Computer Science and Application 计算机科学与应用, 2022, 12(4), 1099-1107
Published Online April 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.124113

Hans汉斯

基于路侧监控的车型分类算法研究

余　波[1]，杨　博[2]，李　健[2]，田　岩[3]

[1]济南金衢公路勘察设计研究有限公司，山东　济南
[2]华北科技学院，河北　廊坊
[3]华中科技大学，湖北　武汉

收稿日期：2022年3月17日；录用日期：2022年4月18日；发布日期：2022年4月25日

摘　要

交通场景中车辆的检测和分类是发展智能交通的应有之义。本文基于利旧的原则，利用路侧交通监控视频数据，设计改进的YOLOv3深度学习网络模型，使用残差单元以保证卷积神经网络收敛损失，算法采取多尺度特征融合预测的策略，直接在多个尺度的特征图上回归预测车辆边界框和类型，实现车辆的目标快速检测和车型判定。将改进前后的模型在测试集中进行测试，实验结果表明本文提出的深度学习网络在实际交通场景中既满足实时性的要求，又具有良好的车型检测和分类效果。

关键词

车辆检测，车型分类，监控视频

Research on Vehicle Classification Algorithm Based on Road Side Monitoring

Bo Yu[1], Bo Yang[2], Jian Li[2], Yan Tian[3]

[1]Jinan Jinqu Highway Survey and Design Research Co. LTD., Jinan Shandong
[2]North China Institute of Science and Technology, Langfang Hebei
[3]Huazhong University of Science and Technology, Wuhan Hubei

Received: Mar. 17th, 2022; accepted: Apr. 18th, 2022; published: Apr. 25th, 2022

Abstract

The detection and classification of vehicles in traffic scenes is an essential part of the development of intelligent transportation. In this paper, based on the principle of benefiting the old, an improved deep learning network is proposed to realize the detection of vehicles and the determina-

文章引用：余波，杨博，李健，田岩. 基于路侧监控的车型分类算法研究[J]. 计算机科学与应用，2022，12(4): 1099-1107. DOI: 10.12677/csa.2022.124113

tion of vehicle models by using the road measurement traffic surveillance video. Experimental results show that the proposed deep learning network has good detection and classification effects in real traffic scenes.

Keywords

Vehicle Detection, Vehicle Classification, Surveillance Video

Copyright © 2022 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

智能交通是交通运输信息化发展的战略目标，信息化是实现智能交通的重要手段，涵盖了人、车辆、道路等多维度信息提取，快速、准确地提取交通参数信息是智能交通系统正常运行的前提和保障。作为一种重要的交通参数，车型的分类研究得到了极大的关注。

传统的交通参数提取主要用到雷达[1] [2]、声波[3]、激光[4]、地感线圈[5] [6]等传感器设备，由于受限于设备安装、维护以及费效比等因素，传统的参数提取设备难以大范围普及。目前路侧的视频监控设施非常完备，其主要目的是为交通管理部门进行远程监测和提供实时路况信息，这些数量庞大的监控摄像数据，是交通参数提取和有效利用的重要应用场景[7] [8] [9] [10]。车型分类的前提是车辆目标的检测，目前广大专家学者已提出许多车辆检测方法，包括：基于背景建模的检测方法[11] [12] [13]，该类方法处理速度快，缺点是容易受到环境因素的扰动，对于交通场景中的静态目标检测效果不佳；基于统计学习的检测方法[14] [15]，该类方法预先训练手工特征目标分类器，采用多尺度滑动窗口搜索目标区域，能够对环境因素有效抑制，减少干扰，不受场景中物体阴影的影响，但处理速度较慢，且泛化能力较差；基于深度学习的检测方法[7] [8] [9] [10] [16]-[28]，随着大规模数据集的出现及计算机软硬件的升级，基于深度学习的目标检测方法的精度得到了快速提升，其中以 Yolo [23]为代表的回归模型在检测精度和速度上都占据了优势。

