文章编号:1674-2974(2024)02-0177-10 DOI:10.16339/j.cnki.hdxbzkb.2024156

基于全景环视的端到端停车位检测方法

秦晓辉[1,2†],殷周平[1],张素英[3],黄圣杰[1],张润邦[1],芦涛[1],刘硕[1],胡满江[1,2]
(1. 湖南大学 机械与运载工程学院,湖南 长沙 410082;
2. 湖南大学 无锡智能控制研究院,江苏 无锡 214072;
3. 潍柴动力股份有限公司 发动机研究院,山东 潍坊 261061)

摘 要:现有停车位检测方案大多将目标检测方案和人工设计的后处理模块进行简单结合,各阶段提取的特征存在大量冗余信息.并且,人工设计的后处理模块通常适应面窄,计算量大,最终导致停车位检测效果难以实用.针对这些问题,本文引入全景视觉,结合现有算法的优点与环视图像的特点,设计端到端的无锚框停车位检测算法.该算法对停车位进行进入线朝向建模,而非单独考虑两个入口点,省去了停车位入口点匹配和朝向判断流程,最终实现完全一体化的停车位位置、朝向和占用情况检测.考虑到实用性,在网络结构设计上进行了速度和精度的平衡、正负样本均衡,以及无后处理等多方面优化.最终,在ps2.0数据集上,本文提出的AFPSD模型以88.7的FPS(每秒帧数)达到68.7%的AP,相较VPS-Net和DMPR-PS方案精度分别提升1.2%和2.1%.由此可知,本文设计的一阶段端到端方案可以代替三阶段方案,在环视图像上实现停车位的稳定检测.

关键词:自主泊车;目标检测;停车位检测;环视图像
中图分类号:TP311 **文献标志码**:A

End-to-end Parking Slot Detection Method Based on Panoramic Surround View

QIN Xiaohui[1,2†], YIN Zhouping[1], ZHANG Suying[3], HUANG Shengjie[1], ZHANG Runbang[1], LU Tao[1], LIU Shuo[1], HU Manjiang[1,2]
(1. College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China;
2. Wuxi Intelligent Control of Research Institute(WICRI) of Hunan University, Wuxi 214072, China;
3. Center of Research and Department, WEICHAI Power Co., Ltd., Weifang 261061, China)

Abstract: Most of the existing parking spot detection solutions simply combine the target detection scheme and with manually designed post-processing modules, and there is a large amount of redundant information in the features extracted at each stage. Moreover, the manually designed post-processing modules are usually narrowly

* **收稿日期**:2023-04-25
基金项目:长沙市自然科学基金资助项目(kq2202162),Changsha Natural Science Foundation of China (kq2202162);国家自然科学基金资助项目(52172384),National Natural Science Foundation of China (52172384);湖南省湖湘青年科技创新人才(2021RC3048),S&T Innovation Program of Hunan Province (2021RC3048);汽车车身先进设计制造国家重点实验室自主研究课题(61775006),Independent Research Program of State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body (61775006)
作者简介:秦晓辉(1988—),男,湖南张家界人,湖南大学副研究员,博士
† 通信联系人,E-mail: qxh880507@163.com

adapted and computationally intensive, which ultimately makes the parking space detection effect difficult to be practical. To address these problems, this paper introduces panoramic vision and combines the advantages of existing algorithms with the characteristics of surround-view images to design an end-to-end anchorless frame parking spot detection algorithm. The algorithm models the entry line orientation of parking spaces instead of considering two entry points separately, eliminating the process of parking space entry point matching and orientation judgment, and finally realizing realizes fully integrated parking space location, orientation, and occupancy detection. Considering the practicality, the network structure design is optimized in many aspects, such as the balance of speed and accuracy, positive and negative sample balance, and no post-processing. Finally, on the ps2.0 dataset, the AFPSD model proposed in this paper achieves 68.7% AP with a FPS(Frames Per Secend) of 88.7, which is 1.2% and 2.1% higher accuracy compared to the VPS-Net and DMPR-PS schemes, respectively. It can be seen that the one-stage end-to-end scheme designed in this paper can replace the three-stage scheme to achieve stable detection of parking slots on the surround-view image.

Key words: autonomous parking; object detection; parking slot detection; surround-view image

自主泊车技术能够通过传感器探测视野盲区,利用算法精准且稳定地控制车辆安全泊入目标车位,逐渐受消费者喜爱.停车位检测技术能够给车辆持续提供周边停车位的位置信息,是自主泊车系统中关键的一环.目前,学者们将许多成熟的目标检测方案应用于停车位检测领域,这些方案通常将停车位检测分为三个步骤.

1)入口点检测:不作区分地检测出图片中所有停车位的入口点;

2)入口点匹配:通过人工设置的规则将属于同一个停车位的入口点进行配对;

3)类别和方向判断:判断停车位的类别(平行、垂直、倾斜)以及停车位的朝向角度.

通过上述三个步骤,可以检测出图片中停车位的位置信息.如果要获得停车位的占用信息,则通常需要再额外增加一个分类模型,将检测出的停车位从原图中截取出来送入停车位分类模型作空闲占用分类.

