

面向自动驾驶的YOLO目标检测算法研究综述*

杨飞帆 李军

(重庆交通大学,重庆 400047)

【摘要】为深入探究YOLO系列目标检测算法及其在自动驾驶领域的发展方向,对YOLO系列算法(YOLOv1~YOLOv8)的研究进程及各阶段的主要成果进行了综述,并就其在自动驾驶领域的车辆、车道线、行人、交通标志相关检测任务中的应用进行分析阐述,最后结合YOLO系列算法的特点对其未来可能的改进思路与研究趋势进行分析展望。

关键词:目标检测 YOLO算法 自动驾驶 车辆检测

中图分类号:TP391.41 **文献标志码:**A **DOI:** 10.20104/j.cnki.1674-6546.20230095

Research Review of YOLO Target Detection Algorithm for Autopilot

Yang Feifan, Li Jun

(Chongqing Jiaotong University, Chongqing 400074)

【Abstract】In order to deeply explore Yolo target detection algorithm and its development direction in the field of autonomous driving, this paper reviewed the research progress and main accomplishments of YOLO algorithms (YOLOv1~YOLOv8), analyzed and expounded its application in relevant detection tasks including vehicles, lane lines, pedestrians and traffic signs in the field of autonomous driving. Combining the characteristics of Yolo series algorithms, the paper analyzed and summarized the future improvement ideas and research trends of Yolo algorithm.

Key words: Target detection, YOLO algorithm, Autopilot, Vehicle detection

【引用格式】杨飞帆,李军.面向自动驾驶的YOLO目标检测算法研究综述[J].汽车工程师,2023(11):1-11.

YANG F F, LI J. Research Review of YOLO Target Detection Algorithm for Autopilot[J]. Automotive Engineer, 2023(11): 1-11.

1 前言

目标检测是计算机视觉领域的重要任务之一,主要处理和检测视觉上特定类(如人、动物或标志等)在数字图像上的类别和位置分布情况^[1]。目标检测通过建立有效的计算机应用模型提供计算机视觉所需要的最基本信息,即何目标在何处。其通常分为目标识别与目标定位,目标识别即特征提取与分类,目标定位则进行边界框回归。作为图像分割、图像识别、目标跟踪等计算机视觉任务的基础,目标检测算法的进步对计算机视觉发展有着重要的意义。

近年来,随着深度学习的不断发展,基于卷积神经网络的两阶段(Two-Stage)和一阶段(One-Stage)算法研究取得了较大进展。两阶段算法在数据特征提取后先生成区域提取(Region Proposal)网络,再进行样本的分类与定位回归,代表性算法有区域卷积神经网络(Region based Convolutional Neural Network, R-CNN)^[2]、快速区域卷积神经网络(Fast Region based Convolutional Neural Network, Fast R-CNN)^[3]、更快的区域卷积神经网络(Faster Region based Convolutional Neural Network, Faster R-CNN)^[4];一阶段算法则直接在网络中提取特征进行分类与回归,大幅提升了目标检测的效率,更容易

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

满足实时检测的要求^[5],代表性算法包括单步多框目标检测(Single Shot MultiBox Detector, SSD)^[6]和YOLO(You Only Look Once)系列算法。

作为一阶段算法的开山之作,YOLOv1^[7]创造性地将检测问题直接作为统一的端到端回归问题,使其相较于同时期的R-CNN算法有更快的检测速度,此后,YOLOv2~YOLOv8算法不断进行优化,为一阶段算法的实际应用提供了重要理论支撑,同时为各领域的深入研究提供了性能卓越的检测方法。

Zou等^[1]从里程碑式检测器、数据集、度量标准和主要技术进展4个方面综述了目标检测领域近20年(1990~2019年)的发展成果,并深入分析了目标检测算法的应用前景;Jiang等^[8]梳理了YOLO系列算法(YOLOv1~YOLOv5)的发展历程,总结了一些目标识别与特征选择方式。本文围绕YOLO系列算法(YOLOv1~YOLOv8)的研究进程、各研究阶段的主要成果及应用于自动驾驶领域的改进YOLO检测算法进行综述,同时对用于自动驾驶领域目标检测的算法性能提升方式进行总结,并展望其发展趋势。

2 YOLO系列算法研究进展

YOLO算法只需要浏览一次便可以识别出图中物体的准确位置和类别。短短数年间,YOLO系列算法已诞生了多个版本,2022年问世的YOLOv7^[9]在5~160帧/s范围内,检测速度和精度远超目前已知的其他检测器,2023年,具备更好兼容性能、可扩展的YOLOv8算法的源代码发布。YOLO系列主流算法的发展进程如图1所示。

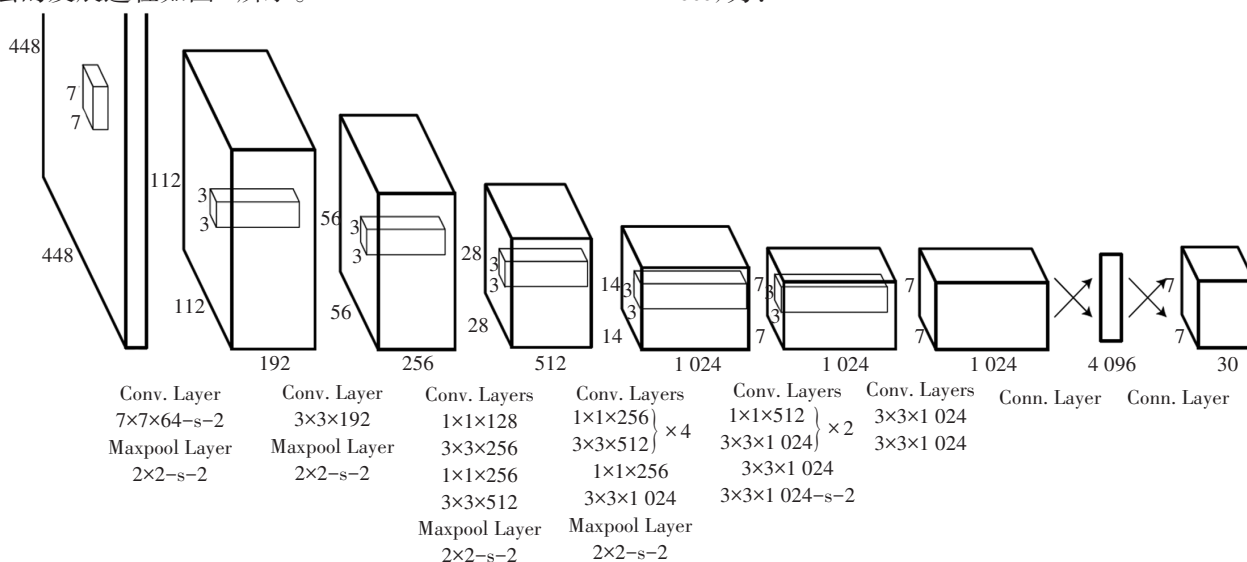


图2 YOLOv1网络结构^[7]

