自动驾驶汽车目标检测算法研究进展*

方文俊[1]　杨燕红[1,2,3]　王昊[1]

（1.西华大学汽车与交通学院，成都 610039；2.汽车测控与安全四川省重点实验室，成都 610039；
3.宜宾西华大学研究院，宜宾 644000）

【摘要】随着深度学习技术的引入，自动驾驶汽车目标检测算法取得了重要进展。从目标检测技术发展的角度，对传统目标检测算法与当前应用于自动驾驶车辆的深度学习目标检测算法进行了梳理，分析了具有里程碑意义的检测器、网络结构以及最新的检测方法，并针对目标检测技术的发展方向进行了探讨。

关键词：自动驾驶　环境感知　深度学习　目标检测　检测器

中图分类号：TP391.41　**文献标志码：**A　**DOI：**10.20104/j.cnki.1674-6546.20240375

Research Progress of Target Detection Algorithm for Autonomous Vehicle

Fang Wenjun[1], Yang Yanhong[1,2,3], Wang Hao[1]

(1. School of Automotive and Transportation, Xihua University, Chengdu 610039; 2. Vehicle Measurement, Control and Safety Key Laboratory of Sichuan Province, Chengdu 610039; 3. Yibin Institute in Xihua University, Yibin 644000)

【Abstract】With the introduction of deep learning technology in recent years, target detection algorithms for autonomous vehicle have made significant progress. This paper analyzes and organizes the traditional object detection algorithms and deep learning object detection algorithms currently applied in autonomous driving from the perspective of the development of object detection technology, analyzes milestone detectors, network structures and the latest detection methods, and explores the development direction of target detection technology.

Key words: Autonomous driving, Environmental perception, Deep learning, Target detection, Detector

【引用格式】方文俊, 杨燕红, 王昊. 自动驾驶汽车目标检测算法研究进展[J]. 汽车工程师, 2025(8): 1-14.

FANG W J, YANG Y H, WANG H. Research Progress of Target Detection Algorithm for Autonomous Vehicle[J]. Automotive Engineer, 2025(8): 1-14.

1 前言

以往，90%的车辆事故是人为失误造成的[1]，而自动驾驶汽车能够有效避免人工驾驶的很多安全问题。自动驾驶过程需要利用精确的定位系统、智能化环境感知技术，以及全面的控制算法来应对各种驾驶场景[2]，因此，目标检测技术在自动驾驶过程中具有重要作用。

准确检测周围车辆的位置、速度、行驶方向等关键信息，可使自动驾驶车辆预判潜在的碰撞风险，并及时作出合理决策[3]。早期，一般通过大量人工提取特征再识别的方法进行目标检测，而随着算法的深入与深度学习技术的出现，自动驾驶汽车使用摄像头和基于深度学习的方法进行

*基金项目：四川省科技厅人才项目（2023JDRC0016）；国家级大学生创新创业训练计划项目（202310623027）；西华大学研究生科创竞赛项目（YK20240238）；西华大学校级教育教学改革项目（xjjg2023092）。

通信作者：杨燕红（1970—），女，硕士研究生，正高级工程师，主要研究方向为新能源与智能汽车控制，yangyh@mail.xhu.edu.cn。

目标检测和分类[4]，极大地提高了车辆的实时检测水平。

本文针对自动驾驶汽车环境感知领域常见的目标检测算法和基于深度学习的目标检测算法进行梳理，分析了车辆目标检测中小尺度漏检、遮挡等问题的解决方案，最后总结目标检测算法面临的挑战及未来可能的发展方向。

2 车辆的目标检测算法

人类视觉系统在面对复杂场景时具有搜索和定位感兴趣目标的能力，可以很容易地注意到图像中的目标物体而忽略背景。这种视觉注意能力受认知生理和心理的控制，是人类处理视觉信息的重要机制[5]。为在车载场景中实现类似的感知能力，计算机视觉系统通过模拟人类注意机制，使自动驾驶车辆能精准识别道路中的目标并过滤冗余信息。计算机视觉识别的任务主要包括[6-7]：图像分类，即将图像分配到某个特定类别；目标检测，对图像进行处理、分析和理解，以识别目标类别，并通过边界框预测每个目标的位置；实例分割，识别不同的对象，并为每个对象分配一个单独的分类像素级掩码，即识别并分割出图像中的每一个特定实例对象；语义分割，将图像分割为多个具有语义的区域，为每个像素分配特定的类别标签，从而提供对图像更丰富的理解。

目标检测是计算机视觉任务的核心难点，且在自动驾驶系统中扮演着关键角色，其技术的发展大致经历了传统目标检测和基于深度学习的目标检测两个阶段[8]。

2.1 传统目标检测算法

传统目标检测方法通常分为3个步骤，即区域提议、特征提取、分类器训练，如图1所示。其中，特征提取可以强调主要信息并降低次要信息的影响，分类器的优劣将极大影响目标检测的准确率。大量研究提出了多种特征提取方法，如方向梯度直方图（Histogram of Oriented Gradient，HOG）[9]、光流法[10]、可变形部件模型（Deformable Part Model，DPM）[11-12]、哈尔（Haar）特征[13]、尺度不变特征变换（Scale-Invariant Feature Transform，SIFT）[14]等，并提出了多种分类器，如支持向量机（Support Vector Machine，SVM）[15]、引导聚集算法（Bootstrap aggregating，Bagging）[16]、非线性SVM、AdaBoost[17]等。

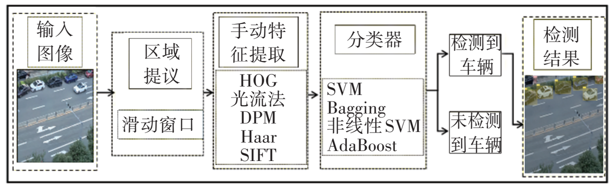

图1 传统目标检测算法的一般步骤

2.1.1 特征提取方法

HOG是Dalal等提出的一种用于行人检测的特征描述子。由于梯度或者边缘方向的分布可有效表达局部物体的外观或形状，所以在HOG中，梯度方向的直方图被视为特征。该方法将图像分为若干个单元（Cell），分别计算每个单元中每个像素的梯度和方向并构建直方图，得到多维特征向量，再将若干个单元组合成块，对块进行归一化，对块内单元的特征向量进行拼接得到最终的特征向量。因其在图像的局部区域操作，所以能够很好地避免图像的几何和光学形变。在此基础上，Tian[18]利用HOG特征和SVM开展夜间车辆检测，在夜间弱光条件下有效增强了移动车辆的检测能力。文献[19]使用直方图扩展和灰度微分值方法（Gray-level Differential Value Method，GDVM）进行车辆检测，结果表明，传统方法在复杂环境下有效，但在雨天和弱光条件下效果不佳，故传统方法在处理显著环境变化时存在局限性。

HOG在处理遮挡重叠问题时存在一定困难，促使Felzenszwalb等提出了DPM。DPM是一种基于多尺度可变形零件模型混合的物体检测系统，DPM算法采用改进的HOG特征及SVM和滑动窗口（Sliding Windows）联合的检测思想对输入的图像进行特征提取，针对特定目标制作相应的激励模板，并在原始图像中计算得到该激励的效果图，再由激励的分布确定目标位置。虽然DPM具有直观简单、运算速度快、能适应变形等特点，但其性能略有不足、激励特征需要人为设计的特点大幅增加了工作量，所以DPM不具有普适性。后续研究[20]针对DPM性能不足的问题，基于颜色融合DPM，通过HSI色彩空间转换进行图像传输，提取色彩空间各通道的信息并训练DPM，再通过自适应融合方法将DPM融合，取得了良好的检测效果，实现了对传统DPM的优化。