Li 等[1]采用 PSD_L 方法构建了4个停车位入口点图案 AdaBoost 分类器,每个分类器负责一定朝向角度范围的停车位入口点图案判断,将图片划分成多个入口点图案大小的区块送入分类器网络中,获得停车位入口点的位置,使用人工设计的距离约束筛选出能够组成进入线的入口点,再使用六个高斯直线模板匹配查询出进入线代表的车位的类别和朝向.DeepPS 方法[2]采用单阶段目标检测网络 YOLO V2[3]作为入口点检测网络,将检测到的入口点提取出来进行基于人工规则的匹配之后,从图片中截取出来送入 AlexNet[4]分类网络中以判断停车位的类别和方向信息.VPS-Net[5]则对此进行改进,使用单阶段锚框检测网络 YOLO V3[6]检测入口点的同时也对停车位头的轴平行外接矩形框一起进行检测,并将位于一个停车位头的入口点进行匹配,最后采用计算交并比的方式进行朝向判断.后续 DeepPS 的作者再对方案进行改进[7],采用点向量描述子代替轴平行矩形描述子对入口点进行建模,同时将类别朝向判断由之前的 AlexNet 网络改为人工设计的模板匹配法.同样是对入口点描述子进行改进,Wu 等[8]提出 PS-Det 方案,使用圆形描述子配合卷积神经网络搭建了一个粗略到精修的方案对入口点进行检测.吕雪杰[9]将 DeepPS 方案中的入口点检测方案由 YOLO V2 换成 YOLO V3,并使用 Focal Loss[10]缓解正负样本不平衡问题.Li 等[11]将停车位检测分为驾驶过程和泊车过程,第一过程使用 VPS-Net 停车位检测网络,第二过程使用 DMPR-PS 停车位检测网络.

上述方法以不同的方式将目标检测方法引入停车位检测任务中,取得一定成效,但是仍然存在以下问题:

1)入口点检测、入口点匹配、类别朝向分类和空闲占用分类模块单独设计,提取的特征存在冗余信息,且简单结合各个模块时效性低;

2)将两个入口点看成独立的目标进行特征提取,没有利用二者之间的关联性;

3)需要垂直和水平停车位之间的入口点距离范

围先验信息;

4)需要入口点形状先验信息;

5)将水平线和分隔线之间的夹角回归问题当作分类问题处理,且需要特定的夹角先验信息;

6)人工设计的入口点匹配方案时效性低、泛化性差;

7)目标检测方案选择不合理,矩形框并不适合描述各种旋转角度以及形状的停车位入口点,点向量模型匹配复杂度高,且适应性差.

在获取这些先验信息时,做了大量假设,包括:垂直停车位和水平停车位之间的距离在特定范围内,进入线和分隔线之间的夹角只包含一些特定的角度,入口点的形状只包含T形状或者L形状.这些假设的来源都是基于对数据集的统计,当数据集中停车位信息变化时,就需要重新人工设计建模和匹配策略,泛化性很低.

针对上述这些问题,本文提出了从停车位建模出发,采用一个箭头来表达停车位,并以无锚框的方式设计端到端的停车位检测方案.

1　模型设计

本文的目标是设计端到端的无锚框停车位检测网络,以此简化停车位检测流程,消除各阶段之间的特征冗余以及人工设计模块,提升模型的时效性和泛化性.本文设计的网络框架如图1和式(1)所示,包含停车位的图片输入到由四个阶段组成的主干网络中进行特征提取,通过瓶颈网络对后三个阶段的输出进行多进单出的融合之后,输出单尺度特征 $F_h \in \mathbb{R}^{(h,w,c)}$ 至检测头,检测头通过回归分支 H_f 和分类分支 H_c 进行域转换,得出特征中编码的停车位回归信息 R_r 和分类信息 R_c.本节后续将依次介绍模型的主干网络、瓶颈网络、检测头的设计.

$$\begin{aligned} F_h &= N_c\left(N_b(I)[1]\right) + N_c\left(N_b(I)[2]\right) + \\ &\quad N_c\left(N_b(I)[3]\right) \\ R_r &= H_f\left(F_h\right) \\ R_c &= H_c\left(F_h\right) \end{aligned} \tag{1}$$

1.1　主干网络

主干网络负责对原始图片的特征提取,这些特征需要包含停车位的全局信息,即朝向、占用以及局部信息(入口点位置);瓶颈网络则负责不同层次的特征融合,得到包含停车位全局信息和局部信息的特征图.

本文采用跨越阶段部分DarkNet53(Cross Stage Partial DarkNet53, CSP-DarkNet53)作为主干网络,该网络在DarkNet的基础上增加了部分跨越层级连接.DarkNet参考了ResNet的残差结构和批标准化结构.

CSPNet[12]将特征图进行分流,能够在保留特征复用的同时避免很多重复梯度信息,并实现丰富的梯度信息融合,使用CSP结构对DarkNet进行改进之后就得到了本文完整的主干网络.该结构由一个基

