



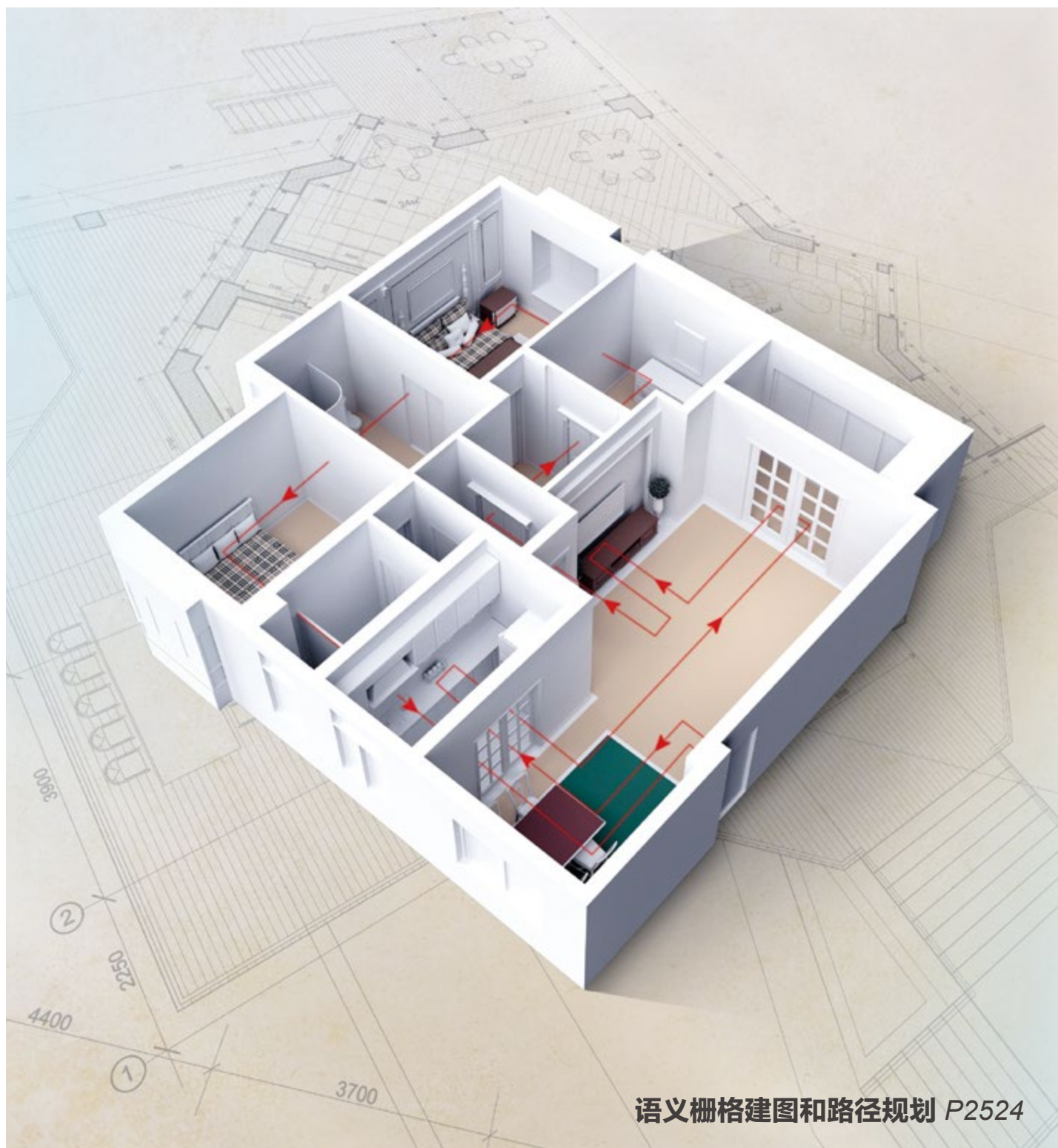
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021
10
VOL.26

ISSN1006-8961
CN11-3758/TB



语义栅格建图和路径规划 P2524



第26卷第10期（总第306期）
2021年10月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

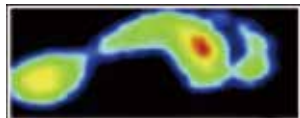
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



穿鞋足迹序列的足迹能量图组表达与识别(第2357页)



时空图卷积网络与注意机制的视频目标分割(第2376页)



双光流网络指导的视频目标检测(第2473页)

