

改进 SSD 算法的道路小目标检测研究

邹慧海¹, 侯进²

(1. 西南交通大学 唐山研究生院, 河北 唐山 063000; 2. 西南交通大学 信息科学与技术学院, 成都 611756)

摘要: 在道路场景中, 因小目标分辨率低且特征不明显, 传统的目标检测算法难以确认其所属类别和位置信息, 导致检测精度低、检测速度慢、漏检率高。提出一种改进 SSD 的道路小目标检测算法 RFG_SSD。在 SSD 网络结构的主干部分和检测部分之间, 通过引入改进的特征金字塔网络结构, 融合浅层和深层感受野的特征信息, 以获得小目标语义信息丰富的特征图。将深层特征提取网络 ResNet 50 作为改进网络的主干特征提取网络, 提高整体网络的检测精度。为加快网络运算速度, 基于检测层结构, 利用全局平均池化层代替全连接层, 减少网络参数量。实验结果表明, 与 SSD、VGG16+SFPN 等算法相比, 该算法能够有效提高小目标检测性能, 且加快检测速度, 其在 BDD100K 数据集上的平均精度和检测速度分别为 98.05% 和 85.56 frame/s, 小目标检测个数相较于 SSD 算法提高 3 倍多。

关键词: 小目标检测; SSD 算法; ResNet50 网络; 特征金字塔网络; 全局平均池化

开放科学(资源服务)标志码(OSID):



中文引用格式: 邹慧海, 侯进. 改进 SSD 算法的道路小目标检测研究[J]. 计算机工程, 2022, 48(5): 281-288.

英文引用格式: ZOU H H, HOU J. Research on road small target detection with improved SSD algorithm[J]. Computer Engineering, 2022, 48(5): 281-288.

Research on Road Small Target Detection with Improved SSD Algorithm

ZOU Huihai¹, HOU Jin²

(1. Graduate School of Tangshan, Southwest Jiaotong University, Tangshan, Hebei 063000, China;

2. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China)

[Abstract] In a road scene, the resolution of small targets is low, and the characteristics are not obvious. Traditional target detection algorithms find it difficult to identify category and location information, resulting in low detection accuracy, slow detection speed, and a high missed detection rate. This study proposes a small target detection algorithm (RFG_SSD) based on an improved SSD. By introducing the improved Feature Pyramid Network (FPN) structure and integrating the feature information of shallow and deep receptive fields, which obtain a feature image with rich semantic information of small targets, to improve the performance of small target detection. The deep feature extraction network ResNet50 is used as the backbone of the improved network to improve detection accuracy of the overall network. To further accelerate the network operation speed based on the detection layer structure, the Global Average Pooling (GAP) layer is used to replace the full connection layer to reduce the number of network parameters. The experimental results on the BDD100K dataset show that compared with the SSD and VGG16+SFPN algorithms, this algorithm can effectively improve the performance of small-target detection and increase the detection speed. Its average accuracy and detection speed are 98.05% and 85.56 frame/s, respectively. The number of small target detections is more than three times higher than that of the SSD algorithm.

[Key words] small target detection; SSD algorithm; ResNet50 network; Feature Pyramid Network (FPN); Global Average Pooling (GAP)

DOI: 10.19678/j.issn.1000-3428.0061499

0 概述

随着深度学习技术的发展, 语义分割、目标跟踪、目标检测、车道线检测在计算机视觉领域得到广泛关注, 尤其是无人驾驶汽车。目标检测任务是对车辆前

方道路环境的目标进行识别检测及定位, 因实际道路环境的复杂多变性, 使得该任务面临诸多挑战。

传统的目标检测方法通常分为区域选择、特征手工提取、分类回归 3 个步骤。文献[1]提出一种差异训练、多尺度的目标检测方法 DPM。该方法通过计算模

基金项目: 四川省科技计划项目“基于深度学习算法研究 LHAASO 高能宇宙线成份鉴别”(2020SYSY0016)。

作者简介: 邹慧海(1996—), 男, 硕士研究生, 主研方向为计算机视觉、目标检测; 侯进(通信作者), 副教授。

收稿日期: 2021-04-28 修回日期: 2021-06-26 E-mail: 2499478454@qq.com

型梯度方向直方图(Histogram of Oriented Gradients, HOG),利用支持向量机(Support Vector Machine, SVM)对模型进行分类训练,从而得到物体的梯度信息。但是,在网络选择区域时因耗费大量时间,导致检测速度慢,而且手工提取的特征语义信息比较少,导致检测精度差。

近年来,将深度学习与目标检测相结合的方法成为研究热点。相比传统目标检测方法,深度学习方法的模型网络层数深、参数多。在目标检测领域中,该方法具有较高的检测精度和较优的实时性。基于此,研究人员将深度学习理论融入到目标检测工作中。深度学习目标检测方法分为^[2]单阶段(one-stage)检测方法和双阶段(two-stage)检测方法。基于区域候选框的two-stage检测方法提取输入图像的候选区域,并对其进行候选区域的分类与修正,从而实现目标检测。代表模型有R-CNN系列,包括R-CNN^[3]、Fast R-CNN^[4]、Faster R-CNN^[5],以及Mask R-CNN^[6]等,都具有较优的检测精度,但是检测速度较慢,从而无法满足实时性要求。与此相反,one-stage检测方法是一种端到端的目标检测方法,基于回归分析的原理,利用卷积神经网络提取图像的特征信息,以识别检测目标的位置和类别,能够有效加快检测速度。代表模型有YOLO系列,包括YOLOv1^[7]、YOLO9000^[8]、YOLOv3^[9]及YOLOv4^[10]算法,SSD^[11]和EfficientDet^[12]等。YOLOv4^[10]检测模型作为单目标检测模型类别中的

