• 重庆交通大学自动驾驶环境感知与定位技术专题 •

基于深度学习的目标检测算法综述*

曾文炳　李军

（重庆交通大学，重庆 400074）

【摘要】介绍了目标检测数据集的发展过程、基本评价指标的设定，并基于此综述了不同类别的目标检测算法，分别对两阶段和单阶段检测算法及相应优化算法进行解析，围绕检测速度和检测精度的迭代过程，阐述了目标检测算法的困难与挑战。最后，就算法本身的提升和算法应用需求下的优化设计提出总结和展望，指出目标检测的训练监督问题、算法对小目标的检测困难问题，同时指出实时检测任务中检测速度与检测精度的协调性问题和多模态融合应用问题，以及算法运行可解释性对算法再提升的重要意义。

关键词：目标检测算法　深度学习　计算机视觉　卷积神经网络

中图分类号：TP341　　**文献标志码：**A　　**DOI：** 10.20104/j.cnki.1674-6546.20230382

Review of Target Detection Algorithms Based on Deep Learning

Zeng Wenbing, Li Jun

(Chongqing Jiaotong University, Chongqing 400074)

【**Abstract**】This paper introduced the development of object detection datasets and the establishment of basic evaluation metrics, and based on this, it reviewed different categories of object detection algorithms. Single-stage and two-stage detection algorithms, as well as corresponding optimization algorithms, were analyzed separately. Highlighting the iterative process of detection speed and accuracy, the paper elaborated the challenges and difficulties in object detection algorithms. A summary and outlook for the improvement of the method itself and the optimization design under the application requirements of the algorithm were proposed in the paper, which indicated training supervision of object detection, the difficulty of detecting small targets by the algorithm. At the same time, the paper also indicated the coordination between detection speed and accuracy in real-time detection tasks and multimodal fusion application, as well as the important significance of the interpretability of algorithm operation for further improving the algorithm.

Key words: Target detection algorithm, Deep learning, Computer vision, Convolution neural network

【引用格式】曾文炳，李军. 基于深度学习的目标检测算法综述[J]. 汽车工程师，2024(1)：1-11.

ZENG W B, LI J. Review of Target Detection Algorithms Based on Deep Learning[J]. Automotive Engineer, 2024(1): 1-11.

1　前言

目标检测是计算机视觉的重要分支，主要用于识别和分析图像中的目标，即其应用方向分为一般目标检测和检测应用两类：一般目标检测以类人的角度参与认知，即模拟人类进行检测与分类，重在探索不同类型对象的检测方法；检测应用具有更为清晰的场景设定，如针对人员密集场所的人脸识别，或针对文件处理的文本检测等。近年来，随着卷积神经网络（Convolution Neural Network，CNN）等深度学习技术的广泛应用，目标检测技术突破传统检测器的技术瓶颈，由复杂向简单化、快速准确的方向发展[1]。

在两阶段（Two-Stage）目标检测算法方面，更快速区域卷积神经网络（Faster Region-based

*基金项目：重庆市研究生联合培养基地项目（JDLHPYJD2018003）。

Convolutional Neural Network,Faster R-CNN)和掩膜区域卷积神经网络(Mask Region-based Convolutional Neural Network,Mask R-CNN)等算法是最先进的方法之一。这些算法在准确率方面表现出色,但相对于单阶段算法,检测速度较慢。在单阶段(One-Stage)目标检测算法方面,YOLO(You Only Look Once)系列算法和单步多框目标检测(Single Shot multibox Detector,SSD)算法已成为最先进的算法之一。这些算法在速度和准确率方面均有很好的表现,且已经得到了广泛应用。

在深度学习算法大规模应用的背景下,目标检测算法已广泛应用于安检防控、自动驾驶以及卫星遥感等领域。各领域不同的检测需求催生了各种检测算法,依据检测过程有无区域建议,可分为两阶段检测器和单阶段检测器。两阶段检测器有区域建议过程,对检测对象的位置信息和边界信息有更为清晰的认知,导致检测精度普遍高于单阶段检测器;单阶段检测器检测过程更为简洁,故检测速度是其优势所在。

2 数据集与评价指标

目标检测算法中,从传统检测器到目前的多数检测算法均以有监督算法为主,数据集也随之发展,并完善了整套评价指标。

2.1 数据集

数据集作为有监督算法的"学习课本",对目标检测算法的训练至关重要。数据集的质量体现在标注的准确性,以及针对算法应用时类的丰富性或同类不同形态的整理完整性。数据集不仅限定了不同算法比较时有相同的初始训练标准,高质量数据集更在深度学习的背景下,对算法模型具有良好指引效果。同一算法采用不同数据集训练后会拥有不同的性能表现,这促使数据集随目标检测算法的进步而迅速发展。目前常用数据集主要有PASCAL VOC[2]和COCO数据集[3]。

PASCAL VOC数据集发源于PASCAL VOC挑战赛,主要包括图像分类(Object Classification)、目标检测(Object Detection)、目标分割(Object Segmentation)和行为识别(Action Classification)几类数据,发展到VOC 2007版本时,已经拥有20个类别的数据[4]。各类图片由官方提供明确的标注,同时,为了更好地发挥目标检测算法的性能,研究人员通常会综合多个版本的检测数据集,利用组合数据集完成目标检测算法的训练和测试。常见的训练与测试数据集组合如图1所示。

