



计算机科学

COMPUTER SCIENCE

基于自注意力与双向特征融合的道路障碍物检测方法

李婷, 赵尔敦, 杨军

引用本文

李婷, 赵尔敦, 杨军. 基于自注意力与双向特征融合的道路障碍物检测方法[J]. 计算机科学, 2024, 51(11A): 240100138-5.

LI Ting, ZHAO Erdun, YANG Jun. Road Obstacle Detection Method Based on Self-attention and Bidirectional Feature Fusion [J]. Computer Science, 2024, 51(11A): 240100138-5.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于加权特征融合的物联网设备识别方法](#)

IoT Devices Identification Method Based on Weighted Feature Fusion

计算机科学, 2024, 51(11A): 240100137-9. <https://doi.org/10.11896/jsjcx.240100137>

[基于相对位置编码转换器模块的深度步态识别网络](#)

Deep Gait Recognition Network Based on Relative Position Encoding Transformer

计算机科学, 2024, 51(11A): 240400064-6. <https://doi.org/10.11896/jsjcx.240400064>

[基于多模态数据与融合深度网络的自动睡眠分期方法](#)

Automatic Sleep Staging Based on Multimodal Data and Fusion Deep Network

计算机科学, 2024, 51(11A): 231100160-6. <https://doi.org/10.11896/jsjcx.231100160>

[融合多尺度特征与位置信息的输电线路山火检测算法](#)

Mountain Fire Detection Algorithm of Transmission Line Based on Multi-scale Features and Coordinate Information

计算机科学, 2024, 51(11A): 230900155-7. <https://doi.org/10.11896/jsjcx.230900155>

[基于注意力机制和双分支网络的胸部疾病分类](#)

Classification of Thoracic Diseases Based on Attention Mechanisms and Two-branch Networks

计算机科学, 2024, 51(11A): 230900116-6. <https://doi.org/10.11896/jsjcx.230900116>

基于自注意力与双向特征融合的道路障碍物检测方法

李 婷 赵尔敦 杨 军

华中师范大学计算机学院 武汉 430079

(liting@mails.cnu.edu.cn)

摘 要 随着科技的飞速发展,辅助驾驶技术已经成为汽车行业未来发展的重要方向。在基于图像的道路障碍物检测中,现有方法对尺度变化大的目标、小目标和存在遮挡目标的检测能力有限,常出现误判和漏判等问题。针对此问题,提出了一种基于自注意力与双向特征融合的道路障碍物检测方法(CoXt-FCOS)。该方法在主干特征提取网络中引入分组的自注意力机制模块 CoXT,以增强网络的全局信息捕获能力;为解决遮挡问题,引入跨阶段金字塔池化模块 SPPCSPC;在特征融合模块中,引入路径增强网络,形成双向特征融合模块 ESPAFPN,提升网络对小目标的感知能力。实验结果表明,CoXT-FCOS 模型的精度较高,在 CODA 数据集上的 mAP 达到了 88%,能够更准确地检测出道路上的障碍物。

关键词: 障碍物检测;自动驾驶;FCOS;自注意力机制;特征融合

中图分类号 TP391

Road Obstacle Detection Method Based on Self-attention and Bidirectional Feature Fusion

LI Ting, ZHAO Erdun and YANG Jun

School of Computer Science, Central China Normal University, Wuhan 430079, China

Abstract With the rapid development of technology, assisted driving technology has become an important direction for the future development of the automotive industry. In image-based road obstacle detection, existing methods have limited detection capabilities for targets with large scale changes, small targets, and targets with occlusion, often resulting in misjudgments and omissions. To address this problem, a road obstacle detection method based on self-attention and bidirectional feature fusion (CoXt-FCOS) is proposed. This method introduces a grouped self-attention mechanism module CoXT in the backbone to enhance the global information capture capabilities of the network. To solve the occlusion problem, a cross-stage pyramid pooling module SPPCSPC is introduced. In the feature fusion module, a path enhancement network is introduced, forming a bidirectional feature fusion module ESPAFPN, to enhance the network's perception of small targets. Experiments show that the CoXT-FCOS model has high accuracy, with an mAP of 88% on the CODA dataset, and can more accurately detect obstacles on the road.