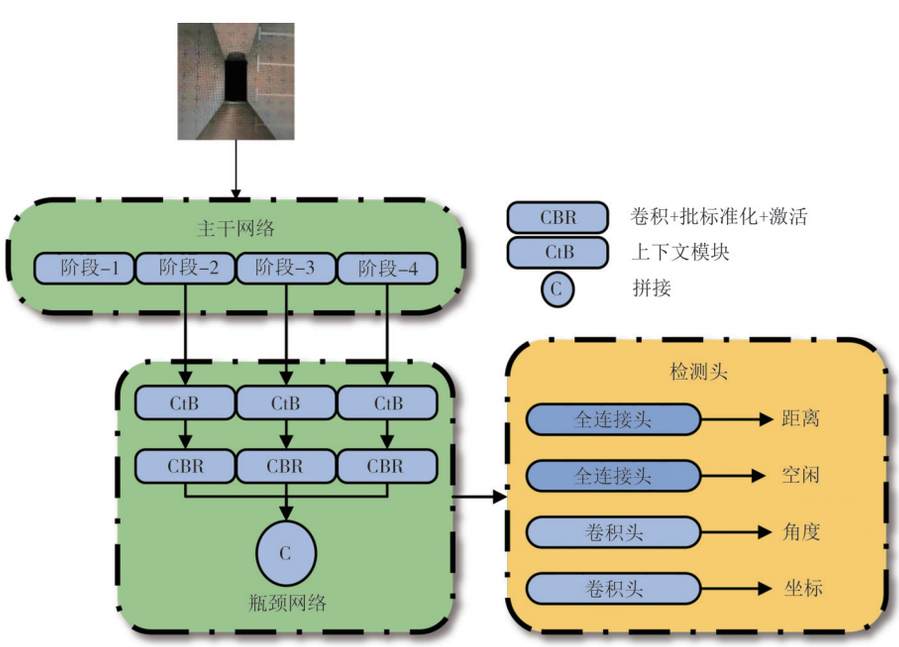


图1　方法结构图

Fig. 1　Architecture of the proposed method

干模块和四个阶段组成.基干模块采用Focus结构,旨在通过切片对图像进行下采样操作,相比于传统的下采样操作,Focus模块能够将原始图片的宽高信息无损地集成到通道维度,并通过卷积进行进一步特征提取.阶段模块则由卷积模块和CSP模块构成,如图2所示,其中包含Conv字样的子模块都是由卷积、批标准化和Swish激活函数构成,而区块则是由两个卷积模块构成的残差模块,C代表通道维度的拼接.由此可见,每个阶段都是由小残差结构堆叠出的大残差结构,且最后一个阶段的第一个Conv模块之后会额外通过SPP模块进一步集成全局信息.

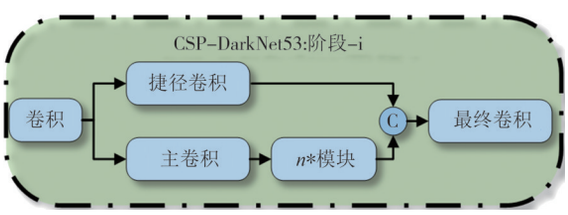


图2 CSP-DarkNet53模块结构

Fig. 2 Architecture of CSP-DarkNet53 Module

1.2 瓶颈网络

卷积层的感受野是逐层增加的,下采样次数越多,语义信息覆盖面越广,位置信息丢失也越严重.为了兼顾语义信息和位置信息,本文设计的瓶颈网络对感受野大的深层特征和感受野小的浅层特征进行融合,结构如图1中瓶颈网络所示.参与融合的特征层来自主干网络中三个阶段的输出,对应原图的下采样倍率分别为8、16、32.下采样倍率越高,特征图分辨率越低、感受野越高、全局信息越丰富、输出的停车位数量越少,反之倍率越低,特征图分辨率越高、感受野越低、局部信息越丰富、输出的停车位数量越多.

检测网络输出的车位数量与网络反向传播时携带的信息量息息相关.如式(2)所示,反向传播时计算的损失由正样本损失和负样本损失组成.正样本需要计算包括位置回归、类别分类和置信度等多个损失,而负样本只需计算置信度损失,当负样本数量N_N远大于N_P时,正样本的损失则可以忽略不计.

$$\begin{aligned}\text{Loss} = &\sum_{i=0}^{N_P}\left(\text{pred}^{x,\,y,\,\cos\theta,\,\sin\theta,\,l,\,\text{vacant},\,\text{dir},\,\text{conf}} - gt\right) + \\ &\sum_{i=0}^{N_N}\left(\text{pred}^{\text{conf}} - gt\right)\end{aligned} \tag{2}$$

$$\text{Loss} \approx \sum_{i=0}^{N_N}\left(\text{pred}^{\text{conf}} - gt\right) \qquad \text{s.t. } N_N \gg N_P$$

以608图片输入为例,最终输出的车位数量为:$7\,581 = (608/8)^2 + (608/16)^2 + (608/32)^2$,结合每个车位的预测参数数量可以得出网络需要输出$68\,229 = 7\,581 \times 9$个预测值.假设一张图片中有10个停车位,那么采用1对1的标签分配时会产生7 501个负样本,10个正样本.对于正样本,需要约束其回归位置的准确度、类别的准确度等共计9个预测值,而对于负样本只能约束其前背景概率一个预测值.也就是说训练一张具有10个停车位的图片,网络输出的68 229个预测值中,只有$7\,591 = 68\,229 \div 10 \times 9 + 68\,229 \div 7\,501 \times 1$个预测值会参与损失计算,只占总输出值的11%,这会使得训练进程十分缓慢.且7 591个参与损失计算的预测值中,只有6个是回归值,这也会使得分类和回归分支的训练进程十分不平衡.为了缓解这种不平衡和缓慢的问题,本文设计的瓶颈网络将三个阶段的输出进行不同程度的下采样之后融合为32倍下采样率的特征图,大大减少网络输出的停车位数量以及负样本数量,为后续设计端到端的模型提供保障.不同阶段的输出结果的感受野不同,将感受野大的特征层和感受野小的特征层直接在通道维度进行拼接,会使得最终得到的特征层在通道层面感受野混叠.所以本文在融合不同阶段之间的特征时,先通过注意力机制对它们的感受野进行调整,使得输出特征层在通道层面的感受野保持一致.