最优模型,在主干部分引入跨阶段局部网络(CSPNet)^[13]和特征融合金字塔网络(PAN)^[14],以解决梯度信息重复问题,从而提高同一物体在不同尺度上的识别精度,还可以通过数据增强等训练方式,提高检测精度和实时性。

在汽车行驶过程中,目标检测方法不仅对近距离的大目标进行精准实时检测,而且还要对远距离的小目标进行位置检测。针对道路前方行人和车辆类别的小目标检测问题,本文提出一种基于改进SSD的道路小目标检测算法。通过引入改进特征金字塔网络,融合不同层感受野特征信息,利用深层特征网络ResNet50代替VGGNet16,同时在残差结构中引入批量归一化,从而提高检测精度并加快收敛速度。

1 相关工作

1.1 SSD 算法

SSD是单阶段多类别的目标检测算法,具有较优的检测精度和较快的检测速度,但是小目标检测效果并不理想。SSD^[11]网络结构如图1所示。SSD网络结构主要分为2个部分:1)VGGNet16特征提取网络,通过提取图片目标的特征信息,得到特征图;2)分类回归层,对每个候选框进行分类与回归,从而识别检测出图片目标。

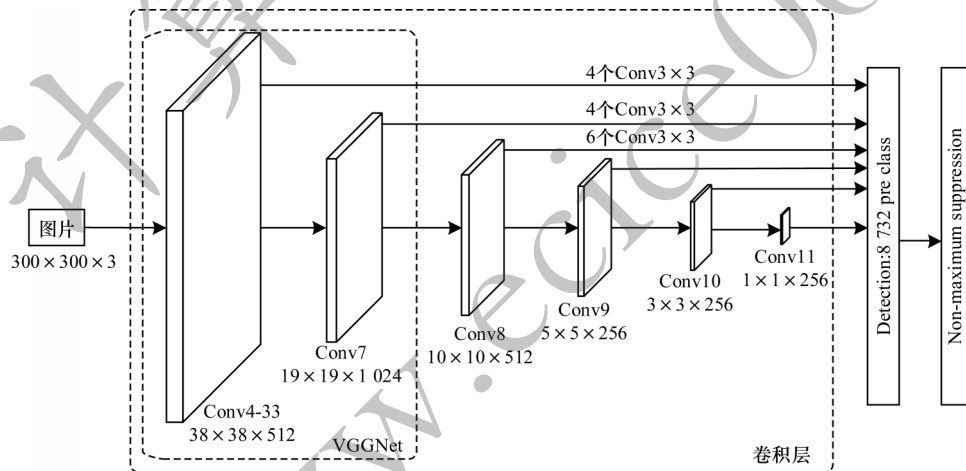


图1 SSD网络结构

Fig.1 Structure of SSD network

SSD算法分别在尺度为38×38、19×19、10×10、5×5、3×3、1×1的特征图上产生候选框,以检测不同尺度的目标大小。本文是在这些尺度特征图后通过连接特征金字塔融合结构进行消融实验,提高小目标的检测性能。SSD算法候选框的数量为38×38×4+19×19×6+10×10×6+5×5×6+3×3×4+1×1×4=8 732。先验框尺度的计算如式(1)所示,位置计算如式(2)所示:

$$h_k^a = \frac{s_k}{\sqrt{\alpha_y}}, w_k^a = s_k \sqrt{\alpha_y} \quad (1)$$

$$d = (d^{cx}, d^{cy}, d^w, d^h) \quad (2)$$

其中: h 、 w 为先验框的长、宽; s_k 为面积; α_y 为系数因素; d 为距离; cx 、 cy 为距离 x 、 y 轴系数。

边界框位置的计算如式(3)所示:

$$b = (b^{cx}, b^{cy}, b^w, b^h) \quad (3)$$

研究表明,匹配先验框与边界框遵循的原理是:首先计算边界框与先验框的所有IoU值,当IoU值大于等于设定的阈值时,说明匹配成功,即为正样本;反之为负样本。若存在一个先验框与多个边界框的IoU值超过设定阈值时,此时匹配其中最大的边界框为正样本,剩余的边界框为负样本。

SSD的损失函数采用多任务损失函数,由分类

损失值和边框回归损失值2个部分相加得到。其中分类损失采用交叉熵函数,回归损失采用Smooth_{L1} loss函数,整体损失函数如式(4)所示:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (4)$$

其中: L_{conf} 为分类损失; L_{loc} 为回归损失; α 为权重系数,系数越大,表明回归损失在整体损失中占比越大,在训练过程中正样本的边框坐标优化效率越高。

