



DOI: 10.12086/oe.2019.180331

融合多尺度上下文卷积特征的 车辆目标检测



高琳*, 陈念年, 范勇

西南科技大学计算机科学与技术学院, 四川 绵阳 621010

摘要: 针对现有的基于卷积神经网络的车辆目标检测算法不能有效地适应目标尺度变化、自身形变以及复杂背景等问题, 提出了一种融合多尺度上下文卷积特征的车辆目标检测算法。首先采用特征金字塔网络获取多个尺度下的特征图, 并在每个尺度的特征图中通过区域建议网络定位出候选目标区域, 然后引入候选目标区域的上下文信息, 与提取的目标多尺度特征进行融合, 最后通过多任务学习联合预测出车辆目标位置和类型。实验结果表明, 与多种主流检测算法相比, 本算法具有更强的鲁棒性和准确性。

关键词: 卷积神经网络; 多尺度特征; 上下文信息; 车辆检测

中图分类号: TP391.41; TB872

文献标志码: A

引用格式: 高琳, 陈念年, 范勇. 融合多尺度上下文卷积特征的车辆目标检测[J]. 光电工程, 2019, 46(4): 180331

Vehicle detection based on fusing multi-scale context convolution features

Gao Lin*, Chen Niannian, Fan Yong

Department of Computing Science and Technology, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China

Abstract: Aiming at the problems of the existing vehicle object detection algorithm based on convolutional neural network that cannot effectively adapt to the changes of object scale, self-deformation and complex background, a new vehicle detection algorithm based on multi-scale context convolution features is proposed. The algorithm firstly used feature pyramid network to obtain feature maps at multiple scales, and candidate target regions are located by region proposal network in feature maps at each scale, and then introduced the context information of the candidate object regions, fused the context information with the multi-scale object features. Finally the multi-task learning is used to predict the position and type of vehicle object. Experimental results show that compared with many detection algorithms, the proposed algorithm has stronger robustness and accuracy.

Keywords: convolutional neural network; multi-scale feature; context information; vehicle detection

Citation: Gao L, Chen N N, Fan Y. Vehicle detection based on fusing multi-scale context convolution features[J]. *Opto-Electronic Engineering*, 2019, 46(4): 180331

1 引言

车辆检测是目标检测中较为重要的领域之一, 它

在智能交通、道路场景监控、无人驾驶等方面有着广泛的应用。由于车辆往往置身于复杂的外部环境, 复

收稿日期: 2018-06-19; 收到修改稿日期: 2018-11-04

基金项目: 四川省教育厅科技项目(18ZA0501); 四川省科技创新苗子工程资助项目基金(2017113)

作者简介: 高琳(1976-), 男, 博士, 讲师, 主要从事计算机视觉, 模式识别的研究。E-mail: 81831283@qq.com

杂背景、拍摄角度造成的尺度差异以及车辆自身形态变化,是车辆检测技术所面临的主要难题。

针对这些问题,传统方法通常采用图像金字塔实现多尺度目标检测,通过对原图像进行梯次下采样得到不同分辨率的图像,以金字塔的形状排列,自底向上分辨率逐步降低,然后对金字塔的每层图像进行特征提取,经典的特征提取方法包括 HOG(histogram of gradient)和 SIFT(scale-invariant feature transform)等,提取的特征最终被用于分类识别,进而实现目标检测。这类方法的代表是著名的 DPM(deformable part model)算法^[1-2],其建立一个 HOG 特征金字塔,从金字塔顶部到底部,分别捕获由粗到细的各个尺度下的梯度信息,能够达到较好的检测结果。但该方法对图像金字塔的每层都要计算特征,时间复杂度高,并且手工特征依赖于设计人员的经验,通常不能适应目标的多样性变化,其鲁棒性和泛化能力较差。

近年来,随着深度学习技术的发展,深度卷积神经网络被逐步应用于车辆检测领域^[3-4],并取得了重大的性能提升。卷积神经网络能够自适应学习目标特征,克服了手工定义特征的不足^[5],学习得到的特征对目标形变、阴影以及光照等干扰因素具有更好的不变性。Faster RCNN 算法^[6]将候选区域选择和卷积特征提取分类融合在一起,能够实现目标检测的端对端训练,在速度上和精度上都具有领先的优势,但在目标定位时仅利用了单层的卷积特征图,单一尺度的感受野无法适应目标在图像上的尺度变化。针对多尺度目标检测问题,出现了许多基于卷积特征金字塔的检测方法。FPN(feature pyramid network)算法^[7]通过引入特征金字塔,将多个尺度下的卷积特征进行融合,以解决目标检测中的多尺度问题。YOLO 算法^[8]将检测问题简化为回归问题,YOLOV2 算法^[9]针对 YOLO 目标定位

