

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354582179>

YOLO v3 Deep-SORT

Article · September 2021

DOI: 10.19818/j.cnki.1671-1637.2021.02.019

CITATION

1

READS

502

4 authors, including:



Yide Ma

Lanzhou University

320 PUBLICATIONS 5,113 CITATIONS

SEE PROFILE

文章编号:1671-1637(2021)02-0222-10

基于改进 YOLO v3 模型与 Deep-SORT 算法的道路车辆检测方法

马永杰¹, 马芸婷¹, 程时升¹, 马义德²

(1. 西北师范大学 物理与电子工程学院, 甘肃 兰州 730070; 2. 兰州大学 信息科学与工程学院, 甘肃 兰州 730030)

摘 要:针对道路车辆实时检测遮挡严重与小目标车辆漏检率高的问题,提出了基于改进 YOLO v3 模型和 Deep-SORT 算法的车辆检测方法;为提高模型对道路车辆的检测能力,采用 K -means++ 聚类算法对目标候选框进行聚类分析,选择合适的 Anchor box 数量,并在网络浅层增加了特征提取层,可提取到更精细的车辆特征;为加强网络对远近不同目标的鲁棒性,在保留原 YOLO v3 模型输出层的同时,增加了一层输出层,将 $52 \text{ 像素} \times 52 \text{ 像素}$ 输出特征图经过上采样后得到 $104 \text{ 像素} \times 104 \text{ 像素}$ 特征图,并将其与浅层同尺寸特征图进行拼接,实现车辆目标的检测;为了降低目标遮挡对检测效果的影响,提高对视频上下帧之间关联信息的关注度,将改进 YOLO v3 模型和 Deep-SORT 算法相结合,以此来弥补两者之间的不足。试验结果表明:改进 YOLO v3 模型有效地提高了车辆检测的性能,与在网络浅层增加特征提取层的模型相比,平均精度提高了 1.4%,与增加一层输出层的模型相比,平均精确度提高了 0.8%,说明改进 YOLO v3 模型提取的特征表达能力更强,增强了网络对小目标的检测能力;改进 YOLO v3 模型在引入 Deep-SORT 算法后,查准率和召回率分别达到 90.16% 和 91.34%,相比改进 YOLO v3 模型,查准率和召回率分别提高了 1.48% 和 4.20%,同时保证了检测速度,对于不同大小目标的检测具有良好的鲁棒性。

关键词:交通图像识别;卷积神经网络;车辆检测;YOLO v3 模型;Deep-SORT 算法; K -means++ 聚类算法

中图分类号:U491.2 文献标志码:A DOI:10.19818/j.cnki.1671-1637.2021.02.019

Road vehicle detection method based on improved YOLO v3 model and deep-SORT algorithm

MA Yong-jie¹, MA Yun-ting¹, CHENG Shi-sheng¹, MA Yi-de²

(1. College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, Gansu, China;
2. School of Information Science and Engineering, Lanzhou University, Lanzhou 730030, Gansu, China)

Abstract: A vehicle detection method based on the improved YOLO v3 model and deep-SORT algorithm was proposed to address the problems of serious occlusion and high misdetection rate of small target vehicles in the real-time detection of road vehicles. To improve the detection ability

收稿日期:2020-11-23

基金项目:国家自然科学基金项目(62066041)

作者简介:马永杰(1967-),男,西北师范大学教授,工学博士,从事图像处理、人工智能、测控技术等研究。

引用格式:马永杰,马芸婷,程时升,等.基于改进 YOLO v3 模型与 Deep-SORT 算法的道路车辆检测方法[J].交通运输工程学报,2021,21(2):222-231.

Citation: MA Yong-jie, MA Yun-ting, CHENG Shi-sheng, et al. Road vehicle detection method based on improved YOLO v3 model and deep-SORT algorithm[J]. Journal of Traffic and Transportation Engineering, 2021, 21(2): 222-231.

of the model for road vehicle, the K -means++ clustering algorithm was used to cluster the target candidate boxes, the appropriate number of anchor boxes was selected, and a feature extraction layer to the shallow layer of the network was added to extract more refined vehicle features. The robustness of the network for different distant targets was enhanced by retaining the original YOLO v3 model's output layer but adding another layer to it. After the $52 \text{ pixel} \times 52 \text{ pixel}$ output feature map was upsampled, a $104 \text{ pixel} \times 104 \text{ pixel}$ feature map was obtained, which was spliced with a shallow layer feature map of the same size to achieve the vehicle target detection. To reduce the influence of target occlusion on the detection and improve the attention to the association information between the upper and lower frames of the video, the YOLO v3 model was improved and combined with the deep-SORT algorithm to compensate for their shortcomings. Experimental results show that the improved YOLO v3 model can enhance the vehicle detection performance. Compared with the model adding feature extraction layer in the shallow layer of the network, the average accuracy improves by 1.4%, and compared with the model adding one output layer, the average accuracy improves by 0.8%. It indicates that the improved YOLO v3 model has a stronger feature expression ability and enhances the network's ability to detect small targets. After the deep-SORT algorithm is introduced into the improved YOLO v3 model, the precision and recall are 90.16% and 91.34%, respectively. Compared with the improved YOLO v3 model, the precision and recall increase by 1.48% and 4.20%, respectively. At the same time, the detection speed is maintained, and the detection of different-sized targets is highly robust. 4 tabs, 5 figs, 32 refs.

Key words: traffic image recognition; convolutional neural network; vehicle detection; YOLO v3 model; deep-SORT algorithm; K -means++ clustering algorithm

Author resume: MA Yong-jie(1967-), male, professor, PhD, myjmyj@nwnu.edu.cn.