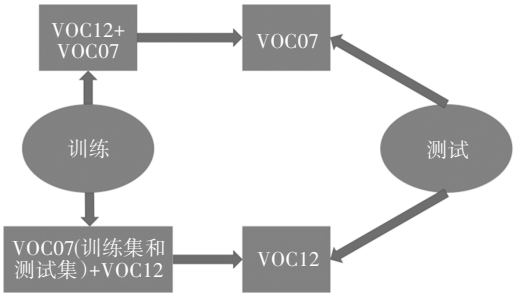


图1 常见的训练与测试数据集组合

COCO数据集由微软于2014年建立,其中的图像取自复杂的日常场景,其内容更丰富,图片类别达91类,图片数量多达32.8万张,其中带有250万个内容标注。与PASCAL VOC数据集相比,能训练出可识别更复杂场景的目标检测器,同时,针对更小的目标识别具有更深的理解。

2.2 评价指标

传统检测器在行人检测的应用研究中[5],以每个窗口的漏检率与误报率(False Positive Per Window,FPPW)作为检测器性能的度量标准。卷积神经网络的应用和检测器检测方法改变后[6],以平均检测精度(Average Precision,AP)作为同类检测准确性的评价指标,以类平均检测精度(mean Average Precision,mAP)作为不同类别间的平均AP,以表现目标检测算法检测精度的综合性能,并引入交并比(Intersection over Union,IoU)来描述目标检测算法的定位准确性。通过IoU阈值的设定判断对象是否成功定位,例如,IoU大于0.5时判定为定位准确,IoU为1时判定为预测对象位置完全正确。同时,以单位时间检测图像的数量表征目标检测的速度。

针对不同应用场景,评价指标可能各有侧重,表1所示为目标检测任务中常用的评价指标。

表1 目标检测常用的评价指标

评价指标	定义
准确率(Accuracy)	正确检测出的目标与实际存在的目标间的比例
精确率(Precision)	检测出的目标中真正为目标的比例
召回率(Recall)	实际存在的目标中被算法正确检测出的比例
F1分数(F1 Score)	精确率和召回率的调和平均值
平均精度(AP)	不同阈值下的精确率和召回率的平均值
漏检率(Miss Rate)	实际存在的目标中被算法漏检的比例
误检率(FPR)	将非目标物体错误地检测为目标的比例
平均漏检率(Average Miss Rate)	在不同目标大小下的漏检率的平均值

3 目标检测算法

目标检测算法因卷积神经网络的加入而注入活力,两阶段和单阶段的检测算法分别代表了研究人员在检测精度和检测速度方面做出的努力。图2所示为目标检测算法的发展及分类。

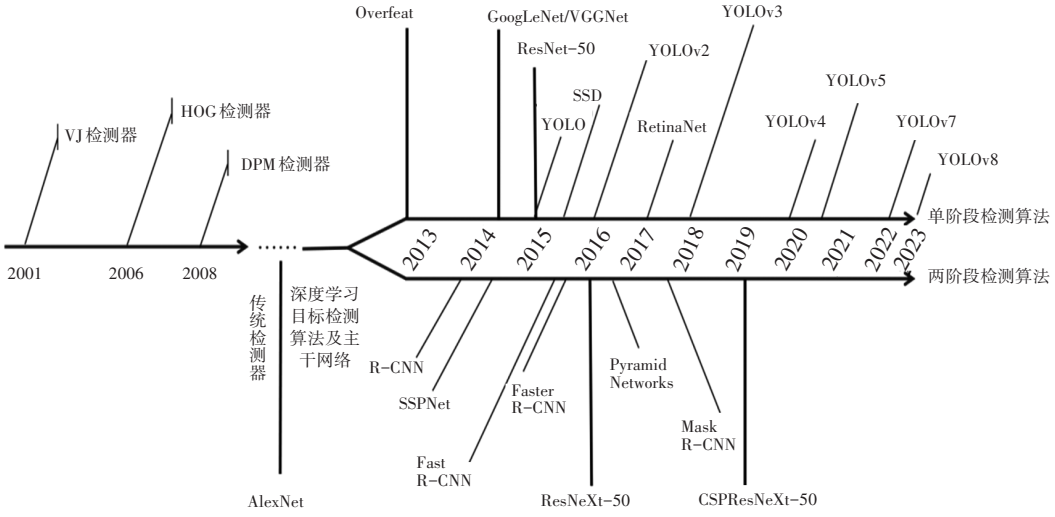


图2 目标检测算法的发展及分类

3.1 传统检测器

传统检测器通过人为设置检测对象的特征,在对象检测的初始阶段即存在一定的复杂性,并由于图像特征缺乏有效表达,特征的设计随描述对象的增多而愈加复杂,采用各种加速技巧仍未能遏制查找复杂特征对计算资源的消耗。同时,滑动窗口的检测机制虽然保证了一定的检测精度,但是巨量的检测滑框进一步提高了对计算能力的要求。

从Viola Jones检测器[7]的“整体图像”“特征选择”“检测级联”三大技术的有机融合,到定向梯度直方图(Histogram of Oriented Gradient,HOG)[8]利用重叠的局部对比度归一化,以及可变形组件模型(Deformable Part Model,DPM)检测器[9]“分而治之”的检测原则,分别表征了传统检测器在检测速度、检测输入图像尺寸多元化和检测精度,以及特征提取难度上做出的努力,并为之后的目标检测算法打下了坚实的基础。

3.2 两阶段检测算法

两阶段算法首先根据输入的图像生成候选框(Region Proposals),候选框的生成方法可分为选择性搜索和基于锚框的方法。选择性搜索首先对图像进行分割,得到一些小的区域,然后通过合并相邻的区域,得到一些更大的候选框,最后对这些候选框进行筛选,保留与目标物体较为相似的候选框。基于锚框的方法先在图像上生成一些固定大小和宽高比的锚框,然后通过卷积神经网络对每个锚框进行分类和回归,得到每个锚框的置信度和位置偏移量,最后根据置信度对锚框进行筛选,保留置信度较高的候选框。两种方法使用不同的逻辑生成区域建议,而后根据建议区域进行检测分类,如图3所示。因此,有两个步骤对输入图像进行处理,其余步骤能够提高目标定位的准确性,进而提高检测精度,但是其检测过程的复杂性导致该类算法检测速度受到影响。