效果差的问题,引入锚框,在速度和精度上有很大的优化,但对于重叠目标、小目标检测效果较差,随后的 YOLOV3 算法^[10]采用多尺度预测,提高了对小目标的检测精度。MS-CNN 算法^[11]利用卷积神经网络建立不同感受野、不同尺度的特征图,然后在各个尺度下设置相应的检测器,通过这些尺度互补的检测器共同检测目标。SSD(single shot MultiBox detector)算法^[12]使用 VGG Nets 作为提取特征的基础网络,该算法在建立特征金字塔时,能够对不同卷积层的多尺度特征进行重复利用,降低了计算和时间代价,具有较好的实时性,但 SSD 的特征金字塔放弃了低层的特征图,损失了高分辨率特征,因此造成小目标检测性能的不足。

本文针对复杂的室外场景,提出一种融合多尺度上下文信息的车辆目标检测方法,采用特征金字塔网络获取目标的多尺度特征,并在每个尺度特征下定位候选目标;引入候选目标的上下文信息,与多尺度特征进行融合,增加特征对目标尺度、光照变化以及遮挡的适应性,进而提高检测与识别的精度。在 PASCAL VOC 公共数据集和自己构建的专用数据集上,分别与多种主流方法进行了实验对比,结果表明本文方法具有更优的检测性能。

2 车辆目标检测算法

本文算法框架如图 1 所示,首先初始化搭建的卷积神经网络模型参数,其中卷积层参数是利用 ImageNet 图像数据集预先训练得到,其他层的参数都是采用随机方式初始化;为了获得图像的多尺度特征表达,利用卷积神经网络提取层次特征,进而建立特征金字塔;然后根据 RPN 网络生成目标候选区域,在特征金字塔各个层级中找到对应的目标区域,从而提取出目标的多尺度特征;为进一步增强目标特征的判

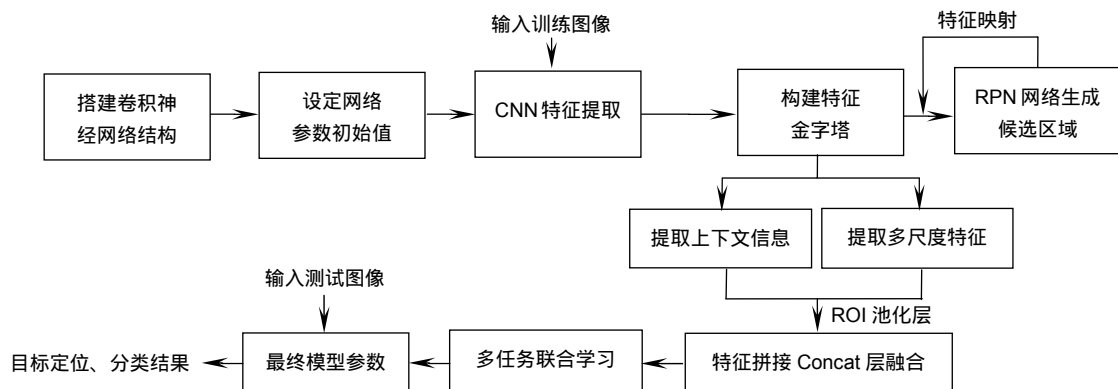


图 1 本文车辆目标检测算法的卷积神经网络模型流程图

Fig. 1 Flow chart of convolutional neural network model of vehicle object detection algorithm

别能力,通过分析目标邻域特征获得上下文信息;最后,将池化后的特征在 Concat 层进行特征融合。整个网络结构采用多任务联合学习,同时实现目标的检测定位与类型识别。

2.1 网络结构

与 Faster RCNN 方法类似,本文方法也分为两个阶段来检测目标,先利用区域建议网络生成候选目标,然后对候选目标区域进行分类,获得最终的目标位置和类型。整个处理网络的结构如图 2 所示。与常规的 Faster RCNN 不同的是,采用 FPN 算法提取的特征金字塔替换了单尺度特征图,多个尺度的特征分别输入至相应的 RPN。通过引入候选区域上下文信息提取模块,将上下文信息与多尺度特征进行融合后,输入至分类器实现车辆的检测和识别。

2.2 多尺度特征提取

2.2.1 特征金字塔

建立特征金字塔网络 FPN(feature pyramid network)提取多尺度图像特征。整个网络通过自下而上、自上而下以及侧边的连接。将不同卷积层中不同语义层次的特征进行融合,从而丰富了各个尺度下的特征信息,网络结构如图 3 所示。