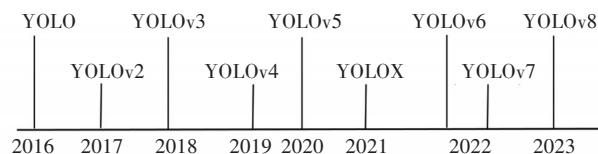


图1 YOLO系列算法发展时间线

2.1 YOLOv1

YOLOv1主干网络类似于GoogleNet^[10],采用卷积层(Conv. Layer)提取特征,最大池化层(Maxpool Layer)进行空间降采样,并由连接层(Conn. Layer)输出预测值,算法将输入图片划分为7×7个栅格,每个栅格形成2个预测框,同时预测20个类别(class)概率,每个预测框包含5个特征的预测结果(x,y,w,h,c),其中(x,y)为预测框相对于栅格的位置,(w,h)为预测框相对于图片的宽和高比例,c为置信度,表示预测框内存在被检物体的概率及预测框的位置准确度。将预测到的类别概率与预测框的置信度相乘,即可获得每个预测框中各类别的概率,以概率最高的结果作为预测结果输出,通过非极大值抑制(Non-Maximum Suppression, NMS)得到预测结果。网络参数信息如图2所示。

监督学习训练通过梯度下降和反向传播迭代进行神经元权重微调使损失函数最小化,算法的目的是使预测结果尽可能与人工标注的标准框(Ground Truth)拟合,最小化损失函数。YOLOv1的损失函数考虑了预测框坐标预测损失、预测框置信度预测损失以及类别预测损失。

预测框坐标损失(Bounding Box Coordinate Loss)为:

$$L_{\text{BBBox}} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \quad (1)$$

目标置信度损失 (Object Confidence Loss) 为:

$$L_{\text{obj}} = \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \quad (2)$$

类别预测损失 (Class Prediction Loss) 为:

$$L_{\text{class}} = \sum_{i=0}^{S^2} l_i^{\text{obj}} \sum_{c \in \text{class}} (p_i(c) - \hat{p}_i(c))^2 \quad (3)$$

式中, l_{ij}^{obj} 为第 i 个栅格的第 j 个预测框, 若负责预测目标则取 1, 否则取 0; l_{ij}^{noobj} 为第 i 个栅格中的第 j 个预测框, 若不负责预测目标则取 1, 否则取 0; l_i^{obj} 为第 i 个栅格, 若包含目标则取 1, 否则取 0; S 为栅格数量; B 为预测框数量; x_i, y_i, w_i, h_i, C_i 分别为预测框位置、宽高比例以及置信度的预测值; $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i, \hat{C}_i$ 分别为预测框位置、宽高比例以及置信度的标签值; $p_i(c), \hat{p}_i(c)$ 分别为第 i 个栅格类别的预测值与标签值; $\lambda_{\text{coord}}, \lambda_{\text{noobj}}$ 为权重; class 表示类别。

YOLOv1 推翻了两阶段算法一直采用的区域提取思路, 直接将整张图片输入神经网络, 经过卷积、池化、全连接层得到一个输出张量, 将目标检测问题转变为简单回归问题。YOLO 在 PASCAL VOC 2007 (Pattern Analysis, Statistical Modeling, and Computational Learning Visual Object Classes 2007) 数据集上测试得到其平均精度均值 (Mean Average Precision, mAP) 达 63.4%^[7], 拥有极快的检测速度和较好的通用特征提取能力, 但同时存在只支持固定尺寸的输入, 对较近、较小或群体目标检测效果较差和定位并不精确等问题。

2.2 YOLOv2

YOLOv2^[11] 在 YOLOv1 基础上, 结合视觉几何组 (Visual Geometry Group, VGG) 网络构建出新的 DarkNet-19 网络, 新网络采用 1×1 的卷积进行通道降维、 3×3 的卷积进行特征提取、5 层最大池化层进行 5 次下采样。DarkNet-19 网络结构如表 1 所示。

表 1 DarkNet-19 网络结构^[11]

类型	参数	核尺寸/步距	输出
卷积层	32	3×3	224×224
最大池化层		$2 \times 2 / 2$	112×112
卷积层	64	3×3	112×112
最大池化层		$2 \times 2 / 2$	56×56
卷积层	128	3×3	56×56
卷积层	64	1×1	56×56
卷积层	128	3×3	56×56
最大池化层		$2 \times 2 / 2$	28×28
卷积层	256	3×3	28×28
卷积层	128	1×1	28×28
卷积层	256	3×3	28×28
最大池化层		$2 \times 2 / 2$	14×14
卷积层	512	3×3	14×14
卷积层	256	1×1	14×14
卷积层	512	3×3	14×14
卷积层	256	1×1	14×14
卷积层	512	3×3	14×14
最大池化层		$2 \times 2 / 2$	7×7
卷积层	1 024	3×3	7×7
卷积层	512	1×1	7×7
卷积层	1 024	3×3	7×7
卷积层	512	1×1	7×7
卷积层	1 024	3×3	7×7
卷积层	1 000	1×1	7×7
平均池化层		全局	1 000
Softmax 层			

YOLOv2 引入 Faster R-CNN 中先验框 (Anchor Box) 思想, 通过先验框同时预测类别和位置, 一个栅格内的 5 个先验框可以分别预测 5 种类别, 预测边界框的数量也由 $7 \times 7 \times 2$ 提高到 $13 \times 13 \times 5$, 提高了不同尺寸、不同长宽比物体预测的准确率。

