计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(12):93−99 [doi: 10.15888/j.cnki.csa.007691]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

多任务学习的车辆结构化信息提取方法①

朱　红[1], 岑跃峰[2], 王思泰[1]

[1](中国电信股份有限公司 浙江分公司 政企客户事业部, 杭州 310001)
[2](浙江科技学院 信息与电子工程学院, 杭州 310000)
通讯作者: 王思泰, E-mail: 15305715006@189.cn

摘　要: 目前, 大部分的车辆结构化信息需要通过多个步骤进行提取, 存在模型训练繁琐、各步骤模型训练数据有限和过程误差累加等问题. 为此, 采用多任务学习将车辆结构化信息提取整合在统一的神经网络之中, 通过共享特征提取结构, 减少过程误差累加, 并构建了一个多任务损失函数用于端到端训练神经网络; 针对训练样本有限的问题, 提出了新的数据整合和增广方法. 在 KITTI 数据集上实验结果表明, VSENet 可以达到 93.82% 的 mAP(均值平均精度), 且能达到实时的处理速度; 与多阶段的车辆结构化特征提取方法对比, 平均运行时间缩减了 60%, 其精度能达到相似或者更好的效果; 实验结果表明, 该方法具有一定的先进性和有效性.

关键词: 多任务学习; 结构化信息; 卷积神经网络; 智慧交通系统

引用格式: 朱红,岑跃峰,王思泰.多任务学习的车辆结构化信息提取方法.计算机系统应用,2020,29(12):93–99. http://www.c-s-a.org.cn/1003-3254/7691.html

Vehicle Structure Information Extraction Based on Multi-Task Learning

ZHU Hong[1], CEN Yue-Feng[2], WANG Si-Tai[1]

[1](Department of Government and Enterprise Customer, Zhejiang Branch, China Telecom Corporation Limited, Hangzhou 310000, China)
[2](School of Information and Electronic Engineering, Zhejiang University of Science & Technology, Hangzhou 310000, China)

Abstract: Currently, most of vehicle structured information was obtained through multiple steps, which caused the problems such as fussy training, limited training data in each step, and the accumulation of error in processing. Therefore, multi-task learning was applied to union the structured information extraction in a single neural network, and shared feature extraction structure can help to reduce the error accumulation in various processing. A loss function of multi-task learning was put forward for end-to-end network training. For solving the training data limitation problem, a new dataset augmentation and combination approach was advanced. The experimental results on the dataset KITTI show that the mAP (mean of the Average Precision) of VSENet achieves 93.82%, and the processing speed can satisfy the real-time request. Compared with multi-step vehicle structure information extraction method, the proposed approach reduces 60% of average processing time and achieves similar or better performance. These results show that the proposed method has certain advancement and is effective.

Key words: multi-task learning; structured information; convolutional neural network; intelligent transportation system

随着城市化建设的不断推进, 针对交通视频的对象行为分析和事件检测日益受到关注[1]. 而交通结构化信息提取技术[2] 作为视频分析和交通控制的基础方法, 其识别精确度决定了视频分析的准确率. 由于人工智

① 基金项目: 浙江省教育厅一般科研项目 (Y201839557)
Foundation item: General Scientific Research Program of Education Bureau, Zhejiang Province (Y201839557)
收稿时间: 2020-04-20; 修改时间: 2020-05-15; 采用时间: 2020-05-26; csa 在线出版时间: 2020-11-30

能在近年来的快速发展, 使得目标检测、目标识别和目标跟踪等图像处理技术在交通领域也得到了广泛的应用. 但基于深度学习的人工智能技术存在标注样本需求大, 系统算力要求较高等问题, 仍对该技术的推广造成了较大的影响. 因此如何高效利用深度学习仍是当前业界关注的热点之一.