Keywords Obstacle detection, Autopilot, Fully convolutional one-stage object detection, Self-attention, Feature fusion

1 引言

避障技术作为辅助驾驶的关键技术之一,通过检测出道路上的障碍物,及时进行规避处理,最大限度地减少交通事故的发生。真实道路上常见的道路障碍物,如护柱、锥型交通路标、水马、交通标志、狗、垃圾桶等。车辆在行驶中对障碍物进行目标检测存在诸多困难:复杂路况和天气等因素造成视觉干扰;密集目标之间存在遮挡;遮挡物的目标种类多、尺度不一、变化较大且存在大量的小目标;辅助驾驶对检测的速度和精度有较高的要求等。

为实现精准的障碍物检测,本文基于 FCOS^[1]提出了一种基于自注意与双向特征融合的道路障碍物检测方法(CoXt-FCOS),其特点如下:在主干特征提取网络 ResNeXt 中引入分组的自注意力机制模块 CoXT (CoXT-block),利用丰富的上下文信息来加强自注意力学习,形成 CoTNeXt101 结构;在主干特征提取网络的最深层中,引入跨阶段结构的金字塔池化模块 SPPCSPC,在减少金字塔池化参数的同时,提升网络的速度和精度;在特征融合时使用增强小目标的双向特

征融合网络 ESPAFPN,提升小目标的检测精度;将中心度分支引入回归分支上,以更好地定位障碍物目标;在回归损失中引入目标尺度比的惩罚项,提高网络的收敛速度。

2 国内外研究现状

目标检测是计算机视觉中一项基本且具有挑战的任务,目标检测算法需要为图像中的每个感兴趣的实例预测其边界框及类别。根据是否生成候选区域可以将目标检测算法分成两大类:二阶段(Two stage)检测网络和一阶段(One stage)检测网络。

二阶段(Two stage)检测网络,通常会采用候选区域生成器来生成一组稀疏的候选区域。然后从每个候选区域中提取特征,并使用分类器预测对象的类别和边界框。这种方法检测精度较高,但速度较慢。2015年, Girshick 等提出了 Fast RCNN 检测器^[2],基于空间金字塔池化(Spatial Pyramid Pooling, SPP)实现了同时训练检测器和边界框回归器。同年, Ren 等基于 Fast RCNN 提出了 Faster RCNN 检测器^[3],摒弃了选择搜索算法,而是引入区域建议网络自动生成候选框。

2017年, Lin等在Faster RCNN检测器上引入了Feature Pyramid Network(FPN)^[4],通过横向链接的自顶向下结构,加强了各种尺度的物体检测精度。

一阶段(One stage)检测网络没有生成候选区域这一阶段,直接在原图上通过回归目标位置的方式进行检测,同时进行分类和回归任务,速度较快,实时性较高。2015年,Joseph等提出了一阶段检测网络YOLO^[5]。YOLO将单个神经网络应用于整张图像,将图像分为多个区域,并预测每个区域的边界框和概率。2017年Lin等提出了RetinaNet^[6]网络,指出了一阶段检测器精度不如二阶段检测器的原因是训练时正负样本不均衡。RetinaNet引入了Focal loss分类损失函数,在训练过程中更加关注难以分类和易被误分类的样本,降低了正负样本不平衡对一阶段检测精度的影响。2020年, AI提出了DETR^[7]网络,将目标检测视为一个集合预测问题。DETR使用了一种基于集合的全局损失,通过二分匹配和编

码器-解码器的架构强制进行唯一预测的网络。基于这种方法,简化了检测的流程,有效地消除了对许多手工设计的组件的需求。2022年,Zhang等基于DETR提出了DINO^[8]网络,使用了去噪声的训练方法和混合查询的选择方法用于锚框的初始化方式,显著减小了模型大小,同时获得了更好的结果。