Foundation item: National Natural Science Foundation of China (62066041)

0 引言

道路安全监控为交通智能控制、解决交通拥堵与交通违法证据提取提供数据支持。道路车辆检测作为道路安全监控的重要组成部分,其研究已经获得了巨大的进展,并成为智能交通系统(Intelligent Transportation System, ITS)的研究热点之一。

传统的车辆检测技术包括预处理、区域选择、提取特征、分类回归4个环节。Kachach等^[1]通过提取滑动窗口内方向梯度直方图(Histogram of Gradient, HOG)特征,并用线性支持向量机(Support Vector Machine, SVM)来进行车辆的检测;Danelljan等^[2]提出了基于相关滤波器估计目标的最佳位置、再估计目标尺度的跟踪方法,对存在大尺度变化的目标取得了较好的跟踪效果;Wei等^[3]提出了一种两步车辆检测算法,首先利用Harr特征分割出感兴趣区域(Region of Interest, ROI),然后从ROI目标区域提取HOG特征,利用SVM进一步提取精确目标,比提取单一特征方法具有更高的检测精度和时

间效率。虽然传统的车辆检测方法可准确的检测出车辆,但选取候选区时是利用不同尺寸的滑动窗口在给定图像上遍历搜索,通过识别每个候选区目标得出检测结果,因此,计算量较大,且存在大量冗余区域,造成时间复杂度高。

近年来,随着深度学习进入快速发展时期,基于深度学习的目标检测方法成为人们研究的热点问题^[4-9]。李琳辉等^[10]将自适应分割方法用于车辆检测,通过降低车底阴影和非车底阴影的干扰,来提高检测准确率;Lee等^[11]将改进的卷积神经网络和Adaboost分类器相结合来进行车辆检测,在检测精度和实时性上有一定的优势;Alireza等^[12]提出了一种基于三维激光雷达和彩色摄像机的多模态车辆检测系统,结合激光雷达和相机的数据,作为卷积神经网络的输入对车辆进行检测,试验结果表明:所提出的融合策略获得了更高的精度;Dai^[13]提出了一种基于两阶段回归的级联目标检测系统,该系统在高分辨率特征图上增加了一个过渡阶段(生成于第1阶段),以获得精确的特征用于第2阶段的决策

细化,实现对车辆的快速检测,试验结果表明:该方法在车辆检测方面比其他先进的方法具有更快、更准确的性能;Luo 等^[14]将 Faster R-CNN 作为网络的基础模型,并结合神经网络架构搜索和特征增强,实现了交通场景中多尺度车辆目标的有效检测;邹伟等^[15]为提高低辨识目标的检测精度,在 Faster R-CNN 模型基础上,融合了其他 3 种模态特征,试验结果表明:多模态特征融合有助于提高低辨识目标的检测性能;汪昱东等^[16]在检测网络中加入注意力机制,通过提升特征提取能力来弥补特殊天气对检测的影响,试验结果表明:该方法有效地提高了网络在雾天环境下检测的平均准确率,且未造成计算量大幅度提升。

目前,基于深度学习的目标检测方法主要分为两类,一类以 RCNN^[17] 为代表的双阶段检测方法^[18-19],先进行区域提取然后再检测,虽然提高了准确率,但速度较慢;另一类为单阶段检测方法,主要代表有 YOLO^[20-22] 和 SSD^[23] 方法,该类方法的回归和类别判定在一个卷积神经网络中完成,简化了网络结构,所以其在速度上有了较大的提升,能够满足对视频处理的实时要求。

YOLO 系列方法在车辆检测任务上的研究取得了很大进展^[24-25]。李珣等^[26]针对道路车辆目标检测率低及鲁棒性差的问题,通过对道路场景的分析,改进了 YOLO-voc 网络模型;黎洲等^[27]为解决车辆检测实时性低的问题,对 YOLO v2 进行了改进,将车辆检测从单一维度检测扩展到了多维度检测,使其能够对车载视频信息进行多维度判断;刘军等^[28]提出了增强 Tiny YOLO v3 模型,并通过匈牙利匹配和卡尔曼滤波方法实现目标车辆的跟

踪,结果表明所提方法使得车辆检测的实时性得到了有效提高,为在嵌入式平台部署提供理论依据;Sri 等^[29]提出了基于 YOLO v3-tiny 网络的深度神经网络,通过对 YOLO v3-tiny 检测网络的特征提取网络进行改进,提高车辆检测的速度和精度;柳长源等^[30]将 YOLO v3 模型的主干网络替换为轻量化模型,并融合不同尺度的特征层与引入 K-means 聚类算法,试验结果表明,该方法相比改进前检测效率有所提升。

借助 YOLO v3 的检测优势,针对车辆检测过程中遮挡严重和小目标车辆漏检率高的问题,提出了基于改进 YOLO v3 模型和 Deep-SORT 算法的车辆检测方法。为了增强网络对远近目标的检测鲁棒性,在保留原 YOLO v3 输出层的同时,增加了对浅层特征图的关注,将 52 像素 \times 52 像素输出特征图经过上采样后得到 104 像素 \times 104 像素特征图,并将其与浅层同尺寸特征图进行拼接;为了提升网络对小目标的敏感性,改变原 YOLO v3 的模型结构,新增 2 个残差单元;为进一步提高检测模型的鲁棒性,在检测模型中引入 Deep-SORT 算法进行多目标跟踪,从而降低了漏检与误检的可能性,降低了遮挡条件下对车辆检测性能的影响。

1 YOLO v3 车辆检测模型

YOLO v3 模型由 Redmon 等^[22]在 2018 年提出,网络结构如图 1 所示,YOLO v3 基本组件由卷积层、批归一化 (Batch Normalization, BN) 和 Leaky RELU 激活函数组成; \approx 表示该下采样块中包含残差单元的数量;concat 为张量拼接函数。该模型在保证实时性的情况下,使用了更深的 Darknet-53 作为基础网络,借鉴了 ResNet 的残差结构,即将低层特征图