目前, 大部分的方法需要将车辆结构化信息提取分为多个子任务进行处理, 包括: 运动检测[3–6]、目标检测[7–12]、目标跟踪[13–19]、车辆颜色识别[20–25] 和车型识别[26–28] 等, 且各个子任务都有较好的解决方案. 因此为了保证系统的准确率, 需要利用多个卷积神经网络进行结构化信息进行提取, 而多个神经网络的训练较为繁琐且各个子任务的样本准备较为困难, 训练数据有限会直接影响网络特征的有效拟合, 导致整体的检测精度下降. 而且多个神经网络的使用也造成了重复的特征提取, 严重影响了整体的系统的运行效率. 将交通结构化信息提取分为多个步骤进行处理, 也使得最终结果的准确率受到各个步骤的准确率影响.

为了同时实现交通图像中机动车的检测和属性识别, 中科院的郭少博[29] 提出了机动车联合检测和识别方法, 实现快速的交通车辆检测和属性识别, 提高了系统的运行效率. 但由于其使用的了双阶段的 Faster RCNN 网络[30] 作为基础结构, 其运行速度较慢, 难以满足实际中的需求. 且该方法针对车辆检测和属性识别, 并未对车辆进行跟踪, 其应用场景有限. 与之不同的是, 格拉茨技术大学的 Feichtenhofer 等[31] 基于 FasterRCNN 的双阶段网络提出了一种结合目标检测和目标跟踪的方法, 通过特征和目标的包围框进行目标跟踪. 该方法同时实现了目标检测和跟踪, 但由于其基于双阶段的目标检测算法, 在检测框架上仍有较大的优化空间. 因此, 清华大学的 Wang 等[32] 将该网络结构进一步改进, 采用基于 YOLOv3[33] 的基础网络结构进行目标检测和特征提取, 结合卡尔曼滤波和匈牙利算法实现目标跟踪, 且在实时的效率下实现较好的检测和跟踪效果. 但由于该问题的完整数据集较少, 其利用行人检测、行人跟踪和行人重识别的数据集进行训练和测试. 与之不同的是车辆的数据集与行人的问题稍有不同, 数据集在联合训练中还有部分的问题需要解决如车辆检测和车辆重识别问题训练样本的场景不相同等. 因此, 针对车辆的结构化信息提取仍是一个尚未较好解决的问题.

因此, 为解决网络训练繁琐、子任务训练数据有限和网络特征重复提取等问题, 本文采用如下方法:

(1) 改进 YOLO 层, 使多个子任务共享特征提取结构. 将多个子任务提取特征部分整合在一个统一的网络 backbone, 并改进 YOLO 层使之同时解决多个任务的输出为问题.

(2) 构造多任务学习损失函数. 将多个子任务整合在统一的损失函数之中, 利用端到端的网络训练, 实现多任务网络的训练.

(3) 融合多种不同结构的数据集. 根据现有的公开数据集, 将不同感受野大小的图像进行有机融合, 并根据子任务需求对其进行扩充和增广, 对缺失的部分任务的样本进行补充标注.

1 车辆结构化信息提取神经网络

1.1 系统流程

系统流程图见图 1 所示, 主要过程包括: (1) 图像输入; (2) 车辆结构化信息提取网络提取结构化信息; (3) 采用多目标跟踪方法 SORT 实现目标跟踪. 本文以 YOLOv3 网络结构为基础, 构建一个端到端提取车辆结构化信息的网络结构, 实现车辆目标检测、车牌定位、车辆颜色识别、车型分类和车辆的特征提取, 结合 SORT 多目标跟踪方法, 可同时实现实时的车辆检测、跟踪和车辆结构化信息提取.