特征提取技术发展过程中，研究人员发现传统的目标检测侧重于目标的静态特征提取，而目标的运动信息难以获取，因此光流法（Optical Flow）应运

而生。光流法基于图像序列在时间域上的变化获取上一帧与当前帧的关系,计算相邻帧之间对象的运动趋势[21]。结合光流法和颜色直方图的方法可以同时对车辆和天气进行检测,并处理复杂天气条件下所需要的细微识别任务。由于光流法受光照噪声影响严重,且车辆检测通常在复杂的光照环境下进行,所以此方法在车辆检测中局限性较大。

受 Papageorgiou 等[22]使用的 Haar 基函数启发,Viola 和 Jones 提出使用积分图计算类哈尔(Haar-Like)特征,Haar-Like 特征是一种用于目标检测或识别的图像特征描述子,通过对图像中每个像素以及由像素组成的相邻矩形区域进行特定操作,来获取能够代表整体图像某些特征的描述信息。整体图像计算完成后,任意一个 Haar-Like 特征都可以在恒定时间内以任意尺度或位置进行计算。Haar-Like 特征分为三类:边缘特征、线性特征,以及中心特征和对角线特征。用两种矩形框联合生成特征模板,利用特征模板内两种矩形像素的差值表示该模版的特征值,特征模板可用于图像中的任一位置,且大小可任意变化,所以 Haar-Like 特征的取值对特征模板中的类别、位置、大小特别敏感,因此在固定大小的图像窗口内可以提取出大量的 Haar-Like 特征,从而大幅提高特征提取效率。

此后,研究人员发现,如果目标存在于不同尺度的图像中会导致之前的特征提取方法失效,SIFT 能够找到对尺度和旋转不变的特征,并进行提取以创建对部分遮挡、杂乱、噪声和照明变化具有鲁棒性的对象识别系统。虽然 SIFT 具有鲁棒性、局部性、可重复性等,但其计算成本高,且在极端情况下性能可能受到影响。

各类特征提取方法的特点如表 1 所示。从 HOG 到 SIFT 的发展过程明确地体现出特征提取方法从简单的基础功能实现逐步向复杂条件下高性能、普适性方向发展的趋势。但是,上述传统算法普遍存在的主要缺点是需要人为制作特征提取器来学习对象的不同描述子,导致工作量增加、时间延长。目标检测任务在从图像序列中生成局部特征后,提取的特征必须由机器学习分类器识别[23]。

2.1.2 分类器

分类器主要用于解决分类问题。最经典的分类器是 SVM,主要解决二分类问题,其核心思想是通过在特征空间中找到一个最优的超平面进行分类,且间隔最大。因需要的存储空间小、训练和分类速度快,SVM 的应用最广泛。针对非线性分类问题,非线性 SVM 将数据映射到高维特征空间,以便找到一个线性可分的超平面来分隔不同类别的数据点。非线性 SVM 对非线性数据的适应能力强,但其对大规模数据集的训练时间较长,且核函数的选择较为困难。为解决非线性 SVM 训练时间较长的问题,Bagging 基于自助采样(Bootstrap Sampling)和偏差-方差分解定理,并引入随机性与个体学习器相互独立的方法,减轻了过拟合问题并可实现并行训练,因此能够大幅度缩短训练时间。该方法首先将一个随机样本放入采集的集合中,再将其放回原本的数据集,重复 N 次采样后可获得一个大小为 N 的样本集合,重复 P 次,得到 P 个含有 N 个样本的采样集合,然后使用该采样集合训练出个体学习器并将其组合起来,最后根据不同的问题采用不同的方法进行预测,如针对回归问题预测平均值、针对分类问题预测多数投票或者平均概率。但是,Bagging 方法存在对异常值十分敏感、参数选择对结果影响较大的问题。

表 1 各类特征提取方法及其特点

名称	优点	缺点
HOG	能较好地捕捉局部形状信息,对几何和光学变化均具有很好的不变性	特征描述子获取过程复杂、维数较高,导致实时性差;很难处理遮挡问题;对噪点相当敏感
DPM	方法直观简单、运算速度快	工作量大
光流法	可准确检测识别运动目标	适用条件不易满足
哈尔特征	简单快捷、计算量小,受光照、旋转和尺度变换等影响小	难以描述图像的纹理和形状特征
SIFT	特征稳定,具有旋转、尺度变换、亮度不变性,可抗噪声	耗时长、实时性弱,内存占用大

随着技术发展和实际需求的变化,AdaBoost 基于 Boosting 集成学习,能够在每一轮迭代后更新样本和弱学习器的权重,并且在每一轮迭代的基础上以同样的方法再训练一个弱分类器,直至到达最大迭代次数或者获得足够小的误差后将所有弱学习器的输出整合为一个整体。AdaBoost 具有准确率高、能自动选择有效特征、灵活性强的优势,但其计算量较大,对异常值和噪声敏感。不同分类器的特

点如表2所示。

表2 各类分类器及其特点

分类器	优点	缺点
线性SVM	所需内存更小、训练和分类速度快	准确率较低
非线性SVM	准确率更高	训练和测试速度较慢,算法复杂
Bagging	可减轻过拟合问题,训练时间短	对异常值敏感,参数选择对结果影响大
AdaBoost	分类速度快、灵活性强、准确率高	训练慢,对异常值和噪声敏感,复杂性随类别数量增加而提高

2.2 基于深度学习技术的目标检测算法

传统目标检测方法大多具有明显的缺点,如依赖手动特征工程,此外,此类模型因无法在不需要大量重新训练的情况下快速纳入新数据,适应恶劣天气或动态环境的能力有限。深度学习在20世纪90年代作为机器学习和人工智能的一个子领域出现[24]。目标检测领域中传统算法对规模化处理特征能力不足,深度学习算法可以较好地弥补这些缺点[25]。基于深度学习的目标检测方法与传统目标检测方法在理论和应用方面存在显著区别,如深度学习可实现更高抽象级别、更高精度和更快执行的能力。近年来,在计算机视觉领域,由于数据量的不断增加和硬件的快速发展,基于深度学习的车辆检测算法的实时性已经可与传统的基于手动特征的车辆检测算法相媲美,而深度学习方法在检测精度方面远超传统算法。这一时期的里程碑检测器如图2所示。

2.2.1 深度学习网络算法的演进

2012 年,世界见证了卷积神经网络(Convolutional Neural Network, CNN)的重生,AlexNet[26]在ImageNet举办的大规模视觉识别挑战赛(ImageNet Large Scale Visual Recognition Challenge, ILSVRC)上首次使用了深度学习方法。该比赛使用ImageNet数据库,包含120×10^4张、1 000个类别的图像。AlexNet在top-1和top-5中分别实现了37.5%和17.0%的错误率,分类效果有了质的提升。随后,OverFeat[27]针对当时检测和分类的瓶颈,首次明确解释了CNN的定位与检测原理。该检测算法使用全卷积网络和滑动窗口技术,将大量的滑动窗口放入图像中,滑动窗口给出对象类别的置信度得分和边界框,与对象有50%重叠的边界框将被合并、累积,并利用回归算法预测最终边界框。然而,该方法所需算力较大,因此效率较低。