除以上改进外, YOLOv2 通过采用批归一化 (Batch Normalization, BN)、高分辨率图像分类器 (High Resolution Classifier)、多尺度训练 (Multi-Scale Training) 等方式对 YOLO 算法进行了全方位改进, 支持少样本对象检测的同时有效提高了召回率和定位准确率。不足之处在于, 没有进行多尺度特征的结合预测, 传递模块 (Pass-Through Module) 的使用在提升细粒度特征的同时也对特征的空间分布产生了一定影响, 以及对小目标的检测能力没有明显提升。

2.3 YOLOv3

YOLOv3^[12]的主干网络DarkNet-53包含卷积层(Convolutional Layer)、残差层(Residual Layer)、特征融合层(Feature Fusion Layer),网络层数的加深提高了检测精度,大量残差网络模块的引入减少了由网络层数加深引起的梯度下降问题,金字塔池化模块的引入可以实现多尺寸的输入和统一尺寸的输出。卷积层与残差层结构对比如图3所示。

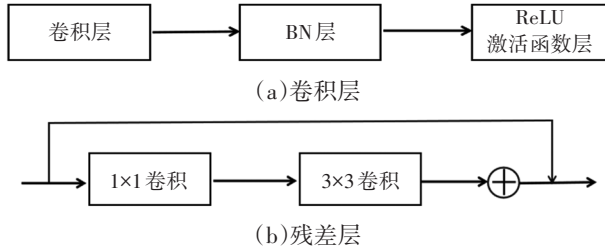


图3 卷积层与残差层对比

借鉴SSD的思想,YOLOv3引入了多尺度特征预测,采用32倍下采样(13×13大小特征图)、16倍下采样(26×26大小特征图)、8倍下采样(52×52大小特征图)3个不同尺寸分别进行大、中、小目标检测,构造出9个先验框,有效提高了尺度特征差异较大情况下的目标检测准确性,尤其是在对小目标的检测能力上产生了质的飞跃。在标签分配策略方面,YOLOv3采用多标签多分类的逻辑(Logistic)层实现复杂环境下多类别物体重叠时良好的识别效果,多个二分类任务的划分使其支持一个物体的多个标签输出。

相较于YOLOv1,YOLOv3的损失函数在置信度损失和类别损失方面进行了优化,采用交叉熵计算缩短收敛时间,其损失函数为:

$$\begin{aligned}
 L = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \\
 & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] - \\
 & \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{obj}} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i) \right] - \\
 & \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{\text{noobj}} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i) \right] - \\
 & \sum_{i=0}^{S^2} l_{ij}^{\text{obj}} \sum_{c \in \text{class}} \left[\hat{p}_i \log(p_i) + (1 - \hat{p}_i) \log(1 - p_i) \right]
 \end{aligned}
 \tag{4}$$

YOLOv3的主要改进体现在主干网络、多尺度特征融合以及标签分配策略方面,虽然模型的召回率相对较低,但是极大提升了整体检测精度,可以

在不提高输入图片分辨率的情况下实现对小目标的精确检测,使YOLOv3成为一阶段算法中的里程碑式算法,在工业领域得到广泛应用。

2.4 YOLOv4

YOLOv4^[13]的主干网络是在DarkNet-53基础上借鉴Wang等^[14]提出的跨阶段局部网络(Cross Stage Partial Network, CSPNet)升级得到的CSP-DarkNet-53,可以有效增强CNN的学习能力、降低模型计算的参数量和内存成本,激活函数采用Mish^[15]替换修正线性单元(Rectified Linear Unit, ReLU),在保证准确度的同时平滑梯度,采用DropBlock^[16]正则化方式提高网络的泛化性能和抗过拟合能力。在输入端借鉴Sang等^[17]提出的CutMix数据增强方式,提出Mosaic数据增强方法将多张图片随机缩放拼接为一张图片,扩充数据的多样性。

在特征提取方面,YOLOv4采用空间金字塔池化网络(Spatial Pyramid Pooling Net, SPP Net)增大感受野,且在不降低运算速度的同时显著分离上下文特征信息,由路径聚合网络(Path Aggregation Network, PANet)代替特征金字塔网络(Feature Pyramid Network, FPN)进行参数聚合,使用张量连接代替捷径拼接。YOLOv4中PANet的结构如图4所示。

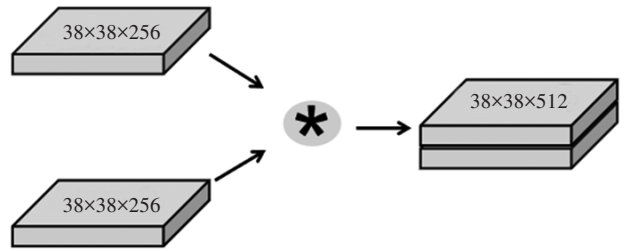


图4 YOLOv4中的PANet结构

YOLOv4的损失函数在预测坐标损失中采用了考虑边框重合度、中心距和宽高比的完全交并比损失(Complete Intersection over Union Loss, CIoU Loss)函数全面体现预测框与真实框的位置关系,有利于预测框的收敛。

YOLOv4算法在YOLO算法的框架上,广泛整合目标检测领域的优化策略进行全方位优化,获得了高效、快速、精确的检测特性,在Tesla V100图形处理器(Graphics Processing Unit, GPU)上以65帧/s的图像处理速度获得COCO数据集上43.5%的平均精度(Average Precision, AP)^[10],但是由于融合了大量的优化策略导致其模型复杂度较高。

2.5 YOLOv5

由YOLOv5的源码可知,其主干网络在YOLOv4

基础上进行了简单改进:将网络的第1层Focus模块替换为6×6的卷积层,在GPU设备上运行更加高效;引入CSPNet优化了主干网络中的梯度重复,在减少参数数量的同时压缩了模型体积;将SPP模块替换为更高效的快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF),大幅提升了运算速度;采用了更多的数据增强方式,如Mosaic、Copy Paste、Random Affine、MixUp,有效提升了模型的泛化能力与鲁棒性^[18]。

YOLOv5采用多尺度训练(Multi-Scale Training),可以包含更多尺寸的图片输入;采用K-均值(K-Means)聚类算法对数据集中的边界框进行聚类分析,进而得到合适的先验框;采用预热和余弦LR调度程序(Warmup and Cosine LR Scheduler)进行学习率的调整。