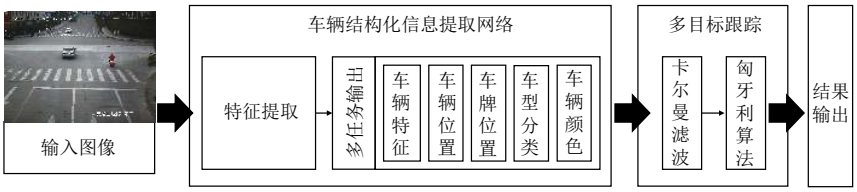

图 1 系统流程图

1.2 车辆结构化信息提取神经网络

YOLOv3 是当前实时目标检测算法中广受工业界欢迎的算法之一, 该网络是 YOLOv2 网络的继承和发展, 主要调整了网络结构, 增加了多尺度的特征提取结

构. 其采用了 Darknet-53 的基础网络结构, 其中包含了 53 个卷积层, 并结合了 Residual Network 的残差网络结构, 在层与层之间加入了 Shortcut 结构, 使得低层特征在深层网络中得到保留. 该网络在第 83 层、95 层和 107 层分别进行特征提取, 实现多尺度的目标检测, 在小目标检测中, 召回率有明显的提高.

本文在 YOLOv3 基础上, 修改了 YOLO 层, 将 YOLO 层输出目标类别和目标位置改进为同时输出车辆特征、车辆位置、车牌位置、车型分类和车辆颜色多个信息, 将 YOLO 层原本的输出特征图大小从 $(2A+4A)\times H\times W$ 改进为 $(2A+4A+4A+2A+D)\times H\times W$, 其中 A 表示该尺度上的锚点模板数量, D 表示 1024 维的车辆特征维度, H 和 W 分别表示该尺度上特征图的高度和宽度. 并通过多任务学习损失函数对多个子任务进行融合. 网络结构见图 2.

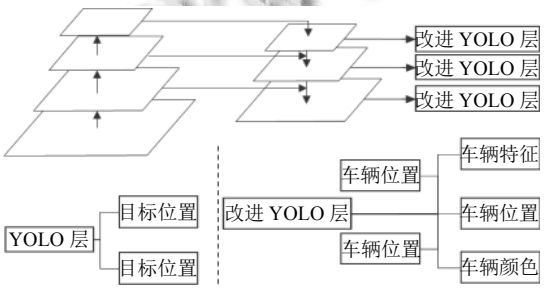

图 2 网络结构图

改进 YOLO 层之后, 目标将 YOLO 网络从单纯的定位和分类的网络拓展为多任务输出的网络, 能同时实现队中任务的训练和使用, 缓解了车辆结构化信息提取因数据量不够而产生的过拟合问题, 且合并了多个子任务, 极大的提高了人工神经网络实际应用价值.

1.3 多任务学习损失函数

本文的损失函数将多任务学习串联在一起, 根据损失函数将多个子任务实现端到端的网络训练, 式 (1) 如下:

$$L_{\text{total}}=\sum_{i=\{\text{obj,Cloc,Lloc,col,type,embed}\}}\frac{1}{2}\left(\frac{1}{e^{s_i}}L_i+s_i\right) \quad (1)$$

其中, L_{obj}, L_{Cloc}, L_{Lloc}, L_{col}, L_{type}, L_{emded} 分别表示前景判断、车辆定位、车牌定位、车辆颜色、车辆类型和车辆特征的损失函数; 前景判断、车辆颜色和车辆类型的损失函数为交叉熵损失函数, 该部分网络后通过 Softmax 对类别进行分类:

$$L_{\text{cls}}=\frac{1}{N}\sum_j L_j=\frac{1}{N}\sum_j\left(-\sum_{c=1}^{M}y_{jc}\log\left(p_{jc}\right)\right) \quad (2)$$

其中, N 表示训练中 batch 的大小, M 表示该子任务的类别数, 其中前景判断分为 2 类, 即前景和背景; 车辆颜色分为 7 类, 即白色、银色、黄色、红色、蓝色、绿色和黑色; 车辆类型分为轿车、载货汽车、客车、挂车和其他车辆 5 种. y_{jc} 表示变量 0 或 1, 如果该类别和 j 的类别相同, 则为 1, 反之则为 0. p_{jc} 表示样本 j 属于类别 c 的预测概率.