3 基于自注意力与双向特征融合的道路障碍物检测方法

CoXt-FCOS的检测框架如图1所示。输入图像经过CoTNeXt101主干网络提取特征后,会将特征图输入到跨阶段的金字塔池化模块,以扩大模型的感受野。每层池化模块的输出会使用双向特征融合模块ESPAFPN进行特征融合,以增强特征图的提取能力。最后,每层输出的特征图使用共享的检测头,来预测类别、中心度和检测框等信息。

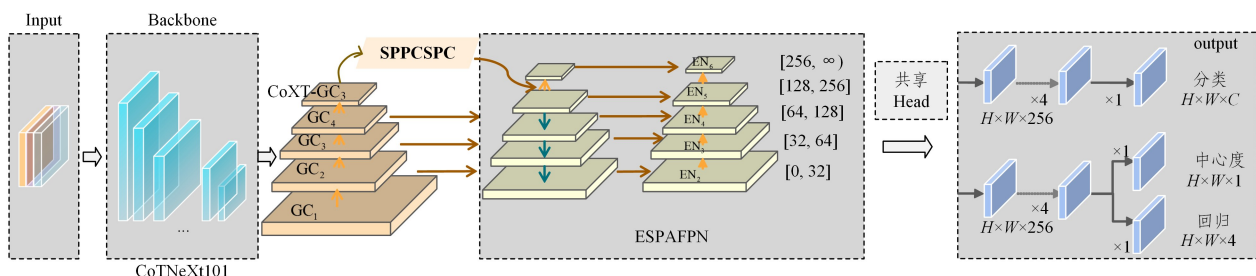


图1 CoXt-FCOS模型图

Fig. 1 Diagram of CoXt-FCOS model

3.1 增强小目标的双向特征融合网络

原始FCOS模型中,特征融合部分使用特征金字塔网络FPN结构,仅进行高层次信息传递,对底层的位置信息利用不足,且障碍物目标作为小目标,目标的信息量较少,容易在深层特征融合时丢失特征信息。本文考虑到FCOS是基于分割思想的检测器,采用路径增强网络PANet^[9]来增强FPN提取底层位置的信息,从而形成CoXt-FCOS模型的双向特征融合模块ESPAFPN。

ESPAFPN将来自GC_i(i=1,2,3,4)层的特征使用最邻近插值进行上采样,将特征图的尺寸扩大一倍,通道数保持不变,再和下层同一尺度的特征图EN_i(i=2,3,4,5)进行对应位置的相加融合,并使用3×3的卷积消除融合后的混叠效应。GC_i则经过下采样与EN_i融合。需要注意的是,EN_i(i=3,4,5,6)均为EN₂下采样得到。ESPAFPN结构将底层的位置信息向上层聚合,可获得丰富的语义信息和位置信息,同时缩短最底层和最高层特征之间的信息路径。该操作增强了特征图的信息提取能力,且能更好地利用底层特征来检测小尺度目标,同时降低了网络参数和训练开销。

3.2 增强空间信息的上下文Transformer的自注意力主干网络CoTNeXt

CoTNeXt网络是在ResNeXt基础上引入了一种新的上下文Transformer结构Contextual Transformer(CoT)^[10]。CoT块可充分利用输入键的上下文信息,将动态上下文信息聚合(dynamic)与卷积的静态上下文信息聚合(static)进行集成,从而形成静态和动态的上下文注意力矩阵。将CoT块迁移到分组的残差网络ResNeXt^[11]结构中,使得Bottleneck结构中3×3的卷积替换成不同结构的CoXT块,代替标准卷积,

这种方法可超越传统的局部自注意力,形成CoTNeXt网络。

Transformer在自然语言处理领域广受欢迎,其中的自注意力机制(self-attention)在计算机视觉领域也取得了较好的成绩。计算机视觉任务中使用卷积运算来提取图像特征,但卷积只能提取局部特征,对于一些需要获得周围信息的依赖关系的特定任务,例如目标检测,需获取全局信息。为了解决这个问题,卷积神经网络通常需要堆叠多层卷积来提升整体感受野,以获取更多的信息。但这种方法仍然会忽略较远像素的全局信息对当前局部信息的贡献,即忽略了远距离关系。自注意力机制计算图像中每个像素与周围其他像素之间的相关性,将其作为该像素的权值,形成了注意力矩阵,从而实现网络全局参考信息,即从图像中捕捉全局的上下文信息。