自下而上的部分就是卷积神经网络的前馈传播过程,特征尺度逐层增加,特征的语义层次也在不断提高。从图 3 中可以看出,本文中特征图的尺度缩放比例为 2,该部分的特征图记为 $\{C_2, C_3, C_4, C_5, C_6\}$,分别由卷积神经网络 ResNet-50^[13]中的卷积层(conv2, conv3, conv4, conv5, conv6)提取得到,特征图与输入图像的尺度比例分别为 $\{4, 8, 16, 32, 64\}$ 。

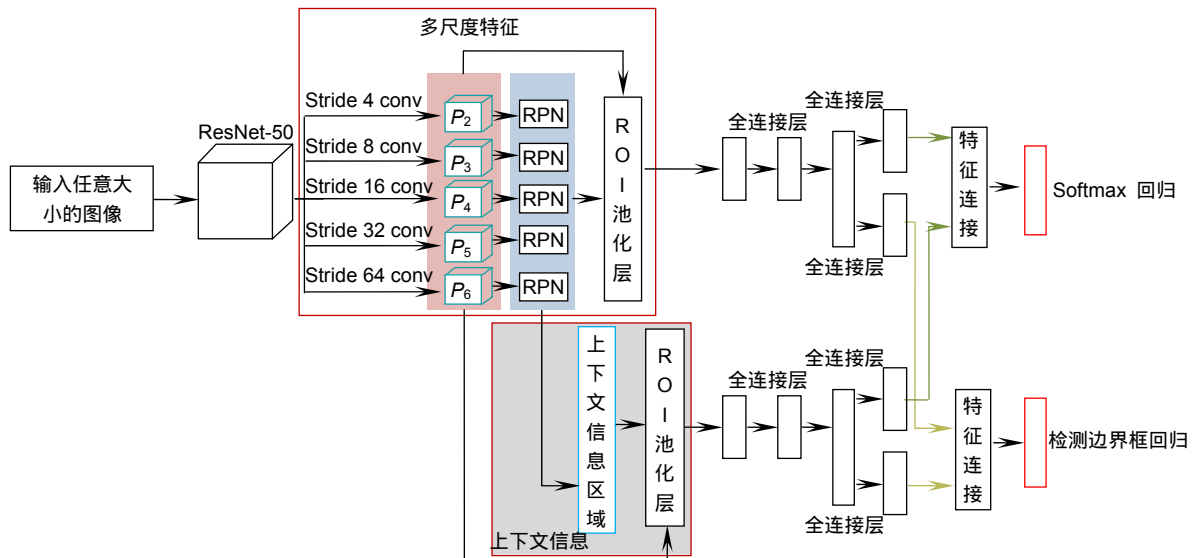


图 2 本文车辆目标检测算法的卷积神经网络模型结构图

Fig. 2 Structure diagram of convolutional neural network model of vehicle object detection algorithm

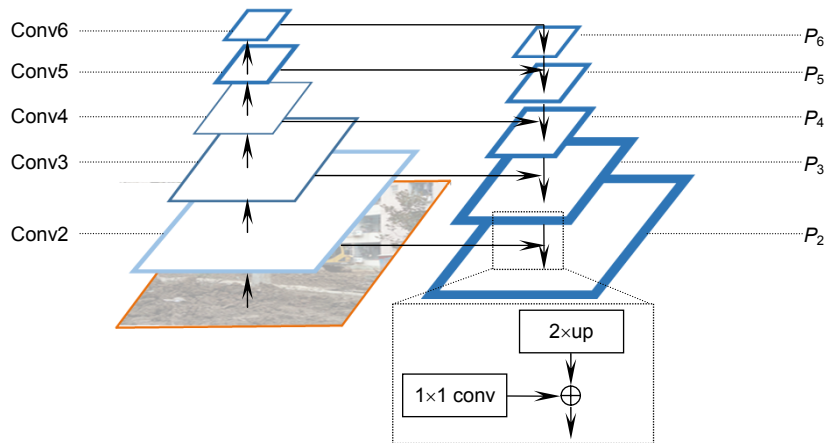


图 3 特征金字塔网络

Fig. 3 Feature pyramid network

自上而下的过程是通过上采样来扩大特征的大小,使得上采样后的特征图不仅与下层特征图具有相同的尺寸,而且还具备了更多的语义信息。

在自底向上和自顶向下的这两个路径中,相对应的两个同等尺度的特征图,前者包含的语义信息较少,但对目标定位较精准,而后的语义信息更多,但经过了多次下采样后,损失了目标的空间信息。因此,通过横向连接可以结合两者的优点,实现特征互补,生成融合后的特征图 $\{P_2, P_3, P_4, P_5, P_6\}$,其空间分辨率与 $\{C_2, C_3, C_4, C_5, C_6\}$ 相对应。