YOLOv5基本沿用了YOLOv4的结构,同时根据通道的尺度缩放提供YOLOv5S、YOLOv5M、YOLOv5L、YOLOv5X 4种不同网络结构,支持在检测精度与速度间进行一定权衡,其中YOLOv5S通常部署于嵌入式设备中构建轻量级网络,相较于YOLOv4,YOLOv5在牺牲部分检测性能的情况下提升了模型的灵活性、检测速度、部署方便性。

2.6 YOLOX

YOLOX^[19]基于YOLOv3的网络结构新增2个3×3卷积将分类和回归任务解耦,达到提升精度和收敛速度的效果,对于FPN特征采用1×1卷积层减小特征通道,然后添加并行分支分别用于分类和回归任务。YOLOv3~YOLOv5耦合头与YOLOX解耦头结构对比如图5所示。

YOLOX采用无先验框(Anchor-Free)方法预先定义预测目标的中心作为原点,基于原点确定网格左上角的预测框高度和宽度2个偏移量作为预测过程取代了YOLOv1~YOLOv5惯用的预测的平移边框过程,避免了聚类引起的时间浪费和检测区域限制,减少了预测框的生成;采用正负样本匹配策略(Advanced Label Assigning Strategy, SimOTA)^[20]为不同目标设定不同的正样本数量,避免了额外的参数优化,提升了检测速度;提供一种通过添加2层卷积、一对一标签分配模块在稍微降低性能的条件下实现NMS Free(Non-Maximum Suppression Free)的端到端检测。

YOLOX是在YOLOv3的基础上融合目标检测领域的无先验框检测器、高级标签分配、端到端检测器得到的检测模型算法,YOLOX-L在Tesla V100上以68.9帧/s的图像处理速度在COCO数据集上实现了50.0%平均精度,同时支持开放神经网络交换(Open Neural Network Exchange, ONNX)、TensorRT、NCNN(Nihao Convolutional Neural Network)、开放视觉推理及神经网络优化(Open Visual Inference and Neural network Optimization, Open VINO)多个部署版本^[19]。

2.7 YOLOv6

YOLOv6^[21]的主干网络Efficient Rep是引入重参数化视觉几何组(Re-parameterization Visual Geometry Group, Rep VGG)^[22]结构设计出的高效可重参数化网络,可以实现速度与精度的平衡,同时优化SPPF得到更高效的SimSPPF(Simplified SPPF)模块,并采取混合通道策略得到更高效的解耦头结

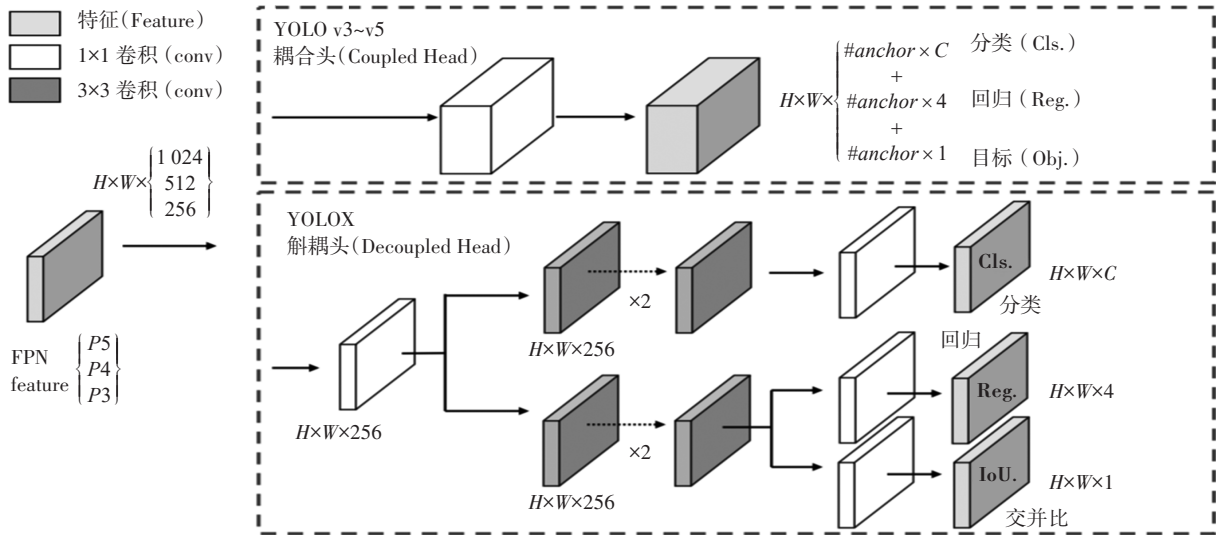


图5 YOLOv3~YOLOv5耦合头与YOLOX解耦头结构对比^[19]

构,在分类和回归过程中均进行了自我提升。

YOLOv6采取了与YOLOX相同的无先验框检测方式和SimOTA样本分配策略,采用标准交并比边界回归损失(Standard Intersection over Union Loss, SIoU Loss)函数进行监督网络学习,提升回归精度,同时引入额外的通用实践性技巧(训练更多轮次、自蒸馏策略),提高了工业领域的应用性能。

YOLOv6在网络设计、标签分配、损失函数、工业便利化方面进行了改进,YOLOv6-nano在COCO数据集上的精度达35.0% AP,在推理专用GPU T4上推理速度达1 242 帧/s,YOLOv6-s在COCO数据集上的精度为43.1% AP,在T4上推理速度为520 帧/s。YOLOv6支持CPU(Open VINO)、ARM(MNN(Mobile Neural Network)、NCNN)等不同平台^[21],为不同工业场景提供了不同规模、各有侧重的目标检测框架网络,但是YOLOv6的优化策略大多针对小型网络结构设计,当网络深度和宽度增加后,检测效果可能下降。

2.8 YOLOv7

YOLOv7^[9]提出了高效层聚合网络(Efficient Layer Aggregation Networks, ELAN)的扩展版本,即扩展高效层聚合网络(Extended ELAN, E-ELAN),通过拓展(Expand)、重组(Shuffle)结构实现不改变原有梯度路径的多次模块堆叠,使网络更好地进行学习和收敛。YOLOv7借鉴YOLOv4、YOLOv5的方式,通过对网络深度和宽度的缩放来扩展效率层网络,利用梯度传播路径的概念分析了重参数化模块的网络搭配,最终提出无恒等映射的重参数化卷积模块。