1.2 ResNet50 特征网络模型

ResNet(Residual Network)^[15-16]是一种用于深度学习目标检测的卷积神经网络主干特征提取网络,因其高效性和实用性,广泛应用于计算机视觉检测、分割、识别等领域。层数越深的网络可以提取到更加复杂的特征信息,效果也越好,但是网络层数的增加导致深层梯度在反向传播过程中越来越小甚至消失,从而阻碍网络的收敛,即网络退化问题。网络退化问题不仅增大了训练误差和测试误差,也降低了网络精度。

ResNet网络中的残差结构(Residual Block)可以解决网络退化问题,残差网络结构对比如图2所示。

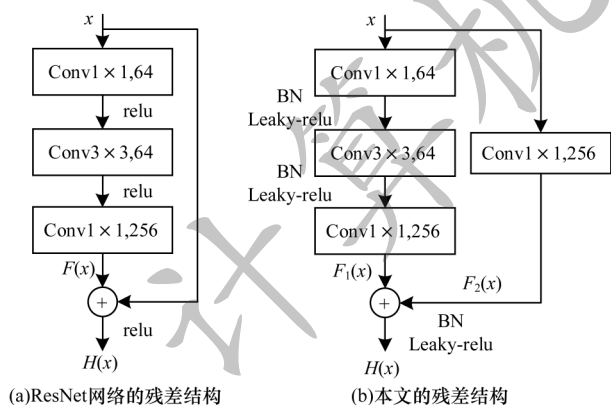


图2 不同的残差网络结构对比

Fig.2 Comparison of different residual networks structure

在残差网络结构中,当输入为 x 时,残差学习特征为 $H(x)$,网络层的原始学习特征为 $F(x)$ 。则残差结构的学习特征 $H(x)$ 如式(5)所示:

$$H(x) = F(x) + x \quad (5)$$

相比原始学习特征 $F(x)$,残差网络能够学习残差特征 $H(x)$,其原因是当卷积网络层数加深时,特征随网络层向下传递,而梯度随网络层向上回传。在此过程中,因网络层数太深导致梯度消失。然而,本文的残差网络可以解决此类问题,即使左边深层网络向上回传的梯度值过小甚至消失,但是右边原始梯度值向上回传,整体梯度值等于两边梯度值相加,从而解决梯度消失的问题。

本文使用的改进残差结构如图2(b)所示,在原残差结构的基础上,将 $\text{relu}^{[17]}$ 激活函数替换为 $\text{Leaky-relu}^{[18]}$ 激活函数,并加入批量归一化(Batch Normalization, BN)^[19]。在网络训练过程中,当 relu

函数在部分输入小于0的情况下,输出恒为0,导致对应权重难以更新,从而无法训练学习特征。 Leaky-relu 函数在输入小于0的情况时,输出持续变化,以更新权重并继续学习。BN是对特征图进行归一化处理,加快收敛速度。同样在训练时,部分输入只通过直线连接前向传播,导致整体结构变成一条直线结构。因此,在残差结构右侧加入 1×1 卷积核操作,可以解决上述整体直线结构问题。

2 本文算法

2.1 特征金字塔结构改进

随着网络层数的加深,主干特征网络提取特征图的感受野逐渐增大,导致小目标的有效信息缺失,从而无法检测出小目标。由于不同深度对应不同层次分辨率的语义特征信息,因此通过学习不同层之间相同的特征信息,提高网络识别目标的精确度。特征金字塔融合结构如图3所示。增加双向融合的特征金字塔网络(Path Aggregation Network, PAN)^[14]结构能够有效解决小目标检测的问题,PAN结构如图3(a)所示。通过自底向上和自顶向下融合不同网络层提取的特征图,整合不同感受野大小和语义强度的特征信息,以得到目标信息更加丰富的特征图,从而提高检测精度。

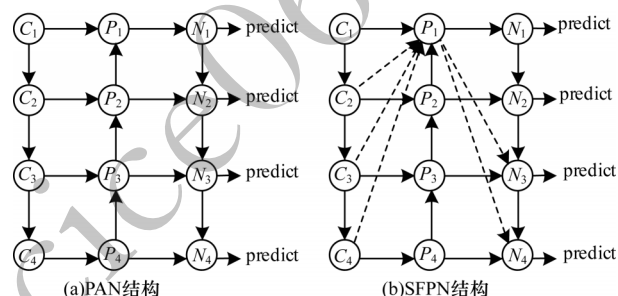


图3 特征金字塔融合结构

Fig.3 Structure of fusion feature pyramid

PAN的提出验证了双向融合的有效性,但是其结构较简单。因此,针对实际小目标检测场景,本文提出更复杂的SFPN特征融合结构,如图3(b)所示。首先进行第1次自底向上的特征融合,将每层特征网络提取的特征图进行相应倍数的上采样,并分别与第1层、上一层特征图级联相加融合得到特征图 P_i (i 表示层数),然后再进行第2次自顶向下特征融合,在此过程中,将第1层特征图 P_1 分别与每层特征图 P_i 级联相加融合得到2次融合特征图 N_i ,其他结构保持不变。SFPN特征融合结构的目的是为了融合浅层特征图感受野小(包含全部小目标)的有效信息,得到小目标语义特征信息更丰富的特征图,从而提高小目标检测性能。