2.2.2 候选区域定位

采用 Faster RCNN 算法中的候选区域定位方法,并加以改进。利用区域建议网络 RPN(region proposal networks)生成候选区域。RPN 的输入是卷积神经网络的 conv4 卷积层特征图。采用 3×3 大小的卷积核作为滑动窗口,在每个滑动窗口中心所对应的原图像位置上,按照尺度比 $\{128, 256, 512\}$ 和长宽比 $\{1:1, 1:2, 2:1\}$ 分别生成 9 种锚框,然后将每个滑动窗口得到的特征矩阵通过 1×1 卷积压缩成低维向量,对其分别进行类型分类和边界回归。

Faster RCNN 算法中的 RPN 是在单张特征图上选择不同大小的锚框来解决多尺度问题,特征图的感受野较为单一,而实际环境中,目标尺度复杂多变,单一感受野不能覆盖所有目标。为此,将 FPN 生成的特征金字塔取代 conv4 特征作为 RPN 的输入,在金字塔每层的特征图上分别增加一个 RPN 网络。基于特征金字塔的 RPN,能够在不同感受野下处理多尺度特征,不同层的 RPN 将生成不同尺度的锚框。如果锚框与真实目标区域之间的 IOU(intersection-over-Union)大于设定的阈值 T_h ,则该锚框标记为正样本;若 IOU 值低于阈值 T_l ,则其被记为负样本。将所有锚框按照 RPN 网络预测的得分由大到小进行排序,对前 N 个锚框进行非极大抑制运算,最终筛选出其中的 N_s 个作为候选目标。

2.3 上下文信息融合

上下文信息反映了目标和背景之间的关联关系,这些信息可以用于判断目标在图像中是否出现。在针对车辆目标检测时,复杂环境如房屋、草坪、道路等,会对检测造成很大影响,结合环境上下文信息有助于从背景中区分出车辆。本文方法中,上下文信息是在特征图中候选目标的周围区域提取,如图 4 所示,候选目标区域及其周围背景分别对应于红色和蓝色的矩形框,从蓝色框区域中提取出上下文信息。目标和背景区域的卷积特征,经过 ROI 池化操作后,分别送入全连接层,将产生的两组固定长度的特征向量进行连接,获得最终的融合上下文信息的多尺度特征。

2.4 损失函数

本文中实现了车辆目标检测网络的端到端训练,并采用多任务学习的方式,同时完成对车辆的定位和车型识别,其损失函数定义如下:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

其中 i 为训练样本在 Batch 中的序号, p_i 是第 i 个锚框预测为目标概率值, p_i^* 为真值, t_i 表示预测边界框的坐标, t_i^* 表示坐标真值。利用 Softmax 损失函数 L_{cls} 进行多种类别目标的识别,用 Smooth L1 损失函数实现目标位置的回归,计算 L_{reg} 时仅考虑正样本的边界框坐标。通过对最小化损失函数来优化模型参数,实现神经网络模型的最终效果。

3 实验与分析

3.1 实验数据

本文实验中用到的数据集分为两个部分,一个是 PASCAL VOC 公开数据集,简称 VOC 数据集,另一

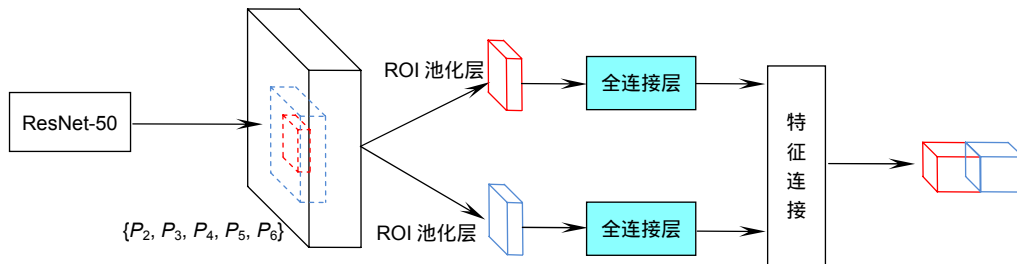


图 4 上下文信息提取

Fig. 4 Context information extraction

个是自建的数据集, 简称 SC 数据集。VOC 数据集包含了街道、公路、停车场等通用场景, 车辆分为公交车(Bus)、轿车(Car)两种类型, 共 3376 张图像。该数据集已含有车辆标注信息。网络训练时, 随机选取其中的 2481 张样本建立训练集, 其余样本建立测试集。SC 数据集有 1924 张图像样本, 均采集于工程施工环境, 主要的检测目标是工程车辆, 包括的车型有吊车(Crane)和挖掘机(Digger)。采用数据增广方法扩充训练样本数量, 以减轻训练模型的过拟合, 增广操作主要包括平移、旋转以及灰度变换等, 通过增广将 SC 数据集的样本数量扩展为 3848 张, 其中随机选择 3078 张图像建立训练集, 其余图像建立测试集。

