

doi:10.11835/j.issn.1000-582X.2020.011

改进深度可分离卷积的 SSD 车型识别

郭 融, 王 芳, 刘 伟

(太原理工大学 电气与动力工程学院, 太原 030024)

摘要:针对现有车辆识别方法对于车型实时识别能力不足的问题,提出一种改进的深度可分离卷积的 SSD(single shot multibox detector)算法用于车型识别研究。首先,利用深度可分离卷积网络进行特征提取,并引入反残差模块来解决因通道数少、特征压缩导致的准确率下降问题。其次,以车辆的刚体特性为依据,重新设计区域候选框,减少模型参数运算量。最后,在 BIT-Vehicle 数据集上进行消融实验来对比不同网络模型性能差异。结果表明:改进的深度可分离卷积的 SSD 车型识别方法有更好的车型识别效果,可以达到 96.12% 的识别精度,检测速度提高至 0.078 s/帧。

关键词:车型识别;深度可分离卷积;反残差模块;区域候选框;卷积神经网络

中图分类号:TP391

文献标志码:A

文章编号:1000-582X(2021)06-043-07

Vehicle type recognition based on improved depthwise separable convolution SSD

GUO Rong, WANG Fang, LIU Wei

(College of Electrical and Power Engineering, Taiyuan University of Technology,
Taiyuan 030024, P. R. China)

Abstract: Aiming at the problem of insufficient real-time recognition capabilities of existing vehicle recognition methods, a single shot multibox detector (SSD) algorithm based on improved depthwise separable convolution is proposed for vehicle type recognition. Firstly, this paper proposes to extract the features using depthwise separable convolution network, and introduces the inverted residuals module to solve the problem of reduced accuracy due to the small number of channels and feature compression. Secondly, based on the rigid body characteristics of the vehicles, the region candidate frame is redesigned to reduce the amount of model parameter calculation. Finally, ablation experiments are performed on the BIT-Vehicle dataset to compare the performance differences of different network models. The results show that the improved depthwise separable convolution SSD vehicle type recognition method can achieve a recognition accuracy of 96.12%, and the detection speed increases to 0.078 s/frame.

Keywords: vehicle type recognition; depthwise separable convolution; inverted residuals model; regional candidate frame; convolutional neural network

收稿日期:2020-03-04 网络出版日期:2020-04-28

基金项目:山西省重点研发(高新领域)项目资助(201903D121132)。

Supported by Key Research and Development (High-tech) Projects of Shanxi Province (201903D121132).

作者简介:郭融(1995—),男,太原理工大学硕士研究生,主要从事视频图像处理与模式识别。

通讯作者:王芳,女,太原理工大学副教授,(E-mail)littlexiaoge@163.com。

随着人民生活水平的提高,汽车成了生活中人们常见的代步工具,在方便人们生活的同时,也造成了道路交通难以得到有效监管的问题。车型识别是电子不停车收费系统(electronic toll collection,ETC)的关键技术之一^[1],对道路上车辆的通畅行驶有重要的意义,在车辆违规行驶和被盜车辆追踪等方面也有很高的应用价值。

传统的车型识别方法依赖人工特征设计,例如利用梯度方向直方图特征进行特征提取^[2],提取车辆目标后使用支持向量机分类器进行分类,应用场景单一且易受背景、光照等因素影响。深度学习作为一种特征学习方法,在经过简单的非线性变换之后把原始数据变成更深层次、高语义的特征表达^[3-4],其特征是可以根据不同对象进行自适应权值调节,完成多种不同检测任务。