车辆定位和车牌定位的损失函数为平滑 $L1$ 损失函数, 如式 (3), 式 (4) 所示:

$$L_{\text{loc}}=\frac{1}{N}\sum_j y_j\cdot smooth_{L1}\left(t_j^{\text{GT}}-t_j^{\text{Pre}}\right) \quad (3)$$

$$smooth_{L1}(x)=\begin{cases}0.5x^2, & \text{if } |x|<1\\ |x|-0.5, & \text{otherwise}\end{cases} \quad (4)$$

车辆特征的获取是一个度量学习的问题, 需要考虑样本数量和收敛速度, 文献 [33] 中对比了三元损失、改进的三元损失和交叉熵损失, 发现相较于其他损失函数, 交叉熵损失函数更适用于多任务学习问题, 因此本文中也用交叉熵损失函数度量车辆特征提取的损失值. 提取的目标特征将被输入到一个全连接层中, 并结合交叉熵损失对其进行训练. 参考文献 [34], S_i 表示各任务的不确定性, 可由训练学习得到. 由于并非所有的样本都进行了完整的标注, 因此需要将样本未标注部分的标签设为–1, 当训练时遇到该部分损失函数计算时, 忽略该部分的损失函数计算.

2 多任务学习训练数据增广

在多任务学习任务中, 收集完整标注的数据集较为困难, 虽然针对单一任务的相关数据集较多, 但其目标大小标注情况都不相同, 比如在车辆检查中, 车辆只占整图中的一小部分, 但车辆重识别的样本中图像大小即为车辆包围框的大小, 在不加整理的情况下同时训练两部分数据会对网络训练的收敛情况造成较大的影响, 尤其是车辆检测和车辆重识别的数据集, 其场景大小差别较大. 车辆检测的任务为检测图中车辆, 并输出车辆在图像中的位置, 其除了车辆之外, 有较多的背景区域; 车辆重识别主要判断多张图中的车辆是否为同一车辆, 其训练数据为了减少其他的干扰, 将车辆目标从原图像中进行截取, 其图像完全包络了车辆目标,

背景区域干扰可以忽略不计, 在同时训练上述数据集时会极大减缓车辆检测的定位损失收敛. 因此需要对训练样本进行增广和整理, 其流程图见图 3. 其主要分为如下几个部分.

图 3 多任务样本准备流程图

(1) 收集数据集, 并将所有的数据集按统一的标注格式进行修改. 由于本文参考了 YOLOv3 的网络, 因此训练数据和 YOLOv3 也较为相似, 样本标注为 (*label*, *cx*/*W*, *cy*/*H*, *cw*/*W*, *ch*/*H*, *type*, *color*, *lx*/*W*, *ly*/*H*, *lw*/*W*, *lh*/*H*, *No*), 其中 *label*, *type* 和 *color* 分别表示是否为车辆、车辆类型和车辆颜色的标注, *cx* 和 *cy* 分别表示车辆包围框中心点的横坐标和纵坐标, *W* 和 *H* 表示图像的宽度和高度, *cw* 和 *ch* 表示车辆包围框的宽度和高度, 类似的, *lx*, *ly*, *lw* 和 *lh* 分别表示车牌包围框的中心点坐标和车牌包围框的宽度和高度, *No* 表示在当前序列中车辆的编号. 在缺省的标签上打上–1, 在训练时忽略该标签的训练. 收集的数据集主要包括车辆检测、车辆类型和车辆重识别的公开数据集.

(2) 数据集整合. 在车辆检测数据集中手工挑选部分目标较少的图像, 将图中的车辆目标抹去, 作为背景样本, 并将车辆重识别的样本在该部分图像中进行填充, 整合成新的车辆重识别数据集. 该数据集在重识别任务中加入了背景干扰, 更符合实际场景中情况, 也解决了车辆重识别和车辆检测数据集在联合训练时产生的检测损失收敛缓慢的问题. 且车辆重识别数据集中除了相同的车辆分为一组之外, 还额外标注了车辆颜色信息, 可以减少部分车辆颜色的手动标注工作.

(3) 数据集补充标注. 由于数据集由各个子任务的数据集组成, 因此需要手工补充标注部分的数据, 使训练数据更接近实际应用场景.