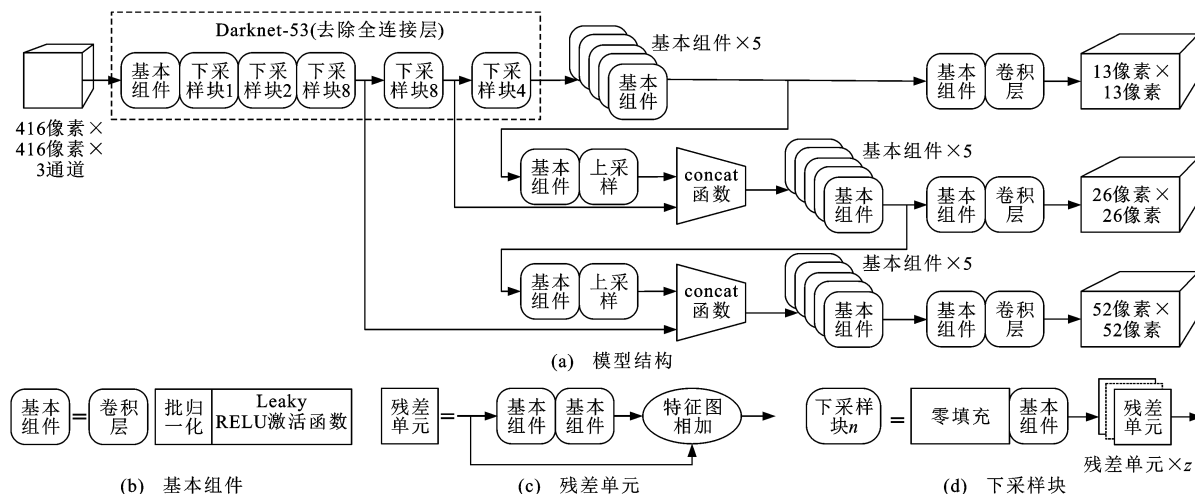


图 1 YOLO v3 网络模型结构

Fig. 1 YOLO v3 network model structure

直接映射到高层网络中可以让网络结构更深,并避免随着层数增加误差也逐层增大的网络退化问题和梯度消失问题。通过输出3个不同尺度的特征图,来对不同大小的目标进行检测,精细的网格单元进一步提高了目标检测精度,YOLO v3 模型相比之前 YOLO 系列虽然速度上没有提升,但准确性能更好。车辆检测需要保证实时性,YOLO v3 模型在目标检测上具有较好的表现,处理速度较快,因此,采用 YOLO v3 模型网络检测车辆。

为了充分利用 YOLO v3 模型的优势以及计算资源,调整 YOLO v3 模型网络参数使其适应车辆检测任务,根据数据集中车辆宽高的先验知识,通过 K -means 聚类算法来获取最优的 Anchor boxes 比例。此外,为了提高网络对远近不同目标检测的鲁棒性,对 YOLO v3 的网络结构进行了修改。

1.1 车辆预选框的聚类分析

YOLO v3 结合了 Faster R-CNN 的 Anchor boxes 机制,在训练过程中,样本框依次与 Anchor box 计算交并比(Intersection over Union, IoU)值,根据最大 IoU 筛选出预选框,通过不断地调整预选框,使其最终接近标注的真实框,因此,目标检测的精度和速度取决于 Anchor boxes 机制的正确选择。针对 COCO 数据集^[31],YOLO v3 设置了9种不同尺度的 Anchor box 框,在 COCO 数据集中,目标所占图片比例都较大,而在道路交通中,车辆目标的成像大小由摄像头距离来决定,因此,Anchor box 需要根据具体数据做出相应地调整,利用 K -means++ 聚类算法对数据集中目标框的宽高进行聚类,以此提高模型的泛化能力。

K -means 聚类方法通过对样本的不断迭代来进行分组,属于无监督学习方法,采用距离作为相似性指标,选择 Anchor box 的宽高作为聚类对象,Anchor box 与标注的样本框的 IoU 越大越好,因此,采用平均重叠度(Average Intersection over Union, Avg IoU)来衡量两个坐标框之间的相似度,聚类的 Avg IoU 目标函数 f 为

$$f = \arg \max \frac{\sum_{p=1}^k \sum_{q=1}^{n_m} I(B_q, C_p)}{n} \quad (1)$$

式中: $I(B_q, C_p)$ 为车辆样本框 B_q 和聚类中心 C_p 的交并比; n 为车辆样本数; k 为聚类中心数量; n_m 为第 m 个聚类中心的车辆样本数; p 为聚类中心编号; q 为在聚类中心 C_p 的车辆样本编号。

该方法采用距离作为相似性指标,通过计算每

个样本框与聚类中心的距离来分配样本框,并通过不断迭代来更新聚类中心,由于 K -means 算法初始聚类中心点是随机选取的,这样会影响簇的划分,因此,选择 K -means++ 算法来对 Anchor box 的宽高进行聚类, K -means++ 算法对初始聚类中心的确定进行了优化,尽可能增大 k 个初始聚类中心之间的距离,对聚类效率的提高有一定的帮助,具体步骤如下。

步骤1:随机选取聚类中心 C_1 。

步骤2:计算样本与聚类中心的距离,根据距离越大,成为聚类中心的概率越大原则计算每个样本成为聚类中心的概率。

步骤3:重复步骤2直到选择出 k 个初始聚类中心。

步骤4:通过计算样本框到 k 个聚类中心的交并比,将其分到交并比最大的簇中。

步骤5:根据所得类别,计算出新的聚类中心。

步骤6:重复步骤4、5,计算相邻两代的 IoU 差值,若差值大于阈值,方法终止。

1.2 YOLO v3 网络结构改进

在 YOLO v3 网络中,网络最后输出为 13 像素×13 像素、26 像素×26 像素、52 像素×52 像素的特征图与上采样得到的同尺寸特征图的拼接,由于不同尺度特征图的感受野(Receptive Field, RF)大小不同,因此,所提取的特征也不一样,感受野是指卷积神经网络每层输出特征图上的像素点映射到输入图片上的区域大小,感受野计算方法为

$$F_l = S_l(F_{l-1} - 1) + K_l \quad (2)$$

式中: F_l 、 S_l 与 K_l 分别为 l 层的感受野、卷积步长与卷积核大小。

由式(2)可知:随着网络的加深,特征图上的感受野也逐渐增大,所提取的特征更关注全局信息^[6]。在道路交通中,摄像头距离决定了车辆目标的成像大小,距离摄像头近的车辆像素点数量较多,距离摄像头远的车辆占用图片比例相对较小,在较深的网络中,远处小尺寸车辆信息丢失严重,因此,对 YOLO v3 网络结构做出以下改进,如图2所示。

改进1:如图2中虚线框中改进1所示,在原有网络3个输出层的基础上,增加了一层大小为 104 像素×104 像素的输出,将上一层 52 像素×52 像素输出特征图经过上采样后得到 104 像素×104 像素特征图,并将其与浅层 104 像素×104 像素的特征图通过 concat 函数进行拼接,该改进方案可以提取到更多网络浅层的特征,为检测网络提供了更多检