本文使用ContextBlock模块[13]对主干网络各个阶段的特征输出进行感受野对齐,该模块采用transformer[14]和SENet[15]的思想对特征进行注意力机制融合,ContextBlock模块结构见图3.通过核大小为1的卷积集成所有输入通道维度的信息,再通过归一化指数函数获得全局的注意力权重,将权重施加在输入特征层上,便获得了全局信息,使用残差结构将全局信息和原特征图进行相加,得到最终对齐感受野的特征图.

1.3 检测头

在将网络提取的特征转换为任务所需的结果时,检测头起到了至关重要的作用.本文设计的一阶

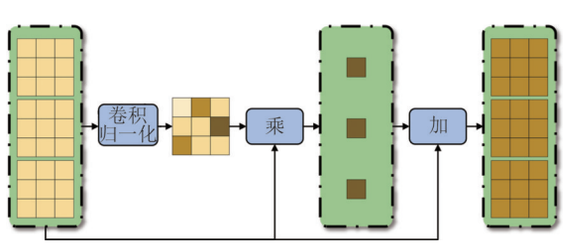


图3 ContextBlock结构图

Fig. 3 Architecture of ContextBlock

段停车位检测方案需要预测建模为箭头的停车位进入线的首尾坐标、停车位的朝向角度,其中坐标需要预测左点和右点的坐标,既包含回归任务也包含分类任务.具体来讲,定位任务和分类任务所需要的特征是不一样的,分类往往关注目标物突出的部分,而定位则关注目标物的边缘[16],如果不加区别地对这两个任务使用同一特征,会出现定位很准确的框分类分数偏低的现象[17].在分类分支,本文使用全连接层对目标物进行特征归纳,全连接层对每个区域使用不同的权重,使得它的空间敏感性更强,更能够区分目标物的局部以及整体.在定位分支则采用滑动窗口设计的卷积层对物体的边界进行定位.

本文设计的二阶段方案检测头结构如图1中检测头所示,检测头模块包含了两个卷积子模块(Conv head),分别负责回归进入线的位置以及停车位相关的角度正弦和余弦值,两个全连接子模块(FC head)分别负责预测停车位是否被占用以及位置精准度.

本文采用回归的方式代替主流方案中的分类模块,以获得准确的停车位进入线和分隔线的夹角预测值.样本中离散角度值数量小,所以本文设计方案不直接回归进入线和分隔线的夹角,而是回归分隔线相对X轴正向的旋转角度,这种回归方式的样本数量不再受到停车位的样式影响,而是取决于停车位整体在图像中的朝向角度,且通过旋转数据增强方法可以很方便地扩充样本数量.直接预测范围在0~360°的角度,值域过大,收敛难度大,本文选择对角度的正弦值和余弦值进行预测,将范围由0~360缩减到0~1,最终通过正弦值和余弦值组合计算出角度的真实值.

检测头输出值的完整解码过程见式(3),其中$(x_p, y_p, l_p, \theta_{p1,p2,p3})$是网络的输出结果,$w_G$和$h_G$是预测特征所在特征层的宽度和高度下采样倍数,也即每个网格的实际像素宽度和高度,x_G和y_G则是网格点左上角的实际像素坐标,使用范围在0~1的sigmoid函数,配合网格左上角点位置和网格宽高即可将c_x和c_y的预测结果控制在该网格点内.解码之后可以获得停车位进入线中点的坐标、进入线的长度、进入线偏转角的cos值,停车位朝向角度的cos值和sin值.

$$\begin{cases} c_x = \sigma(x_p) \cdot w_G + x_G \\ c_y = \sigma(y_p) \cdot h_G + y_G \\ l = e^{l_p} \cdot w_G \\ \theta_{\cos,\sin} = \tanh(\theta_{p1,p2,p3}) \end{cases} \tag{3}$$

至此,上述检测头能够根据每个锚点预测一个代表停车位进入线的箭头,共计19 × 19 = 361个箭头,后处理筛选过程必不可少.而后处理筛选需要一个能够评判进入线质量的依据,如果直接将类别分数作为箭头预测的准度,会使得一些位置精准但是类别分数低的箭头被过滤掉[17].为了保留位置精准的箭头,本文额外设计了一个分支,对箭头与真值的归一化距离进行预测,以此代表箭头的精准度,该距离越接近0,箭头预测越精准,在后处理筛选时重要性排序越高.

2 试 验

2.1 试验准备

本文试验用到的硬件平台为配备12 G显存的英伟达泰坦X显卡的塔式服务器,软件主要包括Pytorch、CUDA、CUDNN等深度学习开发库以及OpenCV图像处理库,具体版本号见表1.

表1 试验配置

Tab.1 Configs of the experiment

类别	详细信息
CPU	Intel Xeon,缓存16 384 kB主频2 594 MHz
GPU	Nvidia TITAN X, 显存12 G显卡驱动515.65.01
GPU库	CUDA==10.2, CUDNN==7 605
深度学习框架	torch==1.6.0, torchvision==0.7.0
图像处理库	Opencv-python==4.7.0.68

数据集采用同济大学公开的ps2.0停车位检测数据集,该数据集由AVM环视影像系统采集,包含训练集图片9 827张、测试集图片2 338张,其中测试集又分为室内、室外正常光照、室外雨天、室外阴天、室外倾斜车位、室外街道光照6个场景,每张环视图片分辨率为600×600像素,代表10 m×10 m的物理长度范围.主流的停车位检测方案在评价算法性能时通常单独计算精确率和召回率,没有考虑平均精确率,且判断正负样本的阈值过大导致正样本仍然存在较大误差.本文重新设计停车位检测算法的平均精确率评价指标AP1:5,据此对各类算法在公开数据集上重新测评.AP1:5是指分别以1~5的5个像素值作为正负样本判断的距离阈值,当检测结果的左点和右点均满足这个阈值时,认定检测结果为正样本,再求出最终的平均AP,得到AP1:5.