(4) 数据增广. 将所有的样本进行适当的透视变换, 并对图像的饱和度和亮度上增加适当的误差扰动, 扰动取值范围设置为当前饱和度和亮度的 0.5~1.5.

3 实验分析

本文结合了多个车辆相关的数据集, 并在整合补充之后, 得到了相应的车辆结构化特征提取数据集 VSFE, 其主要由公开车辆检测数据集 KITTI、UA-DETRAC、车辆重识别数据集 VRID 和部分自己手工收集及标注的数据, 共 100000 张图像, 其中包括 393468 辆车, 其中 80 000 张为训练集, 20 000 张为测试集. 数据集 KITTI 是自动驾驶中较为有名的数据集, 其通过在车辆上安装多个传感器进行数据集的数据获取和自动标注; 数据集 UA-DETRAC 拍摄于京津冀人行天桥的场景, 手动标注了 8250 个车辆, 详细标注了车辆编号、车辆位置、车辆类型和车辆方向; 车辆重识别数据集 VRID 是中山大学 openData 开放平台上的车辆重识别数据集, 其数据来源于某城市道路卡口, 包含 10000 张车辆图像, 可用于车辆重识别、车辆品牌精细识别和车辆颜色识别.

本文在为证明算法的有效性分别在公开数据集 KITTI 和本文提出的数据集 VSFE 上分别进行实验, 实验基于 Darknet 的神经网络框架实现, 运行在配有英特尔酷睿 i7-7700k CPU 和 2 块 1080Ti GPU 的 PC 机上.

3.1 KITTI 数据集结果分析

本文方法在 VSFE 数据集做预训练, 得到预训练网络, 再将 KITTI 数据集作为训练集进行微调. 由于 KITTI 数据集中并不包含本文方法需要的各类标签, 因此需通过额外的数据集对其进行补充, 实验结果见表 1. 相对于其他方法, 本文算法在实验效果上优于其他同类方法, 运行效率上略低于 YOLOv3, 但仍满足实时性要求.

表 1 KITTI 数据集结果比较

方法	输入大小	效率(FPS)	mAP(%)	Car	Van	Truck	Tram
Faster R-CNN(VGG16)	600×—	11.63	76.90	77.18	72.17	79.48	78.77
YOLOv3	416×416	48.60	91.58	90.21	92.92	95.39	87.79
SSD300	300×300	58.32	81.00	82.04	74.59	86.42	80.94
SSD512	512×512	27.69	79.70	84.77	69.96	83.88	80.17
本文方法	416×416	46.31	93.82	92.86	94.51	96.66	91.58

因为本文方法 (VESNet) 较于 YOLOv3 检测网络需要完成更多的子任务,所以其运行速度会略低于 YOLOv3. 由于多任务学习使得各个子任务之间的特征能得到有效的交流, 已补足数据集本身的数据不均衡、训练数据有限等问题, 因此本文的方法在基础网络相同的情况下, 较 YOLOv3 能提升 2.24% 实验效果.

3.2 VSFE 数据集实验分析

由于本文算法 VESNet 是一个多任务学习的网络, 为了证明该算法在多任务学习上的效果, 需要在相应的数据集上进行端到端的训练, 而传统的车辆重识别数据集和车辆检测数据集的感受野存在很大的差异, 因此不能将两种数据集同时进行网络训练. 因此, 需要基于本文的数据增广方法, 将两种不同的感受野的数据统一到相同大小的感受野之中, 实现端到端的网络训练. 增广后的数据集名为 VSFE, 实验结果见表 2.