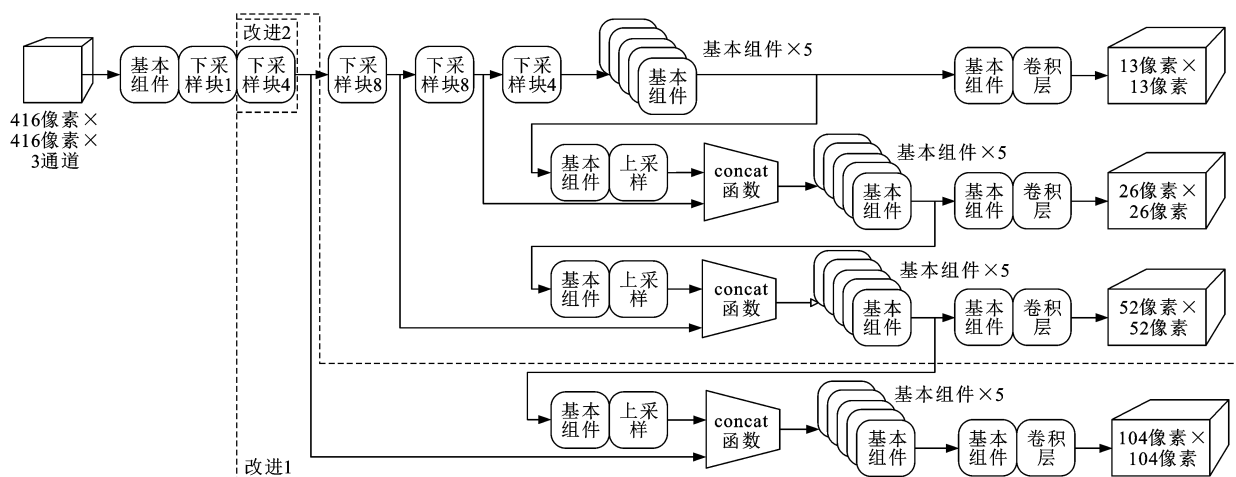


图 2 改进的 YOLO v3 模型网络结构

Fig. 2 Improved YOLO v3 model network structure

测依据,增强了网络的鲁棒性。

改进 2:如图 2 中虚线框中改进 2 所示,在识别的过程中,输入特征提取得越精细,输入图像被检测到的可能性越大,且精度越高,因此,在原网络中的第 2 个残差结构中增加 2 个残差单元,以提取到更多的车辆信息。

2 基于 YOLO v3 模型和 Deep-SORT 算法的道路车辆检测方法

Deep-SORT 算法由 SORT 算法改进而来, SORT 算法实时性较高,但无法保证在状态估计不确定性较高时的准确性,使得 SORT 在有遮挡情况下跟踪失败。为了解决此问题,Deep-SORT 算法在其基础上将目标运动和表面特征信息结合起来,即将基于深度特征向量最小余弦距离的表现信息和基于马氏距离的运动信息进行加权融合成一个新的度量匹配信息,有效地提高了对遮挡的处理能力,减少了目标身份错分类问题。

改进 YOLO v3 模型可以有效地检测道路交通环境下的车辆,但由于改进 YOLO v3 模型在遮挡条件下对目标不够敏感,误检与漏检的情况仍无法避免,因此,目标检测精度受到局限,为了缓解改进 YOLO v3 模型在目标检测中的不足,提高检测模型的鲁棒性,在检测模型中加入 Deep-SORT 算法,从而降低了漏检与误检的可能性,改善了遮挡对车辆检测性能的影响。图 3 为改进 YOLO v3 模型结合 Deep-SORT 算法的整体流程图。

2.1 改进 YOLO v3 模型结合 Kalman 滤波方法

Kalman 滤波器利用检测方法所得的前一帧目标信息,来预测对应目标在最后一帧的跟踪信息。

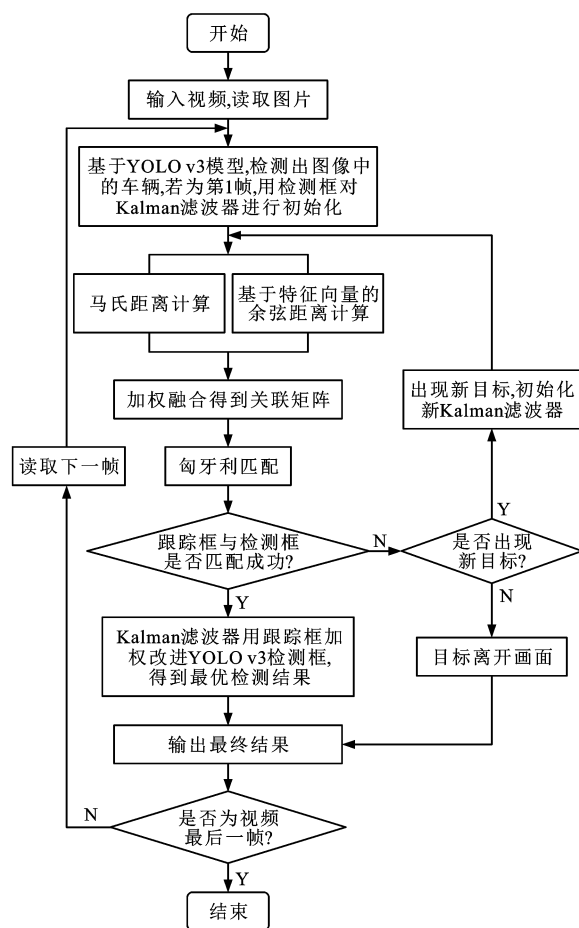


图 3 基于改进 YOLO v3 模型结合 Deep-SORT 算法的检测方法

Fig. 3 Detection method of improved YOLO v3 model combined with deep-SORT algorithm

Kalman 滤波算法主要分预测与更新两步实现跟踪。设置状态向量 X 为 $(x, y, w, h, v_x, v_y, v_w, v_h)$ 初始化 Kalman 滤波器,其中 x, y 为框的中心点横、纵坐标, w, h 为框的宽度与高度,由改进型 YOLO v3 模型检测结果可知, v_x, v_y, v_w, v_h 分别为 x, y, w, h 相

对应的速度分量,系统状态方程和观测方程分别为

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{U}_t + \mathbf{W}_t \quad (3)$$

$$\mathbf{Z}_t = \mathbf{H}\mathbf{X}_t + \mathbf{V}_t \quad (4)$$

式中: \mathbf{X}_t 、 \mathbf{U}_t 、 \mathbf{W}_t 分别为 t 时刻的状态矩阵、控制矩阵、系统噪声; \mathbf{A} 和 \mathbf{B} 分别为系统相关状态转换矩阵和输入控制矩阵; \mathbf{H} 为系统观测矩阵; \mathbf{Z}_t 和 \mathbf{V}_t 分别为 t 时刻的观测状态和观测噪声。

2.1.1 预测过程

由系统 $t-1$ 时刻状态来预测下一时刻状态,系统状态变量的预测方程和协方差预测方程为

$$\mathbf{X}_{t|t-1} = \mathbf{A}\mathbf{X}_{t-1|t-1} + \mathbf{B}\mathbf{U}_t \quad (5)$$

$$\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}^T + \mathbf{Q} \quad (6)$$