2.2 RFG_SSD 网络结构改进

为提高小目标检测性能,改进的SSD网络是通过引入特征融合金字塔SFPN网络,并加载在原SSD特征提取网络之后,对不同尺度分辨率的特征图进

行融合,利用融合浅层和深层的特征图感受野,以提高小目标检测性能,进而分类回归预测目标。同时为提升整体RFG_SSD网络性能,改进RFG_SSD网络结构选用ResNet50替换VGGNet16,并将其作为

改进模型的主干特征提取网络,通过加深网络层数提升性能。改进RFG_SSD网络结构分为主干特征提取结构、SFPN特征融合结构、分类回归检测层结构3个部分,如图4所示。

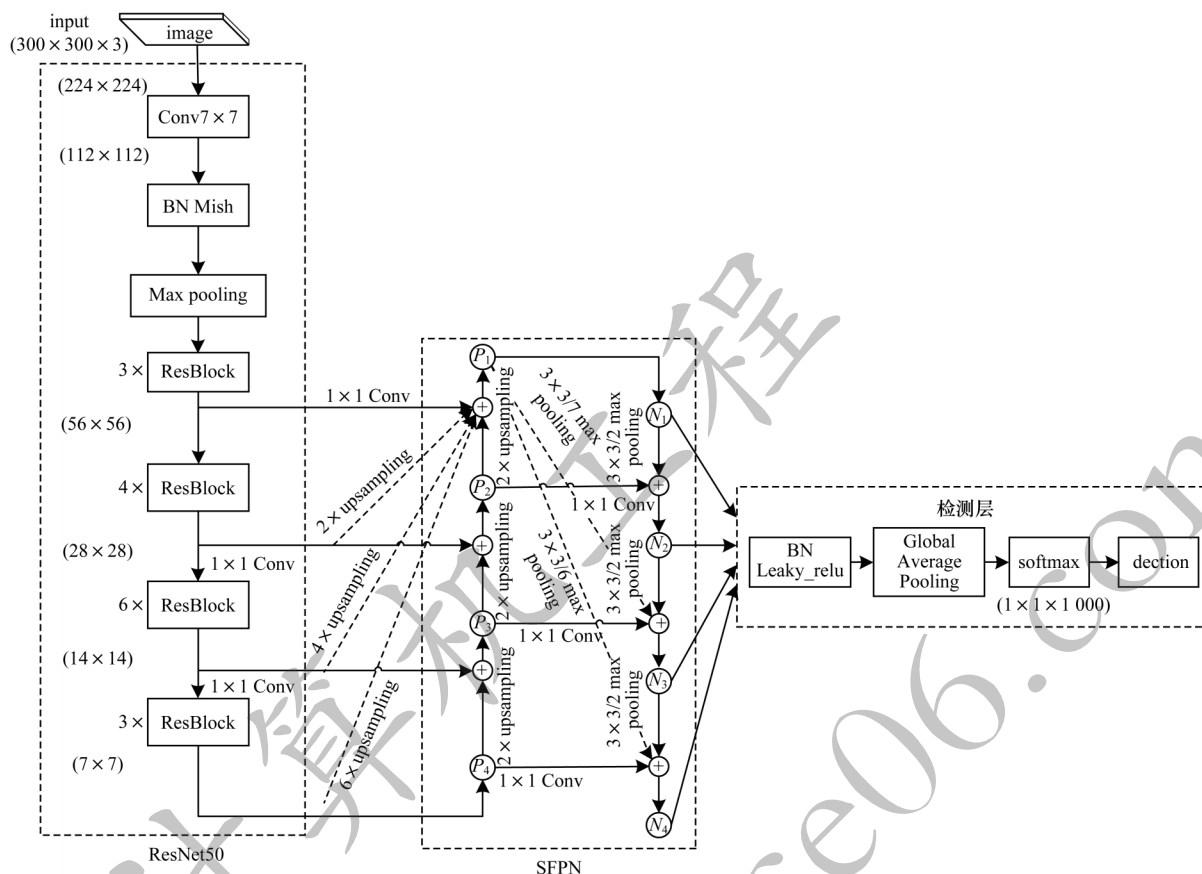


图4 RFG_SSD网络结构

Fig.4 Structure of RFG_SSD network

在主干网络中,输入图片尺寸为 $300 \times 300 \times 3$, 首先通过 ResNet50 网络对输入图片进行特征提取,每经过一个残差块将提取的特征图送入 SFPN 网络中并进行特征金字塔融合操作,尺度分别为 56×56 、 28×28 、 14×14 和 7×7 。在 ResNet50 网络结构中主要使用改进残差结构(ResBlock),引入批量归一化(BN)和 Leaky-relu 激活函数。BN 操作是在特征图的批量大小、长、高 3 个维度上进行归一化处理,目的是通过规范化方法将偏离的分布拉回到标准化,使梯度变大,以加快学习收敛速度,从而解决梯度消失的问题。Leaky-relu 激活函数操作是向上回传梯度,以避免网络退化现象的出现,从而为网络增加了非线性表达能力且提高了网络检测性能。

在 2 次纵向融合过程中,SFPN 特征融合结构将下一层特征图进行 2 倍上采样、上一层特征图进行 1×1 卷积,进而将两者级联相加连接,得到融合不同层语义信息的特征图,使得网络学习特征更加精准,如图 4 中 SFPN 模块实线连接。为提取更多的小目标特征信息,在第