3.2 训练步骤及参数配置

在模型训练的初始化过程中, 首先利用 ImageNet 数据集^[14]对网络结构中用于提取卷积特征的 ResNet-50 部分进行预训练, 网络中的其他参数则为产生的随机数。对整个网络的训练步骤如下:

- 1) 将所有图像数据按照 PASCAL VOC 数据集格式进行转换;
- 2) 根据训练样本数量, 对小样本数据集进行数据增强;
- 3) 利用随机梯度下降法对式(1)的损失函数进行优化, 获得最优的网络模型参数。

训练过程中, 学习率设为 0.001, 迭代次数为 70000 次, 经过约 20 h 的训练, 获得最终的模型参数。

本文算法在深度学习框架 Caffe^[15]下利用 Python 语言实现, 操作系统为 Linux Ubuntu 16.04, 实验的硬件平台为: Intel Xeon E5-1630 v3@3.7 GHz 四核处理器, Nvidia GTX 1080Ti 11 GB GPU 显卡, 16 GB 内存。算法在两个数据集上测试的平均速度为 4 f/s。

3.3 实验结果及讨论

在 VOC 和 SC 数据集上, 与多种主流算法进行了

仿真对比实验, 其中包括 YOLOV2, YOLOV3, SSD, R-FCN^[16]。采用的定量分析指标为平均检测精度 (average precision, AP)、平均检测精度均值(mean average precision, mAP)以及 F1 值, 其计算公式分别如下:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, \quad (2)$$

$$AP = \sum_i (R_i - R_{i-1}) P_i, \quad (3)$$

$$mAP = \frac{1}{S} \sum_C AP_C, \quad (4)$$

$$F1 = \frac{2P \cdot R}{P + R}, \quad (5)$$

$$MacroF1 = \frac{\sum F1}{S}, \quad (6)$$

其中: P 指目标类别的查准率, R 指目标类别的查全率, C 表示目标类别, AP 表示某类目标的平均检测精度, 即 PR 曲线下的面积。 S 表示类别数量, mAP 为所有类别目标的平均检测精度的平均值。 $F1$ 为 P 和 R 的加权平均, $MacroF1$ 是 $F1$ 的平均值。 指标 $MacroF1$ 在评估算法性能时, 能够同时兼顾查全率和查准率。

表 1 给出的是 VOC 数据集上训练得到的检测结果。 实验中 根据检测定位框和标注定位框之间的 IOU 值来确定是否检测到目标, 设置 IOU 阈值为 0.5, 仅将大于该阈值的检测结果视为正确。 从表 1 的对比数据可以看出, 本文算法的 mAP 值和 $MacroF1$ 值均优于其他目标检测算法。 YOLOV3 算法与本文算法的性能最为接近。 YOLOV3 算法是 YOLO、YOLOV2 算法的改进版本, 与本文算法类似, 其通过多尺度预测的方式提高对小目标的检测精度。 在 Bus 目标上, YOLOV3 的 AP 值比本文算法高 0.2%, 但 Car 目标的 AP 值比本文算法低 1.7%。 此外, 本文算法检测不同类型车辆时, Car 的检测结果好于 Bus, 其原因是公交车在不同拍摄角度下的形变较大, 算法中选定锚框的长宽比例没有很好地满足形变条件。

表 1 不同算法在 VOC 数据集下的检测性能

Table 1 Detection performance of different algorithms under VOC data sets

Algorithm	Bus		Car		mAP/%	MacroF1/%
	AP/%	F1/%	AP/%	F1/%		
YOLOV2	79.8	83.3	76.5	84.2	78.1	83.8
YOLOV3	87.6	86.9	87.7	87.2	87.6	87.0
SSD	79.4	86.4	76.1	84.8	77.7	85.6
R-FCN	85.9	86.8	86.1	87.0	86.0	86.9
Ours	87.4	87.0	89.4	87.6	88.4	87.3

对于 SC 数据集,其主要难点在于施工场景图像中的背景比较复杂,并且目标工程车具有比普通车辆更多的姿态和尺度变化,表 2 显示了算法对 SC 数据集的处理性能,图 5 和图 6 为不同算法的检测效果图。