3.3 跨阶段的金字塔池化SPPCSPC模块

为了提高检测准确性,本文采用了跨阶段的金字塔池化SPPCSPC模块^[12],以帮助模型更好地识别被遮挡的目标。该模块可以扩大网络的多尺度感受野,获取更丰富的全局信息,从而降低目标遮挡带来的影响。该模块被添加到CoTNeXt网络的最后一个卷积层之后,ESPAFPN特征融合之前。该模块由空间金字塔池化(Spatial Pyramid Pooling, SPP)^[13]和跨阶段部分网络(Cross Stage Partial Networks, CSP)^[14]组成。

3.4 考虑高宽比值的边界框损失函数F-CIoU

在目标检测框架中,边界框的回归是一个关键步骤。FCOS采用GIoU损失作为边界框损失函数,该函数在IoU的基础上添加了重叠面积指标(边界框和真实框最小包围矩形),解决了IoU边界框的回归损失缺点。但当预测框与真实边界框属于包含关系时,GIoU损失退化为IoU损失,导致收敛速度变慢。为了解决这个问题,文献[15]提出了DIoU

损失,该损失函数添加了中心点距离指标,使用两个中心点之间的欧氏距离来计算,并除以最小包围框对角线的 L2 距离,此举可以加快网络的收敛速度。同时,该文还提出了 CIoU 损失,考虑了预测框与真实框之间的长宽比的一致性对回归损失的重要性。CIoU 损失添加了边界框与真实框之间的纵横比,从而使预测框更贴合真实框的比例。

本章的 CoXt-FCOS 检测网络采用了简化的 CIoU。具体地,在边界框损失 GIoU 中添加了一个长宽比值的惩罚项 v ,形成了 F-CIoU 边界框损失。该损失的定义如式(1)所示。这种改进的边界框损失函数可以提高模型的准确性和收敛速度,从而进一步优化障碍物检测的性能。

假设 $\mathbf{A}(l^*, t^*, r^*, b^*)$ 是真实的边界框信息, $\mathbf{B}(l, t, r, b)$ 是预测的边界框信息, $\mathbf{C}(\max(l^*, l), \max(t^*, t), \max(r^*, r))$ 是 A 和 B 之间的最小包围框。 W^*, H^* 分别是真实框的宽、高, W 和 H 分别是预测框的宽和高。计算式分别为: $W^* = l^* + r^*, H^* = t^* + b^*, H = t + b, W = l + r$ 。 F-CIoU 边界框的损失计算如下:

$$L_{\text{reg}} = L_F - \text{CIoU} = 1 - \text{GIOU} + \alpha v$$

$$= 1 - \left((IoU) \frac{|\mathbf{C} - (\mathbf{A} \cup \mathbf{B})|}{|\mathbf{C}|} \right) + \frac{v}{(1 - IoU) + v + \epsilon} \quad (1)$$

其中, $v = \frac{4}{\pi^2} \left(\arctan \frac{W}{H} - \arctan \frac{W^*}{H^*} \right)^2$ 是度量长宽比一致性的指标; $\alpha = \frac{v}{(1 - IoU) + v}$ 是一个正的权重参数,在回归时可以赋予重叠区域更高的优先级; ϵ 是为了计算的稳定性而引入的微小量,防止出现分母为零的情况; A, B, C 的示意图如图 2 所示, A 为红色矩形的面积, B 为绿色矩形的面积, C 为虚线框的面积。

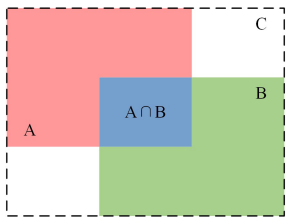


图 2 GIoU 图解(电子版为彩图)