1 次融合过程中,将各层不同感受野大小的特征图 C_i 做 2 倍、4 倍、6 倍上采样后,并与 C_i 特征图融合得到 P_1 ,在第 2 次融合过程中,将 P_1 进行核为 3×3 大小、步长分别为 2、4、6 的 3 种最大池化方式降采样后,并与各层 P_i 融合得到 N_i ,如图 4 中 SFPN 模块虚线连接。最后分别将各层融合得到的高语义信息特征图 N_i 传入到检测层进行分类回归,以预测目标。

在检测层结构中,依次经过批量归一化 BN 和 Leaky-relu 激活函数操作层、全局平均池化层(Global Average Pooling, GAP)^[20]、softmax 层^[21],最终得到目标检测结果。标准目标检测网络模型在特征提取后连接全连接层(Full Connected, FC),其目的是将前面学习得到的特征语义表示映射到样本空间,降维得到一维向量组,并将其输入到 softmax 层进行分类,以得到相应类别目标。由于 FC 层参数冗余,占据整个网络参数数量的 80% 以上,导致运算量过大,且容易过拟合,因此使用代替 FC 层,如图 4 中检测层结构。全局平均池化方式不仅能够降维得到与 FC 相同的结果,而且可以降低 50% 以上的参数量,

同时去除了对输入图片大小的限制,以提高网络运算速度。

3 实验结果与分析

3.1 实验平台配置

本文选用pytorch开源框架进行训练、测试,基于ubuntu16.04操作系统,CPU为Inter Core i7-9700K,3.60 GHz,显卡为NVIDIA GTX2080。编译环境为torch-1.2、torchvision-0.40,cuda10.0、cudnn10.0,python 3.7语言编程。

3.2 BDD100K数据集

针对道路行驶过程中车辆前方的行人和车辆目标,本文选用的数据集必须满足车辆行驶过程中有足够多的行人和车辆,并且是真实的路况环境。因此,本文使用BDD100K数据集^[22],该数据集是2018年伯克利大学发布的大规模、内容多样性的公开驾驶数据集。其中道路目标数据有1×10⁵张图片,包括城市街道、高速公路等路况场景,包含晴天、阴天、雨天等天气环境。这些因素使得数据集具备丰富多样性的行人、车辆路况场景,网络能够学习更丰富的特征,从而适用于道路上各种复杂场景,保障无人驾驶汽车的安全出行。

该数据集的道路目标数据包含训练集7×10⁴张图片、测试集2×10⁴张图片、验证集1×10⁴张图片,包括多种目标类别标签数据。编写代码提取两个数据集的行人、车辆类别标签,并保存得到训练集、测试集、验证集txt格式。同时使用式(1)计算每个类别标签框尺度大小,设定尺度小于或等于19×19的标签框为小目标,统计类别标签个数和小目标标签个数,以得到小目标框个数约占有所有框个数的40%,各标签框数目统计如表1所示。

表1 标签框数目

Table 1 The number of label boxes				
数据集	类别	行人/个数	车辆/个数	小目标框/个数
BDD100K	训练集	90 483	715 300	322 313
	测试集	25 852	204 371	92 089
VOC 2007+2012	训练集	56 721	34 920	44 781
	测试集	18 981	13 840	19 089

3.3 训练与消融实验

本文通过控制变量法(不同模块组合)进行消融实验,以对比各模块作用程度。各网络的结构模块如表2所示。本文使用BDD100K训练集对4种网络进行模型训练实验。训练过程中批次大小为8,迭代次数为500,初始学习率为0.000 1,权值衰减为0.000 5。其中学习率决定得到最优值参数的速度,如果学习率过大,参数可能会跳过最优值,从而导致网络无法收敛甚至发散;如果学习率过小,则优化效率过低,网络长时间无法收

敛,可能得到局部最优。因此,在训练中权值衰减系数的设置是根据模型训练情况来动态改变学习率大小,使得网络在训练迭代中得到最优值参数。

表2 不同网络的模块结构

Table 2 Module structure of different networks	
网络	模块
SSD网络	VGGNet16
ResNet50网络	ResNet50
VGGNet16+SFPN网络	VGGNet16+SFPN
RFG_SSD网络	ResNet50+SFPN

训练时调用原SSD在VOC2007数据集上训练得到的权重并进行初始化,并且在训练中将训练损失值变化绘制成曲线,如图5所示。损失值越小表示模型预测目标越接近真实目标,模型的性能越好。

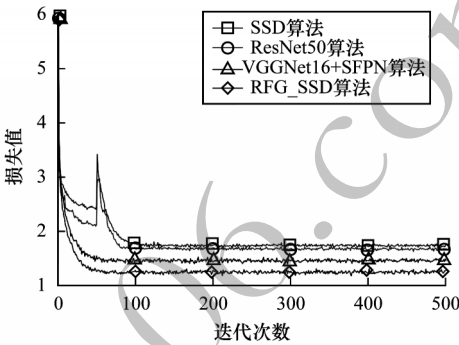


图5 不同算法的损失值对比

Fig.5 Loss values comparison among different algorithms

从图5可以看出,各算法经过迭代100次后损失变化趋于稳定,曲线波动不大。SSD算法使用VGGNet16特征提取网络,损失值最大;只将原SSD结构中特征提取网络替换为ResNet50,损失值排在第2;在原SSD结构中只增加SFPN特征融合网络,损失值排在第3;同时替换ResNet50网络和增加SFPN网络设计得到RFG_SSD网络,损失值最小,相对应的性能最优。