式中: $\mathbf{X}_{t|t-1}$ 为 $t-1$ 时刻预测得到 t 时刻状态; $\mathbf{X}_{t-1|t-1}$ 为 $t-1$ 时刻的最优状态; $\mathbf{P}_{t|t-1}$ 和 $\mathbf{P}_{t-1|t-1}$ 分别为 $\mathbf{X}_{t|t-1}$ 和 $\mathbf{X}_{t-1|t-1}$ 的协方差矩阵; \mathbf{Q} 为对应系统噪声的协方差。

2.1.2 更新过程

结合当前时刻的系统状态结果,利用式(7)得到 t 时刻的最终状态最优估计值,Kalman 增益计算如式(8)所示,即

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \mathbf{K}_t(\mathbf{Z}_t - \mathbf{H}\mathbf{X}_{t|t-1}) \quad (7)$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R})^{-1} \quad (8)$$

式中: \mathbf{K}_t 为 t 时刻的 Kalman 增益矩阵; \mathbf{R} 为对应系统观测噪声的协方差。

状态变量和协方差矩阵需要利用 Kalman 增益对其进行修正。

利用式(9)来更新 t 时刻的协方差,使滤波器重复运行

$$\mathbf{P}_{t|t} = (1 - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1} \quad (9)$$

式中: $\mathbf{P}_{t|t}$ 为 t 时刻状态的协方差更新结果。

2.2 匈牙利匹配

改进 YOLO v3 模型和 Deep-SORT 算法的同时使用,有可能检测到同一目标,为解决重复标记目标区域的问题,在 Deep-SORT 算法中借助匈牙利算法,通过马氏距离和基于特征向量的最小余弦距离加权融合的量值来关联匹配检测框与跟踪框的信息,得出检测结果。

2.2.1 运动匹配

马氏距离为检测框和跟踪框之间相似度的评价指标,对于样本向量 \mathbf{a} 和 \mathbf{b} ,协方差矩阵为 \mathbf{S} ,其马氏距离 \mathbf{M} 为

$$\mathbf{M} = \sqrt{(\mathbf{a} - \mathbf{b})^T \mathbf{S}^{-1} (\mathbf{a} - \mathbf{b})} \quad (10)$$

通过计算 Kalman 滤波预测跟踪框和检测框之间的马氏距离,得出两者的运动匹配程度,即

$$\mathbf{d}_{(i,j)}^{(1)} = (\mathbf{d}_j - \mathbf{y}_i)^T \mathbf{S}_i^{-1} (\mathbf{d}_j - \mathbf{y}_i) \quad (11)$$

式中: $\mathbf{d}_{(i,j)}^{(1)}$ 为跟踪框 i 和检测框 j 之间的马氏距离; \mathbf{S}_i^{-1} 为跟踪框 i 的协方差预测矩阵; \mathbf{y}_i 、 \mathbf{d}_j 分别为跟踪框 i 和检测框 j 的状态向量。

2.2.2 表观匹配

检测框与跟踪框的正确匹配程度直接决定了车辆检测网络的性能优劣,因此,引入表观匹配的方法来增强匹配准确度,采用最小余弦距离来评价表征特征相似度,跟踪框 i 和检测框 j 之间的表征相似度距离 $\mathbf{d}_{(i,j)}^{(2)}$ 为

$$\mathbf{d}_{(i,j)}^{(2)} = 1 - \frac{\mathbf{G}_j \mathbf{N}_i}{\|\mathbf{G}_j\| \|\mathbf{N}_i\|} \quad (12)$$

式中: \mathbf{G}_j 、 \mathbf{N}_i 分别为当前帧检测框 j 的特征和上一帧跟踪框 i 的特征。

最后,将式(11)的运动特征相似度距离和式(12)的表观相似度距离进行加权融合,超参数为 λ ,得到最终的匹配度量值 $\mathbf{D}_{(i,j)}$,越小则表示两者越相似,即

$$\mathbf{D}_{(i,j)} = \lambda \mathbf{d}_{(i,j)}^{(1)} + (1 - \lambda) \mathbf{d}_{(i,j)}^{(2)} \quad (13)$$

2.2.3 更新策略

更新过程由检测框和跟踪框的匹配结果决定。首先通过匈牙利算法得到匹配结果,用匹配度量值对代价矩阵进行修正,然后用 Kalman 滤波器依据匹配结果进行迭代更新。

3 试验结果分析

3.1 试验数据与运行环境

试验采用的数据集分为 UA-DETRAC 公开数据集和手动采集制作的数据两种。UA-DETRAC 数据集主要拍摄于北京和天津的 24 个不同道路过街天桥,测试集包含 40 个视频序列,训练集包含 60 个视频序列,其中手动标注了 8 250 辆车和 121 万目标对象边界框。手动采集的数据集图片总数 1 500 张,包含小轿车、公交车、SUV 等,统一为标签 car,使用开源图像标注工具 LabelImg 对数据集中车辆进行了人工标注,生成相对应的 xml 文件,该文件记录了图片名称、路径、目标标签名称及位置坐标信息。包含标注信息的 xml 文件不能直接用来训练 YOLO 网络,使用 python 脚本 voc_label 将 xml 文件转换成 YOLO 支持的 txt 文件。

从查准率 p 、召回率 r 、平均精确度 A 、每秒检测帧数这 4 个客观评价指标,以及视觉效果来评价基于改进 YOLO v3 和 Deep-SORT 的道路车辆检测方法。查准率 p 反映了检测的准确性,体现检测模

型拒绝误检的能力,召回率 r 反映了检测的全面性,体现检测模型正确检测的能力,平均精确度 A 为以 r 为横坐标、 p 为纵坐标所绘制曲线下的面积,从查准率和召回率两个角度来衡量方法的准确性,反映了模型对车辆目标的综合检测性能。

3.2 道路车辆检测网络的训练与测试

3.2.1 网络训练参数的确定

随着 Anchor box 的数量增加,检测过程中可将大小、距离相似的目标框分配到不同格子上,使得尺寸划分更为精细,能够提高网络的检测性能,降低漏检情况的发生率。

采用 K -means++ 聚类算法对目标候选框进行聚类分析,选择合适的 Anchor box 数量,并且该方法所选出的 Anchor box 更适用于道路车辆目标的检测,提高检测精度。目标聚类分析采用上文提到的 Avg IoU 来度量,对训练数据集进行聚类分析。图 4 为 Avg IoU 随着聚类中心数量 k 变化的关系曲线,可知:随着聚类中心数量 k 的增加,平均交并比也逐渐增大,当聚类中心数量大于 9 之后,平均交并比上升幅度变小且逐渐趋于平稳,因此,以平均交并比作为指导,通过 K -means++ 聚类算法分析,得到 Anchor box 数量为 9,其对应聚类中心的宽度和高度分别为 (11,13)、(19,33)、(34,21)、(31,64)、(62,48)、(61,117)、(121,89)、(158,196)、(262,215)。