从表 2 的结果可以看出,本文算法的 mAP 值仍然高于其他算法。YOLOV3 算法在检测 Crane 目标的 AP 值比本文算法高 0.1%,但 Digger 目标检测的 AP 值比本文算法低 0.9%,其他几种算法的性能指标则明显低于本文算法;此外,MacroF1 值也高于其他 4 种

算法,表明本文算法在查全率与查准率的综合性能上是最优的。而且,从图 5 和图 6 中可以看到,在复杂的施工场景中,本文算法能够准确检测到目标,而其他算法存在明显的漏检现象,且 R-FCN 算法在图 6(d)中出现了误检。

为了更好地比较不同算法,采用 PR 曲线评估算法性能。PR 曲线指的是精度-召回率曲线 (precision-recall),当算法的精度和召回率都最大的时候,说明算法的性能最好。由图 7 所示,本文算法在

表 2 不同算法在 SC 数据集上的检测性能

Table 2 Detection performance of different algorithms on SC data sets

Algorithm	Crane		Digger		$mAP/\%$	MacroF1/ $\%$
	AP/ $\%$	F1/ $\%$	AP/ $\%$	F1/ $\%$		
YOLOV2	89.3	91.1	91.2	92.3	90.2	91.7
YOLOV3	91.1	92.4	91.8	93.2	91.4	92.8
SSD	82.7	84.6	86.6	85.6	84.7	85.1
R-FCN	88.0	84.8	89.7	87.3	88.8	86.0
Ours	91.0	92.6	92.7	93.4	91.9	93

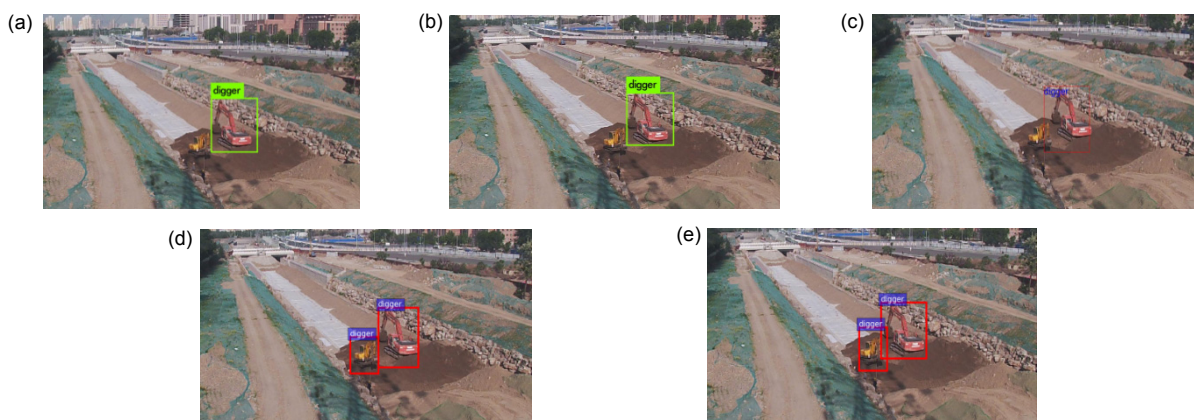


图 5 五种算法在场景一的检测效果对比图。(a) YOLOV2; (b) YOLOV3; (c) SSD; (d) R-FCN; (e) 本文算法
Fig. 5 Comparison of the detection effects of five algorithms in the first scene. (a) YOLOV2; (b) YOLOV3; (c) SSD; (d) R-FCN; (e) Ours