3.4 实验结果

3.4.1 消融实验精度评估

在目标检测领域中,本文选用准确率和召回率来评估系统的性能。准确率是指在所有正样本中,正确目标所占的比例,衡量查准率;召回率是指在所有真实的目标中,被模型正确检测出来的目标所占的比例,衡量查全率。精确率和召回率如式(6)和式(7)所示:

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}} \times 100\%$$
 (6)

$$R = \frac{T_{TP}}{T_{TP} + F_{FN}} \times 100\%$$
 (7)

其中: P 为准确率; R 为召回率; T_{TP} 为模型正确检测的目标个数; F_{FP} 表示模型错误检测的目标个数; F_{FN} 为模型漏检的正确目标个数。

在实际过程中,准确率和召回率相互关联,因此将计算得到的 P 、 R 值绘制成P-R曲线,综合考虑P-R曲线下的面积 A_{AP} 值来评估各类别检测性能, A_{AP} 值越大,表示模型检测精度越高,即性能越好,如式(8)所示:

$$A_{AP} = \int_0^1 P dR$$

(8)

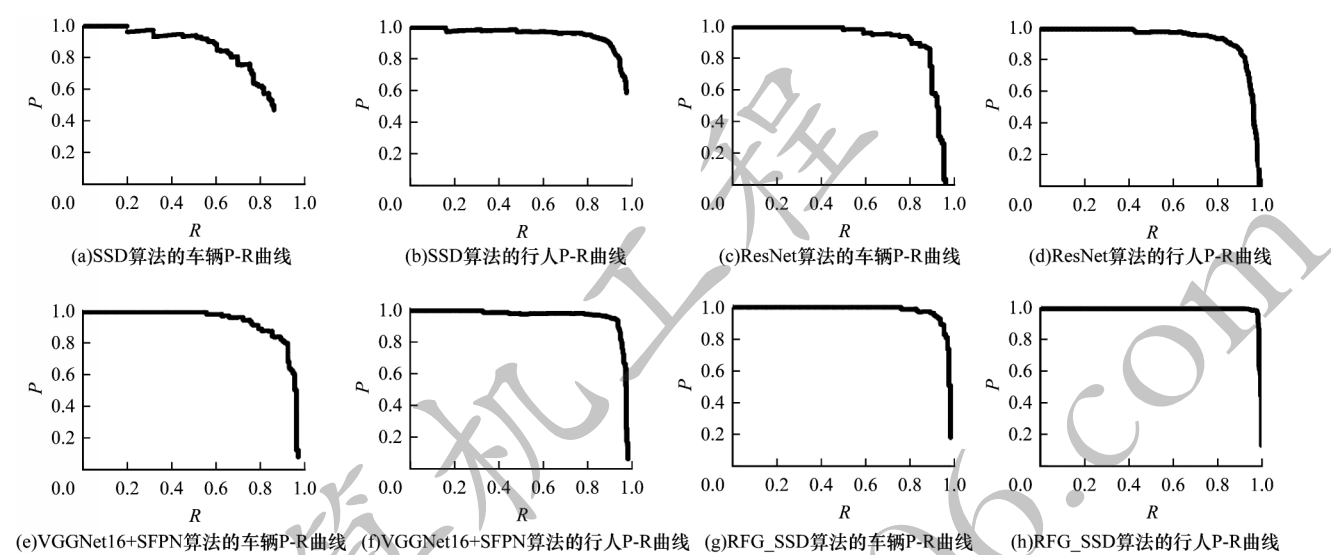


图6 不同算法的P-R曲线对比

Fig.6 P-R curves comparison among different algorithms

不同算法的检测精度对比如表3所示。从表3可以看出,ResNet50算法相对于SSD算法 m_{mAP} 提高了6.42个百分点,表明深层特征提取模块ResNet50能够有效提高模型的检测精度。VGG16+SFPN算法相对于SSD算法 m_{mAP} 提高8.6个百分点,表明SFPN结构有效提高模型的检测精度。最终改进算法RFG_SSD是将ResNet50模块和SFPN模块相结合得到,与SSD相比,其 m_{mAP} 提高12.69个百分点,与单独使用ResNet50模块或SFPN模块相比,其 m_{mAP} 分别提高6.72和4.09个百分点,表明2个模块结合使用能够有效提高模型的检测精度。

表3 不同算法的检测精度对比

Table 3 Detection accuracy comparison among different algorithms			%
算法	车辆 A_{AP}	行人 A_{AP}	m_{mAP}
SSD 算法	76.90	93.82	85.36
ResNet50 算法	90.07	93.49	91.78
VGGNet16+SFPN 算法	92.07	95.86	93.96
RFG_SSD 算法	96.98	99.12	98.05

3.4.2 不同算法的性能对比

算法性能的评估不仅要考虑平均检测精度 m_{mAP} 值,同时也需要评估检测速度,即实时性(F_{FPS})的快慢情况。实时性表示目标检测网络检测一张图片需要的时间,