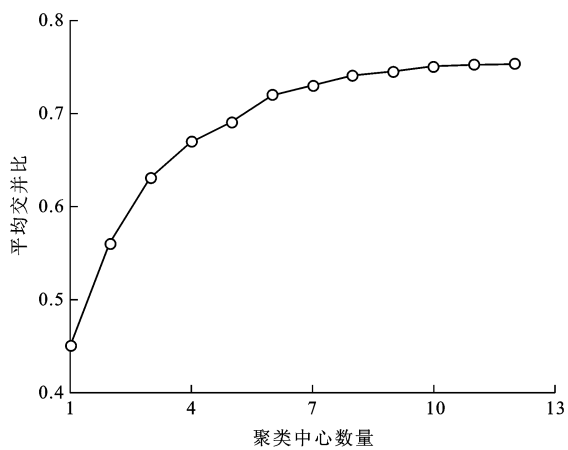


图 4 K -means++ 聚类算法分析曲线
Fig. 4 Analysis curve of K -means++ clustering algorithm

经过反复试验,最终确定网络训练参数见表 1。

3.2.2 改进网络结构模型性能测试

YOLO v3 模型采用多个尺度对目标进行检测,可同时兼顾大小目标,适用于道路交通环境中的车辆检测,因此,设定只采用改进条件 1 为对比模型 1,设定只采用改进条件 2 为对比模型 2,并与所改进的 YOLO v3 模型以及 YOLO v3 模型进行对比试验,

表 1 网络训练参数设置

Tab. 1 Network training parameter settings

名称	参数
训练样本数量	32
权重衰减系数	0.000 5
学习速率变化因子	0.9
模型最大迭代次数	30 000
初始学习率	0.001

结果见表 2。由表 2 可知:对比模型 1 相比 YOLO v3 模型平均精确度提升了 1.4%,对比模型 1 对 YOLO v3 模型进行了四尺度预测,增大了模型对远处小目标车辆的敏感度;对比模型 2 相比 YOLO v3 模型平均精确度提升了 0.8%,对比模型 2 在原网络的基础上在第 2 个残差模块中增加了 2 个残差单元,增加了低层网络层数,获取了更多的低层局部信息,进一步增强了 YOLO v3 对遮挡车辆和远处车辆的检测能力。

表 2 改进 YOLO v3 模型与 YOLO v3 模型性能比较

Tab. 2 Performance comparison between improved YOLO v3 and YOLO v3 model

模型	平均精确度/%	每秒检测帧数
YOLO v3	82.7	39
对比模型 1	84.1	33
对比模型 2	83.5	36
改进 YOLO v3	85.4	32

改进 YOLO v3 模型相比原模型平均精确度有了 2.7% 的提升,因为改进 YOLO v3 模型结合了 2 种改进方法的优点,使得测量精度有所提升,但由于改进模型加强了对浅层特征的关注,增加了车辆候选框的数量,使得模型复杂度提高,导致检测速度变慢,每秒检测帧数为 32,相比于 YOLO v3 模型有所降低,但仍满足实时性要求。

3.3 改进 YOLO v3 模型+Deep-SORT 算法检测方法结果分析

3.3.1 客观分析

为了进一步提高车辆检测性,在改进 YOLO v3 模型的基础上加入 Deep-SORT 算法。设计一组对比试验来验证加入 Deep-SORT 算法的性能,结果如表 3 所示。

表 3 检测方法性能比较

Tab. 3 Performance comparison of detection methods

方法	实际车辆数	正确检测数量	误检数量	漏检数量	$p/\%$	$r/\%$
改进 YOLO v3 模型	163	142	18	21	88.68	87.14
改进 YOLO v3 模型+Deep-SORT 算法	163	149	16	14	90.16	91.34

改进 YOLO v3 模型通过增加对低层特征的关注来改善对远处小目标的检测问题,从而提高了模型的整体性能。单一的改进 YOLO v3 模型查准率和召回率分别是 88.68%和 87.14%,而结合改进 YOLO v3 和 Deep-SORT 方法在最终的车辆目标检测上查准率和召回率提升至 90.16%和 91.34%,这是因为检测配合跟踪方法缓解了“掉帧”现象的影响,相比于单一的改进 YOLO v3 模型,检测能力有了一定提高。

3.3.2 主观分析

图 5 为实际交通环境车辆的检测跟踪效果对比,图 5(a)为 SSD 方法的车辆检测效果,图 5(b)为改进 YOLO v3 模型的检测效果,图 5(c)为融合方法的检测效果。由图 5 可以看出:SSD 方法丢失了较多车辆目标,且无法检测到远处的小目标,改进 YOLO v3 模型可同时兼顾远、近的大小目标,但丢失了一辆远处遮挡较为严重的车辆小目标,可以看出结合改进 YOLO v3 模型+Deep-SORT 算法可

以根据置信度估计车辆目标的位置,减少误检和漏检现象,且提升了远处小目标车辆的检测效果,很好地保留了车辆目标,而且检测跟踪得到的目标位置和大小都大致与真实目标区域一致,对于视频序列的车辆检测具有更好的连续稳定性能。

3.3.3 与其他检测方法对比

为进一步评估改进 YOLO v3 模型+Deep-SORT 算法的性能,选择 Liu 等^[23,32]提出的车辆检测方法进行比较,结果见表 4,可知:改进 YOLO v3 模型+Deep-SORT 算法比其他 2 种检测方法的查准率和召回率更高。分析表明,Liu 等^[23]在检测多个紧密邻近的小目标上表现一般,其对应特征图上的像素难以得到充分训练,只有通过增大输入图像的大小来提高对小目标的检测效果;王宇宁等^[32]提出的方法采用 YOLO 作为基础网络,相比 YOLO v3 模型缺少了多尺度特征的输出,得到的特征信息不够全面,对远处车辆检测能力有限;改进 YOLO v3 模型+Deep-SORT 算法为改善对不同大小目标的鲁棒性,增强了对 YOLO v3 模型底层特征的关注度,低层特征更适用于检测小目标,因此,改进 YOLO v3 模型+Deep-SORT 算法具有更好的效果。