2.2 主干网络试验

本文采用上述试验配置,验证停车位检测算法的性能.首先对主干网络进行试验,使用一些常用的主干网络构建停车位检测网络,进行训练之后在ps2.0数据集上测试其性能.试验结果见表2,其中ResNet50由He等[18]于2016年提出,在网络结构设计中引入残差结构应对退化问题,此后被许多不同领域的算法作为主干网络使用.hourglass[19]由多个包含了上采样、下采样、跳层连接的沙漏模块组成,反复的上采样再下采样,使其能够捕获特征之间的关联性,因而常被用于关键点检测任务中,而本文设计的停车位检测方案也包含两个入口点之间关联信息的提取,所以该主干网络也被选为对比网络之一.表2中,P(Precision)表示精确率,R(Recall)表示召回率,FPS(Frames Per Second)表示每秒帧数.

表2 主干网络试验结果

Tab.2 Results of backbone

网络名称	P	R	AP1:5	FPS
ResNet50	0.463	0.429	0.658	52.2
hourglass	0.451	0.442	0.661	26.8
本文	0.516	0.502	0.687	88.7

网络设计时需要考虑的往往有三个因素:宽度、深度和基数,式(4)表示的是一个卷积层,该卷积层由C路分支卷积结构Convs构成,最终该卷积层的输出是C路分支之和,其中C就是基数.基数反映的是网络正向传播时特征层的结合以及反向传播时梯度的结合.

本文所使用的CSP-DarkNet53相比于ResNet丰富了特征和梯度的融合.停车位检测任务中,网络最需要关注的是车位线的特征,而这些特征都是单一的颜色,且形状相对固定,所以在前向传播时会存在大量的特征冗余.CSP-DarkNet53结构(见图2)的外部残差和内部残差相结合,增加了冗余特征的复用,使得后续层无需学习冗余的特征.同时这种基数的增加能够给网络提供隐式的正则化深度监督,因为每个小残差首尾都会通过大残差相连来缩短损失传递距离.所以相比ResNet50,CSP-DarkNet53能够在将参数缩小至35.89%的同时,在停车位检测任务上取得更好的效果.

$$\mathrm{Layer}(\boldsymbol{X}) = \sum_{i=1}^{C} \mathrm{Convs}_i(\boldsymbol{X}) \tag{4}$$

hourglass反复的堆叠沙漏结构大大增加了网络的参数量,其网络结构设计的原理在于通过反复的上下采样融合定位和语义信息,这和本文设计的瓶颈网络结构相冲突,所以速度降低的同时在精度方面并没有取得优势.

2.3 瓶颈网络试验

瓶颈网络模块是不同层级特征融合的关键部分,本文在停车位检测网络中使用不同的瓶颈网络进行试验.FPN[20]以主干网络的最后三个阶段的输出特征层为输入,将深层的语义信息融合到浅层,输出三个不同尺寸的特征图.PAN[21]在FPN的基础上额外增加了从浅层到深层的定位信息与语义信息的融合.MergeFPN是本文提出的特征融合方案.试验结果见表3.

由表3可知,FPN和PAN的AP1:5精度和速度在本文的停车位检测框架上表现并不理想.本文输入的维度为(3,600,600),进行reshape之后变为(3,608,608)的图片输入主干网络,输出主干网络最后三个阶段的特征图,并以该尺度划分网格进行锚点预测,FPN和PAN划分的锚点数量为:7 581 = 19 × 19 + 38 × 38 + 76 × 76,图4对38 × 38的网格划分进行可视化,网络会在每个网格内预测一个锚点偏移.

表3 瓶颈网络试验结果

Tab.3 Results of neck

网络名称	P	R	AP1:5	FPS
FPN	0.435	0.414	0.583	58.9
PAN	0.493	0.441	0.600	55.4
MergeFPN	0.516	0.502	0.687	88.7