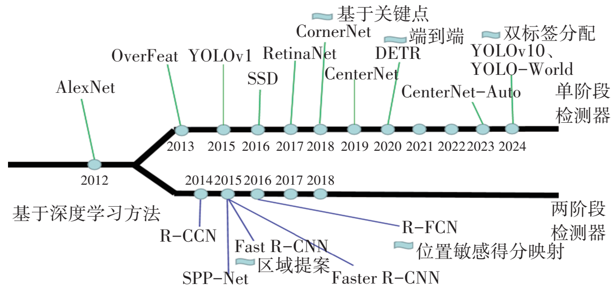

图2 目标检测发展进程

在2014年的ILSVRC中,视觉几何组网络(Visual Geometry Group Network, VGGNet)[28]在分类任务中错误率达到7.32%,在定位任务中的错误率为25.32%。为减少所需参数量,该模型通过堆积叠加,使用多个3×3卷积核代替大尺度卷积核,并多次使用2×2的池化,使结构简洁明了,同时利用多个卷积层增加了非线性,提高了模型的性能,如图3所示。VGGNet至今仍在图像特征提取中广泛应用。

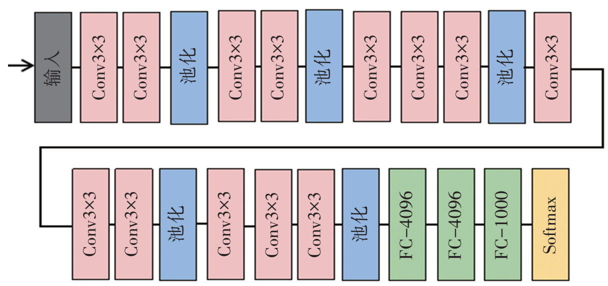

图3 VGGNet模型

2014年,IanGoodfellow等首次提出生成对抗网络(Generative Adversarial Network, GAN)[29],该网络包括一个名为生成器的神经网络和一个名为判别器的神经网络,开创了许多新的研究方向。但是基于GAN的模型训练和优化仍然较为繁琐,模型易崩溃,结果不易收敛且不稳定,故基于GAN的变体陆续出现,如深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Networks, DCGAN)[30]、信息生成对抗网络(Information Maximizing Generative Adversarial Networks, infoGAN)[31]等。

通常,深度学习模型层数越多,能捕获的信息就越多,但是层数过多会导致梯度爆炸或退化问题。为减轻网络的训练负担、解决深层网络的退化问题,2015年,He等提出了ResNet[32],此后,模型层数提高到了百层以上。ResNet通过残差学习的方法提出残差结构,并使用批归一化(Batch Normalization, BN)加速训练。在2015年的ILSVRC

比赛中，152层ResNet的top-5任务错误率仅为3.57%，准确率较此前的模型大幅提升。尽管ResNet通过快捷连接降低了训练难度，但它并未充分利用先前层的功能。在逐元素运算中，浅层中的原始特征丢失，因此不能在以后直接使用。不同于ResNet，密集连接卷积网络（Densely Connected Convolutional Network，DenseNet）[33]相互连接了所有层，保留了浅层特征，并通过将输入与残差输出连接实现特征重用，提升了效率，在参数大小相同的条件下，DenseNet的性能优于ResNet。在DenseNet中，来自浅层的大多数新开发的特征是重复的，并产生了很高的计算成本。结合ResNet与DenseNet的优点，研究人员提出了一种双路径网络（Dual-Path Network，DPN）[34]，该网络能共享公共特征，且其双路径架构可保留灵活性以获取新的特征。它与DenseNet、ResNet以及次代残差网络（Residual Network with Next，ResNeXt）这类网络在模式分析、统计建模和计算学习视觉对象分类（Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes，PASCAL VOC）挑战赛的检测与分割数据集中，相比DPN具有更优良的性能。2016年，沙漏网络（Hourglass Network，HourglassNet）[35]被提出，其最早用于人类姿势识别，是具有沙漏模块序列的完全卷积结构。沙漏模块能够同时捕获局部和全局信息，非常适合目标检测任务。

受各网络的复杂程度和设备算力的影响，深度学习的目标检测很难搭载在移动端。2017年，移动网络（Mobile Network，MobileNet）[36]诞生，该网络利用深度可分离卷积构建轻量级的深度神经网络，为移动设备和嵌入式设备的性能提升提供了可能，且其显著降低了计算成本和参数数量，而分类精度没有明显损失。小尺度物体自身的像素信息较少，在下采样过程中极易被遗弃，而许多网络利用单个高层特征，从而忽略了这一问题，文献[37]构造了不同尺度的特征信息并进行融合，旨在提升检测算法对不同尺寸检测目标的鲁棒性。可变形卷积网络（Deformable Convolutional Network，DCN）[38]解决了卷积核权重的固定导致的同一CNN在处理一幅图像的不同位置时感受野大小相同的问题。而DetNet（Detection Network，DetNet）[39]通过在多尺度特征图上采用扩张卷积、保持高分辨率特征图等措施改进感受野，解决了多尺度物体识别问题。2017年，检测驱动增强网络（Detection-driven Enhancement Network，DeNet）[40]描述了一个用CNN进行稀疏估计的框架，并提出了一种感兴趣区域（Region of Interest，ROI）检测器和分类模型，提高了检测性能，使用在语义分割背景下描述的反卷积和跳过层，中小型物体的检测率得到了提升。2024年，Xu等[41]提出一种高阶特征关联网络（High-Order Feature Association Network，HOFA-NET），该网络可有效捕捉信道和空间维度特征之间的相互依赖关系，产生更可分的特征，有效提升对密集物体、小尺寸物体的检测性能。为解决小尺度物体缺乏特征信息的问题，Dai等[42]设计了自适应样本分配策略（Adaptive Sample Assignment Strategy，ASAS）和微小目标聚焦增强模块，有效增强了高层次语义信息对微小物体的传播能力。这些网络奠定了现代深度学习的发展基础。

2.2.2 目标检测器

目标检测器分为两阶段检测器和单阶段检测器，前者在检测中表现出更高的精度，而后者具有更快的速度。两种检测器都生成建议，主要区别在于两阶段检测器生成仅具有前景或背景信息的稀疏建议集，而单阶段检测器将图像中的每个区域均视为潜在建议，并相应地估计每个位置的潜在对象的类和边界框坐标。常见的两阶段检测器有区域卷积神经网络（Region-Convolutional Neural Network，R-CNN）及其变体快速区域卷积神经网络（Fast R-CNN）、更快速区域卷积神经网络（Faster R-CNN）、空间金字塔池化网络（Spatial Pyramid Pooling Network，SPP-Net）、基于区域的全卷积网络（Region-based Fully Convolutional Network，R-FCN）等。从深度学习技术提出开始到目前最为有效的单阶段检测器有YOLO（You Only Look Once）、单步多框检测器（Single Shot MultiBox Detector，SSD）、视网膜网络（RetinaNet）、DETR（DEtection TRansformer）等。

2.2.2.1 两阶段检测器

两阶段检测器包括两个任务阶段：提案生成，识别输入中可能是目标对象的区域；对生成的提案进行预测，采用基于深度学习的模型为提案设置正确的分类标签。典型的两阶段检测器及其特点如表3所示。