Fig. 2 GIoU illustration

3.5 总损失函数

目标检测任务输出目标的类别和在图像中的位置坐标, CoXt-FCOS 对于其他检测器多了一个预测目标中心度的值,总损失应包含分类损失、回归损失、中心度损失,因此采用多任务损失函数进行模型优化。CoXt-FCOS 检测网络的总损失为特征图上的所有点 (x, y) 的回归分支损失、分类分支损失和中心度分支损失之和。总损失的计算式如式(2)所示:

$$L(\{p_{x,y}, t_{x,y}, s_{x,y}\}) = \frac{1}{N_{\text{POS}}}_{x,y} \sum L_{\text{cls}}(p_{x,y}, c_{x,y}^*) +$$

$$\frac{1}{N_{\text{POS}}}_{x,y} \sum \mathbb{I}_{\{c_{x,y}^* > 0\}} L_{\text{reg}}(t_{x,y}, t_{x,y}^*) +$$

$$\frac{1}{N_{\text{POS}}}_{x,y} \sum \mathbb{I}_{\{c_{x,y}^* > 0\}} L_{\text{cen}}(s_{x,y}, s_{x,y}^*) \quad (2)$$

其中, L_{cls} 是分类损失, L_{reg} 是边界框回归损失, L_{cen} 是中心度损失, $p_{x,y}, t_{x,y}, s_{x,y}$ 分别为网络预测框的类别概率、回归目标和预测的 centerness, $c_{x,y}^*, t_{x,y}^*, s_{x,y}^*$ 为真实框的类别概率、回归目标、centerness, N_{POS} 是正样本的数量。 $\mathbb{I}_{\{c_{x,y}^* > 0\}}$ 为指示函数,当点 (x, y) 对应在真实框的类别 $c_{x,y}^* > 0$ 时,取值为 1,否则为

0,代表只计算正样本的回归损失和中心度损失。

分类的损失采用 Focal Loss,计算式如式(3)所示:

$$L_{\text{cls}} = L_{\text{focal_loss}} = -\alpha (1 - P_t)^\gamma \log P_t$$

$$P_t = \begin{cases} p, & \text{if } y=1 \\ 1-p, & \text{otherwise} \end{cases} \quad (3)$$

其中, $\alpha \in [0, 1]$ 是权重因子,用来调节正负样本的权重; $(1 - P_t)^\gamma$ 调节系数,用来调节难分类样本的权重; y 是真实标签, $y \in \{\pm 1\}$; P_t 是类别概率 $p \in [0, 1]$,当检测样本的 P_t 较低时,该样本为困难样本,在调节因子作用下,该样本的权重会增大。相反,当 P_t 较高时,该样本为容易样本,损失权重会相应降低。本文实验中 α 取值为 0.25, γ 取值为 2。

中心度损失采用交叉熵损失(BCE), $s_{x,y}$ 为预测的中心度值, $s_{x,y}^*$ 为真实的中心度值,中心度损失的计算式如式(4)所示:

$$L_{\text{cen}} = s_{x,y}^* \log(s_{x,y}) + (1 - s_{x,y}^*) \log(1 - s_{x,y}) \quad (4)$$

4 实验

实验方案如下:首先构建道路障碍物数据集,为达到足够的支撑,使用离线数据增强对数据集进行扩充,增加样本数量,再对数据集进行划分,训练过程中使用验证集对模型的超参数进行调整,按照训练验证集与测试集的比例为 9:1,训练集与验证集为 9:1 进行划分,在迭代一定次数时,使用验证集上的最小误差作为最优的检测模型,最后对划分出的测试集送入模型中并通过非极大值抑制(NMS)得到输出。

4.1 实验环境与参数设置

本实验训练和验证均使用基于 PyTorch 的深度学习开发框架,在搭载 NVIDIA Tesla V100-16GB 和 Intel(R) Xeon (R) Platinum 8160T CPU@2.10GHz 的 Linux 系统上进行。总训练次数为 100 次,使用自适应 batch size,以充分使用 GPU 内存,优化器采用 Adam 进行迭代优化,设置动量 momentum=0.9,学习率初始值设置为 3×10^{-4} ,使用余弦学习率下降方式,余弦退火最小学习率为 3×10^{-4} 。