本文的主要工作聚焦于如何稳定、鲁棒地对交通场参数进行实时提取，基于利旧的原则，利用部署在道路龙门架上的原有交通监控视频数据，结合改进后的 YOLOv3 的深度学习模型，实现交通场景中车辆的实时检测和车型的分类。

2. 算法模型

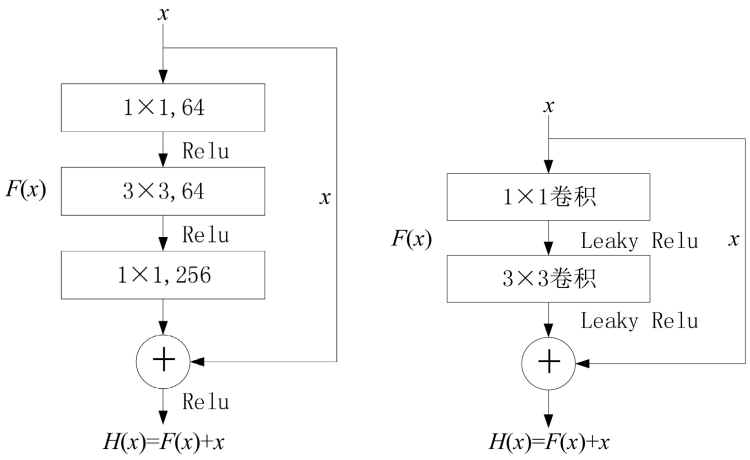
目前基于深度学习的目标检测算法主要分为两大类：一是基于区域提名的方法，因其主要包含两个过程，因此又称为两阶段方法，以 R-CNN (Regions with CNN features) [16] [20]为代表，基本思路为针对图像中目标物体位置，预先提出候选区域，再利用卷积神经网络提取图像深度特征并判断区域内物体类型，该类算法的检测精度普遍较高，但是耗时比较严重；另一种是基于端到端学习的方法，又称为单阶段方法，以 YOLO (You Only Look Once) [24] [25]和 SSD (Single Shot MultiBox Detector) [26] [28]为代表，其主要思路是均匀地在图片的不同位置进行密集抽样，抽样时可以采用不同尺度和长宽比，然后利用卷积神经网络提取特征并直接进行分类与边框回归，整个过程只需要一步，因此该类方法检测速度普遍较快。

本文的研究基于实时监控视频，对目标检测算法的实时性要求十分严苛。上述深度学习检测算法中只有 SSD 和 YOLO 系列能够达到实时检测，其中 YOLOv2 算法的检测精度与 SSD 相当，但检测速度却是后者的 2 倍，而 YOLOv3 是对 YOLOv2 的进一步改进，因此本文的算法模型建立在 YOLOv3 的基础上。

由于交通监控视频分辨率普遍较大，当归一化至 YOLOv3 所需的尺寸后，会使得图像中尺寸较小的摩托车、行人等目标变得更小，导致对这些交通场景中小目标的检测效果下降，因此本文针对交通道路监控场景，对 YOLOv3 算法进行改进，设计了一种更适用于交通道路监控场景的深度学习车辆检测与车型分类算法。

2.1. 网络结构

本文采用的特征提取网络为 Darknet53，包含 52 个卷积层和 1 个全连接层，借鉴了 ResNet 的残差学习的思想，在网络结构中大量使用残差单元，而且性能要比 ResNet101 和 ResNet152 两种深层残差网络更好。残差单元结构如图 1 所示。



(a) 标准残差单元　　(b) Darknet 53 中残差单元

Figure 1. Residual element structure
图 1. 残差单元结构

使用残差单元的目的在于保证卷积神经网络在很深的情况下仍然能够使损失收敛，从而使网络可以不断加深，表达出更好的特征并提升分类与定位的精度。残差单元中的 1 × 1 卷积核可以对特征图每个像素在不同通道上进行线性组合，在不改变特征图原有平面结构的情况下调节其深度，从而实现拓宽通道数的升维或减少参数的降维功能，也在一定程度上减少了网络计算量。