表 2 VSFE 数据集结果比较

方法	目标检测(%)		车牌定位(%)		车型分类(%)		颜色识别(%)		目标跟踪跟丢率(%)	平均运行时间(ms)
	准确率	召回率	准确率	召回率	准确率	召回率	准确率	召回率		
本文方法	**95.2**	**96.3**	**5.5**	89.1	92.3	**94.2**	**90.4**	**95.9**	**94.6**	**25.5**
多阶段方法	94.9	95.1	11.4	**90.4**	**95.1**	86.3	84.9	92.8	91.6	20.6+23.2+20.7

在 VSFE 数据集实验中, 本文方法利用本文提出的多任务学习神经网络 VESNet 结合多目标跟踪方法 SORT 进行车辆的结构化信息提取和多目标跟踪; 多阶段方法由几个部分组成, 先通过目标检测神经网络进行车辆检测, 提取车辆具体位置, 再根据车辆当前位置利用 KCF 跟踪算法[14] 进行目标跟踪, 最后再用神经网络对车辆子图像进行车牌定位、车型分类和颜色识别. 由表 2 可知, 本文提出的方法在大部分的子任务处理中都优于多阶段的方法, 且运行时间远少于多阶段方法. 在该实验中, 目标跟踪通过判断目标自被检测至最终一帧是否仍在跟踪, 来判断目标是否跟丢, 从而计算跟丢率.

在目标检测子任务上, 本文方法和多阶段方法由于都用的是 YOLOv3 的基础网络结构, 因此结构较为相近. 因此在相同的训练数据情况下, 目标检测的检测结果相近, 因为多任务学习将多个子任务进行同时训练, 使得网络的收敛效果更好, 因此准确率和召回率略高于单纯的 YOLOv3 目标检测网络. 目标跟踪子任务中, 由于本文方法提取了目标特征, 并维护了跟踪队列, 所以即使在中间某一帧未匹配成功, 也并不影响后续的跟踪结果, 因此效果远好于依赖单目标跟踪的多阶段方法. 而单目标跟踪方法容易受到目标形变、遮挡等问题的影响, 且单目标跟踪的运行时间受图像中目标数量影响, 相较而言本文方法运行时间稳定, 鲁棒性更强, 尤其适应容易出现目标遮挡的复杂场景. 车牌定位子任务中, 由于多阶段方法利用检测出车辆的子图像进行进一步的车牌定位, 因此其结果较好, 且因为远处的车辆车牌较小, 使得本文方法会出现部分漏检的问题, 影响了该部分的成绩. 在车型分类和颜色识别上, 多阶段方法受到多步骤的误差累加问题影响, 其结果明显低于本文的方法. 检测效果图如图 4 所示.

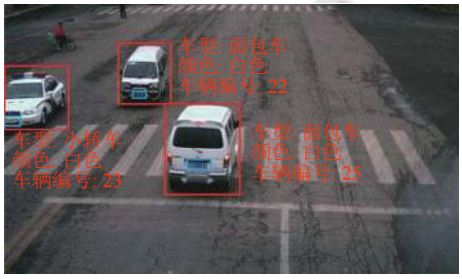

图 4 多任务检测效果图

对比本文方法和多阶段的车辆结构化信息提取方法, 如表 2 所示, 本文方法通过端到端的神经网络计算, 结合快速多目标跟踪的卡尔曼滤波和匈牙利算法实现相邻帧之间的目标跟踪配对, 仅需要 25.5 ms 便完成了一帧图像的检测、跟踪和结构化信息提取; 而多阶段的方法检测、结构化特征提取和目标跟踪分离开来, 分为 3 步进行计算, 其运行时间也由 3 部分组成, 即目标检测时间 (20.6 ms)、结构化特征提取时间 (23.2 ms) 和目标跟踪时间 (平均 20.7 ms). 其中, 特征提取部分进行了多次的重复计算, 使得计算量远高于本文提出的方法.

再者, 使用多次单目标跟踪算法完成多目标跟踪任务, 会在图像中存在多个跟踪目标时, 大量占用计算机资源, 不利于系统稳定, 其鲁棒性较差. 在多阶段方法中, 前一阶段方法的准确率直接对后一阶段的效果有明显的影响, 其误差会随着阶段增多而不断累积, 影响最终结果.