YOLOv7提供了全新的标签分配策略:使用引导头的预测结果引导生成由粗到细的层次标签用作引导头和辅助头的学习,使引导头具备更强的残余信息学习能力,为不同的分支输出动态目标。

YOLOv7设计出几种可以训练的免费包,使检测器在不提升成本的情况下大幅提高检测精度,为实时探测器提出扩展、缩放策略从而对参数进行高效利用,为重参数化模块代替原有模块、动态标签处理输出层的分配提出了处理方法,并优化了网络架构和训练过程。在COCO数据集上,YOLOv7-E6E的精确度为56.8% AP,相较于YOLOv6-s(43.1% AP)取得了较大提升^[9]。

2.9 YOLOv8

YOLOv8和YOLOv5均由Ultralytics团队发布,

提供了一个全新的SOTA(State Of The Art)模型并基于缩放系数提供N、S、M、L、X尺度的不同大小模型以满足不同场景需求,在YOLOv5的基础上使用基于YOLOv7中E-ELAN模块的C2f模块代替YOLOv5中的C3模块,增加了更多的梯度流信息,并参考YOLOX,采用了解耦头结构将分类和回归解耦。

YOLOv8采用无先验框检测方式及任务对齐学习(Task Alignment Learning, TAL)动态匹配的方式,从任务对齐的角度出发,依据所设计的指标动态选择高质量的先验框作为正样本融入损失函数的设计,使用DFL(Distribution Focal Loss)和CIoU Loss作为回归损失^[23]。

YOLOv8是在YOLOv5的基础上参考了YOLOX、YOLOv6、YOLOv7算法而提出的包括图像分类、无先验框物体检测和实例分割的高效算法,同时具备可扩展性,与YOLO算法框架兼容,可以切换、比较不同版本的性能。

2.10 YOLO系列算法的主要改进方向

YOLO系列目标检测算法专注于平衡检测速度与精度,旨在不牺牲检测精度的同时提升检测的实时性,YOLO的每个版本都在维持甚至提升框架实时性的同时提升精确性,YOLOv4之后的模型实现了速度与精度的进一步权衡,提供不同比例的模型以适应特定应用环境,更适合使用者进行权衡选择。YOLO系列主要改进方向如表2所示,其未来的发展将致力于获得更快、更高效的网络架构,更有效的特征积分方法,更准确的检测方式,鲁棒性更强的损失函数,更有效的标签分配方法,更有效的训练方式以及更优良的兼容性能。

3 YOLO算法在自动驾驶领域中的应用

YOLO系列算法以其准确、高效的检测效果被应用到各检测领域,包括医疗口罩检测^[24]、工程安全帽检测^[25]、农业作物检测^[26]等,在自动驾驶领域,YOLO算法也凭借其实时性和较高的准确性在车辆、车道线、行人和交通标志等检测任务中得到了广泛应用,并在相关学者的不断改进下取得了良好的效果。

3.1 车辆检测

随着城市交通的不断发展,机动车、非机动车等车辆检测的压力不断增加,良好的车辆检测系统可以有效缓解交通压力。

霍爱清等^[27]对YOLOv3算法的残差结构单元和

表2 YOLO系列主要改进方向

算法	主要改进	优缺点	使用场景
YOLOv1	Anchor、高效的框架结构	网络简单、检测快,但定位差、小目标检测效果差	目标检测
YOLOv2	BN层、DarkNet-19、聚类选择先验框、引入传递模块	减少定位错误、分类精确度高,但准确度不高	目标检测
YOLOv3	DarkNet-53、多尺度特征融合、在3个尺寸上进行预测	精确度提高,速度是SSD算法的3倍以上,但在边界框预测严格的情况下精度不足	多尺度目标检测
YOLOv4	CSP-DarkNet-53、FPN+PANet、Mish激活函数、Mosaic数据增强	小目标检测效果提高、模型复杂度高	高精度实时目标检测
YOLOv5	提升模型的性能、SPPF、提供多种模型	模型尺寸小、部署成本低、灵活性高、检测速度快,性能有待提高	目标检测
YOLOX	解耦头、Anchor-Free、高级标签分配、端到端检测器	改善基于锚点的管道过度优化问题,有多种网络框架选择	高精度实时目标检测
YOLOv6	Efficient Rep、Anchor-Free、SimSPP	检测速度快、更偏向工业应用领域	工业应用目标检测
YOLOv7	重参数化模块、辅助头检测、动态标签分配策略	高精度、检测速度快	高精度实时目标检测
YOLOv8	解耦头、Anchor-Free、TAL	可扩展性、兼容YOLO系列框架	工程领域目标检测

特征提取网络进行改进,采用深度可分离卷积代替了传统卷积,改用Swish激活函数减小网络模型的计算量,引入CIoU作为损失函数,缓解了目标回归框的发散,改进后算法的mAP提升了3.68%,参数量下降了67%,拥有较高的准确性、适用性、鲁棒性。Wu等^[28]通过调整YOLOv5s的网络层结构使其更适合嵌入式设备,应用于车辆和距离检测,在mAP略微降低(3%)的情况下检测速度提升了19帧/s。章程军等^[29]在YOLOv5s网络基础上通过引入一次性聚合模块优化网络加强特征提取,引入非局部注意力机制增强特征,采用加权非极大值抑制的方法进行预测框筛选,在自制的车辆数据集上mAP提升了3%,可以在车辆密集和各类光照条件下完成检测。赵帅豪等^[30]通过改进YOLOX网络结构,采用深度可分离卷积代替传统卷积,引入卷积块注意力模块(Convolutional Block Attention Module, CBAM)融合特征,采用CIoU优化损失函数,提升定位精度,得到的模型在自建数据集上测试时mAP提升了2.01%。Zhang等^[31]提出使用翻转拼接算法加强YOLOv5网络对小目标的感知,在自建的数据集上的测试结果表明,该数据增强方式可以有效提高车辆的检测准确率。

在车辆检测任务中,通过改善损失函数可以提升定位精度,采用翻转拼接算法可以提升网络对小目标特征的提取能力,采用可分离卷积或调整网络深度可以在牺牲部分精度的条件下拥有更快的检测速度以适用于嵌入式设备,采用聚合模块可以有

