 副教授, 从事无人机协同智能优化与控制、目标追踪与探测、无人驾驶汽车智能决策与控制、预测控制等研究, E-mail: cfhu@tju.edu.cn.

(责任编辑: 郑晓蕾)