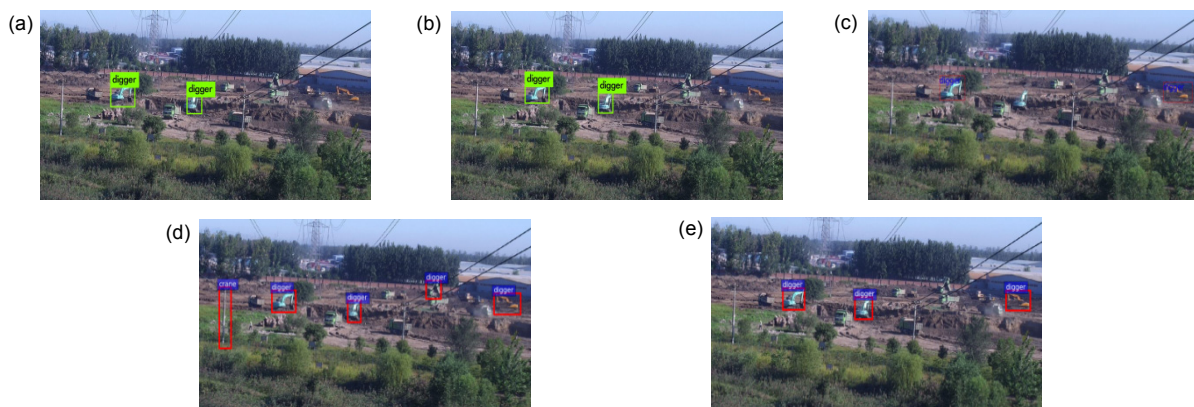


图 6 五种算法在场景二的检测效果对比图。(a) YOLOV2; (b) YOLOV3; (c) SSD; (d) R-FCN; (e) 本文算法
Fig. 6 Comparison of the detection effects of five algorithms in the second scene. (a) YOLOV2; (b) YOLOV3; (c) SSD; (d) R-FCN; (e) Ours

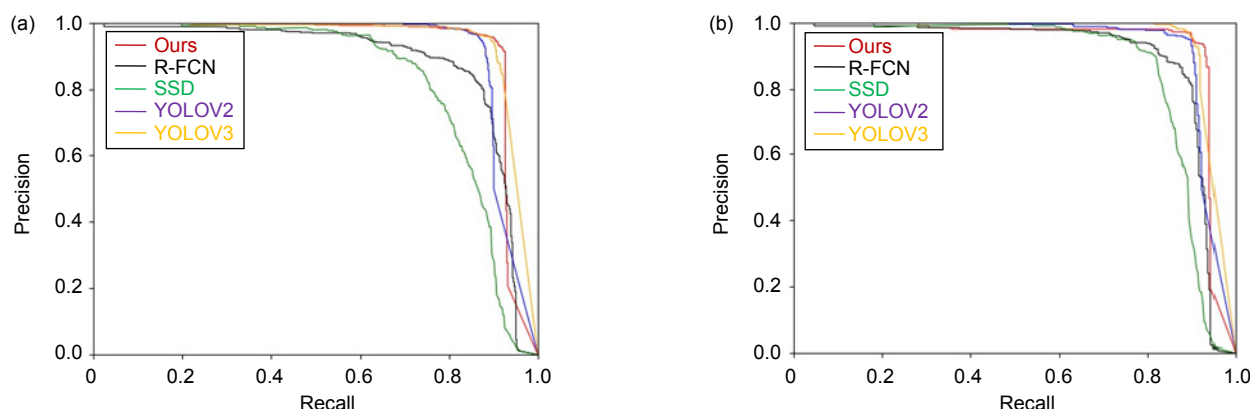


图7 不同算法在工程车数据集上的PR曲线。

Fig. 7 PR curves of different algorithms on the engineering vehicle dataset. (a) Crane; (b) Digger

这些指标上均优于所对比的其他算法。从以上实验结果可以看出，本文算法通过融合目标的多尺度特征以及上下文信息，有效地提高了检测模型的精确度和召回率，并能够适应不同场景中的车辆检测。

4 结 语

本文将多尺度特征和上下文信息融合实现车辆检测任务，通过在工程车数据集和 PASCAL 数据集上训练测试，表明了本文算法在精度和召回率方面优于现有的目标检测算法，对车辆的尺度、形态变化以及复杂背景等影响因素具有良好的鲁棒性。本文算法的模型结构较为复杂，计算量较大，因此通过进一步优化模型结构来提高运行效率，将作为下一步的研究内容。

参考文献

[1] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//*Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008: 24–26.

[2] 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.

[3] Manana M, Tu C L, Owolawi P A. A survey on vehicle detection based on convolution neural networks[C]//*Proceedings of the 3rd IEEE International Conference on Computer and Communications*, 2017: 1751–1755.

[4] Cao S Y, Liu Y H, Li X Z. Vehicle detection method based on fast R-CNN[J]. *Journal of Image and Graphics*, 2017, **22**(5): 671–677.
曹诗雨, 刘跃虎, 李辛昭. 基于 Fast R-CNN 的车辆目标检测[J]. *中国图象图形学报*, 2017, **22**(5): 671–677.

[5] Gu Y, Xu Y. Fast SAR target recognition based on random convolution features and ensemble extreme learning ma-

chines[J]. *Opto-Electronic Engineering*, 2018, **45**(1): 170432.

谷雨, 徐英. 基于随机卷积特征和集成超限学习机的快速 SAR 目标识别[J]. *光电工程*, 2018, **45**(1): 170432.

[6] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91–99.

[7] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 936–944.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779–788.

[9] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6517–6525.

[10] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. arXiv:1804.02767[cs.CV].