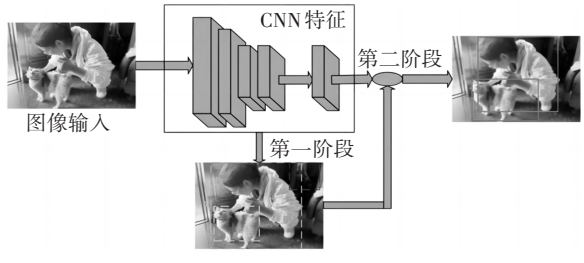


图3 两阶段算法结构表征

3.2.1 区域卷积神经网络

区域卷积神经网络(Region-based Convolutional Neural Network,R-CNN)[10]是第一个引入感兴趣区域(Region Of Interest,ROI)的目标检测算法,是机器学习时代里程碑式的检测算法之一。R-CNN采用选择性搜索(Selective Search,SS)算法生成候选框,如图4所示。通过对候选框进行图像裁剪和尺寸缩放获得固定大小的图像块,并对图像块进行卷积神经网络的特征提取和分类,得到每个图像块的置信度和位置偏移量,对置信度较高的图像块进行非极大值抑制,得到最终的检测结果。

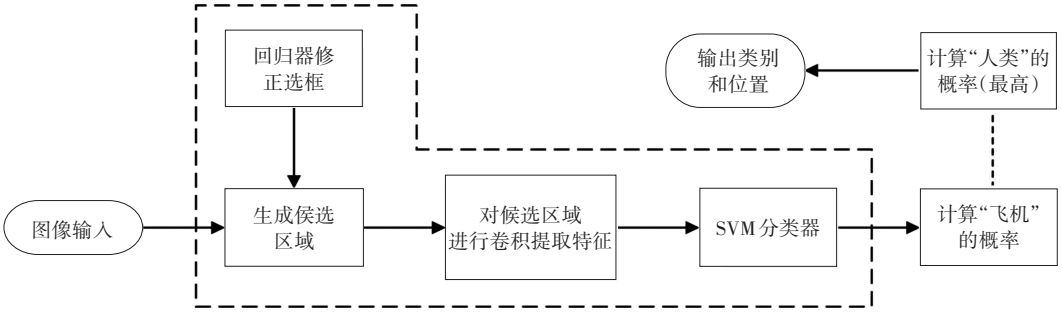


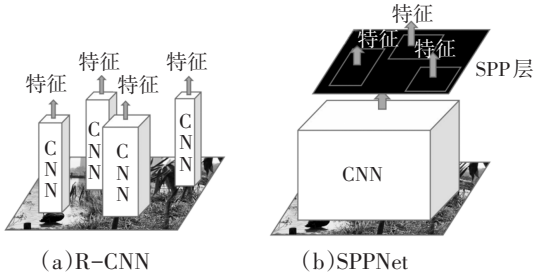
图4　R-CNN算法流程

OverFeat模型[11]同样采用卷积神经网络进行特征提取和分类,却与传统检测器一样通过滑动窗口的方式获得系列候选框,再进行分类和回归。

相较于滑动窗口所采取的穷举式检测方式,候选区域的引入大幅降低了计算消耗,并解决了固定窗口造成的检测不准确的情况,使检测速度得以突破,且在卷积神经网络的加持下,检测精度得以大幅提高。但是作为R-CNN的特色建议框,其大量重复成为制约检测算法进步的主要因素,特征的冗余计算严重影响了检测速度。这一问题在之后的金字塔池化网络(Spatial Pyramid Pooling Networks,SPPNet)中得以解决。

3.2.2　金字塔池化网络

SPPNet[12]用于改进R-CNN因对图像进行2 000个大小不一的候选区域分割而造成的区域特征提取进程缓慢问题。首先,SPPNet直接对图像进行不考虑候选区域的卷积操作,得到一次卷积的特征图,而后将2 000个候选区域在特征图上实现映射。相较于R-CNN对候选区域的逐一卷积提取特征,SPPNet从卷积次数上实现了检测速度的优化。图5所示为R-CNN和SPPNet的网络结构对比。



(a)R-CNN　(b)SPPNet

图5　R-CNN与SPPNet网络结构比较

其次,算法增加了金字塔池化(Spatail Pyramid Pooling,SPP)层,SPP层能将特征图转化为固定尺寸的特征向量,其将每个候选区域分成4×4、2×2、1×1大小的3个子图,针对子图各区域进行最大池化处理,而后进行支持向量机(Support Vector Machine,SVM)分类操作,进一步优化检测速度,使其检测速度高于R-CNN算法检测速度20倍以上。

3.2.3　快速区域卷积神经网络

快速区域卷积神经网络(Fast Region-based Convolutional Neural Network,Fast R-CNN)[13]作为R-CNN的升级版本,是R. Girshick在2015年提出的综合R-CNN和SPPNet而改进的检测算法。首先,该算法沿用SPPNet的一次卷积操作,避免特征提取的冗余操作;其次,摒弃之前的SVM分类器改用归一化指数(Softmax)函数进行分类处理,并在网络末端并行不同连接层,实现在不提供额外特征存储空间的同时进行端到端(End-to-End)的多任务训练,以及分类结果和定位框的回归反馈;最后,作者设计感兴趣区域池化(Region Of Interest Pooling,ROI Pooling)板块,使得在特征图上不同尺寸的候选框在进入全连接层分类、回归之前池化为固定尺寸。图6所示为Fast R-CNN算法网络模型。该检测算法实现了检测精度不变的前提下,以超过200倍的检测速度优于R-CNN算法。