4.2 数据集

本文研究道路上特殊的障碍物进行检测,使用 CODA2022^[6]中提出的新颖类别,即护柱(bollard)、锥型交通路标(traffic cone)、交通标志(traffic sign)、水马(barrier)、施工车辆(construction vehicle)、垃圾桶(dustbin)、狗(dog)7 类作为道路常见障碍物进行检测的研究。CODA2022 数据集中的障碍物图像如图 3 所示。本文在公开的 CODA 数据集上进行特定类别的筛选,在 29 个类别中筛选出目前道路交通行驶最容易忽略的 7 种障碍物,获得 4022 张图像和对应的 xml 格式的标注信息,并使用离线数据增强方式增加数据集样本量,再进行训练集、验证集、测试集的构建。



图 3 CODA 数据集中障碍物图像样例

Fig. 3 Sample of obstacle images in CODA dataset

通过数据集扩充,数据集总量达到 8010 张,符合神经网络对数据集数量的标准。实验采用 9:1 比例划分训练验证集和测试集,其中训练集和验证集按 9:1 比例划分,即训练集为 6488 张,验证集为 721 张,测试集为 801 张。

4.3 实验评价指标

根据障碍物检测器输出的分类结果和坐标信息,检测器的评价指标采用:各类别的平均精度(Average Precision, AP)、所有类别的平均精度均值(mean of Average Precision, mAP)和 mF1(micro F1-score)、对数平均漏检率(Log-average miss rate, MR^{-2})^[17]。其中,mF1,AP,mAP 的值越高代表检测性能越好, MR^{-2} 的值越低代表检测性能越好,本文通过这 4 个指标来综合评估障碍物检测器的性能。

4.4 实验结果对比

本文在相同的数据集上进行不同检测模型的对比,将 CoXt-FCOS 与 SSD, YOLOv3, YOLOv5, YOLOv8, CenterNet 等一阶段检测算法进行对比, IoU 阈值均为 0.5, 采用 mAP, MR^{-2} , mF1 来衡量性能。性能指标如表 1 所列。可见 CoXt-FCOS 在 mAP, MR^{-2} , mF1 等指标上均优于其他算法。由于 CoXt-FCOS 使用了自注意力机制,其推理速度较慢,但也基本满足实时使用的要求。

表 1 不同算法的检测效果对比(IoU 阈值 0.5)

Table 1 Comparison of detection effectiveness of different algorithms(@IoU 0.5)

算法	主干网络	$MR^{-2}/\%$	mF1/ $\%$	mAP/ $\%$	FPS
SSD	VGGNet	49.43	60.14	66.71	94
YOLOv3	Darknet53	45.00	68.14	72.71	100
CenterNet	ResNet50	56.14	50.43	59.29	125
YOLOv5	CSPDarknet53	43.57	66.29	72.29	64
YOLOv8	CSPDarknet53	60.71	54.43	51.24	57
FCOS	ResNet50	39.86	77.14	78.43	62
FCOS*	ResNet101	25.43	81.57	84.25	41
CoXt-FCOS	CoTNeXt101	21.86	86.43	88.00	36

4.5 消融实验

为证明 CoXt-FCOS 中几种改进方法对检测的精度均有提升,在增强后的 CODA2022 数据集上进行消融实验,对提出的方法进行对比,使用 AP0.5 的评价指标以及对应的对数平均漏检率、mF1 值,实验结果如表 2 所列。

表 2 CoXt-FCOS 检测算法的消融实验结果

Table 2 Results of ablation experiments of CoXt-FCOS

序号	1	2	3	4	5
ResNet50	✓				
ResNet101		✓			
CoTNeXt101			✓	✓	✓
cen_on_reg				✓	✓
ESPAFPN					✓
$MR^{-2}/\%$	34.43	30.86	25.43	25.14	21.86
mF1/ $\%$	77.14	78.43	81.57	81.57	86.43
mAP/ $\%$	78.43	80.86	84.25	84.57	88.00