2015年,Ren等^[5]提出基于区域建议网络(region proposal networks,RPN)的Faster R-CNN目标检测算法,获得了相较传统方法更高的检测精度,但其检测速度较慢,只有5 fps。2016年,Redmon等^[6]在CVPR会议上提出YOLO(you only look once)目标检测算法,同年,Liu等^[7]在ECCV会议上提出了SSD目标检测算法。YOLO与SSD将预测边框的任务变成回归问题,在检测速度方面有明显提升,基本满足实时要求。深度学习目标检测算法已在车型识别领域有一定研究,文献[8]将Faster-RCNN分别与ZF,VGG-16以及ResNet-101 3种卷积神经网络相融合,可以针对不同的数据集选用最佳网络进行车型识别。由于该方法对候选框进行了预分类,虽然车型识别的精度有所提高,但是检测速度较慢。文献[9]融合循环神经网络和卷积神经网络嵌入二级框架,增强了内部特征信息的完整性,在对视频监控图像中车型识别的精确性上表现良好,但是其存在模型过大、运行时内存占用高的问题。

SSD目标检测算法在检测精度和速度方面综合性能更好,但仍存在参数多导致模型臃肿的缺点^[10]。结合上述研究,笔者提出使用改进的深度可分离卷积作为SSD算法基础网络,对特征提取过程进行简化,重新设计区域候选框,减少了模型参数量,实现对车辆类型准确的识别,并且提升了模型的检测速度。

1 经典 SSD 模型

经典SSD深度模型用VGG16分类网络作为特征提取器,并将VGG16后两层全连接层FC6,FC7改为卷积层,随后,SSD在一系列不同尺度的卷积层上做特征提取,其尺度分别为 38×38 , 19×19 , 10×10 , 5×5 , 3×3 和 1×1 。在融合不同尺度的特征信息后,对目标的位置和类别进行预测。

目标检测损失函数由位置损失和分类损失加权求和获得,如式(1)所示。

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)), \quad (1)$$

式中: N 是匹配正样本的总量; x 表示默认框与真实框的匹配结果, $x=0$ 表示匹配失败, $x=1$ 表示匹配成功; c 是softmax函数分别对任务中每一类别的置信度; l, g 分别是预测框和真实框; α 是位置损失的权重。

2 改进的 SSD 车型识别模型设计

经典SSD算法的模型参数运算繁多,运行时占用内存大,为了在有限的计算资源平台上完成图像识别任务并做出及时的反应,对SSD算法进行“减重”势在必行。

笔者提出改进的深度可分离卷积网络作为SSD前端特征提取的基础网络,避免卷积运算过程所带来的过大消耗,降低算法运算复杂度,适应未来实时识别车型的应用需求。

2.1 深度可分离卷积

深度可分离卷积(depthwise separable convolution,DSC)计算如式(2)所示。

$$G_{k,l,m} = \sum_{i,j} K_{i,j,m} \cdot F_{k+i-1,l+j-1,m}, \quad (2)$$

式中: G 为输出特征图; K 为卷积核; F 为输入特征图; i, j 为特征图像素位置; k, l 为输出特征图分辨率; m 为通道数。

深度可分离卷积把卷积分为深度卷积DW(depthwise convolution)和点卷积PW(pointwise convolution)2个阶段。DW是深度可分离卷积的滤波阶段,每个通道对应卷积核进行卷积操作;PW是深度可分离卷积的组合阶段,整合多个特征图信息,串联输出。

模型训练中,输入 M 通道大小为 $D_F \times D_F$ 的图像,经过 $D_K \times D_K$ 大小的卷积核进行卷积操作后,输出大小为 $D_I \times D_I$ 的 N 通道特征图。

对于一般卷积来说,滤波器大小为 $D_K \times D_K$,输入大小为 $D_F \times D_F$ 的 M 通道图像,参数运算量为 $D_F \times D_F \times D_K \times D_K \times M \times N$ 。而深度可分离卷积中特征提取只用了一个大小为 $D_K \times D_K \times 1$ 卷积核,运算量为 $D_F \times D_F \times M \times D_K \times D_K$,组合阶段用了 N 个点卷积,运算量为 $D_F \times D_F \times M \times N$,因此总计算量为 $D_F \times D_F \times M \times N + D_F \times D_F \times M \times D_K \times D_K$ 。