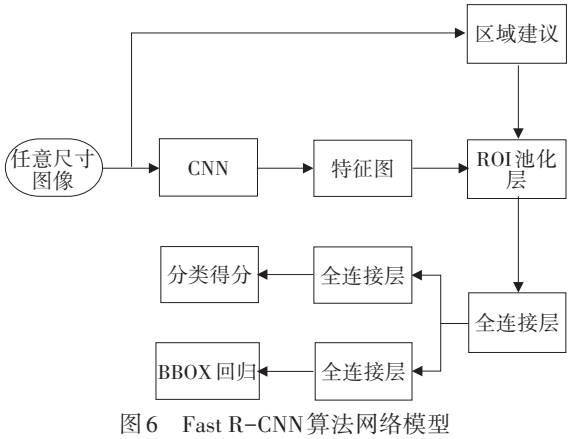


图6　Fast R-CNN算法网络模型

值得注意的是,R-CNN和Fast R-CNN都具有任意尺寸图像输入的功能。但是从尺寸修改的时机和对象来看,R-CNN在网络初始就对“图像”进行尺寸的重定义,而Fast R-CNN在感兴趣区域池化时对“特征图”进行尺寸修改。

3.2.4　Faster R-CNN

Faster R-CNN[14]解决了该系列算法一系列检测步骤上的问题，提出的区域建议网络（Regional Proposal Network，RPN）[15]取代了选择性搜索算法，解决了两阶段检测器不能实现端到端完成任务的问题，同时，区域建议算法相较于传统检测器性能有所提升。区域建议一直是深度学习算法性能受限的原因之一，直到Faster R-CNN解决了这一问题，实现了检测速度的大幅提升。

Faster R-CNN通过RPN将区域建议简化为二分类的过程，如图7所示：首先，生成不同尺寸的锚框[16]随滑动窗口移动；然后，依据设定的阈值对锚框是否含物体进行正负的标定；最后，由RPN整理后的二分类标签即为锚框所在坐标以及物体类概率的数据。端到端的训练显著提高了区域建议的质量，取代了之前广泛应用的区域建议而后进行逐一的完整检测，大幅提高了检测速度，实现了近乎零代价的区域建议。Faster R-CNN是一个将RPN和Fast R-CNN有机融合的整体。

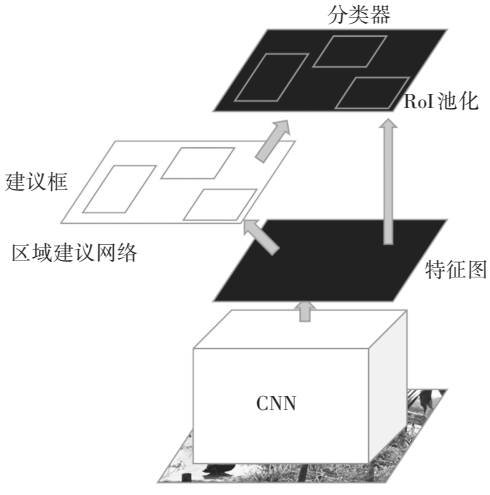


图7　Faster R-CNN算法网络结构

Faster R-CNN重点解决了前代算法检测过程复杂的问题，突破了检测速度的瓶颈，使实时检测在两阶段检测器上成为可能。但是由于RPN和ROI池化层缺少适应性的调整，基于锚框的区域建议在ROI池化层丧失了平移不变性，致使定位精度受到影响，以及锚框的尺度及大小固定，导致不在锚框范围内的物体检测失真，因而该检测算法不适用于所有目标的检测，尤其对小目标的检测性能缺失明显。

3.2.5　两阶段检测器总结

两阶段检测算法的流程中，第一阶段是确定检测候选区域，第二阶段对建议区域目标进行分类和回归定位。事实上，诸多检测算法也是围绕这两个阶段进行改进优化实现检测算法的性能升级。如RPN带来的高效区域建议获取效果，也有针对特征的高效利用算法，如采用级联架构的Cascade R-CNN[17-18]和使用特征金字塔[19]的Libra R-CNN[20]，都为特征的高效利用提供了可能。但是这些改进没有从本质上解决问题，反而通过增加模块的方法优化检测过程中的某一步骤，进一步增加了整个算法的复杂程度，使得网络在一定程度上捉襟见肘。各种优化均无法兼顾检测精度和检测速度，即便同时实现了检测精度和检测速度的提升，也将面临训练难度的提高。所以研究人员将目光转移到单阶段检测算法的研究中。

3.3　单阶段检测算法

单阶段检测算法通过简单的一次网络处理即可成功输出检测分类和预测框的边界，如图8所示。因此该类检测算法具备良好的检测速度，适合移动端使用，同时为算法模块的添加保留了足够的结构空间，用以实现检测应用的各种需求。

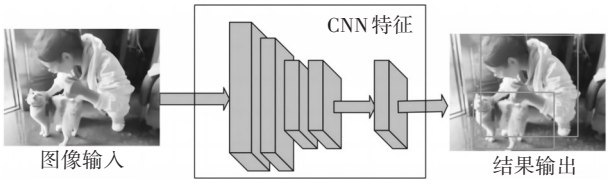


图8　单阶段结构表征

3.3.1　YOLOv1

一改两阶段算法“先选后测”的思路，2015年，R. Joseph等[21]带着“回归问题”的思路，提出了YOLO系列的初代算法。YOLO采用图像均匀分割的思路，对输入图像进行7×7的切割，针对每一个切割网格确定其中心落点，并对区域内目标进行检测。区别于RPN在高效解决区域建议效率低下问题的同时提高了网络训练成本，YOLO算法在图像切割中相当于生成了49个检测区域，数量少、检测效率高，固定的切割方式也不会导致网络训练成本的提高。