R-CNN[43]是一种简单且可扩展的检测算法，是CNN之后目标检测领域极具特点的里程碑之一。R-CNN具有优良的特征提取和分类性能，其

流程如图4所示:首先采用区域建议生成2 000个候选框,再将每个候选框处理成227像素×227像素大小后放入网络中提取特征,获得4 096维特征后将其送入SVM分类器,通过非极大值抑制(Non-Maximum Suppression,NMS)筛选出最可能的框,即对目标进行定位,最后修正候选框位置并使用最小二乘法解决回归问题完成检测。该算法简单且可拓展,在PASCAL VOC 2012数据集中的平均精度均值(mean Average Precision,mAP)达到53.3%。但是R-CNN简单的操作带来了检测速度慢、占用存储空间大、特征提取容易重复计算的问题。

表3　典型的两阶段检测器

检测器	发表时间	贡献	优缺点
R-CNN	2014年	解决了CNN的定位问题,提出了大型CNN的有效训练方法	优点:算法简单、可拓展 缺点:检测速度慢、占用存储空间大、特征提取容易重复计算
Fast R-CNN	2015年	实现了端到端训练,引入了ROI池化层	优点:共享卷积计算、端到端训练简化了流程 缺点:依赖选择性搜索算法生成候选区域,速度慢
Faster R-CNN	2015年	提出了区域提案网络(RPN),将卷积层共享	优点:通过RPN生成候选区域,大幅减少计算量,提高了速度 缺点:处理密集和小尺寸目标仍存在问题,RPN仍需耗费大量时间
SPP-Net	2015年	解决了图像畸形导致的信息消失和存储问题,提出了空间金字塔变换层	优点:可以避免重复计算卷积特征 缺点:没有解决存储空间消耗问题,训练慢、效率低
R-FCN	2016年	提出了位置敏感得分映射	优点:速度较Faster R-CNN更快,精度略有提升 缺点:速度仍较慢

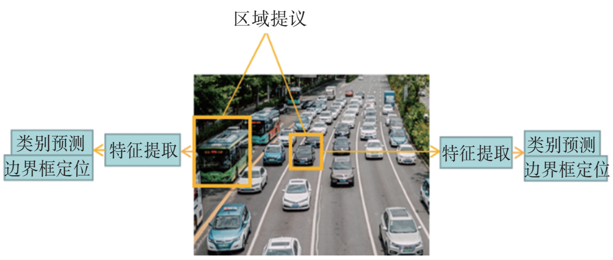

图4　R-CNN流程

SPP-Net[44]是一种改进的R-CNN算法。通常,CNN模型需要将输入数据的尺寸固定,如AlexNet需输入224像素×224像素的图像,而候选区域尺度各异,难以直接输入CNN中实现区域的特征提取。SPP-Net在R-CNN的基础上提出了两项重大改进:一是去除裁剪/扭曲图像归一化过程,解决了图像畸形导致的信息消失和存储问题[45];二是在R-CNN的基础上提出了以3种池化方式对特征进行池化并对结果进行拼接的空间金字塔池化层。通过上述两项改进措施,SPP-Net极大提高了R-CNN的训练速度和推理速度,在PASCAL VOC 2007数据集中的mAP达到59.2%。但是,SPP-Net依旧存在多级检测和忽略所有先前的层、只微调其他全连接层的问题。候选框的选择非常耗时,且候选框的提取、CNN特征的计算、SVM分类、边界框回归均需要独立训练,并保存大量中间结果,无法实现训练参数的全局优化,而Fast R-CNN可以解决上述问题,提高速度和精度。

Fast R-CNN[46]作为R-CNN的变体,融入了SPP-Net的特点,并解决了SPP-Net的局限性问题。Fast R-CNN将要输入的图像归一化后再进行特征提取,并在最后一层特征图上映射候选框,提高了算力的利用率。与R-CNN不同,Fast R-CNN在分类前不需要将特征存储在硬盘中。在PASCAL VOC 2007数据集上,Fast R-CNN的mAP达到了70.0%。

SPP-Net和Fast R-CNN等为缩短检测网络运行所需时间而采用区域提案方法,但区域提案需要大量计算资源,成为其发展的瓶颈。Faster R-CNN[47]在使用Fast R-CNN框架的基础上,提出了区域提案网络(Region Proposal Network,RPN),RPN可通过端对端训练生成高质量候选框,并在训练时采用与目标检测网络共享的卷积层,大幅提高了检测速度,但其在生成候选框时仍需花费大量时间。Faster R-CNN在PASCAL VOC 2007数据集上的mAP达到73.2%。

R-FCN[48]是一个两阶段检测器,为解决目标检测的位置敏感性和分类位置不敏感性的矛盾问题,提出位置敏感得分映射(Position-Sensitive Score Map)。该检测器进行图像预处理后,将其送入分类网络,在分类网络的最后一个卷积层有3

个分支,第一种是在特征图上通过RPN生成候选框,第二种是在特征图上进行位置敏感得分映射,映射维度是$K \times K \times (C+1)$,第三种是在特征图上计算$4 \times K \times K$维的位置敏感度得分映射,其中,K、C分别为位置敏感得分映射的网格大小(将候选区域划分为$K \times K$子区域)、目标类别数。最后,在计算得到的映射上实现了ROI池的功能,可识别并获取所需要的地址和类别信息。该检测器与Faster R-CNN相比,具有较高的速率,精确度略有提升。

两阶段方法存在对锚点的多次细化,故与单阶段方法相比,检测结果更为精准。

2.2.2.2 单阶段检测器

单阶段检测器通过一个阶段完成目标的类别预测和边界框回归,侧重于检测速度,精度尚可。相对于两阶段检测器,单阶段检测器不需要生成候选区域后再进行预测,因此有效提升了检测速度。部分典型的单阶段检测器及其特点如表4所示。

表4 典型的单阶段检测器

检测器	发表时间	贡献	优缺点
YOLOv1	2015年	提出了一种实时检测算法,将物体检测作为回归问题进行求解	优点:速度快,使用全局图像进行预测,迁移能力强,可在其他新领域使用 缺点:一个网格只预测2个框,且属同一类,遮挡、小物体检测效果较弱,不常见角度的目标泛化性能偏弱
SSD	2016年	提出一种快速的实时检测算法,引入单阶段检测器,使用小的卷积滤波器	优点:实时性好,能够多尺度预测 缺点:精度略显不足
RetinaNet	2017年	提出焦点损失,解决了目标检测中正负样本不平衡和难易样本不平衡的问题	优点:解决了类别不平衡问题,处理不同尺寸目标时性能较优 缺点:需要微调超参数,算力要求大
CornerNet	2018年	提出了角点池化(Corner Pooling),设计了一种角点的无锚框检测方法	优点:训练速度快,可以更好地定位角点 缺点:性能略有不足
CenterNet	2019年	提出了一种新的检测算法	优点:误检率相对于CornerNet更低,模型更为轻量化 缺点:中心点和关键点的预测对准确度影响较大,物体重叠时易预测错误
ExtremeNet	2019年	提出一种自下而上、无锚框的关键点预测和分组方式实现目标检测	优点:能够减少干扰导致的信息丢失,从而提升精度 缺点:算法复杂度高,所需时间长
FSAF	2019年	提供一种引导特征选择和基于重叠的锚点采样方法	优点:通用性强、性能优异 缺点:锚点设置和分配方式需要改进,需要大量的资源和时间
YOLOv10	2024年	提出了一种双标签分配和无NMS的一致双分配策略	优点:速度快,解决了后处理中的冗余预测问题 缺点:在小目标检测方面表现稍逊于YOLOv8和YOLOv9
YOLO-World	2024年	提出了一种可重新参数化的视觉-语言路径聚合网络(PAN)来连接视觉和语言特征	优点:实时性强,能开放词汇表识别,无需预定义类别 缺点:技术困难、存在错误识别问题