Darknet53 前面的 52 层是一个全卷积网络，执行下采样时没有使用池化层而是使用步长为 2 的卷积核，总共执行了 5 次，即对输入图像实现最多 32 倍下采样。本文的车辆检测算法在主网络的基础上又增加了 2 个更大尺度的卷积层，与原来 3 个卷积层组成 5 个不同尺度的特征金字塔，整个网络结构如图 2 所示。

第一个卷积层使用 32 个大小为 3 × 3 的卷积核过滤归一化至 416 × 416 大小的彩色三通道输入图像，得到 416 × 416 × 32 的特征图；接着用步长为 2 的 64 个 3 × 3 的卷积核对之前层的输出进行滤波，实现下采样操作，得到 208 × 208 × 64 的特征图；接下来依次执行残差单元数量为 1、2、8、8、4 的 5 组残差网络，每组残差网络之间使用步长为 2 的卷积核构成的卷积层执行下采样操作，分别输出 104 × 104 × 128、

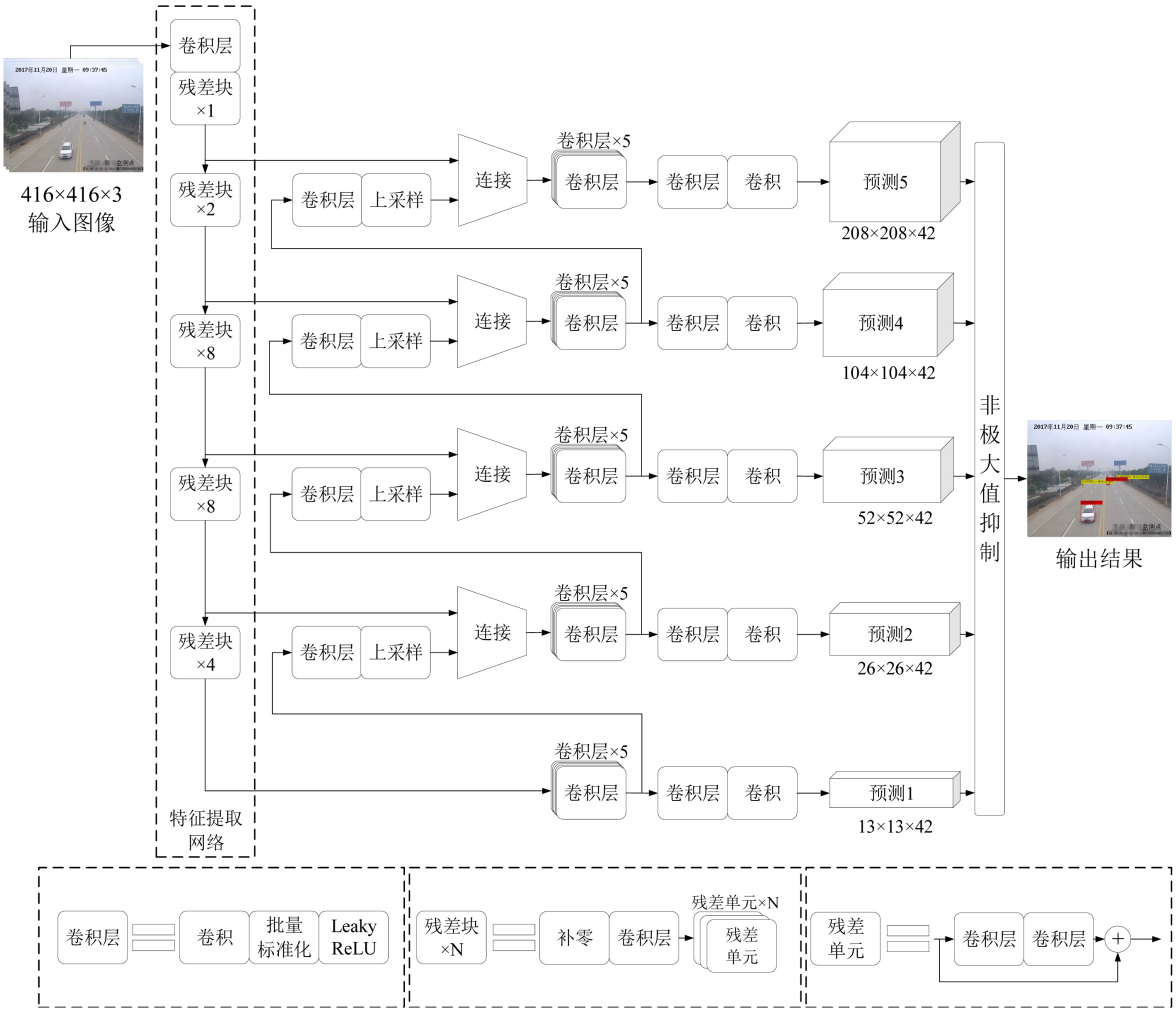


Figure 2. The network structure
图 2. 网络结构

52 × 52 × 256、26 × 26 × 512 和 13 × 13 × 1024 的特征图。此外，所有卷积层卷积后都执行批量标准化(Batch Normalization)操作以规范网络，加快收敛速度，同时用非线性函数 Leaky ReLU 进行激活。

特征金字塔的 5 个卷积层尺寸分别与特征提取网络中 5 种尺寸的特征图对应，这样做的目的在于充分利用特征提取网络获得的每一种尺寸的特征图，更多地考虑到图像的全局信息与局部信息，让网络同时学习深层特征和浅层特征，使得网络对图像具有更好的表达能力。

2.2. 锚盒机制

锚盒机制最早由 Faster R-CNN 提出，用于替代图像金字塔，实现对不同尺度与宽高比例的边界框的高效预测，同时降低网络模型训练复杂度，加快检测速度。本文检测算法延用 YOLOv3 中的锚盒机制，每种尺度特征图的网格预测 3 种不同尺寸的边界框，5 种尺度特征图共计 15 种尺寸的边界框，使用 K-means 聚类[28]的方法从训练数据集中确定了 15 个锚盒，即 15 种边界框的先验。聚类中的距离定义如下：