图9所示为YOLOv1网络框架：从输入的图像（尺寸为448×448×3）开始，分割出的网格由中间不同深度的卷积层和最大池化层处理，提取图像抽象特征；而后，由2个全连接层完成目标位置预测和分类概率计算；最后，输出尺寸为7×7×30的预测结果。YOLOv1凭借简单的结构以及端到端的图像处理方式，使检测速度满足实时检测需求，达到了同期最

快的45 帧/s,且YOLO快速版本拥有150 帧/s的图像处理速度。

但是YOLO算法检测目标的普适性有待提高,每个网格只能识别一个类别,粗放的定位落点对损失函数的影响较大,都决定了该算法针对小目标检测效果较差,同时对非常规目标泛化能力弱。

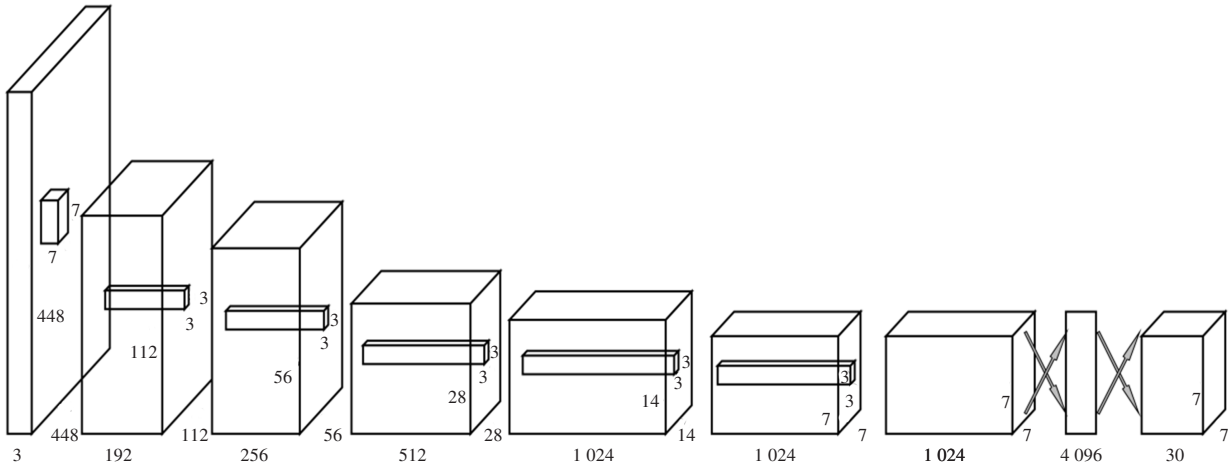


图9 YOLOv1网络结构

3.3.2 单步多框目标检测

由Liu等提出的SSD[22]是一种基于多基准和多分辨率的探测方法。与YOLO检测器不同,SSD在卷积阶段即进行检测,其结构如图10所示,SSD的卷积引入了多尺度的卷积操作,对不同尺度的目标可进行不同尺度的预测,很好地解决了目标尺寸造成的检测结果不理想问题。针对小目标,使用浅层的特征图解析以保留更多细节,针对大目标,使用深层特征图解析以深挖语义。但该操作也存在局限性:不同尺度特征检测重复,提高了检测计算难度;针对小尺寸目标,虽然浅层特征图卷积能保留更多细节,但浅层检测丢失了目标的语义信息,对小目标的检测优化效果不佳。

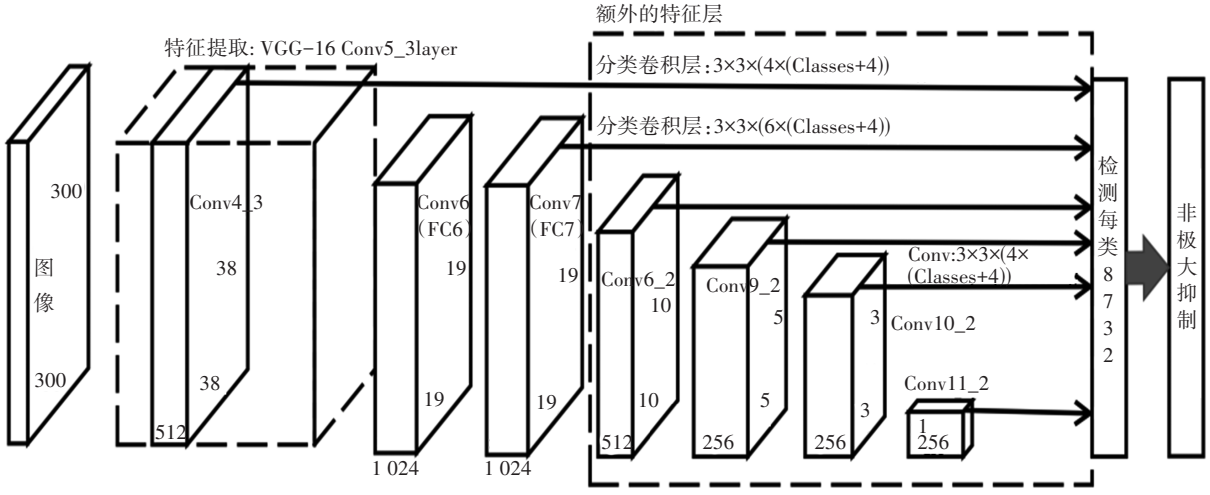


图10 SSD网络结构

3.3.3 YOLO系列算法总结

YOLO作为单阶段检测算法的代表,检测速度大幅提高,同时算法的网络结构简单,使其迭代升级成为可能,截至2022年底,YOLO系列算法已经从YOLOv1升级到YOLOv7。表2所示为YOLO系列算法的迭代升级过程,对检测算法的研究逐渐向网络架构、特征集成方法、检测方法、优化损失函数、标签分配方法以及高效的训练方法调整。