此外, 本文中的数据增广方法使得本文的方法能较好地适应场景中光照变化的情况, 如夜间的车辆检测问题, 也能有较为准确的检测结果, 如图 5 所示. 收到光照变化影响, 夜间车辆颜色检测存在一定的误差, 但其他车辆结构化信息提取仍有较好的实验结果.

图 5 夜间车辆结构化信息提取

当图像中存在多辆车辆同时进行检测跟踪时, 本方法依然能实现较好的检测跟踪效果, 如图 6 所示. 图中大部分的车辆都能较好的完成检测和特征提取, 但由于输入图像大小有限, 部分过于小的目标仍存在丢失的情况, 尤其是远处车辆的车牌位置. 在场景中车辆间会出现相互遮挡的情况下, 本文方法利用车辆前后帧特征进行对应的车辆匹配, 使其在遮挡之后仍然能重新被跟踪到, 能较好地用于实际的视频分析应用之中.

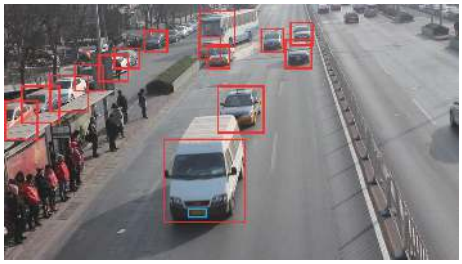

图 6 多车辆场景检测效果图

4 结论与展望

针对车辆结构化特征提取中存在的重复特征提取、过程误差累加和训练样本有限的问题, 本文提出了端到端的车辆结构化特征提取网络. 网络基于 YOLOv3 进行了改进, 通过多任务学习损失函数将车辆检测、车辆特征提取、车辆颜色识别、车辆类型识别和车牌定位任务整合在同一网络之中, 在检测车辆的同时同步输出相应的属性. 在结合多目标跟踪之后, 可对视频中的车辆同时进行检测跟踪, 在智慧交通领域有着广泛的应用.

考虑到车辆的各种数据集存在不兼容的问题, 本文提出了数据整合和增广的方法, 在不增加标注工作量的情况下, 通过将车辆检测数据集和其他数据集进行组合, 实现数据整合和增广. 在本文中, 车辆跟踪基于车辆位置和车辆特征实现, 需要在网络之后增加多目标跟踪方法 SORT 实现目标跟踪, 未来希望能将目标跟踪部分整合到网络中, 实现真正的端到端车辆结构化特征提取.

参考文献

1 才薇. 视频结构化技术在智慧交通领域的新发展与深入应用. 中国安防, 2019, (9): 50–53.
2 李春红. 基于图像结构化描述的车辆识别平台应用研究 [硕士学位论文]. 苏州: 苏州大学, 2016.
3 Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK. 2004. 28–31.
4 Kim K, Chalidabhongse TH, Harwood D, *et al*. Real-time foreground-background segmentation using codebook model. Real-Time Imaging, 2005, 11(3): 172–185. [doi: 10.1016/j.rti.2004.12.004]
5 Liu YZ, Yao HX, Gao W, *et al*. Nonparametric background generation. Journal of Visual Communication and Image Representation, 2007, 18(3): 253–263. [doi: 10.1016/j.jvcir.2007.01.003]
6 Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences. IEEE Transactions on Image Processing, 2011, 20(6): 1709–1724. [doi: 10.1109/TIP.2010.2101613]
7 Patwardhan K, Sapiro G, Morellas V. Robust foreground detection in video using pixel layers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 746–751. [doi: 10.1109/TPAMI.2007.70843]
8 Zhou Y, Liu L, Shao L, *et al*. DAVE: A unified framework for fast vehicle detection and annotation. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 278–293.
9 Shen ZQ, Liu Z, Li JG, *et al*. DSOD: Learning deeply supervised object detectors from scratch. Proceedings of 2017 IEEE International Conference on Computer Vision.