针对多个类别 N ,本文使用平均值 m_{mAP} 评估模型整体性能,如式(9)所示:

$$m_{mAP} = \frac{1}{N} A_k^{AP}$$

(9)

本文分别使用原SSD算法、ResNet50算法、VGG16+SFPN算法和RFG_SSD算法对BDD100K测试集测试检测精度。编写代码利用各个算法训练得到的权重模型计算 P 、 R 值,并绘制成P-R曲线,得到各个算法类别AP值,如图6所示。

假设网络检测一张图片需要 s , F_{FPS} 值越大,表示检测时间越少,速度越快。 F_{FPS} 如式(10)所示:

$$F_{FPS} = \frac{1}{s}$$

(10)

本文分别使用SSD算法、YOLOv4算法及改进SSD算法对VOC2007+2012测试集与BDD100K测试集进行测试评估,并对比各类算法的检测精度和速度。其中文献[23]算法在原SSD浅层网络结构上设计浅层特征增强模块,以提高浅层小目标的特征提取能力,文献[24]算法在检测识别时利用小目标强化检测模块与原SSD级联方式,提高小目标的检测效果。这2种改进算法都只是针对小目标部分结构,虽然小目标检测效果有所提高,但是整体检测性能并没有显著提高。本文改进算法RFG_SSD不仅能够提高小目标检测性能,还提高了整体性能。

本文利用编写代码计算各类算法在两个数据集上的平均精度和实时性表现,同时使用式(1)计算模型检测的目标框尺度大小,并统计尺度小于19×19的小目标框检测个数。在VOC2007+2012和BDD100K测试集上不同方法的检测结果及适用场景对比如表4所示。从表4可以看出,在VOC2007+2012和BDD100K测试集上,VGG16+SFPN算法相对文献[22]、文献[23]及

SSD 算法的检测精度和速度都有所提高,并且小目标检测个数增多,验证了 SFPN 模块结构提升小目标检测性能的有效性,但是与 YOLOv4 算法相比,VGG16+SFPN 算法的性能优势并不明显。在 BDD 100K 数据集上,本文提出的改进算法 RFG_SSD 在检测精度和速度上相对于其他算法都具有明显优势,与目前性能最优的检测算法 YOLOv4 相比,其整体性能相差不大,精

度降低了 0.97 个百分点,检测速度加快 1.8 frame/s,小目标检测数目相差约 600 个。与原 SSD 算法相比,RFG_SSD 算法精度提高了 12.69 个百分点,检测速度提高 34.33 frame/s,小目标框检测数高出 3 倍以上。与文献[23]、文献[24]算法相比,RFG_SSD 算法的检测效果得到明显提高,表明 ResNet50+SFPN 模块能够有效提高检测准确度。

表 4 不同算法的目标检测结果与适用场景对比

Table 4 Target detection results and applicable scenarios comparison among different algorithms

算法	数据集	车辆 $A_{AP}/\%$	行人 $A_{AP}/\%$	$m_{mAP}/\%$	检测小目标框数	检测速度 /(frame·s ⁻¹)	适用场景
文献[22] 算法	BDD100K	84.94	72.21	78.57	33 051	50.12	遥感图像目标检测
	VOC2007+2012	82.41	71.23	76.82	11 097	49.56	
文献[23] 算法	BDD100K	83.63	80.05	81.84	39 359	49.88	多类别目标检测
	VOC2007+2012	85.70	79.50	82.60	10 231	48.11	
SSD 算法	BDD100K	76.90	93.82	85.36	19 054	51.23	多类别目标检测
	VOC2007+2012	80.80	72.50	76.65	8 909	46.00	
YOLOv4 算法	BDD100K	98.46	99.58	99.02	60 532	83.76	多类别目标检测
	VOC2007+2012	95.66	97.83	96.75	18 565	76.51	
VGG16+SFPN 算法	BDD100K	92.07	95.86	93.96	45 533	62.12	行人、车辆多尺度目标检测
	VOC2007+2012	86.83	95.63	91.23	13 127	58.44	
RFG_SSD 算法	BDD100K	96.98	99.12	98.05	59 904	85.56	行人、车辆多尺度目标检测
	VOC2007+2012	94.14	98.45	96.29	17 304	72.13	