3.4 基于深度学习的经典检测算法总结

基于前文的分析可知,目标检测网络结构决定了检测算法的初始优势,如两阶段检测算法具有定位准确、检测精度高的特点,单阶段检测算法检测速度更快。然而,根据统一的评价指标,两种类型的检测算法都在弥补结构上的不足,向着更高精度和更快检测速度的目标改进。表3总结了基于深度学习的经典检测算法在统一的评价指标下的性能表现。

表2　YOLO系列算法整理

YOLO版本	提出时间	升级/改进	成绩表现	优点	缺点
YOLOv1[23]	2015年	回归处理目标检测赋予目标检测网络以简单的结构;网格划分图像降低网络训练成本	45帧/s VOC 2007数据集上的mAP为66.4% VOC 2012数据集上的mAP为57.9%	检测速度快;迁移能力强	群体检测效果差;小目标检测效果差;目标泛化性能弱(不常见角度)
YOLOv2[24]	2016年	主干网络由VGG16替换为DarkNet-19;批量归一化应用于卷积层;高分辨率分类网络预训练;引入聚类锚框机制;加入透传层融合特征	40帧/s VOC 2007数据集上的mAP为78.6% VOC 2012数据集上的mAP为73.4% COCO数据集上的mAP为21.6%	检测精度和速度较YOLOv1有所提高;模型泛化能力增强	密集检测效果差;主干网络较深,小目标的召回率不高
YOLOv3[25]	2018年	主干网络替换为DarkNet-53;引入残差网络ResNet;分类器由Softmax替换为逻辑(Logistic)分类;加入特征金字塔网络(FPN)结构	78帧/s COCO数据集上的mAP为33%	密集锚框提高召回能力、增强小目标检测能力	锚框尺度长宽比设计困难;锚框存在冗余问题
YOLOv4[26]	2020年	主干网络替换为CSP-DarkNet53;采用SPP+路径聚合网络(PAN)替代FPN;加入数据增强	66帧/s COCO数据集上的mAP为43.5%	精度提升明显;提升感受野;降低算法对计算机的要求	训练耗时增加;锚框尺寸固定,限制检测算法泛化能力
YOLOv5[27]	2020年	主干网络替换为聚焦(Focus)结构+跨阶段部分网络(CSPNet)的组合;特征融合(FPN+PAN+CSPNet结构);加权非极大值抑制(NMS);输入端马赛克(Mosaic)数据增强;定位损失采用完全交并比损失(CIoU Loss)	140帧/s	提升网络特征提取和特征融合能力;边界筛选清晰	数据增强使小目标进一步减小,造成检测困难,导致模型泛化能力弱
YOLOX[28]	2021年	主干网络加入Fcous结构;Yolo头部(Head)修改为解耦合头(Decoupled Head);引入动态样本匹配,即简化最优传输分配(SimOTA)和去除锚框操作	57.8帧/s COCO数据集上的mAP为51.2%	网络收敛速度提升、精度提高;缓解正、负样本不平衡;减少额外参数优化	增加计算量和内存消耗,SimOTA也会增加训练时间和计算量;去除锚框导致小目标检测效果不佳
YOLOv7[29]	2022年	动态标签分配策略;模块重参化	5~160帧/s,检测速度和精度均超越已知检测算法	可训练的赠品包在不增加计算成本的基础上提升检测准确性;通过“扩展”和“复合缩放”提高参数利用效率	增加了训练成本,“缩放”中提升主要来自固定的缩放因子,“复合”导致了更多参数量和计算量

表3　基于深度学习经典检测算法性能对比

类别	算法	主干网络	检测速度/帧·s^{-1}	图形处理器（GPU）	mAP/%		
					VOC 2007	VOC 2012	COCO（IoU∈[0.50,0.95]）
两阶段检测算法	R-CNN	AlexNet	0.03	Titan X	58.5（ILSVRC 2012+VOC 2007）		
		VGG16	0.5	Titan X	66.0（ILSVRC 2012+VOC 2007）		
	SPPNet	ZF-5	2	Titan X	59.2（ImageNet 2012）		
	Fast R-CNN	VGG16	3	K40	70.0（VOC 2007+VOC 2012）	68.4（VOC 2007+VOC 2012）	19.7
	Faster R-CNN	VGG16	7	Titan X	73.2（VOC 2007+VOC 2012）	70.4（VOC 2007+VOC 2012）	21.9
单阶段检测算法	YOLOv1	VGG16	45	Titan X	66.4（VOC 2007+VOC 2012）	57.9（VOC 2007+VOC 2012）	
	SSD300	VGG16	46	Titan X	74.3（VOC 2007+VOC 2012）	72.4（VOC 2007+VOC 2012）	23.2
	SSD512	VGG16	19	Titan X	76.8（VOC 2007+VOC 2012）	74.9（VOC 2007+VOC 2012）	26.8
	YOLOv3	DarkNet-53	78	Titan X			33.0