深度可分离卷积与传统卷积参数运算量对比计算如式(3)所示。

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

式中深度可分离卷积使用 3×3 的卷积核,即 D_K 取值为 3。深度可分离卷积在准确率较小下降的情况下运算量减少 $1/8 \sim 1/9$ ^[11]。在对模型轻量化过程中,引入了宽度参数 α ,压缩输入输出通道分别为 $\alpha M, \alpha N$, $\alpha < 1$,计算量为 $D_F \times D_F \times \alpha M \times \alpha N + D_F \times D_F \times \alpha M \times D_K \times D_K$ 。

2.2 反残差模块

宽度参数通过压缩通道数进一步减少模型参数量,但是也造成了特征压缩的问题。而修正线性单元(ReLU)对于负输入,全部输出为零,会导致更多的特征信息丢失,深度卷积 DW 无法改变输入通道,这种情况下,特征提取效果不理想,会导致目标检测准确率降低。

研究在 DW 之前引入反残差(inverted residuals, IR)^[12] 模块,实现过程如图 1 所示。

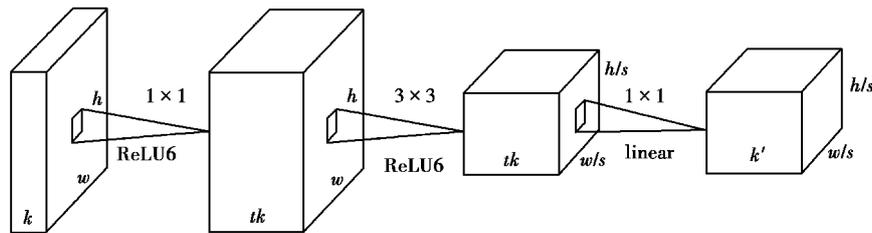


图 1 反残差模块

Fig. 1 Inverted residuals module

图 1 中,输入为 $h \times w$ 的 k 维特征图,将 ReLU 改为 ReLU6 以限制输出范围,输出 $h/s \times w/s$ 的 k' 维特征图。 s 是步长, t 为扩张系数。首先用一个大小为 1×1 的扩展核把输入的低维特征图扩展到高维,在高维空间进行深度卷积,最后使用点卷积将特征提取结果降维映射到低维空间并用线性激活函数(linear)进行输出。对通道数扩增之后再收缩,提高了模型特征提取的能力,有效防止由于通道数较小时因非线性激活导致特征信息丢失的情况,可以提高目标检测准确率。结合反残差模块的深度可分离卷积 SSD 结构如图 2 所示。