从表 2 中可以看出,将 FCOS 主干网络更换为其他类型网络的效果对比中,主干网络采用基于分组卷积形式的残差网络 ResNeXt101 提取特征效果最好,与 ResNet50 相比,mAP 提升了 5.82%,F1 提升了 4.43%, MR^{-2} 降低了 9%。这说明了本文模型使用的分组卷积主干有明显的检测效果提升。在实验 3 的基础上添加 cen_on_reg 结构,作为实验 4,其中 cen_on_reg 表示将中心度分支与回归分支

共享前四个卷积参数,并采用在回归特征层上计算中心度的值,平均检测精度提升了 0.32 个百分点,F1 的值保持不变, MR^{-2} 降低了 0.29%。因回归分支与中心度分支参数的相关性更强,cen_on_reg 在检测精度上有了少许提升,由于该改动涉及参数较少,故提升有限;在实验 4 的基础上添加增强小目标的双向特征融合模块 ESPAFP,作为实验 5,平均检测精度提升了 3.43%,F1 提升了 4.86%,漏检率 MR^{-2} 下降了 3.28%。

4.6 检测效果示例

从图 4 中可以看出,相比 FCOS,CoXt-FCOS 对小目标与被遮挡的目标具有较好的检测能力,解决了 FCOS 中对小目标的漏检问题和误检问题,提升了网络的检测能力。



注:第一行图为 CoXt-FCOS,第二行图为 FCOS。

图 4 两种算法的效果对比

Fig. 4 Effect comparison of two algorithms

结束语 本文提出了一种较为安全可靠的基于自注意力与双向特征融合的道路障碍物检测方法 CoXt-FCOS,提高道路上的护柱、锥形交通路标、水马、狗等障碍物的检测精度,该方法主要使用含自注意机制的 CoTNeXt101 作为主干特征提取网络,并且引入跨阶段结构的空间金字塔池化模块 SPPC-SPC 以更准确地预测障碍物位置和类别,同时引入增强小目标的路径增强路径,形成双向特征融合模块网络 ESPAFP,使得融合更多的浅层信息,增强小目标的检测能力,最后在回归损失中引入目标尺度比的惩罚项,可以更精确地预测目标位置和形状,从而提升网络的检测能力。本文在公开的 CODA2022 道路障碍物数据集上进行了实验,实验结果表明,本文提出的 CoXt-FCOS 能够更好地满足道路障碍物检测对精度的需求。

由于本文目标检测模型只涉及到障碍物的分类和定位,而真实行车场景下还需要对障碍物目标进行距离估计。为此,后续可以在该模型的基础上添加基于视觉的深度估计模块,形成端到端的多任务模型,以提高检测模型的应用范围。但深度估计网络往往需要较高的算力支持,如何轻量化网络、提升检测速度也是需要考虑的问题。同时,本文基于自注意力机制的模块在训练和推理时需要较多计算资源。为提高训练效率和推理速度,后续希望可以使用更高效的算法来减少训练和使用时所需的资源。

参考文献

- [1] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:9627-9636.
- [2] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015:1440-1448.
- [3] REN S, HE K, 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.
- [4] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [5] 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, 2016: 779-788.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [7] CARION N, MASSA F, SYNNAEVEG, et al. End-to-end object detection with transformers[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [8] ZHANG H, LI F, LIU S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. arXiv: 2203.03605, 2022.
- [9] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [10] LI Y, YAO T, PANY, et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489-1500.
- [11] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1492-1500.
- [12] 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, 2023: 7464-7475.
- [13] HE K, ZHANG X, REN S, 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.
- [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]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.
- [15] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993-13000.
- [16] LI K, CHEN K, WANG H, et al. Coda: A real-world road corner case dataset for object detection in autonomous driving[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 406-423.
- [17] ZHAO S, CHEN N. Review on Small-scale Pedestrian Detection Technology for Complex Pavement[J]. Computer Systems and Applications, 2022, 31(7): 1-11.



LI Ting, born in 1997, postgraduate. Her main research interests include computer vision, object detection and autopilot.



ZHAO Erdun, born in 1972, Ph.D, associate professor. His main research interests include computer vision, object detection and autopilot.