$$d(\text{box},\text{centroid}) = 1 - \text{IOU}(\text{box},\text{centroid}) \tag{1}$$

其中 IOU (Intersection over Union)为两个边界框交集与并集的比。适当的 IOU 可以使得网络模型在复杂度和召回率之间取得较好的平衡，本文选取的 IOU > 0.5 作为阈值。

2.3. 多尺度预测

本文检测算法采取多尺度特征融合预测的策略，直接在多个尺度的特征图上回归预测车辆边界框和类型。如图 2 所示，首先将最后一组残差网络输出 13 × 13 的特征图用于第一次预测；其次执行 2 倍上采样得到 26 × 26 的特征图，并与特征提取网络中相应尺寸的特征图进行拼接后进行第二次预测；再次继续进行上采样和拼接操作，最后总共执行 5 次不同尺度的特征融合和预测。

当使用一个尺度为 $N \times N$ 的特征图进行检测时，检测算法会将输入图像划分为 $N \times N$ 个网格(Cell)，每个网格包括的区域称为特征图的感受野，因此越小的特征图其感受野越大，蕴含更多全局信息和语义层次更高的特征；越大的特征图其感受野越小，包含更多的局部信息和特征。每个网格的任务是负责检测中心位于该网格内的目标，即预测目标边界框和目标类别，如图 3 所示。

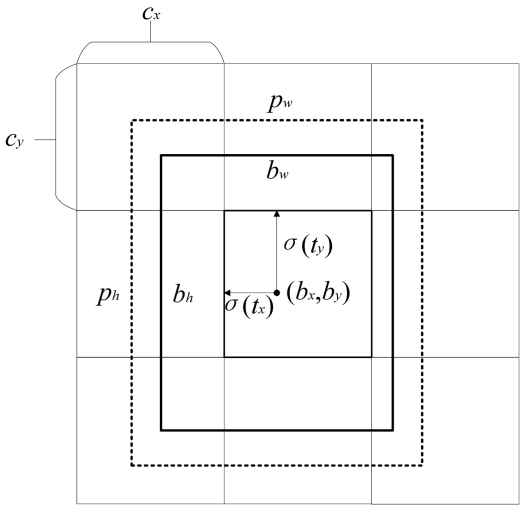


Figure 3. Bounding box prediction
图 3. 边界框预测

每个边界框需要预测 4 个参数(t_x, t_y, t_w, t_h)，记负责该边界框的网格左上角在全图的坐标为(c_x, c_y)，边界框对应的先验框宽为 p_w，高为 p_h，logistic 函数 $\sigma(t)$负责将坐标归一化到 0 到 1 之间，则边界框信息可以通过式(1)计算得出。训练时对边界框坐标的损失计算方式采用平方和距离误差损失。