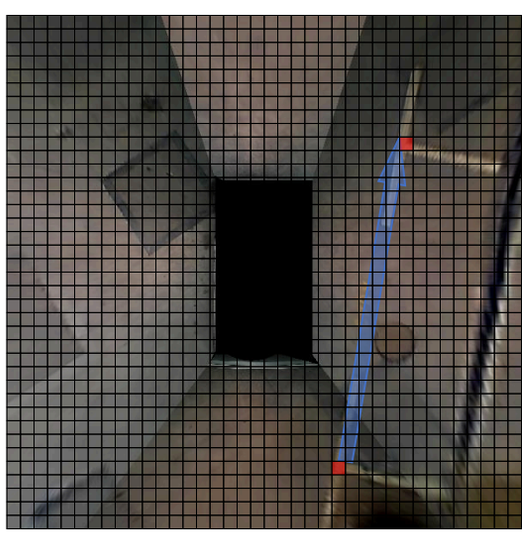


图4 预测密度

Fig.4 Density of predictions

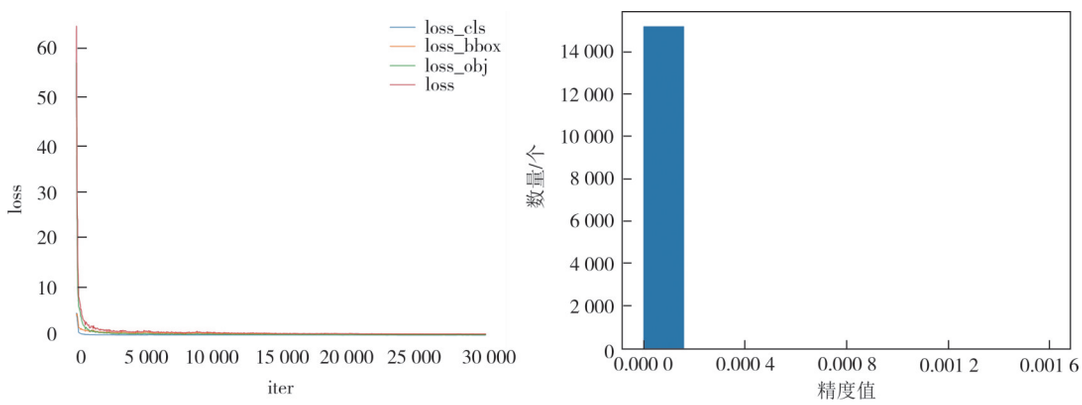
FPN和PAN在停车位检测任务上表现欠佳的原因可以归纳为:

1)网络使用1对1的标签分配,一张图片中7 000多个预测结果,只有5个以内的预测结果需要计算位置损失.

2)预测的锚点放置太密集,如图4所示,即使是网络随机输出位置偏移,也会有比较准确的锚点落在入口点上(图中红色网格),使得位置损失非常小.

3)7 000多个预测结果中每个都需要计算位置精度二元交叉熵损失,但是只有5个以内的标签为1,其他标签均为0,网络只需要将所有样本的位置精度都预测为0,准确度就高达:99.93%=(7 581−5)/7 581×100%.

也就是说FPN和PAN网络只需要随机输出位置,再输出全0的位置精度就能够在训练时得到一个较小的损失.图5是以FPN为瓶颈网络搭建的停车位检测网络的训练结果.图5(a)为损失下降曲线,可以看出损失函数中,位置损失(入口点的L1距离损失)和类别损失(交叉熵损失)始终都维持在一个较低的值,因为这两个损失只需要在正样本上计算,而正样本数量十分少.位置精度损失pos_conf(交叉熵损失)在初始阶段特别大,之后再维持在一个稳定的值.图5(b)为训练后期,网络推理时的位置精度值显示,可以发现该值全部都接近0,验证了上述的说法,也就是说在FPN和PAN这类网格划分十分细密的网络预测中,网络倾向于什么事都不做,就可以获得一个较低的损失.而本文设计的MergeFPN瓶颈网络将输出锚点的数量降低为FPN/PAN的4.76%,有效缓解了正样本损失被淹没的现象.



(a)FPN损失下降曲线　　(b)位置精度值分布直方图

图5　不同瓶颈网络训练信息

Fig. 5　Training statistics of different necks

2.4　检测头试验

检测头以瓶颈网络的输出特征图作为输入,输出最终的结果,本文对解耦头和耦合头分别进行试验,其中耦合头采用的是全卷积头,一个卷积模块直接将特征层维度转变为(19,19,2+1+3),试验结果见表4.

表4　检测头试验结果

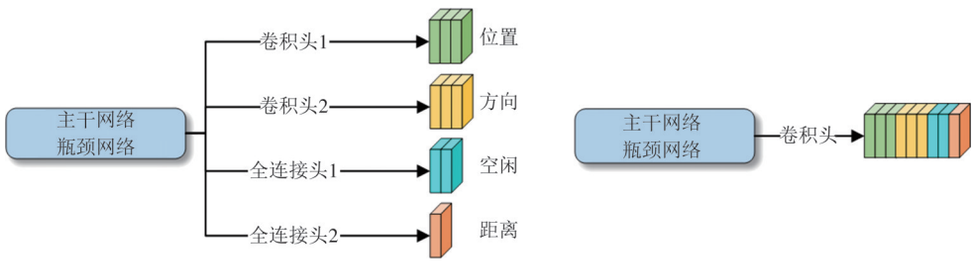
Tab.4　Results of detection head

网络名称	P	R	AP1:5	FPS
Coupled	0.461	0.453	0.634	96.7
Decoupled	0.516	0.502	0.687	88.7

本文设计的解耦头对各个任务分开处理,而耦合头将停车位的位置、角度、空闲占用、精准度统一使用一个卷积层进行预测,这其中混杂着不同种类的分类信息和回归信息,如图6(a)和图6(b)所示.耦合检测头没有充分利用卷积层和全连接层的特性,全连接层在特征层的每个区域都使用不同的权重,能够捕捉停车位整体的信息,所以在空闲占用和精度分类任务上优势更大,而卷积层使用滑动窗口进行特征提取,不同区域使用共享的权重,所以特征视野更小,更适合位置和角度这类回归任务,所以统一使用卷积层的耦合检测头在AP1:5精度方面不如将回归和分类信息使用卷积层和全连接层分开处理的解耦头.

2.5　整体性能

本文将上述改进应用于停车位检测算法中,形成了AFPSD(Anchor-Free Parking Slot Detection)停车位检测方案.为了以统一的标准对各个算法进行测评,本文对VPS-Net和DMPR-PS的代码进程重



(a)解耦头 (b)耦合头

图 6 不同检测头结构

Fig. 6 Architecture of different detection heads

构,增加了以 AP1:5 作为性能指标的评测代码,试验结果见表 5.