效提高特征提取能力,适用于车辆密集场合。

3.2 车道线检测

随着城市交通的发展与智能驾驶的普及,车道线信息的采集可以为车辆的导航提供参考、保证车辆的安全,是辅助驾驶的重要组成部分。

崔文靛等^[32]在YOLOv3基础上采用K-Means++代替K-Means进行聚类分析,通过改进网络架构得到共101层新网络架构,该网络结构可以有效降低小目标检测场景下卷积层过深导致的特征消失情况,在美国加利福尼亚理工大学车道线数据集上的测试结果表明,算法的mAP提高了11%。Dong等^[33]在YOLOv4的基础上提出YOLOP,构建了高效的多任务网络,采用1个特征提取编码器和3个解码器同时处理物体检测、车道线检测和可行驶区域分割3个任务。丁玉玲等^[34]提出采用K-Means++聚类对YOLOv4算法进行参数优化,使其更便于检测车道线,同时根据车道线特征增加了纵轴的检测密度,取得了更好的检测效果。Ji等^[35]通过分析车道线特征,在YOLOv3的基础上调整网络结构的卷积层数、划分车道线图网格、调整检测尺寸使之适合车道线小目标检测,同时改进簇中心距以及损失函数参数,获得了92.03%的mAP。

在车道线检测任务中,车道线检测数据集通常为彩色图像,算法在信息复杂、图像信息不多的情况下训练效果有待提高。通过K-Means++聚类分析对算法进行参数优化可以使算法适用于车道线检测,通过改变特定网络结构可以提升网络针对车

道线目标的特征提取能力,通过增加解码器也可以将车道线检测并入其他检测任务。关于车道线的检测,还需要更多的数据集支持。

3.3 行人检测

交通领域的行人作为最广泛的特殊参与群体,是安全检测中最主要的检测对象之一,检测行人的体态语言并预测其行为动作对自动驾驶领域有着重要意义。

Liu等^[36]在YOLOv3网络中引入压缩与激励网络(Squeeze and Excitation Network, SENet)和距离-交并比损失函数(Distance-Intersection over Union Loss, DIoU Loss),考虑真实框与预测框的中心点距离,提升了对小尺寸的非机动车、行人的检测精度。李翔等^[37]在YOLOv3基础上通过上采样提升金字塔分辨率并引入浅层特征增强相邻特征层的差异,生成可以区分高重叠度的深层特征,同时引入优化预测框方差均值降低假阳率的聚类损失算法,有效提升遮挡情况下行人的检测准确率。Yan等^[38]基于YOLOv5提出通过融合多尺度感受野捕捉图像的信息,引入注意力分支增强特征的表达能力,解决了行人等小目标特征容易消失的问题。李克文等^[39]通过引入一种边界增强算子,从边界的极限点中自适应地提取边界特征来增强已有点特征,增加目标检测尺度,基于目标检测中目标实例特性及改进网络模型引入完全交并比函数对YOLOv3损失函数进行改进,可以有效提高拥挤环境下目标检测的准确性。Liu等^[40]利用MobileNetv3优化特征提取网络,在YOLOv4的基础上加入CBAM提高检测效率,采用可分离卷积替换3×3卷积减小参数量,有效提升了行人检测的精度与速度,在PASCAL VOC 2007+2012数据集上进行测试,与YOLOv4的测试结果相比,mAP提升了5.00%。方康等^[41]通过在YOLOX网络中融入多尺度特征增强模块丰富特征图中的语义信息,将多模态特征提取解耦为特性与共性特征提取,在KAIST多光谱行人检测基准数据集(KAIST Multispectral Pedestrian Detection Benchmark Dataset)和FLIR热成像数据集(FLIR Thermal Dataset)上的验证结果表明,该方法可以有效提高多光谱行人的识别。

在行人检测任务中,引入SENet模块根据损失函数学习特征权重可以使有效特征图所占权重重大,无效或效果小的特征图所占权重小,使训练模型达到更好的效果;通过提升特征提取与融合能力可以

在进行小目标的特征提取时融合目标图像的各种特征信息,实现精度与速度平衡。在进行行人等易遮挡的小目标检测时,引入上下文信息、注意力模块,以及加强特征提取与融合往往会提高模型的复杂程度,增加运算成本,故融合更多轻量化网络模型设计至关重要。

3.4 交通标志检测

交通标志的检测作为智能驾驶的关键任务,直接关系到驾驶的安全性,有效且可靠的交通标志检测算法可以在很大程度上减少安全事故的发生。

Dewi等^[42]通过生成式对抗网络生成交通标志进行增强数据集的训练,在YOLOv3、YOLOv4模型上的mAP分别为84.9%、89.33%。Gao等^[43]在道路交通信号图像上进行了检测试验,证明了YOLOv4拥有较高的交通标志识别性能。潘惠萍等^[44]在YOLOv4的基础上结合SPP模块消除网络的尺寸限制,修改原特征图尺寸的分辨率增大感受野,利用行车记录仪采集的交通标志进行检测,平均检测时间为0.449 s,mAP达到99.0%。钱伍等^[45]在YOLOv5算法基础上使用复合数据增强对模型输入进行增强处理,设计多尺度代替固定尺度训练,构建多尺度特征融合网络,直接提升了模型对交通信号的检测能力,在自采集数据集上检测时间可以达到9.5 ms,mAP为99.8%。

在交通标志检测任务中,通过生成式对抗性网络可以生成更多的交通标志图像,结合真实图像以增强数据集,通过该策略扩充训练数据集规模,可以增强检测模型的鲁棒性;通过修改特征图尺寸的分辨率增大感受野、构建多尺度特征融合网络可以有效提升交通标志检测算法的性能。

车辆领域的YOLO系列目标检测优化方式如表3所示,此外,YOLO系列算法应用于汽车牌照检测^[46]、汽车零部件检测^[47]和车辆徽标检测^[48]等领域,在极大解放人力、缓解交通压力的同时取得了良好的检测效果,为安全行车提供了便利。

4 研究展望

YOLO系列算法历经一代代的发展,不断融合目标检测领域的研究成果,提升检测性能,但距离实现自动驾驶领域的广泛应用还面临一些挑战,仍需在以下几个方面开展进一步研究:

a. YOLO系列在进行复杂场景下的小目标检测时,细节特征的捕捉难以达到较高的检测精度,国

表3 车辆目标检测常见的优化方式

算法优化方向	优化模式	优势	局限
基础网络	轻量化卷积网络	缩短检测时间	检测性能有所下降
多尺度特征融合	卷积层特征融合	精度提高	检测时间延长
引入注意力模型	引入注意力模型	深层特征提取加强	计算量增加
损失函数	分类、回归损失	鲁棒性更好	计算量增加
非极大值抑制	去除冗余框	定位更加精确	计算量增加