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases} \tag{2}$$

检测算法对每个边界框通过逻辑回归预测一个目标的得分，若预测的边界框与真实的边界框大部分重叠且比其它所有的预测要好，则目标得分为 1。若重叠率未达到阈值 0.5，则该边界框将会被忽略，也就是会显示为没有损失值。

对于目标类别，每个边界框使用多标记分类来预测框中可能包含的类，分类时单独的逻辑分类器代替 Softmax，并在训练过程中采用二元交叉熵损失。

每种尺度的特征图划分的网格预测 3 种不同尺寸的边界框，每个边界框预测 4 个坐标、1 个目标得分和 C 个分类得分，因此每种尺度的特征图预测的张量为 $N \times N \times [3 \times (4 + 1 + C)]$。以本文面向的交通场景为例，类别数量为 9，因此对一幅图需要预测$(13 \times 13 + 26 \times 26 + 52 \times 52 + 104 \times 104 + 208 \times 208) \times 3 = 172,887$ 个边界框，预测总张量为$(13 \times 13 + 26 \times 26 + 52 \times 52 + 104 \times 104 + 208 \times 208) \times 3 \times (4 + 1 + 9) = 2,420,418$。最后采用非极大值抑制(NMS)检查同类型里高度重叠的边界框并丢弃除最高置信度以外的其它预测框。

3. 实验结果与分析

3.1. 实验环境

实验的硬件环境为 PC 机，CPU 为 IntelCore i7-7700K4 核处理器，内存为 16 GB，GPU 为 NVIDIA GeForce GTX 1080，显存为 8GB，软件环境为 64 位 Windows10 操作系统，CUDA 版本为 8.0，cuDNN 版本为 8.0，深度学习框架为 Darknet。