本文 RFG_SSD 算法和 SSD 算法对数据集图片进行目标检测对比,如图 7 所示。

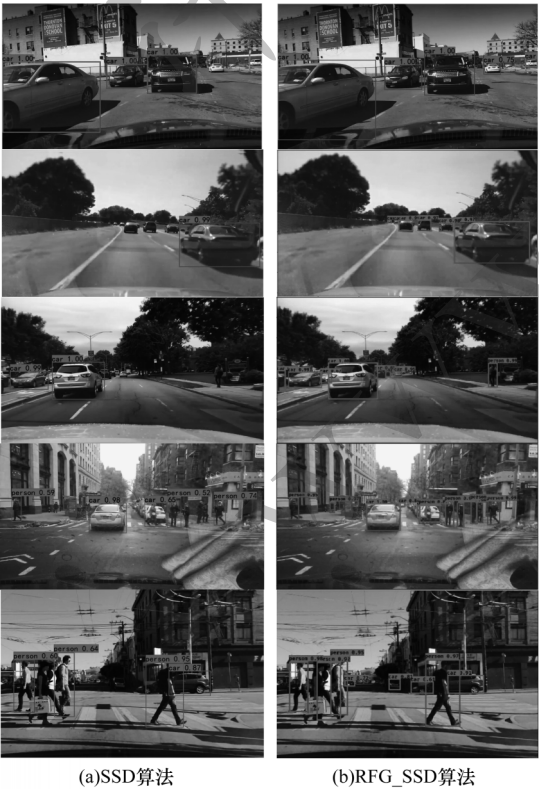


图 7 不同算法的目标检测结果

Fig.7 Target detection results of different algorithms

从图 7 可以看出,对于图片中行人和汽车目标,SSD 算法未识别检测出小目标,只检测出部分大目标,RFG_SSD 算法能够精准检测出大、小目标。结果表明,RFG_SSD 算法能够精准检测出图片中任意尺度的目标,并且速度更快,性能得到显著提升,以达到预期行人、汽车多尺度检测的效果。

4 结束语

本文提出一种改进 SSD 算法的道路小目标检测算法。在 SSD 网络结构基础上,通过引入改进特征金字塔结构 SFPN 融合不同层特征的语义信息,将特征网络 VGGNet16 替换为网络层数较多的 ResNet50,以提高整体网络性能,并通过增加批量归一化、全局平均池化等结构,从而降低参数量。实验结果表明,相比 SSD、VGG16+SFPN 算法,该算法能够显著提高检测精度,并加快检测速度,实现多尺度目标检测,其在 BDD100K 数据集上的精确度达到 98.05%。由于本文算法仅对道路行人、车辆进行检测,因此后续将改进检测模型,使其适用于道路多类别、多尺度的检测。

参考文献

[1] DAVID F. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 47(7): 6-7.
[2] 段仲静,李少波,胡建军,等. 深度学习目标检测方法及其主流框架综述[J]. 激光与光电子学进展, 2020, 57(12):

- 1-16.
DUAN Z J, LI S B, HU J J, et al. Summary of deep learning target detection methods and mainstream frameworks[J]. Laser and Optoelectronics Progress, 2020, 57(12): 1-16. (in Chinese)
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2015: 1440-1448.
- [5] 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, 2015, 39(6): 1137-1149.
- [6] HE K, GKIOXARI G, DOLLÁR D, et al. Mask R-CNN [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2017: 1-10.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2017: 6517-6525.
- [9] JOSEPH R, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2021-03-25]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. [2021-03-25]. <https://arxiv.org/abs/2004.10934>.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of European Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2016: 21-37.
- [12] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2020: 1-10.
- [13] 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 IEEE Conference on Computer Vision and Pattern Recognition, Workshops. Washington D. C. , USA; IEEE Press, 2020: 390-399.
- [14] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018: 1-10.
- [15] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016: 1-10.
- [16] 田强, 贾小宁. 基于深度残差网络的车标识别[J]. 吉林大学学报(理学版), 2021, 59(2): 319-324.
TIAN Q, JIA X N. Vehicle logo recognition based on deep residual network [J]. Journal of Jilin University (Science Edition), 2021, 59(2): 319-324. (in Chinese)
- [17] LI Y Z, YUAN Y. Convergence analysis of two-layer neural networks with ReLU activation [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA; ACM Press, 2017: 597-607.
- [18] XU J, LI Z S, DU B, et al. Reluplex made more practical: Leaky ReLU [C]//Proceedings of IEEE Symposium on Computers and Communications. Washington D. C. , USA; IEEE Press, 2020: 1-7.
- [19] 刘建伟, 赵会丹, 罗雄麟, 等. 深度学习批归一化及其相关算法研究进展[J]. 自动化学报, 2020, 46(6): 1090-1120.
LIU J W, ZHAO H D, LUO X L, et al. Research progress on batch normalization of deep learning and its related algorithms [J]. Acta Automatica Sinica, 2020, 46(6): 1090-1120. (in Chinese)
- [20] HSIAO T Y, CHANG Y C, CHOU H H, et al. Filter-based deep-compression with global average pooling for convolutional networks [J]. Journal of Systems Architecture, 2019, 95(12): 9-18.
- [21] LIU W Y, WEN Y D, YU Z D, et al. Large-margin softmax loss for convolutional neural networks [EB/OL]. [2021-03-25]. <https://arxiv.org/pdf/1612.02295.pdf>.
- [22] YU F, CHEN H F, WANG X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning [EB/OL]. [2021-03-25]. <https://arxiv.org/pdf/1805.04687.pdf>.
- [23] 史文旭, 谭代伦, 鲍胜利. 特征增强 SSD 算法及其在遥感目标检测中的应用 [J]. 光子学报, 2020, 49(1): 154-163.
SHI W X, TAN D L, BAO S L. Feature enhancement SSD algorithm and its application in remote sensing images target detection [J]. Acta Photonica Sinica, 2020, 49(1): 154-163. (in Chinese)
- [24] 雷华迪, 陈东方, 王晓峰. 基于级联 SSD 的目标检测算法 [J]. 计算机工程与设计, 2020, 41(12): 3517-3524.
LEI H D, CHEN D F, WANG X F. Target detection algorithm based on cascaded SSD [J]. Computer Engineering and Design, 2020, 41(12): 3517-3524. (in Chinese)