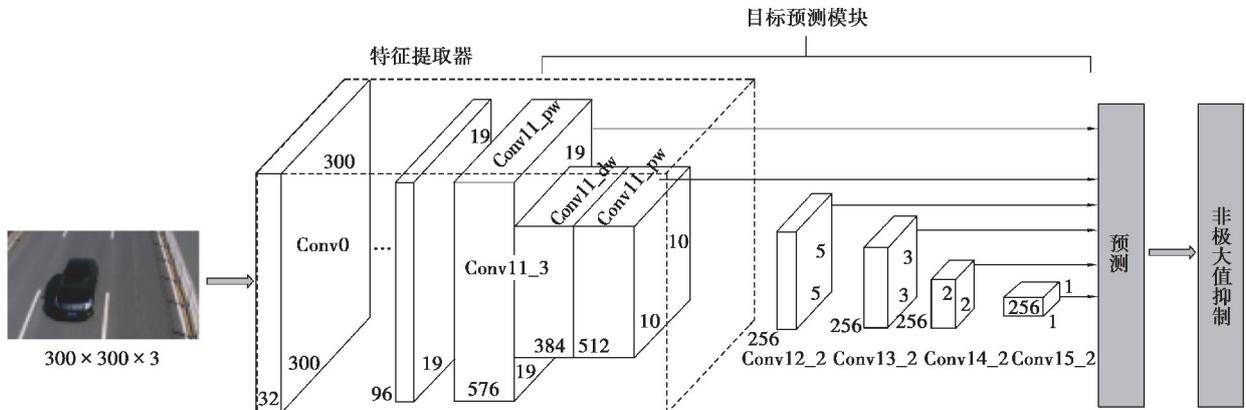


图 2 改进 SSD 模型结构图

Fig. 2 The model structure of improved SSD

图 2 中,特征提取部分除第 1 层之外全部为结合反残差模块的深度可分离卷积,扩张系数 t 为 6,并且在 conv1, conv3, conv6, conv11 中采用步长 s 为 2 的卷积,在 conv2, conv4, conv5, conv7, conv8, conv9, conv10 中步长 s 为 1,进行残差信息传递。模型选取包括特征提取器后两层在内的 6 种不同尺度的特征图,其尺度分别为 $19 \times 19, 10 \times 10, 5 \times 5, 3 \times 3, 2 \times 2$ 和 1×1 ,并计算目标的分类信息和位置信息。

2.3 区域候选框重构

图 3 为 SSD 多尺度特征图及区域候选框示意图,网络的输入为训练集所包含图像及手工标定的目标的真实标签框,如图(a)所示。在特征图中的每一单位点处,预测多个具有不同尺度、不同宽高比的候选框,如图(b)(c)所示,对候选框与真实框做匹配来进行位置损失的计算。区域候选框与目标真实位置交并比大小是区分正负样本的条件,训练效果是由区域候选框决定的。

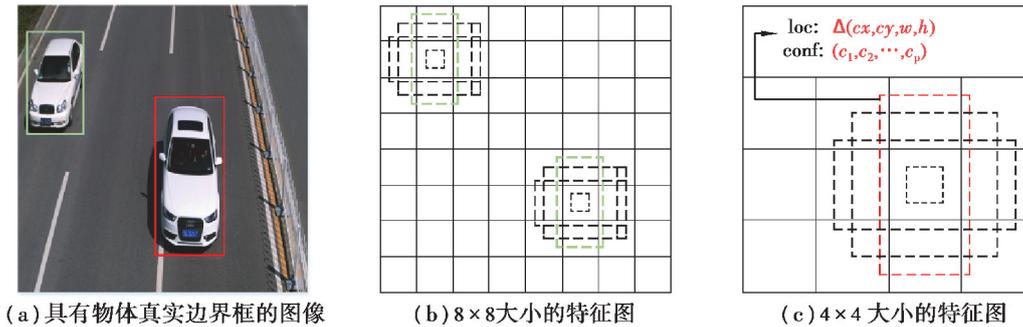


图 3 多尺度特征图及区域候选框示意图

Fig. 3 Sketch of multi-scale feature and regional candidate frame

区域候选框参数计算如式(4)所示。

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1}(k - 1), k \in [1, m], \quad (4)$$

式中: s_k 是第 k 个特征图的 minsize 参数; s_{\min} 是设计好的最小归一化尺寸,取值为 0.2; s_{\max} 为设计好的最大归一化尺寸,取值为 0.9; m 是特征图的总数。

SSD 区域候选框有不同的宽高比 a_r (aspect ratio),候选框的宽和高分别为 $w_k^a = s_k \sqrt{a_r}, h_k^a = s_k / \sqrt{a_r}$,对于宽高比为 1 的情况添加一个缩放为 $s'_k = \sqrt{s_k s_{k+1}}$ 的候选框,使得每个特征图有 6 个候选框,设定默认框中心为 $(\frac{i+0.5}{|f_k|}, \frac{j+0.5}{|f_k|})$, $|f_k|$ 为第 k 个正方形特征图大小, $i, j \in [0, |f_k|)$ 。