内外学者通常通过整合上下文语意信息、联系局部与整体信息提升检测精度,但会增加模型的复杂度与计算量,因此需要一种更好的特征融合提取结构来提升算法性能。

b. YOLO算法当前主要面向计算机端,在边缘计算需求增加的背景下,面向嵌入式AI设备的应用要求YOLO算法在保证检测效果的前提下更轻、更快。

c. YOLO算法在极端环境条件下的检测效果还有待提高,面对风霜雨雪等条件干扰时如何保证其实际的检测效果仍需不断地研究与突破。

5 结束语

深度学习的发展推动着目标检测算法的进步,从YOLOv1到YOLOv8,检测精度与速度、框架的兼容性能逐步提高,YOLOv8所具备的可扩展性使得YOLOv1~YOLOv7可以实现切换进行性能比较。同时,YOLO系列算法的发展应用于智能驾驶领域也取得了一定进展,在进行车辆、车道线、行人、交通标志等不同检测任务时通常对检测模型采取不同的优化方式:引入上下文信息、增强特征表示可以在大多数场景下提升检测效果;在进行行人、车辆等密集且背景复杂的目标检测任务时可以通过改善损失函数提升训练效果,进而提升模型鲁棒性;在交通灯、车道线检测时通常通过聚类进行精度提升。随着YOLO系列算法的不断发展,更多适用于自动驾驶领域的YOLO系列改进算法将推动该领域的扎实进步。

参 考 文 献

[1] ZOU Z X, CHEN K Y, SHI Z W, et al. Object Detection in 20 Years: A Survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.

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

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

[4] 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: 91-99.

[5] 薛松, 钱立志, 张航, 等. 多源末制导弹载融合图像目标检测研究进展[J]. 火箭与制导学报, 2021, 41(3): 67-75.

XUE S, QIAN L Z, ZHANG H, et al. Review of Key Technologies for Target Detection of Multi-Source Terminal Missile-Based Fusion Image[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2021, 41(3): 67-75.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector[C]// European Conference on Computer Vision. Berlin: Springer International Publishing, 2016: 21-37.

[7] 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, USA: IEEE, 2016: 779-788.

[8] JIANG P Y, ERGU D, LIU F Y, et al. A Review of Yolo Algorithm Developments[J]. Procedia Computer Science, 2022, 199: 1066-1073.

[9] 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]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 7464-7475.

[10] SZEGEDY C, LIU W, JIA Y Q, et al. Going Deeper with Convolutions[C]// IEEE Conference on Computer Vision & Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2015: 1-9.

[11] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6517-6525.

[12] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. (2018-04-08)[2023-04-15]. <https://>

- arxiv.org/abs/1804.02767.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: Scaling Cross Stage Partial Network[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 13029–13038.
- [14] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 390–391.
- [15] MISRA D. Mish: A Self Regularized Non-Monotonic Neural Activation Function[J]. Machine Learning, 2019, 18 (5): 5–21.
- [16] GHIASI G, LIN T Y, LE Q V. DropBlock: A Regularization Method for Convolutional Networks[C]// Proceedings of 33rd Conference on Neural Information Processing Systems. Montréal, Canada: NIPS, 2018: 10750–10760.
- [17] YUN S, HAN D, CHUN S, et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features[C]// 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019: 6023–6032.
- [18] JOCHER G, STOKEN A, BOROVEC J, et al. Ultralytics/YOLOv5: v3.1–Bug Fixes and Performance Improvements [EB/OL]. (2020–10–29) [2023–04–15]. <https://zenodo.org/record/4154370>.
- [19] GE Z, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. (2021–08–06)[2023–04–15]. <https://arxiv.org/abs/2107.08430>.
- [20] GE Z, LIU S T, LI Z M, et al. OTA: Optimal Transport Assignment for Object Detection[C]// 2021 IEEE/CVF Conference on Computer Vision & Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 303–312.
- [21] LI C Y, LI L L, JIANG H L, et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications[EB/OL]. (2022–09–07) [2023–04–15]. <https://arxiv.org/abs/2209.02976>.
- [22] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-Style ConvNets Great Again[C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 13733–13742.
- [23] TERVEN J R, CORDOVA-ESPARAZA D M, et al. Ultralytics/YOLOv8: YOLOv8 Docs[EB/OL]. (2023–01–10) [2023–04–15]. <https://ultralytics.com/yolov8>.
- [24] 陈继平, 陈永平, 谢懿, 等. Ghost-YOLO: 轻量化口罩人脸检测算法[J]. 信号处理, 2022, 38(9): 1954–1964.
- CHEN J P, CHEN Y P, XIE Y, et al. Ghost-YOLO: Lightweight Masked Face Detection Algorithm[J]. Journal of Signal Processing, 2022, 38(9): 1954–1964.
- [25] 金雨芳, 吴祥, 董辉, 等. 基于改进YOLO v4的安全帽佩戴检测算法[J]. 计算机科学, 2021, 48(11): 268–275.
- JIN Y F, WU X, DONG H, et al. Improved YOLOv4 Algorithm for Safety Helmet Wearing Detection[J]. Computer Science, 2021, 48(11): 268–275.
- [26] 卢伟, 邹明莹, 施浩楠, 等. 基于YOLO v5-TL的褐菇采摘视觉识别-测量-定位技术[J]. 农业机械学报, 2022, 53 (11): 341–348.
- LU W, ZOU M X, SHI H N, et al. Technology of Visual Identification – Measuring – Location for Brown Mushroom Picking Based on YOLO v5-TL[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(11): 341–348.
- [27] 霍爱清, 杨玉艳, 谢国坤. 基于改进YOLOv3算法的车辆目标检测[J]. 计算机工程与设计, 2022, 43(7): 1981–1989.
- HUO A Q, YANG Y X, XIE G K. Vehicle Target Detection Based on Improved YOLOv3 Algorithm[J]. Computer Engineering and Design, 2022, 43(7): 1981–1989.
- [28] WU T H, WANG T W, LIU Y Q. Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network [C]// 2021 3rd World Symposium on Artificial Intelligence. Guangzhou, China: IEEE, 2021: 24–28.
- [29] 章程军, 胡晓兵, 牛洪超. 基于改进YOLOv5的车辆目标检测研究[J]. 四川大学学报(自然科学版), 2022, 59(5): 79–87.
- ZHANG C J, HU X B, NIU H C. Vehicle Object Detection Based on Improved YOLOv5 Method[J]. Journal of Sichuan University (Natural Science Edition), 2022, 59(5): 79–87.
- [30] 赵帅豪. 基于YOLOX的车辆检测[J]. 计算机系统应用, 2022, 31(12): 195–202.
- ZHAO S H. Vehicle Detection Based on YOLOX[J]. Computer Systems & Applications, 2022, 31(12): 195–202.
- [31] ZHANG Y, GUO Z Y, WU J Q, et al. Real-Time Vehicle Detection Based on Improved YOLO v5[J]. Sustainability, 2022, 14(19): 12274–12274.
- [32] 崔文靓, 王玉静, 康守强, 等. 基于改进YOLOv3算法的公路车道线检测方法[J]. 自动化学报, 2022, 48(6): 1560–1568.
- CUI W L, WANG Y J, KANG S Q, et al. Road Lane Line Detection Method Based on Improved YOLOv3 Algorithm [J]. Acta Automatica Sinica, 2022, 48(6): 1560–1568.
- [33] WU D, LIAO M W, ZHANG W T, et al. YOLOP: You Only Look Once for Panoptic Driving Perception[J]. Machine Intelligence Research, 2022, 19(6): 550–562.