YOLO[49]是Joseph Redmon等于2015年提出的一种深度学习时代最具代表性的单阶段检测器,经多次版本迭代,目前已能实现优秀的检测效果。与两阶段检测不同,这种算法不采用滑动窗口技术和区域建议技术,而是将神经网络应用到整个图像中,将输入的图像分割为几个格子区,每个格子区预测边界框和对象在其中的置信度,再通过对比置信度来判断相对特定类的置信度得分,但每个网格只能分类和定位一个对象,使其很难检测到小尺度的物体和被遮挡的物体,且仅应用最后一个特征图进行预测,也难以用于预测多尺度和多纵横比的对象。YOLO已经陆续发展出多个版本,如:YOLOv5着力于减小模型的复杂度和提高实时性;YOLOv7[50]通过引入优化的结构(如动态标签分配和模型结构重新参数化),在速度和准确性方面优于大多数检测器;YOLOv8[51]抛弃基于锚的方法,引入了无锚分裂Ultralytics头,它有助于提高检测过程的准确性和效率,并

且也采用了先进的骨干网络（Backbone）架构提高特征提取和检测的性能。YOLOv10[52]为解决之前版本因计算冗余导致参数利用率低、效率次优的问题和改进模型性能，首先提出了一种双标签分配和无NMS的一致双分配策略，用于解决后处理中的冗余预测问题，随后改进模型的体系结构，设计出了新的轻量化分类头、空间信道解耦下采样和秩引导块，以减轻显式的计算冗余，实现更为高效的模型架构，最后，通过研究大卷积核提出部分自关注模块来提升模型性能，参数、计算量大幅度减少，并且在COCO数据集上取得了54.4%的平均精度（Average Precision，AP）。2024年，或许受实时检测转换器（Real-Time DEtection TRansformer，RT-DETR）[53]能够在实时检测中击败当时的YOLO算法的影响，YOLOv11横空出世，相比于之前的版本，其更新了更适用于特征提取的主干网络和颈部（Neck）结构，并优化了训练流程，使训练更为迅速，可在速度与精度间很好地保持平衡，更满足实际需求。YOLO算法在自动驾驶场景中也存在一定的局限性，首先是存在小目标难以检测的情况，其次是版本较早的YOLO算法对硬件资源的要求较高，难以在自动驾驶车辆上部署，最后是在复杂交通环境中以及恶劣天气条件下实时检测精度急剧下降，影响实时检测的可靠性。为解决上述问题，YOLO系列后期的优化重心可以放在针对提升图像远端小目标的检测精度以及通过提升模型抗干扰能力提高恶劣环境下的召回率，保证其在自动驾驶领域中的适用性。

SSD[54]是Liu提出的一种强大且应用广泛的单阶段目标检测算法。在PASCAL VOC 2007测试集上，SSD在300像素×300像素图像的检测任务中，mAP达到74.3%，且实现了59帧/s的帧速率。SSD是一种多尺度的检测方法，且以VGG16网络为基础，通过增加卷积层来获得更多的特征图。SSD使用的特征图包括38像素×38像素、19像素×19像素、10像素×10像素、5像素×5像素、3像素×3像素、1像素×1像素共6种。大尺度特征图用于检测小尺度的目标，包括远处的车辆、交通灯等，小尺度特征图用于检测大尺度的目标，如近处车辆等，模型如图5所示。与之前的检测器只在顶层进行检测不同，SSD具有多种检测层来检测不用尺度的对象。

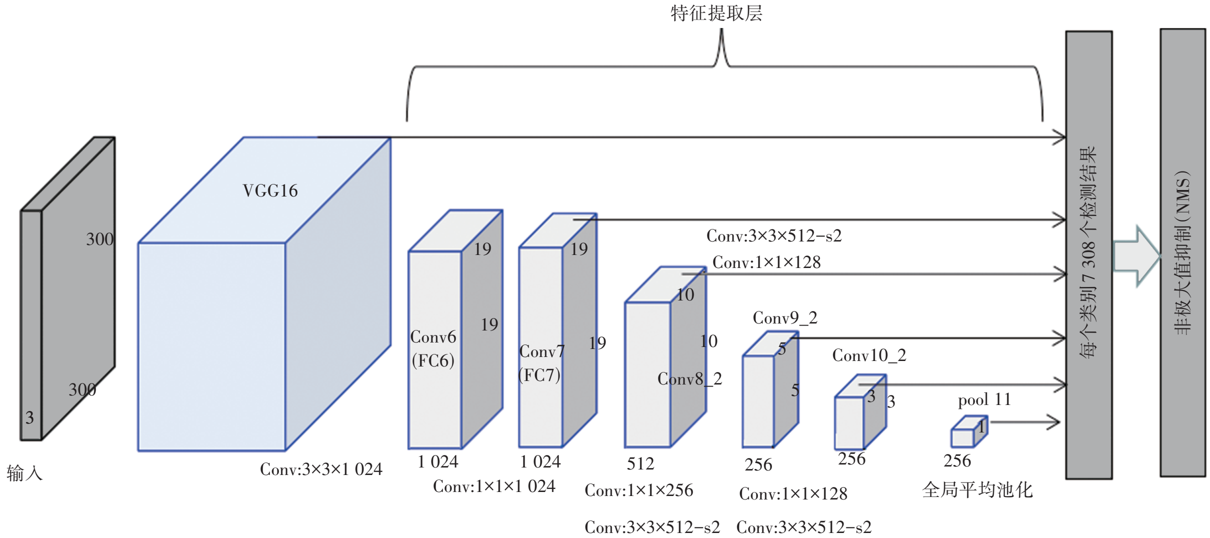

图5　SSD模型

RetinaNet[55]通过修正正负样本不平衡实现与两阶段方法同等的检测准确度。在RetinaNet提出之前，相对于两阶段目标检测，单阶段目标检测的精确度相差甚远。焦点损失的提出使单阶段方法具有与两阶段方法相差无几的准确性以及更快的检测速度。

2018年后，目标检测出现了新的方向，即无锚框（Anchor Box）检测。CornerNet、中心点网络（CenterNet）、极值点网络（ExtremeNet）、代表性点（Representative Points，RepPoints）、网格区域卷积神经网络（Grid Region Convolutional Neural Network，Grid R-CNN）等是基于关键点的无锚框检测。无特征选择锚（Feature Selective Anchor-Free，FSAF）模型、导向锚定区域提案（Region Proposal by Guided Anchoring，GARPN）、自适应样本选择（Adaptive Training Sample Selection，ATSS）、Foveabox等是基于中心点的无锚框检测。CornerNet[56]开创性地抛弃锚