表 4 与其他检测方法对比

Tab. 4 Comparison with other detection methods			
方法	$p/\%$	$r/\%$	每秒检测帧数
Liu 等 ^[23] 的检测方法	78.92	79.56	35
王宇宁等 ^[32] 的检测方法	89.30	81.03	60
改进 YOLO v3 模型+Deep-SORT 算法	90.16	91.34	38

4 结 语

(1)提出了将 YOLO v3 模型+Deep-SORT 算法应用于道路车辆检测中,实现了对道路车辆的实时检测。

(2)对数据集样本进行聚类分析,选择合适的 Anchor box 参数,当 Anchor box 数量为 9 时,可以有效地提高网络的精度,增强 YOLO v3 模型对未知对象的检测能力。

(3)以 YOLO v3 作为骨架网络,通过对浅层特征图感受野的分析,发现特征图上的感受野随着网络的加深逐渐增大,所提取的特征更关注全局信息,容易丢失小目标信息,因此,在网络浅层增加了特征提取层,使平均精确度达到 84.1%,增强了网络对小尺寸目标的检测能力;通过增加一层输出层,使平均精确度达到 83.5%,加强了网络对远近不同目标

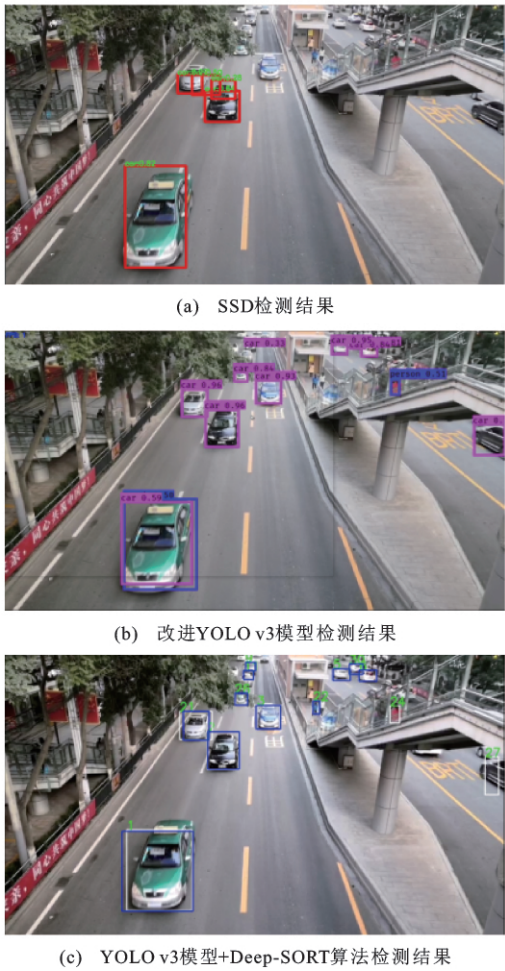


图 5 同一视频的部分检测结果对比
Fig. 5 Comparison of partial detection results of same video

的鲁棒性。相比 YOLO v3 模型,改进模型的平均精确度有了 2.7% 的提升,说明对 YOLO v3 模型的改进合理有效。

(4)在改进 YOLO v3 模型基础上加入 DeepSORT 算法,弥补了 YOLO v3 模型在视频目标检测中忽略上下帧关联信息的不足以及出现目标“掉帧”的现象,同时缓解了目标遮挡对检测率的影响,查准率和召回率分别提升至 90.16% 和 91.34%,而单一的改进 YOLO v3 模型分别为 88.68% 和 87.14%,说明该方法的检测能力相比于单一模型有了一定提高。

(5)改进方法虽然提高了检测性能,但增加了计算量,每秒识别帧数降为 32,相比于 YOLO v3 模型有所降低,因此,网络结构需要继续优化。

参考文献:

References:

- [1] KACHACH R, CAÑAS J M. Hybrid three-dimensional and support vector machine approach for automatic vehicle tracking and classification using a single camera[J]. Journal of Electronic Imaging, 2016, 25(3): 033021.
- [2] DANELLJAN M, HAGER G, KHAN F, et al. Discriminative scale space tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [3] WEI Yun, TIAN Qing, GUO Jian-hua, et al. Multi-vehicle detection algorithm through combining Harr and HOG features[J]. Mathematics and Computers in Simulation, 2019, 155: 130-145.
- [4] 杨 娟,曹浩宇,汪荣贵,等.基于语义 DCNN 特征融合的细粒度车型识别模型[J].计算机辅助设计与图形学学报,2019, 31(1):141-157.
YANG Juan, CAO Hao-yu, WANG Rong-gui, et al. Fine-grained car recognition model based on semantic DCNN features fusion[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(1): 141-157. (in Chinese)
- [5] 余 烨,傅云翔,杨昌东,等.基于 FR-ResNet 的车辆型号精细识别研究[J/OL].自动化学报,(2019-04-03)[2020-12-22]. DOI:10.16383/j.aas.c180539.
YU Ye, FU Yun-xiang, YANG Chang-dong, et al. Fine-grained car model recognition based on FR-ResNet[J/OL]. Acta Automatica Sinica, (2019-04-03)[2020-12-22]. DOI: 10.16383/j.aas.c180539. (in Chinese)
- [6] 凌 艳,陈 莹.多尺度上下文信息增强的显著目标检测全卷积网络[J].计算机辅助设计与图形学学报,2019, 31(11): 2007-2016.
LING Yan, CHEN Ying. Salient object detection with multiscale context enhanced fully convolution network[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(11): 2007-2016. (in Chinese)
- [7] 鞠默然,罗海波,王仲博,等.改进的 YOLO v3 算法及其在小目标检测中的应用[J].光学学报,2019,39(7):253-260.
JU Mo-ran, LUO Hai-bo, WANG Zhong-bo, et al. Improved YOLO v3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 253-260. (in Chinese)
- [8] LI Su-hao, LIN Jin-zhao, LI Guo-quan, et al. Vehicle type detection based on deep learning in traffic scene[J]. Procedia Computer Science, 2018, 131: 564-572.
- [9] 曹 磊,王 强,史润佳,等.基于改进 RPN 的 Faster-RCNN 网络 SAR 图像车辆目标检测方法[J].东南大学学报(自然科学版),2021,51(1):87-91.
CAO Lei, WANG Qiang, SHI Run-jia, et al. Method for vehicle target detection on SAR image based on improved RPN in Faster-RCNN[J]. Journal of Southeast University (Natural Science Edition), 2021, 51(1): 87-91. (in Chinese)
- [10] 李琳辉,伦智梅,连 静,等.基于卷积神经网络的道路车辆检测方法[J].吉林大学学报(工学版),2017,47(2):384-391.
LI Lin-hui, LUN Zhi-mei, LIAN Jing, et al. Convolution neural network-based vehicle detection method[J]. Journal of Jilin University (Engineering and Technology Edition), 2017, 47(2): 384-391. (in Chinese)
- [11] LEE W J, DONG S P, DONG W K, et al. A vehicle detection using selective multi-stage features in convolutional neural networks[C]//IEEE. 17th International Conference on Control, Automation and Systems. New York: IEEE, 2017: 1-3.
- [12] ALIREZA A, LUIS G, CRISTIANO P, et al. Multimodal vehicle detection: fusing 3D-LIDAR and color camera data[J]. Pattern Recognition Letters, 2018, 115: 20-29.
- [13] DAI Xue-rui. HybridNet: a fast vehicle detection system for autonomous driving[J]. Signal Processing: Image Communication, 2019, 70: 79-88.
- [14] LUO Ji-qing, FANG Hu-sheng, SHAO Fa-ming, et al. Multi-scale traffic vehicle detection based on faster R-CNN with NAS optimization and feature enrichment[J]. Defence Technology, 2021, DOI: 10.1016/j.dt.2020.10.006.
- [15] 邹 伟,殷国栋,刘昊吉,等.基于多模态特征融合的自动驾驶车辆低辨识目标检测方法[J/OL].中国机械工程,(2020-06-24)[2020-12-22]. https://kns.cnki.net/kcms/detail/42.1294.TH.20200624.1308.008.html.
ZOU Wei, YIN Guo-dong, LIU Hao-ji, et al. Low-observable targets detection method for autonomous vehicles based on multi-modal feature fusion [J/OL]. China Mechanical Engineering, (2020-06-24)[2020-12-22]. https://kns.cnki.net/kcms/detail/42.1294.TH.20200624.1308.008.html. (in Chinese)
- [16] 汪昱东,郭继昌,王天保.一种改进的雾天图像行人和车辆检测算法[J].西安电子科技大学学报,2020,47(4):70-77.
WANG Yu-dong, GUO Ji-chang, WANG Tian-bao. Algorithm for foggy-image pedestrian and vehicle detection[J]. Journal of Xidian University, 2020, 47(4): 70-77. (in Chinese)
- [17] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature

- hierarchies for accurate object detection and semantic segmentation[C]//IEEE. 27th IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE, 2014: 580-587.
- [18] GIRSHICK R. Fast R-CNN[C]//IEEE. 15th IEEE International Conference on Computer Vision, New York: IEEE, 2015: 1440-1448.
- [19] REN Shao-qing, HE Kai-ming, 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.
- [20] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//IEEE. 29th IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE, 2016: 779-788.
- [21] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//IEEE. 30th IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 6517-6525.
- [22] REDMON J, FARHADI A. YOLOv3: an incremental improvement[R]. Ithaca: Cornell University, 2018.
- [23] LIU WEI, ANGUELOV D, ERHAN D, et al. SSD: single shotmultibox detector[C]//Springer. 14th European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [24] 周 苏,支雪磊,林飞滨,等.基于车载视频图像的车辆检测与跟踪算法[J]. 同济大学学报(自然科学版), 2019, 47(S1): 191-198.
- ZHOU Su, ZHI Xue-lei, LIN Fei-bin, et al. Research on vehicle detection and tracking algorithm based on onboard video images [J]. Journal of Tongji University (Natural Science Edition), 2019, 47(S1): 191-198. (in Chinese)
- [25] 金立生,郭柏苍,王芳荣,等.基于改进YOLOv3的车辆前方动态多目标检测算法[J/OL]. 吉林大学学报(工学版), (2020-12-17)[2020-12-22]. <https://doi.org/10.13229/j.cnki.jdxbgxb20200588>.
- JIN Li-sheng, GUO Bai-cang, WANG Fang-rong. Dynamic multiple object algorithm for vehicle forward based on improved YOLOv3 [J/OL]. Journal of Jilin University (Engineering and Technology Edition), (2020-12-17)[2020-12-22]. <https://doi.org/10.13229/j.cnki.jdxbgxb20200588>. (in Chinese)
- [26] 李 珣,刘 瑶,李鹏飞,等.基于Darknet框架下YOLO v2算法的车辆多目标检测方法[J]. 交通运输工程学报, 2018, 18(6): 142-158.
- LI Xun, LIU Yao, LI Peng-fei, et al. Vehicle multi-target detection method based on YOLO v2 algorithm under darknet framework[J]. Journal of Traffic and Transportation Engineering, 2018, 18(6): 142-158. (in Chinese)
- [27] 黎 洲,黄妙华.基于YOLO_v2模型的车辆实时检测[J]. 中国机械工程, 2018, 29(15): 1869-1874.
- LI Zhou, HUANG Miao-hua. Vehicle detections based on YOLO_v2 in real-time[J]. China Mechanical Engineering, 2018, 29(15): 1869-1874. (in Chinese)
- [28] 刘 军,后士浩,张 凯,等.基于增强Tiny YOLOV3算法的车辆实时检测与跟踪[J]. 农业工程学报, 2019, 35(8): 118-125.
- LIU Jun, HOU Shi-hao, ZHANG Kai, et al. Real-time vehicle detection and tracking based on enhanced Tiny YOLOV3 algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(8): 118-125. (in Chinese)
- [29] SRI J S, ESTHER R P. Little YOLO-SPP: a delicate real-time vehicle detection algorithm[J]. Optik, 2021, 225: 165818.
- [30] 柳长源,王 琪,毕晓君.多目标小尺度车辆目标检测方法的研究[J/OL]. 控制与决策, (2020-09-03)[2020-12-22]. <https://doi.org/10.13195/j.kzyjc.2020.0635>.
- LIU Chang-yuan, WANG Qi, BI Xiao-juan. Research on multi-target and small-scale vehicle target detection method[J/OL]. Control and Decision, (2020-09-03)[2020-12-22]. <https://doi.org/10.13195/j.kzyjc.2020.0635>. (in Chinese)
- [31] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Springer. 13th European Conference on Computer Vision, 2014. Berlin: Springer, 2014: 740-755.
- [32] 王宇宁,庞智恒,袁德明.基于YOLO算法的车辆实时检测[J]. 武汉理工大学学报, 2016, 38(10): 42-46.
- WANG Yu-ning, PANG Zhi-heng, YUAN De-ming. Vehicle detection based on YOLO in real time[J]. Journal of Wuhan University of Technology, 2016, 38(10): 42-46. (in Chinese)