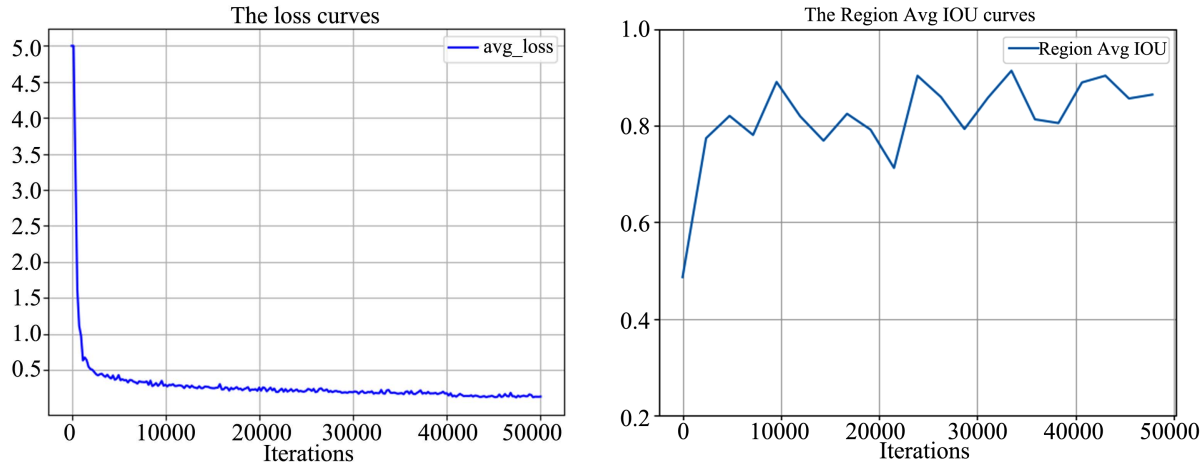
3.2. 实验方法

本文使用的数据集图像样本尺寸有 1920 × 1080、1280 × 960 等尺寸，在训练时统一归一化至 416 × 416。通过 K-means 聚类算法在训练集中产生了 15 个锚盒，大小分别为(7 × 13)，(11 × 18)，(14 × 28)，(21 × 35)，(19 × 52)，(30 × 53)，(31 × 70)，(41 × 81)，(46 × 116)，(65 × 97)，(58 × 152)，(67 × 196)，(136 × 143)，(96 × 253)，(145 × 266)。

这里的网络模型在训练时进行了 50,200 次迭代，批量(batch)大小设为 64，动量设为 0.9，衰减设为 0.0005，学习率初始为 0.001，依次在迭代 40,000 和 45,000 次时下调至原来的 0.1 倍。

3.3. 实验结果与分析

训练过程中的平均损失和平均 IOU 如图 4 所示。



(a) 平均损失　　(b) 平均 IOU

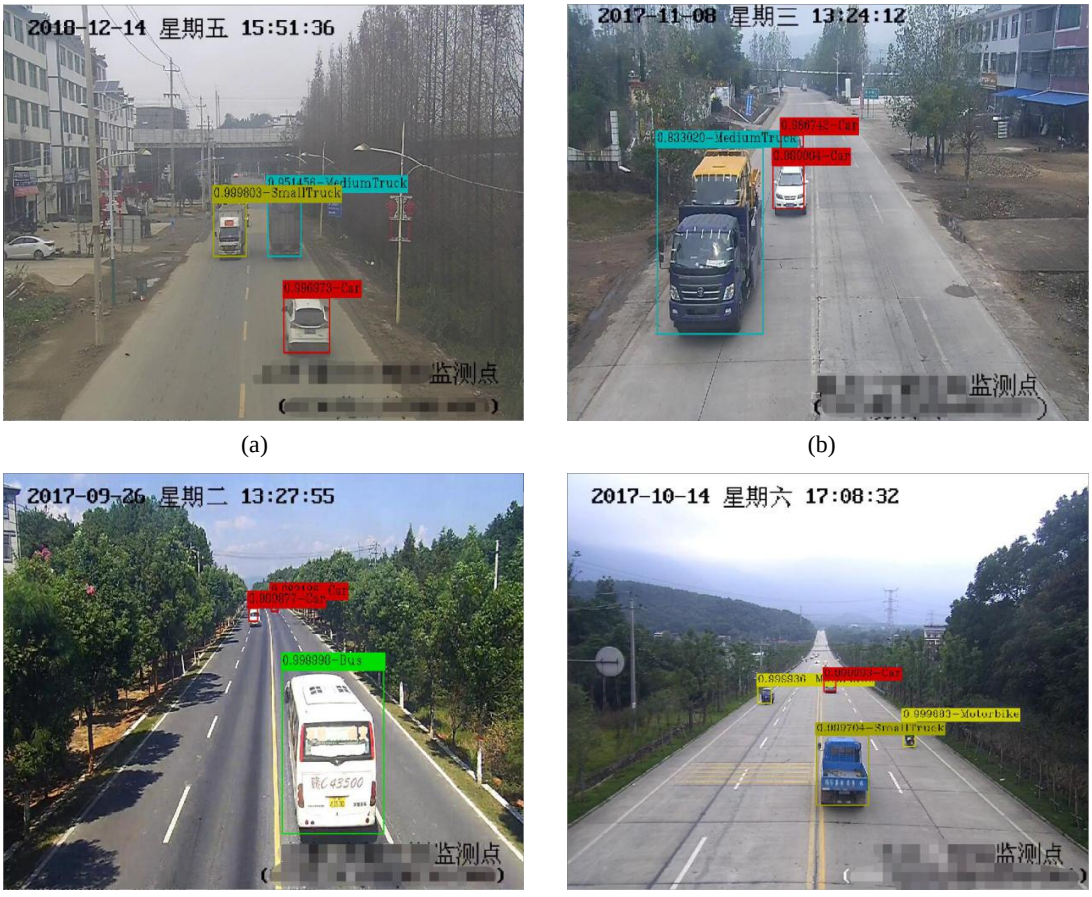
Figure 4. Training set average loss and average IOU
图 4. 训练集平均损失和平均 IOU