表 5 不同停车位检测算法结果

Tab.5 Results of different parking slot detection algorithm

网络名称	P	R	AP1:5	Param/M	FPS
VPS-Net	0.513	0.492	0.675	61.85	48.8
DMPR-PS	0.507	0.493	0.666	30.31	83.3
AFPSD	0.516	0.502	**0.687**	**12.22**	**88.7**

从表 5 中 AP1:5 和 FPS 两列可以看出,本文设计的 AFPSD 卷积神经网络锚点端到端停车位检测算法在精度上高于 VPS-Net 和 DMPR-PS,分别高出 1.2% 和 2.1%. 精度上的领先得益于本文使用统一的箭头建模方式,将停车位两个入口点作为一个整体,而 VPS-Net 和 DMPR-PS 均将同一个停车位的两个入口点作为孤立的个体分开进行检测,忽略了二者之间的关联性特征. 在特征融合方面,本文设计了多进单出的瓶颈网络,将训练时的负样本数量降低为原来的 4% 左右,极大缓解了训练时的不平衡现象,同时考虑到分类任务和回归任务的区别,使用了独立的检测分支,最终 AFPSD 的 AP1:5 达到了 0.687. 在速度方面,本文设计的 AFPSD 是完全端到端的停车位检测网络,无需进行非极大值抑制后处理和入口点匹配以及朝向判断流程,在流程上进行了极大的简化,在网络结构上采用了跨越阶段连接轻量化设计,多进单出融合等设计达到了 12.22 M 的参数量和 88.7 的 FPS.

2.6 特点分析

2.6.1 旋转敏感性

使用矩形描述子对入口点进行建模的方式对停车位角度的变化很敏感,即,当测试集中出现了朝向不一致的车位时,网络可能无法准确对其进行识别,而本文使用的进入线朝向建模法检测的是点的位置,具有旋转不变性. 为了验证这一特性,本文将一个车位进行四次随机角度的旋转之后对其进行检测. 图 7 第一行是采用矩形描述子的 VPS-Net 的检测结果,第二行是本文提出的 AFPSD 的检测结果,每列是同一个旋转角度的图片,第一列是未旋转的原图.

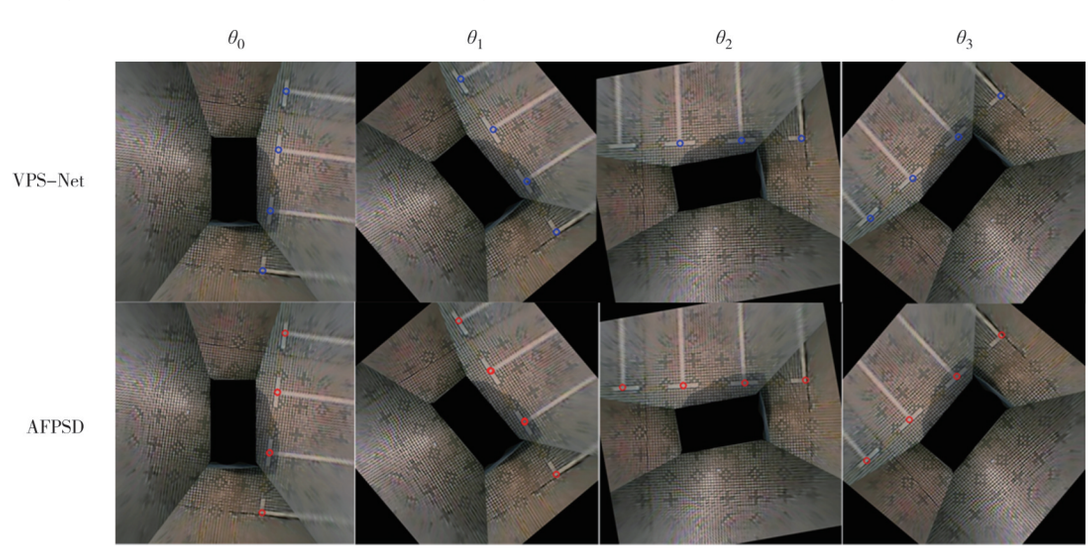


图 7 旋转车位检测结果

Fig. 7 Results of rotating parking slot detection

从图中可以看出,网络都能够将未经旋转的原图中的车位检测出来,但是当车位发生不同角度的旋转时,VPS-Net网络的检测结果开始发生变化,第二列可以看出蓝色检测结果已经发生了轻微的偏移,在第三列时出现了漏检现象;而当旋转角度进一步加大,第四列时蓝色检测结果出现了更大程度的偏移.而本文的算法提取的是箭头的点特征,当车位朝向出现旋转时,从点特征的层面来讲,只是点的位置发生了变化,并没有角度上的不同,所以本文对各种旋转角度的图像的检测表现几乎没有差异.

2.6.2　空闲占用

本文设计的停车位检测网络能够端到端地直接预测停车位是否被占用,其他方案则需要额外训练一个停车位空闲占用网络,并制作分类数据集对其进行训练.图8是本文空闲占用预测的效果,其中黑色进入线表示该车位处于空闲状态,红色进入线表示该车位已经被其他车辆占用.由图8可以看出,在不同地面上,本文算法均能检测出不同状态下的车位.