区域候选框与目标真实位置框的匹配度是模型性能重要的指标之一,而不同的区域候选框与真实框的匹配操作是十分复杂的。根据车辆的刚体特性,且图像采集于固定角度,对区域候选框重新设计,利用先验知识判定车辆宽高比在特定范围,针对不同采样角度、公路交通状况可统分为 $a_r > 1$ 或 $a_r < 1$ 的不同类,在参数设定时删除不需要的候选框,通过减少候选框匹配运算,提升 SSD 深度学习网络模型速度。实验车辆图片为路面车辆正面俯视图,宽高比参数 $a_r < 1$,将区域候选框取值范围选定为 $\{1, 1/2, 1/3\}$ 。

表 1 区域候选框重构前后总数对比

Table 1 Comparison of the number of regional candidate frame before and after improvement

实验算法	候选框数量生成计算	区域候选框总数
SSD	$38 \times 38 \times 4 + 19 \times 19 \times 6 + 10 \times 10 \times 6 + 5 \times 5 \times 6 + 3 \times 3 \times 4 + 1 \times 1 \times 4$	8 732
改进 SSD	$19 \times 19 \times 4 + 10 \times 10 \times 4 + 5 \times 5 \times 4 + 3 \times 3 \times 4 + 2 \times 2 \times 4 + 1 \times 1 \times 4$	2 000

SSD 采用 6 个特征层来生成不同尺寸大小的候选框,假设特征图大小为 $n \times n$,在每单元点设置 k 个候选框,则单层特征图生成候选框的个数为 $n \times n \times k$ 。笔者在 SSD 中使用反残差模块的深度可分离卷积进行特征提取,并且删除了宽高比例为 2 和 3 的框。由表 1 可以看出,经典 SSD 选取 6 个特征图的 k 值分别为 4, 6, 6, 6, 4 和 4,计算可以得到 8 732 个候选框;改进后的 SSD 选取 6 个特征图的 k 值均为 4,计算可以得到候选框 2 000 个,与原 SSD 的区域候选框生成总数相比减少了约 3/4。

3 实验分析

3.1 训练数据及平台

使用北京理工大学建立的 BIT-Vehicle Dataset^[13-14]数据集来验证改进算法有效性,数据集中的所有车辆都被分为 6 类:Bus、Microbus、Minivan、Sedan、SUV 和 Truck,每种车型分别有 558, 883, 476, 5 922, 1 392 和 822 辆。图像分辨率为 1 920×1 080 和 1 600×1 200。数据集按比例 6:2:2 分为训练集、验证集和测试集。

实验采用的软件和硬件设备:Linux 系统、GPU 采用 NVIDIA RTX 2060、Tensorflow、AMD Ryzen 52600 型号的 CPU、CUDA9.0 加 cuDNN7 版本深度学习加速库,python 版本为 3.6.5。算法设置每批次训练样本为 32 张,初始学习率设置为 0.001,迭代 80 000 步后设置为 0.000 1,在 Loss 变化差值基本不变,大概 150 000 步停止模型训练,得到车型识别模型。

3.2 实验结果及分析

实验选用 imageNet^[15]数据集上预训练的模型来初始化模型参数,可以加速模型收敛,获得更好的识别效果。笔者使用不同基础网络、不同区域候选框设计方法的目标检测算法进行消融实验,对比模型性能,验证改进算法有效性,结果如表 2 所示。

表 2 不同实验算法模型性能对比

Table 2 Performance comparison of different experimental algorithm models

实验算法	基础网络类型			区域候选框 重构应用	识别 精度/ %	参数量/ MB	平均检测 时间/ (s·帧 ⁻¹)
	VGG16	深度可分离 卷积	反残差 模块				
SSD	√				94.54	497.3	0.119
		√			93.69	28.1	0.083
		√	√		96.33	31.7	0.091
		√	√	√	96.12	25.0	0.078
Faster R-CNN	√				96.46	145.7	0.635