[11] Cai Z W, Fan Q F, Feris R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 354–370.

[12] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21–37.

[13] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.

[14] 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*, 2012: 1097–1105.

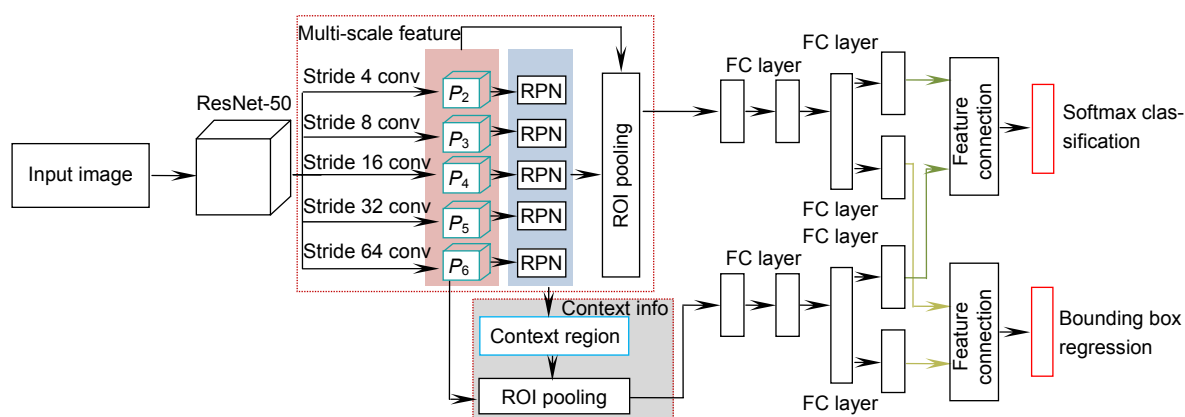
[15] Jia Y Q, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding[C]//*Proceedings of the 22nd ACM international conference on Multimedia*, 2014: 675–678.

[16] Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks[EB/OL]. arXiv:1605.06409.

Vehicle detection based on fusing multi-scale context convolution features

Gao Lin*, Chen Niannian, Fan Yong

Department of Computing Science and Technology, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China



Structure diagram of convolutional neural network model of vehicle object detection algorithm

Overview: Aiming at the problems of the existing vehicle object detection algorithm based on convolutional neural network that cannot effectively adapt to the changes of object scale, self-deformation and complex background, a new vehicle detection algorithm based on multi-scale context convolution features is proposed. In real scenes, the scale of the object is often changeable, and it is difficult to distinguish all objects based on single scale image features. In order to obtain multi-scale feature representation of images, hierarchical features are extracted by convolutional neural network, and then FPN (feature pyramid network) is established. FPN is composed of convolutional layers. The feature maps of different scales are outputted from different convolutional layers. The information of FPN is propagated in three directions: bottom-up, top-down and transverse. In the bottom-up and top-down paths, the feature map of the former contains less semantic information, but it is more accurate for object location, while the latter has more semantic information. However, after several downsampling, most spatial information of the object is lost. Through transverse connection, feature complementarity and multi-scale fusion can be realized. The object candidate regions are generated by RPN network, and the corresponding object regions are located in each level of feature pyramid. Then, the object multi-scale features are extracted. Since the object usually does not exist independently, the background has more or less influence on the object. The structural relationship between the object and the background produces context information. Context information is introduced into the algorithm and fused into the multi-scale feature representation of the object to further enhance the discriminant ability of the object features. The contextual features are extracted around the candidate targets in the multi-scale feature map, and then, like the object features, are pooled by ROI and sent to the full-connectivity layer, respectively. The two sets of fixed-length feature vectors are connected to obtain the multi-scale features fused with the contextual information. The whole convolutional neural network can be trained end-to-end. In order to realize vehicle detection and type recognition simultaneously, multi-task loss function is defined to learn network parameters. In order to verify the validity of the proposed algorithm, the performance of several current mainstream algorithms is compared, including YOLOV2, YOLOV3, SSD, R-FCN. Through training and testing on PASCAL VOC data set and self-made engineering vehicle data set, it is shown that the proposed algorithm is superior to the existing object detection algorithm in precision and recall rate, and has good robustness to the influence factors of vehicle scale, shape change and complex background.

Citation: Gao L, Chen N N, Fan Y. Vehicle detection based on fusing multi-scale context convolution features[J]. *Opto-Electronic Engineering*, 2019, 46(4): 180331

Supported by the Foundation of Sichuan Provincial Education Department (18ZA0501) and Science and Technology Innovation Talents of Sichuan Province (2017113)

* E-mail: 81831283@qq.com