图 4(a)和图 4(b)分别显示了本文网络模型的平均损失在训练过程中的下降趋势和平均 IOU 的上升趋势。其中损失在前 2000 次迭代中下降很快，在迭代 30,000 次之后基本收敛，在迭代 40,000 次时由于学

习率自动调整为原来的 1/10，损失又有小范围的降低；IOU 在整个训练过程中整体上呈上升趋势，但波动范围较大，说明网络模型可以对目标边界框的主体进行较好的拟合，但是边界不够稳定，会在一定范围内波动。

本文将原版 YOLOv3 在本文训练集上进行训练，并将改进前后的模型在测试集中进行测试，效果对比如下：

如表 1 所示，在输入图像尺寸相同的情况下，本文方法在测试集上的 mAP (Mean Average Precision) 达到了 83.24%，比原版 YOLOv3 的 82.54%有 0.7 个百分点的提升，同时本文方法的平均 IOU 达到 73.92%，比原版 YOLOv3 的 72.15%提高了 1.77 个百分点，说明本文提出的目标检测与分类方法在分类平均精度和定位平均准确度方面均优于原版 YOLOv3。此外，本文方法的检测速度为 35.51 帧/秒，虽然低于原版的 38.92 帧/秒，但是仍可以满足实时性需求，本文检测算法在实际监控视频中的检测效果如图 5 所示。



(a) (b)

(c) (d)

Figure 5. This paper detects the algorithm effect diagram
图 5. 本文检测算法效果图

Table 1. The detection results of the improved methods on the test set in this paper
表 1. 改进前后的方法在本文测试集上的检测结果

方法	输入	Fps	mAP/%	Ave IOU/%
YOLOv3	416 × 416	38.92	82.54	72.15
本文方法	416 × 416	35.51	83.24	73.92

图 5 中每个车辆目标对应一个矩形边界框，边界框左上角显示了该目标的分类置信度和英文类标签，其中类标签包括：Car (中小客车)、SmallTruck (小型货车)、Bus (大客车)、MediumTruck (中型货车)、LargeTruck (大型货车)、HugeTruck (特大型货车)、ContainerTruck (集装箱车)、Motorbike (摩托车)以及 Person (行人)。

综合以上实验结果，可以确认本文提出的算法，能够利用路侧摄像机的视频数据进行交通参数的提取，分类精度与定位精度相较于 YOLOv3 都有提升，且依然能够满足道路交通参数提取系统所需的实时性。

4. 结论

本文针对交通场景中的车型分类问题，基于交通监控视频，设计并实现了一种基于 YOLOv3 的改进型车辆检测和分类方法。本文设计的方法通过将深度特征图与浅层特征图进行融合形成多尺度预测，充分利用了特征提取网络中每一种尺寸的特征图，增强了对图像特征的表达能力。本文还采取 K-means 聚类算法在训练集中自动生成锚盒，在提高网络模型定位精度的同时也加快了检测速度。实验结果表明，本文的检测和分类算法不仅具有很好的实时性，而且在车辆检测与分类精度方面优于传统的 YOLOv3 算法。该算法可以有效利用路侧原有监控设备进行实时交通参数提取，提高交通场景能效比，辅助智能交通发展。

基金项目

廊坊市科学技术研究与发展计划项目(编号：2021013071，2021011066)。

参考文献

[1] 郝兴伟. 基于车速传感器的雷达测速仪检定装置[D]: [硕士学位论文]. 长春: 吉林大学, 2014.

[2] 曹林, 李佳, 张鑫怡, 王东峰, 付冲. 一种基于微波雷达回波信号的车型分类方法[J]. 电讯技术, 2020, 60(5): 542-548.