4　总结与展望

本文简要介绍了传统检测器发展中遇到的问题，分析了基于卷积神经网络的目标检测算法的发展历程，根据检测算法分类综述了具有代表意义的检测器，包括R-CNN系列算法、YOLO系列算法和SSD算法等。在目标检测算法发展过程中，通过模型结构改进（如增加卷积层、改变激活函数等）提高了模型的准确率和速度，通过数据增强（如旋转、翻转、裁剪等）[30]增加了数据样本的多样性并提高了模型的泛化能力，利用损失函数改进（如使用聚焦损失（Focal Loss）[31]等）提高了模型对难样本的处理能力，通过将多个目标检测器进行融合提高了目标检测的准确率和鲁棒性，通过对GPU、现场可编程门阵列（Field Programmable Gate Array，FPGA）等硬件进行优化提高了目标检测算法的速度和效率。在目标检测算法研究中，如何针对应用需求改进和创新检测算法仍值得思考，未来可能的研究方向包括以下几个方面：

a. 更快的检测速度。随着物联网和边缘计算的发展，越来越多的应用场景需要实时目标检测。因此，目标检测算法需要更快的速度，以满足实时性要求。未来的研究方向包括更轻量级的网络结构、更高效的计算方法、更优化的硬件设备等。

b. 更高的定位精度。目标检测算法的定位精度对于一些应用场景非常重要，如医学影像分析、工业质检等。未来可以通过研究更深、更宽的神经网络结构、更有效的特征提取方法、更精细的目标分类方法等实现定位精度和检测效率的提升。

c. 轻量级目标检测算法。从YOLOv3开始，首次体现了移动设备对轻量级算法的需求。检测速度和精度得以保证之始，应用级的需求就开始向更轻、更便捷的方向转变，如移动机器人的目标识别抓取[32]、农业应用等。尽管目前的检测算法在检测精度方面已经超越人类，但在细节和功能性上仍不及人眼，尤其是足够轻足够小的检测算法。

d. 弱监督检测[33]。从传统检测器开始，研究人员们就苦于对目标的特征标定，深度学习检测算法的训练仍旧大量依赖良好的图像内容标定。而图像标定过程费时费力，从成本和发展的角度考虑，都应该以弱监督甚至无检测为目标。减少或者部分使用边界框进行注释都将对检测器的运用灵活性带来极大帮助。

e. 小目标检测[34]。YOLO等结构简单的算法对小目标的检测失真，导致泛化能力低下，解决了小目标检测问题的算法检测速度较低而应用受限。

f. 视频检测[35]。针对连贯视频中的实时目标检测需求，如视频监控、自动驾驶[36]或用户上传视频审核，目标检测虽然可针对视频逐帧检测，但忽略了视频帧之间的连贯性，浪费了过多的计算资源却不能满足探测要求。针对连贯性对检测算法进行时间和空间连续的适应性调整，是算法应用性的改进

方向之一。

g. 目标检测的多模态融合[37]。自动驾驶汽车凭借多种传感器实现对驾驶环境的机器识别,目标检测不仅可以通过图像进行,还可以通过声音、红外线、雷达等多种传感器进行。将来自不同传感器的数据进行融合,从而实现对目标更加准确和全面地进行检测,可以使信息互补,实现误差校正、多样性增强以及实时性提高。

h. 目标检测的场景自适应[38]。目标检测算法在不同场景下的表现可能存在差异,通过自适应学习等方法,提高目标检测算法在不同场景下的准确率和鲁棒性,将极大提升目标检测在各行业中的应用泛化性能。

i. 算法运行可解释性。目标检测的可解释性包括解释其检测结果和决策过程,通过分析算法数据传递过程,能更好地理解和优化算法。算法可解释性的发展方向包括:可视化[39],即将算法的特征提取前向传递过程以图像或者其他形式呈现;可解释性[40],即通过算法中具有解释性的特征与目标检测建立联系,以方便用户理解算法特征与目标检测之间的关系;人机交互,即将算法结合人类知识和经验,利用人类对检测特征的深刻理解,更好地优化算法的特征处理过程。

j. 注意力机制[41]。检测算法区域划分、锚框选定等操作经常出现检测算法复杂性徒增的问题,注意力机制的引入能够大幅降低检测成本,高效关注图像有用信息,使现有检测算法检测性能快速提升成为可能。

参 考 文 献

[1] DHILLON A, VERMA G K. Convolutional Neural Network: A Review of Models, Methodologies and Applications to Object Detection[J]. Progress in Artificial Intelligence, 2020, 9(2): 85-112.

[2] SHETTY S. Application of Convolutional Neural Network for Image Classification on Pascal VOC Challenge 2012 Dataset[EB/OL]. (2016-07-13)[2023-10-25]. https://arxiv.org/abs/1607.03785.

[3] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context[C]// 13th European Conference on Computer Vision. Zurich, Switzerland: Springer International Publishing, 2014: 740-755.

[4] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, et al. The PASCAL Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2015, 111: 98-136.

[5] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005: 886-893.

[6] VIOLA P, JONES M. Rapid Object Detection Using a Boosted Cascade of Simple Features[C]// Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA: IEEE, 2001.

[7] VIOLA P, JONES M J. Robust Real-Time Face Detection [J]. International Journal of Computer Vision, 2004, 57: 137-154.

[8] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A Discriminatively Trained, Multiscale, Deformable Part Model[C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008: 1-8.

[9] VAN DE SANDE K E A, UIJLINGS J R R, GEVERS T, et al. Segmentation as Selective Search for Object Recognition [C]// 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 1879-1886.

[10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.

[11] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: Integrated Recognition, Localization and Detection Using Convolutional Networks[EB/OL]. (2014-02-24)[2023-10-25]. https://arxiv.org/abs/1312.6229.

[12] HE K M, ZHANG X Y, REN S Q, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[13] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1440-1448.

[14] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. Advances in Neural Information Processing Systems, 2015, 28.

[15] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2961-2969.