框，采用关键点来检测目标，用一对角点表示目标的位置，在特征图的每个位置预测热图（Heatmaps）、嵌入向量（Embeddings）和角偏移（Offsets），热图用于计算角点的概率，嵌入向量用于匹配成对的角点，角偏移用于回归角点的位置。该检测器采用HourglassNet作为骨干网络，搭配新提出的角点池化（Corner Pooling）来预测热图。CornerNet在COCO数据集上的AP达到了42.1%，超过了当时所有的单阶段模型。2019年，CenterNet[57]将基于中心的方法和基于角点的方法相结合，通过关键点（中心点）检测物体，回归到物体的尺寸、维度、三维范围、姿态等属性，将物体检测问题有效转化为关键点估计问题。CenterNet通过预测中心点热图来定位目标中心，生成边界框，同时利用中心点热图过滤掉低置信度的检测结果，且在推断时只需一次前向传播，无需NMS后处理。2019年，ExtremeNet[58]被提出，其通过检测目标的4个极点进行目标定位，然后通过几何关系将其组合起来生成最终的边界框，性能与其他主流检测算法相当。ExtremeNet的检测方法十分独特，但是也借鉴了CornerNet的设计思路，但其依赖于较多的后处理方法，故仍有许多优化空间。与此同时，文献[59]提出了Grid R-CNN，它用基于角点的定位机制取代了线性边界框回归器。综上，CornerNet、CenterNet等基于关键点的算法本质上与基于锚框的算法相似，通过预测目标中心点及边框距中心点的距离来检测目标。文献[60]提出了一种FSAF框架，是一种可以在结构中加入特征金字塔网络（Feature Pyramid Network，FPN）的单阶段检测器。FSAF将在线特征选择应用于多级无锚框分支的训练，训练过程中，FSAF将每个对象动态分配到最合适的特征级别，以训练基于中心的分支，取得了较好的检测效果。RepPoints[61]提出使用点集的方式来表示目标，该方法基于全卷积网络，在分类和最后一次回归时使用DCN，在准确度与计算量之间实现了非常好的平衡。

2.3 深度学习在车辆目标检测中的应用

自动驾驶车辆的目标检测算法需满足2个条件：实现道路上目标的高精确度检测；检测速度能够满足车载控制器实时控制车辆的要求。基于深度学习的算法可以在使用较少计算资源的前提下实现较高的检测精度，成为自动驾驶系统必不可少的目标检测方法。如图6所示，目标检测存在一些普遍的困难：当车辆在弱光照条件下或夜间行驶时，因光线不充足，感知系统难以获取足够的信息；车辆高速行驶时，车载相机拍摄到的照片会变得模糊；在雨、雪、雾天气条件下，因受到遮挡，拍摄的照片也会不清晰；车辆行驶时，常出现小目标隐藏在复杂环境中导致漏检的情况，从而影响行驶安全性。

夜间光线问题　遮挡问题　小目标检测困难　雨雾等复杂天气

图6　车辆目标检测的一般难点

针对车辆夜间行驶条件下的检测困难问题，一些研究[62-63]通过改进和设计新的网络结构提升弱光环境下的检测准确率，如文献[64]通过对图像增强算法进行改进并与YOLO算法融合，使夜间条件下的检测精度大幅提升，平均检测精度达95.92%，比YOLOv5s精度提升了1.19百分点，这表明图像增强技术可以很好地优化车载目标检测任务中获取的信息，解决夜间环境等车辆边界难以区分的情况。然而，在实际应用中也发现了一些问题，如当遇到一些车辆轮廓微弱并且车灯等泛光严重的情况时，增强处理后的图像仍然难以准确区分车辆边界和特征，导致检测准确率有所下降。这表明在复杂多变的现实环境中，单一的图像增强技术无法完全应对所有不利因素，当前的技术瓶颈之一就是缺乏对多种恶劣天气和复杂路况的适应性。文献[65]、文献[66]提出的方法也能够很好地改善夜间弱光环境下车辆检测的准确率。针对雨、雪、雾天气下照片不清晰以及目标遮挡的问题，Hasan Abbasi等[67]提出了一种雾感知自适应YOLO算法，通过使用雾化评估器算法配合YOLO算法提高了雾化条件下的目标检测性能。文献[68]通过将去雾模块与多尺度视网膜皮层理论（Multi-Scale Retinex，MSR）相结合，增强了雨雾天气等复杂环境条件下的检测性能。另一项研究[69]通过将卷积注意力模块（Convolutional Block Attention Module，CBAM）加入YOLOv5的跨阶段部分连接（Cross Stage Partial，CSP）层来改善雾条件下的检测准确率。文献[70]利用多层次特征融合的去雨雾模块对图像进行优化，通过对6 000张含雨水图片与6 000张含雾气图片进行实际测试，其搭载的算法对真实测试集雨雾条件下车辆检测的准确率接近89%。这些方法均通过对图像进行去雨雾操作提升检测效果，但是在实际的自动驾驶任务中，大雨会导致信息的严重冗余，多种类型的雾气

会造成不同程度的图像模糊，类似于团雾天气使场景中对比度显著下降，导致检测算法难以准确获取实际信息，影响其对检测目标的判断。这反映出当前车辆检测算法在应对不同类型恶劣天气时的适应性有待提高，无法准确区分被雨滴或雾气遮挡的车辆目标轮廓，这是需要进一步攻克的技术难题。

车辆目标检测对实时性要求严苛，所以搭载精度高的单阶段检测器十分必要，迄今为止，单阶段检测器中性能最为突出的毫无疑问是YOLO，因此车辆目标检测领域中基于YOLO的算法层出不穷，研究人员为解决小物体难以检测、漏检率高等问题，提出了更为有效的小目标检测方法[71-73]。Zhang等[74]针对漏检率高、远景小物体特征提取能力和感知能力差、对目标定位能力较弱等问题，通过将YOLOv7的骨干网络替换为Res3Unit以及在空间金字塔池化跨阶段部分连接（Spatial Pyramid Pooling Cross Stage Partial Connection，SPPCSPC）层后增加即插即用的混合注意力机制模块ACmix，并在特征融合和检测头连接处使用感受野局部注意力（Receptive Field Local Attention，RFLA）中的高斯感受野方案（Gaussian Receptive Field Scheme），相较于YOLOv7，成功将平均准确检测率提升2.4百分点，帧速率达到了96帧/s。这意味着在实时性上能够更好地满足交通流量较大时对车辆目标快速准确检测的需求。Qiu等[75]为在检测精度和计算代价之间取得平衡，从无人机的角度提出了一种新的特殊车辆检测算法YOLO-GNS。该算法基于单级无头上下文结构的特征提取算法针对幻影网络（Ghost Network，GhostNet）的缺点进行改进，与YOLOv7相比，mAP提高了4.4百分点，帧速率提高1.6%。Cai等[76]基于YOLOv4将CSPDarknet（Cross Stage Partial Darknet）最后的输出层替换为DCN，然后设计了特征融合模块路径聚合网络++（Path Aggregation Network++，PAN++），并提出一种网络剪枝算法，相较于原算法，在各类数据集上成功提升了mAP，并能以高于66帧/s的速度进行检测。为了处理移动车辆的遮蔽问题，文献[77]使用U型网络（U-Net）和深度卷积网络，通过将卷积网络在不同尺度上的信息整合到不同分辨率的图像中，使检测准确率显著提高，Li等[78]通过将深度信息与YOLOv5相结合，降低了漏检和误检率，并获得了距离信息。虽然上述方法能够在漏检、误检方面取得良好的效果，但是在车辆密集与遮挡程度严重的情况下，其漏检和误检问题依然颇为突出，因为遮挡的程度不同会导致车辆不同程度的特征丢失，使算法难以准确识别，需要在实际交通情况下克服这一难点。

2023年，Wang等[79]提出了CenterNet-Auto算法，将特征提取网络RepVGG与CenterNet相结合，在不增加参数的情况下，对FPN进行优化以提高其特征提取能力，并提出了一种新的ABMin检测头，充分利用边界特征辅助中心点回归到其他物体属性，在不损失检测精度的前提下提高了检测速度。BDD数据集上的验证结果表明，其准确性和速度有较大优势，推进了实时车辆检测技术。