Venice, Italy. 2017. 1919–1927.
10 Zhu CC, He YH, Savvides M. Feature selective anchor-free module for single-shot object detection. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 840–849.
11 Pang JM, Chen K, Shi JP, *et al*. Libra R-CNN: Towards balanced learning for object detection. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 821–830.
12 张富凯, 杨峰, 李策. 基于改进 YOLOv3 的快速车辆检测方法. 计算机工程与应用, 2019, 55(2): 12–20.
13 Henriques JF, Caseiro R, Martins P, *et al*. Exploiting the circulant structure of tracking-by-detection with kernels. Proceedings of the 12th European Conference on Computer Vision. Florence, Italy. 2012. 702–715.
14 Henriques JF, Caseiro R, Martins P, *et al*. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583–596. [doi: 10.1109/TPAMI.2014.2345390]
15 Danelljan M, Häger G, Khan F, *et al*. Accurate scale estimation for robust visual tracking. Proceedings of 2014 British Machine Vision Conference. Nottingham, UK. 2014. 1–11.
16 Vojir T, Noskova J, Matas J. Robust scale-adaptive mean-shift for tracking. Pattern Recognition Letters, 2014, 49: 250–258. [doi: 10.1016/j.patrec.2014.03.025]
17 Bertinetto L, Valmadre J, Golodetz S, *et al*. Staple: Complementary learners for real-time tracking. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 1401–1409.
18 Danelljan M, Robinson A, Khan FS, *et al*. Beyond correlation filters: Learning continuous convolution operators for visual tracking. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 472–488.
19 Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 4293–4302.
20 Fu HY, Ma HD, Wang GY, *et al*. MCFF-CNN: Multiscale comprehensive feature fusion convolutional neural network for vehicle color recognition based on residual learning. Neurocomputing, 2020, 395: 178–187. [doi: 10.1016/j.neucom.2018.02.111]
21 Zhang Q, Zhuo L, Li JF, *et al*. Vehicle color recognition using multiple-layer feature representations of lightweight convolutional neural network. Signal Processing, 2018, 147: 146–153. [doi: 10.1016/j.sigpro.2018.01.021]
22 Zhang Q, Li JF, Zhuo L, *et al*. Vehicle color recognition with vehicle-color saliency detection and dual-orientational dimensionality reduction of cnn deep features. Sensing and Imaging, 2017, 18(1): 20. [doi: 10.1007/s11220-017-0173-8]
23 Hu CP, Bai X, Qi L, *et al*. Vehicle color recognition with spatial pyramid deep learning. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2925–2934. [doi: 10.1109/TITS.2015.2430892]
24 Dong YM, Pei MT, Qin XM. Vehicle color recognition based on license plate color. Proceedings of the 10th International Conference on Computational Intelligence and Security. Kunming, China. 2014. 264–267.
25 Chen P, Bai X, Liu WY. Vehicle color recognition on urban road by feature context. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5): 2340–2346. [doi: 10.1109/TITS.2014.2308897]
26 Li T, Li DM, Dou YM. A novel vehicle type recognition based on template matching. Advanced Materials Research, 2014, 945–949: 1856–1860.
27 Nguyen QA, Irhebhude ME, Ali MA, *et al*. Vehicle type recognition using multiple-feature combinations. IS&T International Symposium on Electronic Imaging (EI 2016). San Francisco, CA, USA. 2016. 1–7.
28 Hu B, Lai JH, Guo CC. Location-aware fine-grained vehicle type recognition using multi-task deep networks. Neurocomputing, 2017, 243: 60–68. [doi: 10.1016/j.neucom.2017.02.085]
29 郭少博, 刘旭, 王子磊. 交通图像中机动车联合检测与识别. 中国图象图形学报, 2017, 22(11): 1503–1511.
30 Ren SQ, He KM, Girshick RB, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 28. Montreal, QB, Canada. 2015. 91–99.
31 Feichtenhofer C, Pinz A, Zisserman A. Detect to track and track to detect. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 3038–3046.
32 Wang ZD, Zheng L, Liu YX, *et al*. Towards real-time multi-object tracking. arXiv: 1909.12605v1, 2019.
33 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.
34 Cipolla R, Gal Y, Kendall A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7482–7491.