[16] CHEN Y H, LI W, SAKARIDIS C, et al. Domain Adaptive Faster R-CNN for Object Detection in the Wild[C]//

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, IEEE, 2018: 3339-3348.

[17] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into High Quality Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, IEEE, 2018: 6154-6162.

[18] 方钧婷, 谭晓阳. 注意力级联网络的金属表面缺陷检测算法[J]. 计算机科学与探索, 2021, 15(7): 1245-1254.
FANG J T, TAN X Y. Defect Detection of Metal Surface Based on Attention Cascade R-CNN[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(7): 1245-1254.

[19] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, IEEE, 2017: 2117-2125.

[20] GUO H Y, YANG X, WANG N N, et al. A Rotational Libra R-CNN Method for Ship Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(8): 5772-5781.

[21] PANG J M, CHEN K, SHI J P, et al. Libra R-CNN: Towards Balanced Learning for Object Detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 821-830.

[22] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// - 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer International Publishing, 2016: 21-37.

[23] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779-788.

[24] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 7263-7271.

[25] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. (2018-04-08)[2023-10-25]. https://arxiv.org/abs/1804.02767.

[26] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[EB/OL]. (2020-04-23) [2023-10-25]. https://arxiv. org/abs/2004.10934.

[27] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, BC, Canada: IEEE, 2021: 2778-2788.

[28] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. (2021-08-06)[2023-10-25]. https://arxiv.org/abs/2107.08430.

[29] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors[EB/OL]. (2022-06-06)[2023-10-25]. https://arxiv.org/abs/2207.02696.

[30] HAO X, LIU L, YANG R, et al. A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition[J]. Remote Sensing, 2023, 15(3): 827.

[31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2980-2988.

[32] LIU G, HU Y, CHEN Z, et al. Lightweight Object Detection Algorithm for Robots with Improved YOLOv5[J]. Engineering Applications of Artificial Intelligence, 2023, 123.

[33] ZHANG D, HAN J, CHENG G, et al. Weakly Supervised Object Localization and Detection: A Survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 5866-5885.

[34] LIU Y, SUN P, WERGELES N, et al. A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection[J]. Expert Systems with Applications, 2021, 172.

[35] SHI P, ZHANG C, XU S, et al. MT-Net: Fast Video Instance Lane Detection Based on Space Time Memory and Template Matching[J]. Journal of Visual Communication and Image Representation, 2023, 91.

[36] 高扬, 陈士伟, 刘进渊, 等. 基于深度学习的无人驾驶汽车车道跟随方法[J]. 汽车技术, 2022(3): 14-20.
GAO Y, CHEN S W, LIU J Y, et al. Lane Following Method for Autonomous Vehicles Based on Deep Learning[J]. Automotive Technology, 2022(3): 14-20.

[37] QIN L, SHI Y, HE Y, et al. ID-YOLO: Real-Time Salient Object Detection Based on the Driver's Fixation Region[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 15898-15908.

[38] OZA P, SINDAGI V A, SHARMINI V V, et al. Unsupervised Domain Adaptation of Object Detectors: A

Survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.

[39] LINARDATOS P, PAPASTEFANOPOULOS V, KOTSIANTIS S. Explainable AI: A Review of Machine Learning Interpretability Methods[J]. Entropy, 2020, 23(1): 18.

[40] ARRIETA A B, DÍAZ-RODRÍGUEZ N, DEL SER J, et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI[J]. Information Fusion, 2020, 58: 82-115.

[41] BORJI A, CHENG M M, HOU Q, et al. Salient Object Detection: A Survey[J]. Computational Visual Media, 2019, 5: 117-150.

（责任编辑　斛　畔）

修改稿收到日期为2023年10月25日。

《汽车技术》征稿启事

《汽车技术》杂志是中国第一汽车集团有限公司主办的国内外公开发行的汽车前瞻与应用技术类月刊，为我国高质量科技期刊分级目录入选期刊、中国科学引文数据库（CSCD）来源期刊、中文核心期刊、中国科技核心期刊、RCCSE中国核心学术期刊（A）、俄罗斯《文摘杂志》（AJ）收录期刊。

《汽车技术》杂志以报道汽车整车及其零部件设计、研究、试验等方面的前瞻与应用技术为主，并兼有理论研究内容，是中国汽车行业核心学术和知识传播与共享的平台。

《汽车技术》将在国家提出的“创新、协调、绿色、开放、共享”发展理念的指引下，把握《节能与新能源汽车技术路线图》和“低碳化、信息化、智能化”的汽车技术主流发展趋势，努力在传统内燃机汽车高效动力系统、轻量化、低阻力领域，新能源汽车和互联智能汽车技术领域，大力吸收优质稿源，为广大科研和工程技术人员服务，为我国汽车工程技术创新能力提升贡献力量。

《汽车技术》欢迎高等院校师生、研发工程技术人员、技术管理人员及相关人员不吝赐稿，反映国家重点扶持项目、自然科学基金项目和其他重点项目等研究成果的稿件将被优先选择刊登。

投稿要求：

1. 文章字数最好控制在6 000~8 000字范围之内；

2. 请按科技论文要求撰写文章摘要，摘要中文字数控制在180字左右；

3. 文章必须附有公开发表的、体现本领域最新研究成果的参考文献，且在文中应标注文献引用处；

4. 文章主要作者应提供其简介，包括出生年、性别、职称、学历、研究方向及技术成果等；

5. 来稿的保密审查工作由作者单位负责，确保署名无争议，文责自负；

6. 请勿一稿多投；

7. 本刊使用网站投稿，请先登陆网站注册成功后投稿，详细投稿要求见本刊网站中“下载中心”栏目的“作者指南”，网址：http://qcjs.cbpt.cnki.net。

《汽车技术》编辑部