[3] 周博, 马戎, 李岁劳, 等. 基于多普勒的车辆测速仪[J]. 机械与电子, 2014(2): 70-73.

[4] 蔡常青, 孙桥, 张跃, 等. 机动车激光测速仪校准技术的研究[J]. 计量学报, 2008, 29(4): 339-343.

[5] 叶青, 刘剑雄, 刘铮, 陈众, 李靓. 基于电磁感应的道路车辆车型在线分类方法研究[J]. 湖南大学学报(自然科学版), 2019, 46(12): 41-49.

[6] 王树欣, 伍湘彬. 地感线圈在交通控制领域中的应用[J]. 电子世界, 2005(8): 48.

[7] 石鑫, 赵池航, 林盛梅, 李彦伟, 薛善光, 钱子晨. 基于深度学习网络模型的车辆类型识别方法研究[J]. 筑路机械与施工机械化, 2020, 37(4): 67-73+78.

[8] 晏世武, 罗金良, 严庆. 基于卷积神经网络车型分类的研究[J]. 智能计算机与应用, 2020, 10(1): 67-70.

[9] 吕磊, 李文彬, 王晓鸣, 胡隆基. 基于深度卷积神经网络的小样本车型分类方法[J]. 兵器装备工程学报, 2020, 41(8): 193-200+221.

[10] 杜青, 苗艳华, 沈花玉. 基于机器视觉的车型自动分类算法设计[J]. 电子测试, 2019(1): 56-58.

[11] Lipton, A., Fujiyoshi, H. and Patil, R. (1998) Moving Target Classification and Tracking from Real-Time Video. *Proceedings of the* 1998 *DAPA Image Understanding Workshop* (*IUW*'98), Princeton, 19-21 October 1998, 8-14.

[12] Stauffer, C. and Grimson, W.E.L. (1999) Adaptive Background Mixture Models for Real-Time Tracking. *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, Vol. 2, 246-252.

[13] Barnich, O. and Droogenbroeck, M.V. (2009) ViBE: A Powerful Random Technique to Estimate the Background in Video Sequences. 2009 *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, 19-24 April 2009, 945-948. https://doi.org/10.1109/ICASSP.2009.4959741

[14] Viola, P. and Jones, M. (2001) Rapid Object Detection Using a Boosted Cascade of Simple Features. *Proceedings of the* 2001 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, 511.

[15] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, 20-25 June 2005, 886-893.

[16] Szegedy, C., Toshev, A. and Erhan, D. (2013) Deep Neural Networks for Object Detection. *Proceedings of the* 26*th International Conference on Neural Information Processing Systems*, Volume 2, 2553-2561.

[17] 吴乐平, 窦祥星. 基于 CNN 的车辆目标检测与车型分类研究[J]. 电子测试, 2021(6): 37-39.

[18] Girshick, R., Donahue, J., Darrell, T., *et al*. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81

[19] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[20] Ren, S., He, K., Girshick, R., *et al*. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Proceedings of the* 28*th International Conference on Neural Information Processing Systems*, Volume 1, 91-99.

[21] 李大湘, 王小雨, 刘颖. 监控视频中的车型分类方法[J]. 西安邮电大学学报, 2018, 23(4): 40-47.

[22] Dai, J., Li, Y., He, K., *et al*. (2016) R-FCN: Object Detection via Region-Based Fully Convolutional Networks. *Proceedings of the* 30*th International Conference on Neural Information Processing Systems*, Barcelona, 5-10 December 2016, 379-387.

[23] Redmon, J., Divvala, S., Girshick, R., *et al*. (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[24] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. *IEEE Conference on Computer Vision & Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525. https://doi.org/10.1109/CVPR.2017.690

[25] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement.

[26] Liu, W., Anguelov, D., Erhan, D., *et al*. (2016) SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision*, Amsterdam, 11-14 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[27] Fu, C.Y., Liu, W., Ranga, A., *et al*. (2017) DSSD: Deconvolutional Single Shot Detector.

[28] Hartigan, J.A. and Wong, M.A. (1979) Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, **28**, 100-108. https://doi.org/10.2307/2346830