3 总结与展望

传统的目标检测方法中特征提取算法工作量大、易受弱光和恶劣图像质量的影响，深度学习算法在网络结构及后处理方法等方面已被证明具有比传统检测算法更佳的性能，且更易研究，但仍受复杂环境和算力的巨大影响。深度学习算法被认为是一种黑箱子，其工作过程难以解释，而且在早期发展过程中因中央处理器（Central Processing Unit，CPU）、图形处理器（Graphics Processing Unit，GPU）和内存的性能不足而受到制约，其在车辆目标检测领域极易出现小目标、遮挡导致的漏检和误检，在复杂环境下容易出现检测率下降的问题，目前诸多研究正在努力克服这些局限性并已取得了一些成果，还有更多的研究旨在提升整体检测准确度。未来，针对车辆目标检测，需要在以下几个方面开展更为深入的研究：

a. 小目标检测。小目标检测一直存在背景复杂情况下特征信息少的问题，可以深入研究注意力机制如何聚焦小目标特征区域，以更好地捕捉其关键信息，也可以通过研究更适合小目标检测的检测头或高分辨率轻量级网络的架构搭建和参数优化等，保证较高分辨率以精准检测小目标。

b. 多模态融合。在自动驾驶领域中，激光雷达点云数据与视觉数据融合、声波雷达数据与视觉数据融合等是当前的研究热点，模态融合在一定程度上提升了车辆对周围环境的感知能力，但是也面临一些挑战：不同硬件导致所采集数据的形式和频率不同，需要进行校准；模态的增多意味着数据量增大，会导致计算资源需求量急剧增大。未来，可通过新型传感器研发提供更为优质的数据或者针对计算资源不足的环境进行轻量化模型设计，在保证

性能的前提下降低计算需求量。同时也需加强推进文本特征、语音特征、图像特征的融合，以及类似Transform的模型从自然语言处理（Natural Language Processing，NLP）到计算机视觉（Computer Vision，CV）领域的应用，以满足更为复杂环境条件下的车辆目标检测。

c. 精度与召回率之间的平衡。部分检测算法检测准确度较高，但召回率略有欠缺，严重制约了算法在复杂检测任务中的效能，需要新的方法在精度与召回率之间取得平衡，如通过有效融合多种数据源、优化特征提取架构、优化后处理方式提升算法在复杂环境下对各类目标的检测能力。

d. 高品质数据资源库。基于监督的深度学习模型需要通过数据驱动，所以需要足够大量的准确数据支持其训练以保证模型的准确性，未来可着力于制作各领域的高质量数据集。

参 考 文 献

[1] 茅智慧, 朱佳利, 吴鑫, 等. 基于YOLO的自动驾驶目标检测研究综述[J]. 计算机工程与应用, 2022, 58(15): 68-77.
MAO Z H, ZHU J L, WU X, et al. Review of YOLO Based Target Detection for Autonomous Driving[J]. Computer Engineering and Applications, 2022, 58(15): 68-77.

[2] LIU J. Application of Object Detection in Autonomous Driving[C]// Proceedings of the 2023 2nd International Conference on 3D Immersion, Interaction and Multi-Sensory Experiences (ICDIIME). Madrid, Spain: IEEE, 2023.

[3] 段伟. 汽车自动驾驶技术简述[J]. 中国自动识别技术, 2024(2): 66-68.
DUAN W. A Brief Overview of Autonomous Vehicle Technology[J]. China Auto-ID, 2024(2): 66-68.

[4] HNEWA M, RADHA H. Object Detection under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques[J]. IEEE Signal Processing Magazine, 2021, 38(1): 53-67.

[5] ITTI L, KOCH C, NIEBUR E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[6] HUANG T S. Computer Vision: Evolution and Promise[C]// 1996 CERN School of Computing. Geneva, Switzerland: CERN, 1996.

[7] WU X W, SAHOO D, HOI S C H. Recent Advances in Deep Learning for Object Detection[J]. Neurocomputing, 2020, 396: 39-64.

[8] ZOU Z X, SHI Z W, GUO Y H, et al. Object Detection in 20 Years: A Survey[J]. Proceedings of the IEEE, 2019: 111(3): 257-276.

[9] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// 2025 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005.

[10] BEAUCHEMIN S S, BARRON J L. The Computation of Optical Flow[J]. ACM Computing Surveys, 1995, 27(3): 433-466.

[11] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[12] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D A. Cascade Object Detection with Deformable Part Models[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010.

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

[14] LOWE D G. Object Recognition from Local Scale-Invariant Features[J]. Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV). Kerkyra, Greece: IEEE, 1999.

[15] HEARST M A, DUMAIS S T, OSUNA E, et al. Support Vector Machines[J]. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18-28.

[16] BREIMAN L. Bagging Predictors[J]. Machine Learning, 1996, 24: 123-140.

[17] FREUND Y, SCHAPIRE R. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[18] TIAN Q, ZHANG L, WEI Y X, et al. Vehicle Detection and Tracking at Night in Video Surveillance[J]. International Journal of Online Engineering, 2013, 9(S6): 60-64.

[19] WU B F, JUANG J H. Adaptive Vehicle Detector Approach for Complex Environments[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(2): 817-827.

[20] ZHANG D B. Vehicle Target Detection Methods Based on Color Fusion Deformable Part Model[J]. Eurasip Journal on Wireless Communications and Networking, 2018, 2018(1): 94.

[21] PADILLA D A, VILLAVERDE J F, MAGDARAOG J J T, et al. Vehicle and Weather Detection Using Real Time

Image Processing Using Optical Flow and Color Histogram [C]// 2019 5th International Conference on Control, Automation and Robotics (ICCAR). Beijing, China: IEEE, 2019.

[22] PAPAGEORGIOU C P, OREN M, POGGIO T. A General Framework for Object Detection[J]. Sixth International Conference on Computer Vision. Bombay, India: IEEE, 1998.

[23] KARANGWA J, LIU J, ZENG Z. Vehicle Detection for Autonomous Driving: A Review of Algorithms and Datasets [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(11): 11568-11594.

[24] ASIM M, WANG Y, WANG K Z, et al. A Review on Computational Intelligence Techniques in Cloud and Edge Computing[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 4(6): 742-763.

[25] 陈晓, 杨琪, 姚海洋, 等. 基于细粒化特征感知的水下目标检测算法[J]. 陕西科技大学学报, 2024, 42(4): 177-183.
CHEN X, YANG Q, YAO H Y, et al. Underwater Target Detection Algorithm Based on Fine-Grained Feature Perception[J]. Journal of Shaanxi University of Science and Technology, 2024, 42(4): 177-183.

[26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA: Curran Associates Inc., 2012.

[27] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: Integrated Recognition, Localization and Detection Using Convolutional Networks[EB/OL]. (2014-02-24)[2025-02-21]. https://arxiv.org/abs/1312.6229.

[28] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL]. (2015-04-10) [2025-02-21]. https://arxiv. org/abs/1409.1556.

[29] WANG H, YU Y J, CAI Y F, et al. A Comparative Study of State-of-the-Art Deep Learning Algorithms for Vehicle Detection[J]. IEEE Intelligent Transportation Systems Magazine, 2019, 11(2): 82-95.

[30] MEHRALIAN M, KARASFI B. RDCGAN: Unsupervised Representation Learning with Regularized Deep Convolutional Generative Adversarial Networks[C]// 2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium. Kish Island, Iran: IEEE, 2018.

[31] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets[C]// 30th Conference on Neural Information Processing Systems (NIPS). Barcelona, Spain: Curran Associates Inc., 2016.

[32] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016.

[33] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely Connected Convolutional Networks[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2016.