对表 2 分析可知,Faster R-CNN 分步目标检测算法精度较高,但其模型参数量大,平均检测时间为 0.635 s/帧,不满足实时性要求。表中加粗部分为笔者提出的改进深度可分离卷积的 SSD 车型识别方法所得结果,其识别精度达到 96.12%,相较于原 SSD 算法有显著提升,参数量减少 472 MB,平均检测时间由 0.119 s/帧下降到 0.078 s/帧。与区域候选框重构前的模型相比,在识别精度仅下降 0.21%的前提下,经过区域候选框重构后的模型参数量减少 21.14%,平均检测时间减少 14.29%,综合性能最优,可以实现对车型的实时识别。

图 4 为经典 SSD 和改进 SSD 的车型识别效果结果对比图,其中,第 1、第 3 列为经典 SSD 车型识别结果,第 2、第 4 列为改进 SSD 车型识别结果。通过识别结果对比可以看出,算法改进使得模型在边框的回归有更好的效果,准确率相较于经典深度模型也有一定的提升。

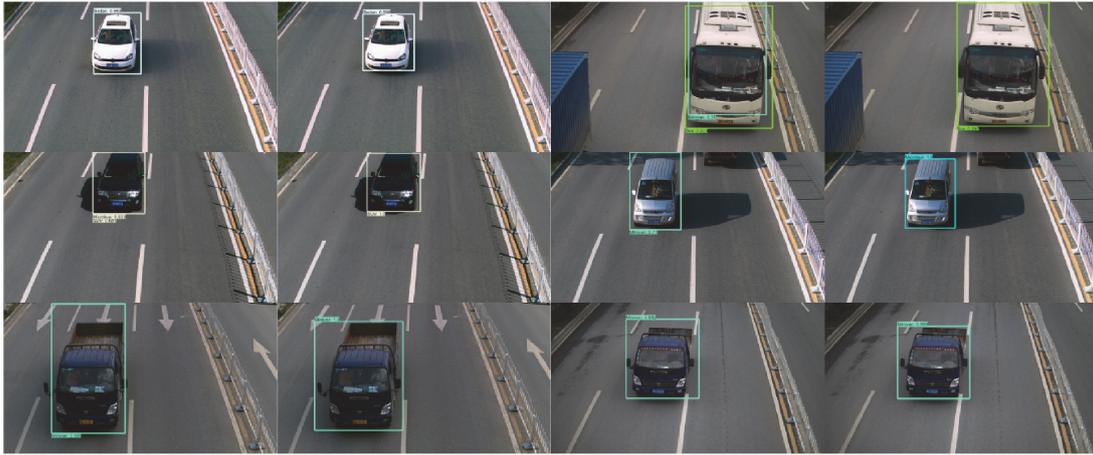


图 4 经典 SSD(第 1、第 3 列)与改进 SSD(第 2、第 4 列)车型识别结果对比

Fig. 4 Comparison of vehicle type recognition results between the classical SSD(the first column and the third column) and improved SSD(the second column and the fourth column)

4 结 语

针对目前车型实时识别能力不足的问题,笔者提出了基于改进深度可分离卷积的 SSD 模型对道路车辆类型进行识别。该算法大幅减少了模型参数量,识别精度达到 96.12%,同时平均检测时间达到 0.078 s/帧,各项指标均优于原 SSD 模型,满足车型实时识别要求,为交通场景中高速口缴费、车辆追踪及套牌检查等应用提供了参考。未来的研究方向是增加复杂场景道路图像数据,并对文中模型进行网络微调,提高模型在不同场景下车型识别能力。

参考文献:

- [1] 朱立峰, 赫连浩博. 电子不停车收费(ETC)专利技术综述[J]. 中国发明与专利, 2019, 16(S2): 95-99.
Zhu L F, Helian H B. Patent survey on electronic toll collection(ETC) technologies[J]. China Invention & Patent, 2019, 16(S2): 95-99. (in Chinese)
- [2] 刘操, 郑宏, 黎曦, 等. 基于多通道融合 HOG 特征的全天候运动车辆检测方法[J]. 武汉大学学报·信息科学版, 2015, 40(8): 1048-1053.
Liu C, Zheng H, Li X, et al. A method of moving vehicle detection in all-weather based on melted multi-channel HOG feature[J]. Geomatics and Information Science of Wuhan University, 2015, 40(8): 1048-1053. (in Chinese)
- [3] Miglani A, Kumar N. Deep learning models for traffic flow prediction in autonomous vehicles: a review, solutions, and challenges[J]. Vehicular Communications, 2019, 20: 100184.
- [4] 余成波, 田桐, 熊递恩, 等. 中心损失与 Softmax 损失联合监督下的人脸识别[J]. 重庆大学学报, 2018, 41(5): 92-100.
Yu C B, Tian T, Xiong D E, et al. Joint supervision of center loss and softmax loss for face recognition[J]. Journal of Chongqing University, 2018, 41(5): 92-100. (in Chinese)
- [5] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE; 1137-1149.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 779-788.

(下转第 83 页)

- gear[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2019, 233(2): 578-592.
- [9] Deng X Q, Wang S S, Hammi Y, et al. A combined experimental and computational study of lubrication mechanism of high precision reducer adopting a worm gear drive with complicated space surface contact[J]. Tribology International, 2020, 146: 106261.
- [10] 马振群, 王小椿, 沈兵. 对称弧形齿线圆柱齿轮的真实齿面接触分析研究[J]. 西安交通大学学报, 2005, 39(7): 722-725, 761.
Ma Z Q, Wang X C, Shen B. Real tooth contact analysis of the cylindrical gears with symmetrical arcuate tooth trace[J]. Journal of Xi'an Jiaotong University, 2005, 39(7): 722-725, 761. (in Chinese)
- [11] Deng X Q, Wang J L, Wang J G, et al. Parametric analysis of the end face engagement worm gear[J]. Chinese Journal of Mechanical Engineering, 2015, 28(6): 1177-1185.

(编辑 罗 敏)

~~~~~  
**(上接第 48 页)**

- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]// Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [8] 桑军, 郭沛, 项志立, 等. Faster-RCNN 的车型识别分析[J]. 重庆大学学报, 2017, 40(7): 32-36.  
Sang J, Guo P, Xiang Z L, et al. Vehicle detection based on faster-RCNN[J]. Journal of Chongqing University, 2017, 40(7): 32-36. (in Chinese)
- [9] 陈立潮, 卜楠, 潘理虎, 等. 基于改进卷积神经网络的车型识别[J]. 计算机工程与设计, 2019, 40(11): 3331-3336, 3348.  
Chen L C, Bu N, Pan L H, et al. Vehicle identification based on improved convolutional neural network[J]. Computer Engineering and Design, 2019, 40(11): 3331-3336, 3348. (in Chinese)
- [10] 吴天舒, 张志佳, 刘云鹏, 等. 基于改进 SSD 的轻量化小目标检测算法[J]. 红外与激光工程, 2018, 47(7): 0703005.  
Wu T S, Zhang Z J, Liu Y P, et al. A lightweight small object detection algorithm based on improved SSD[J]. Infrared and Laser Engineering, 2018, 47(7): 0703005. (in Chinese)
- [11] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-02-12]. <https://arxiv.org/abs/1704.04861>.
- [12] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 4510-4520.
- [13] Dong Z, Wu Y W, Pei M T, et al. Vehicle type classification using a semisupervised convolutional neural network[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4): 2247-2256.
- [14] Dong Z, Pei M T, He Y, et al. Vehicle type classification using unsupervised convolutional neural network[C]// 2014 22nd International Conference on Pattern Recognition, August 24-28, 2014, Stockholm, Sweden. IEEE, 2014: 172-177.
- [15] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

(编辑 张 苹)