- [34] 丁玉玲. 基于视觉的车道线检测算法研究与应用[D]. 北京: 北方工业大学, 2022.
DING Y L. Research and Application of Lane Line Detection Algorithm Based on Vision[D]. Beijing: North China University of Technology, 2022.
- [35] JI G Q, ZHENG Y C. Lane Line Detection System Based on Improved YoloV3 Algorithm[EB/OL]. (2021-10-13)[2023-04-15]. <https://doi.org/10.21203/rs.3.rs-961172/v1>.
- [36] LIU S S, GENG Y N, SONG Y H, et al. Research on Small Target Pedestrian Detection Algorithm Based on Improved YOLOv3[C]// 14th International Conference on Genetic and Evolutionary Computing. Singapore: Springer, 2021: 203-214.
- [37] 李翔, 何森, 罗海波. 一种面向遮挡行人检测的改进YOLOv3算法[J]. 光学学报, 2022, 42(14): 160-169.
LI X, HE M, LUO H B. Occluded Pedestrian Detection Algorithm Based on Improved YOLOv3[J]. Acta Optica Sinica, 2022, 42(14): 160-169.
- [38] YAN B N, LI J X, YANG Z Z, et al. AIE-YOLO: Auxiliary Information Enhanced YOLO for Small Object Detection[J]. Sensors, 2022, 22(21).
- [39] 李克文, 杨建涛, 黄宗超. 基于边界极限点特征的改进YOLOv3目标检测[J]. 计算机应用, 2023, 43(1): 81-87.
LI K W, YANG J T, HUANG Z C. Improved YOLOv3 Target Detection Based on Boundary Limit Point Features [J]. Journal of Computer Applications, 2023, 43(1): 81-87.
- [40] LIU L S, KE C Y, LIN H, et al. Research on Pedestrian Detection Algorithm Based on MobileNet-YoLo[J]. Computational Intelligence and Neuroscience, 2022, 2022.
- [41] 方康, 黄琴, 王克琪, 等. 基于改进YOLOX的多光谱行人检测算法[J/OL]. 小型微型计算机系统 (2022-11-28)[2023-04-15]. <https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C45S0n9fL2suRadTyEV12pW9UrhTDCdPD64UUbHyA2NF-emH4E0YhenJi0k-HZWQvvASrFfVXrkX6JXZ2D8kDSQd&uniplatform=NZKPT>.
FANG K, HUANG Q, WANG K Q, et al. Multispectral Pedestrian Detection Based on Improved YOLOX[J/OL]. Journal of Chinese Computer Systems (2022-11-28)[2023-04-15]. <https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C45S0n9fL2suRadTyEV12pW9UrhTDCdPD64UUbHyA2NF-emH4E0YhenJi0k-HZWQvvASrFfVXrkX6JXZ2D8kDSQd&uniplatform=NZKPT>.
- [42] DEWI C, CHEN R C, LIU Y T, et al. Yolo V4 for Advanced Traffic Sign Recognition with Synthetic Training Data Generated by Various GAN[J]. IEEE Access, 2021, 9: 97228-97242.
- [43] GAO H, WANG W H, YANG C J, et al. Traffic Signal Image Detection Technology Based on YOLO[C]// Journal of Physics: Conference Series, 2021, 1961(1).
- [44] 潘惠苹, 王敏琴, 张福泉. 基于优化YOLO-V4的交通标志检测识别方法[J]. 计算机科学, 2022, 49(11): 179-184.
PAN H P, WANG M Q, ZHANG F Q. Traffic Sign Detection and Recognition Method Based on Optimized YOLO-V4[J]. Computer Science, 2022, 49(11): 179-184.
- [45] 钱伍, 王国中, 李国平. 多尺度YOLOv5的交通灯检测算法[J]. 软件导刊, 2022, 21(9): 19-25.
QIAN W, WANG G Z, LI G P. Traffic Light Detection Algorithm Based on Multi Scale YOLOv5[J]. Software Guide, 2022, 21(9): 19-25.
- [46] HENDRY, CHEN R C. Automatic License Plate Recognition via Sliding-Window Darknet-YOLO Deep Learning[J]. Image and Vision Computing, 2019, 87: 47-56.
- [47] 彭露露, 朱媛媛, 金文倩, 等. 基于改进YOLOv4的汽车钢铁零件表面缺陷检测[J]. 计算机与现代化, 2022(9): 32-39+50.
PENG L L, ZHU Y Y, JIN W Q, et al. Surface Defect Detection of Automotive Steel Parts Based on Improved YOLOv4[J]. Computer and Modernization, 2022(9): 32-39+50.
- [48] YANG F, YANG D M, HE Z M, et al. Automobile Fine-Grained Detection Algorithm Based on Multi-Improved YOLOv3 in Smart Streetlights[J]. Algorithms, 2020, 13(5).
(责任编辑 斛 畔)
- 修改稿收到日期为2023年4月16日。