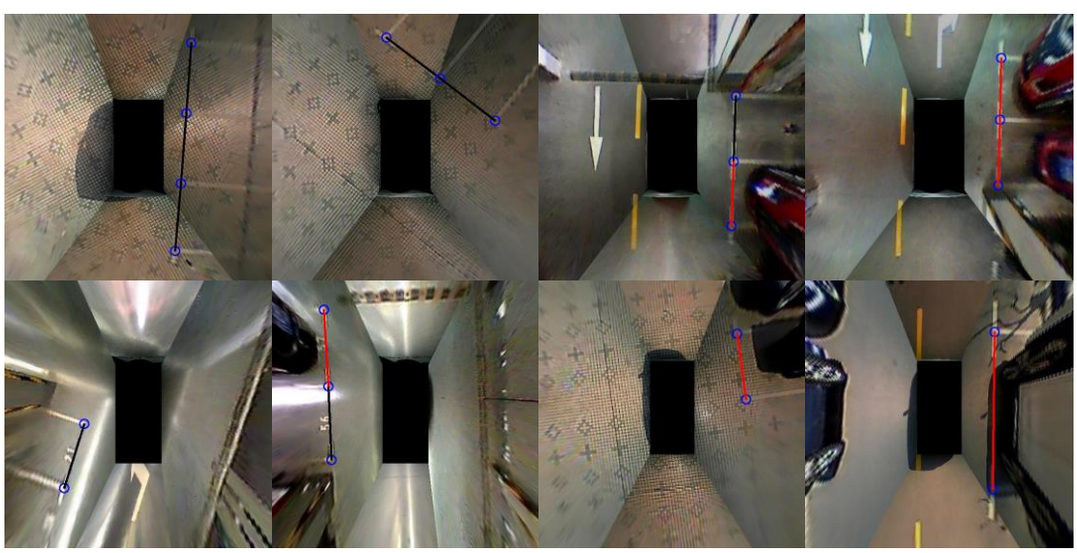


图8　占用分类结果

Fig. 8　Results of occupancy classification

3　结论

针对主流停车位检测方案对入口点检测和匹配分开进行处理,提取的特征存在冗余,并需额外设计空闲占用分类网络等问题.本文提出了基于全景环视的端到端停车位检测方案,该方案对停车位进行整体性建模,利用属于同一停车位的两个入口点之间的关联信息一体化生成停车位的位置信息,去除入口点匹配、朝向判断和空闲占用分类步骤.在结构设计上,本文以无锚框检测方案为基础,对瓶颈网络进行正负样本平衡化改进,同时设计了能够缓解特征不匹配问题的检测头,最终实现了端到端的停车位检测方案的设计.本文的主要贡献如下:

1)对停车位进行整体性建模,而非单独处理两个入口点,简化了停车位入口点匹配和朝向判断模块;

2)设计了多进单出的瓶颈网络进行特征融合,降低了输出样本数量以缓解正负样本不平衡问题;

3)设计了分类回归解耦检测头,缓解特征不匹配问题;

4)设计了端到端的无锚框停车位检测算法,能够一体化地检测出停车位位置和占用信息,无需额外进行入口点匹配和朝向判断.

本文方法对于下斜坡这种不符合地平面假设的场景,检测结果无法回投到世界坐标系,将在后续的研究中进一步优化.

参考文献

[1] LI L, ZHANG L, LI X, et al. Vision-based parking-slot detection: a benchmark and a learning-based approach[C]//2017 IEEE International Conference on Multimedia and Expo (ICME). Hong Kong, China, IEEE, 2017: 649-654.

[2] ZHANG L, HUANG J, LI X, et al. Vision-based parking-slot detection: A DCNN-based approach and a large-scale benchmark

dataset [J]. IEEE Transactions on Image Processing, 2018, 27(11): 5350-5364.

[3] 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.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[5] LI W, CAO L, YAN L, et al. Vacant parking slot detection in the around view image based on deep learning [J]. Sensors, 2020, 20(7): 2138.

[6] REDMON J, FARHADI A. Yolov3: an incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.

[7] HUANG J, ZHANG L, SHEN Y, et al. DMPR-PS: a novel approach for parking-slot detection using directional marking-point regression [C]//2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai, China: IEEE, 2019: 212-217.

[8] WU Z, SUN W, WANG M, et al. Psdet: efficient and universal parking slot detection [C]//2020 IEEE Intelligent Vehicles Symposium (IV). Las Vegas, NV, USA: IEEE, 2020: 290-297.

[9] 吕雪杰. 基于环视视觉的自动泊车系统[D]. 北京: 北京交通大学, 2020.
LÜ X J. Automatic parking system based on surround view vision [D]. Beijing: Beijing Jiaotong University, 2020. (in Chinese)

[10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2020,42(2): 319-327.(in Chinese)

[11] LI W, CAO L, YAN L, et al. Vacant parking slot detection and tracking during driving and parking with a standalone around view monitor [J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2021, 235(6): 1539-1551.

[12] 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. Seattle, WA, USA: IEEE, 2020: 390-391.

[13] CAO Y, XU J, LIN S, et al. Gcnet: non-local networks meet squeeze-excitation networks and beyond [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul, Kovea: ICCV, 2019.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speach and signal processing (ICASSP), Toronto, ON, Canada, 2021.

[15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.

[16] SONG G, LIU Y, WANG X. Revisiting the sibling head in object detector [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 11563-11572.

[17] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich Germany: ECCV, 2018: 784-799.

[18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.

[19] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation [C]//Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14. Springer International Publishing, 2016: 483-499.

[20] 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. Honolulu, HI, USA: 2017: 2117-2125.

[21] 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. Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.