[34] SHI H M, ZHOU Q, NI Y H, et al. DPNET: Dual-Path Network for Efficient Object Detectioj with Lightweight Self-Attention[J]. 2022 IEEE International Conference on Image Processing. Bordeaux, France: IEEE, 2021.

[35] NEWELL A, YANG K Y, DENG J. Stacked Hourglass Networks for Human Pose Estimation[M]// LEIBE B, MATAS J, SEBE N, et al. Computer Vision - ECCV 2016. Cham, Switzerland: Springer International Publishing, 2016: 483-499.

[36] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[EB/OL]. (2017-04-17)[2025-02-21]. https://arxiv.org/abs/1704.04861.

[37] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017.

[38] DAI J F, QI H Z, XIONG Y W, et al. Deformable Convolutional Networks[C]// 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017.

[39] LI Z M, PENG C, Yu G, et al. DetNet: A Backbone network for Object Detection[EB/OL]. (2018-04-19)[2025-02-21]. https://arxiv.org/abs/1804.06215.

[40] TYCHSEN-SMITH L, PETERSSON L. DeNet: Scalable Real-Time Object Detection with Directed Sparse Sampling [C]// 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017.

[41] XU Y P, WU X, WANG L, et al. HOFA-Net: A High-Order Feature Association Network for Dense Object Detection in Remote Sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 1513-1522.

[42] DAI H H, GAO S S, HUANG H, et al. An Adaptive Sample Assignment Network for Tiny Object Detection[J]. IEEE Transactions on Multimedia, 2024, 26: 2918-2931.

[43] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and

Semantic Segmentation[J]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014.
[44] 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, 2014, 37(9): 1904-1916.
[45] WANG J, ZHANG T J, CHENG Y, et al. Deep Learning for Object Detection: A Survey[J]. Computer Systems Science & Engineering, 2021, 38(2): 165-182.
[46] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015.
[47] REN S Q, HE K M, 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.
[48] DAI J F, LI Y, HE K M, et al. R-FCN: Object Detection via Region-Based Fully Convolutional Networks[M]// 30th Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016.
[49] 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. Las Vegas, NV, USA: IEEE, 2016.
[50] 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[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023.
[51] VARGHESE R, SAMBATH M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness[C]// 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India: IEEE, 2024.
[52] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-Time End-to-End Object Detection[C]// Annual Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2024.
[53] ZHAO Y A, LV W Y, XU S L, et al. DETRs Beat YOLOs on Real-Time Object Detection[C]// Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024.
[54] WEI L, DRAGOMIR A, DUMITRU E, et al. SSD: Single Shot MultiBox Detector[M]// LEIBE B, MATAS J, SEBE N, et al. Computer Vision - ECCV 2016. Cham, Switzerland: Springer, 2016.
[55] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017.
[56] LAW H, DENG J. CornerNet: Detecting Objects as Paired Keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.
[57] DUAN K W, BAI S, XIE L X, et al. CenterNet: Keypoint Triplets for Object Detection[C]// 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019.
[58] ZHOU X Y, ZHUO J C, KRHENBÜHl P. Bottom-up Object Detection by Grouping Extreme and Center Points [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019.
[59] LU X, LI B Y, YUE Y X, et al. Grid R-CNN[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2018.
[60] ZHU C C, HE Y H, SAVVIDES M. Feature Selective Anchor-Free Module for Single-Shot Object Detection[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019.
[61] YANG Z, LIU S H, HU H, et al. RepPoints: Point Set Representation for Object Detection[C]// 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019.
[62] XU Y Q, CHU K B, ZHANG J. Nighttime Vehicle Detection Algorithm Based on Improved Faster-RCNN[J]. IEEE Access, 2024, 12: 19299-19306.
[63] YU Y, CHEN W X, CHEN F X, et al. Night-Time Vehicle Model Recognition Based on Domain Adaptation[J]. Multimedia Tools and Applications, 2024, 83: 9577-9596.
[64] 袁子强. 图像增强与夜间车辆识别算法研究[D]. 成都: 电子科技大学, 2023.
YUAN Z Q. Research on Image Enhancement and Night Vehicle Recognition Algorithm[D]. Chengdu: University of Electronic Science and Technology of China, 2023.
[65] ZHANG H, YANG K F, LI Y J, et al. Night-Time Vehicle Detection Based on Hierarchical Contextual Information[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(10): 14628-14641.
[66] XUE T, ZHANG Z Q, MA W N, et al. Nighttime Pedestrian and Vehicle Detection Based on a Fast Saliency and Multifeature Fusion Algorithm for Infrared Images[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23: 16741-16751.
[67] ABBASI H, AMINI M, YU F R. Fog-Aware Adaptive YOLO for Object Detection in Adverse Weather[C]//

Proceedings of the 2023 IEEE Sensors Applications Symposium (SAS). Ottawa, ON, Canada: IEEE, 2023.
[68] LI W. Vehicle Detection in Foggy Weather Based on an Enhanced YOLO Method[C]// 2022 International Conference on Machine Vision, Automatic Identification and Detection. Nanjing, China: IOP Publishing Ltd., 2022.
[69] WANG X J, WANG C. Vehicle Multi-Target Detection in Foggy Scene Based on Foggy env-YOLO Algorithm[C]// Proceedings of the 2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE). Beijing, China: IEEE, 2022.
[70] 孙再鸣. 雨雾天气下的车辆目标检测关键技术研究[D]. 北京: 华北电力大学, 2023.
SUN Z M. Research on Key Technologies of Vehicle Detection in Haze and Rain[D]. Beijing: North China Electric Power University, 2023.
[71] CARRASCO D P, RASHWAN H A, GARCÍA M Á, et al. T-YOLO: Tiny Vehicle Detection Based on YOLO and Multi-Scale Convolutional Neural Networks[J]. IEEE Access, 2023, 11: 22430-22440.
[72] HU S, ZHAO B X, DING T J, et al. Prior-YOLO: Enhancing Intelligent Vehicle Small Object Detection with Driving Status-Informed YOLOv8[C]// 2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI). Wuhan, China: IEEE, 2023.
[73] YAN X, SHEN B B, LI H. Small Objects Detection Method for UAVs Aerial Image Based on YOLOv5s[C]// Proceedings of the 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT). Qingdao, China: IEEE, 2023.
[74] ZHANG Y H, SUN Y P, WANG Z, et al. YOLOv7-RAR for Urban Vehicle Detection[J]. Sensors, 2023, 23(4).
[75] QIU Z, BAI H, CHEN T. Special Vehicle Detection from UAV Perspective via YOLO-GNS Based Deep Learning Network[J]. Drones, 2023, 7(2).
[76] CAI Y F, LUAN T Y, GAO H B, et al. YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70.
[77] NATTE V S, KADICHERLA J, KATUKOJWALA S D, et al. Semantic Segmentation Using U-Net for Autonomous Driving[C]// 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE). Ballari, India: IEEE, 2024.
[78] LI L H, FANG M H, YIN Y H, et al. A Traffic Scene Object Detection Method Combining Deep Learning and Stereo Vision Algorithm[C]// 2021 IEEE International Conference on Real-Time Computing and Robotics (RCAR). Xining, China: IEEE, 2021.
[79] WANG H, XU Y S, WANG Z N, et al. CenterNet-Auto: A Multi-Object Visual Detection Algorithm for Autonomous Driving Scenes Based on Improved CenterNet[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(3): 742-752.

（责任编辑　解　晔）

修改稿收到日期为2025年1月6日。