综述

图像分类的深度卷积神经网络模型综述

张珂, 冯晓晗, 郭玉荣, 苏昱坤, 赵凯, 赵振兵, 马占宇, 丁巧林 2305

中国图像工程25年

章毓晋 2326

数据集论文

FGSC-23: 面向深度学习精细识别的高分辨率光学遥感图像舰船目标数据集

姚力波, 张筱晗, 吕亚飞, 孙玮玮, 李孟洋 2337

图像处理和编码

并行生成网络的红外—可见光图像转换

余佩伦, 施佳, 王晗 2346

图像分析和识别

穿鞋足迹序列的足迹能量图组表达与识别

王新年, 于丹, 张涛 2357

时空图卷积网络与注意机制的视频目标分割

姚睿, 夏士雄, 周勇, 赵佳琦, 胡伏原 2376

多支路协同的RGB-T图像显著性目标检测

蒋亭亭, 刘昱, 马欣, 孙景林 2388

改进R-FCN模型的小尺度行人检测

刘万军, 董利兵, 曲海成 2400

边缘与区域不一致性引导下的图像拼接篡改检测网络

蒋小玉, 刘春晓 2411

图像理解和计算机视觉

各向异性导向滤波的红外与可见光图像融合

刘明威, 王任华, 李静, 焦映臻 2421

卷积稀疏与细节显著图解析的图像融合

杨培, 高雷阜, 訾玲玲 2433

计算机图形学

利用距离与内能极小平滑链接Bézier曲线

李军成, 刘成志, 刘珊君 2450

NCIG 2020

结合置信度加权融合与视觉注意机制的前景检测

成科扬, 孙爽, 王文杉, 师文喜, 李鹏, 詹永照 2462

双光流网络指导的视频目标检测

尉婉青, 禹晶, 史薪琪, 肖创柏 2473

生成对抗网络驱动的图像隐写与水印模型

郑钢, 胡东辉, 戈辉, 郑淑丽 2485

融合语义—表观特征的无监督前景分割

李熹, 马惠敏, 马洪兵, 王弈冬 2503

场景视点偏移的激光雷达点云分割

郑阳, 林春雨, 廖康, 赵耀, 薛松 2514

激光—相机系统语义栅格建图和路径规划

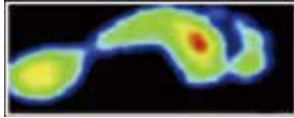
丁梦远, 郭迟, 黄凯 2524

融合词袋连通图的图像检索特征选择

李国祥, 王继军, 马文斌 2533

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Shoeprint sequence representation and recognition using shoeprint energy map set (P2357)



Spatial-temporal video object segmentation with graph convolutional network and attention mechanism(P2376)



Dual optical flow network-guided video object detection(P2473)

Review

Overview of deep convolutional neural networks for image classification

Zhang Ke, Feng Xiaohan, Guo Yurong, Su Yukun, Zhao Kai, Zhao Zhenbing, Ma Zhanyu, Ding Qiaolin 2305

Twenty-five years of image engineering in China

Zhang Yujin 2326

Dataset

FGSC-23: a large-scale dataset of high-resolution optical remote sensing image for deep learning-based fine-grained ship recognition

Yao Libo, Zhang Xiaohan, Lyu Yafei, Sun Weiwei, Li Mengyang 2337

Image Processing and Coding

Infrared-to-visible image translation based on parallel generator network

Yu Peilun, Shi Quan, Wang Han 2346

Image Analysis and Recognition

Shoeprint sequence representation and recognition using shoeprint energy map set

Wang Xinnian, Yu Dan, Zhang Tao 2357

Spatial-temporal video object segmentation with graph convolutional network and attention mechanism

Yao Rui, Xia Shixiong, Zhou Yong, Zhao Jiaqi, Hu Fuyuan 2376

Multi-path collaborative salient object detection based on RGB-T images

Jiang Tingting, Liu Yu, Ma Xin, Sun Jinglin 2388

Small-scale pedestrian detection based on improved R-FCN model

Liu Wanjun, Dong Libing, Qu Haicheng 2400

Edge and region inconsistency-guided image splicing tamper detection network

Jiang Xiaoyu, Liu Chunxiao 2411

Image Understanding and Computer Vision

Infrared and visible image fusion with multi-scale anisotropic guided filtering

Liu Mingwei, Wang Renhua, Li Jing, Jiao Yingzhen 2421

Image fusion method of convolution sparsity and detail saliency map analysis

Yang Pei, Gao Leifu, Zi Lingling 2433

Computer Graphics

Smoothing linked Bézier curves by distance and internal energy minimization

Li Juncheng, Liu Chengzhi, Liu Shanjun 2450

NCIG 2020

Foreground detection via fusing confidence by weight and visual attention mechanism

Cheng Keyang, Sun Shuang, Wang Wenshan, Shi Wenxi, Li Peng, Zhan Yongzhao 2462

Dual optical flow network-guided video object detection

Yu Wanjing, Yu Jing, Shi Xinqi, Xiao Chuangbai 2473

End-to-end image steganography and watermarking driven by generative adversarial networks

Zheng Gang, Hu Donghui, Ge Hui, Zheng Shuli 2485

Semantic-apparent feature-fusion-based unsupervised foreground segmentation method

Li Xi, Ma Huimin, Ma Hongbing, Wang Yidong 2503

LiDAR point cloud segmentation through scene viewpoint offset

Zheng Yang, Lin Chunyu, Liao Kang, Zhao Yao, Xue Song 2514

Semantic grid mapping and path planning combined with laser-camera system

Ding Mengyuan, Guo Chi, Huang Kai 2524

Feature selection method for image retrieval based on connected graphs and bag of words

Li Guoxiang, Wang Jijun, Ma Wenbin 2533

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)10-2400-11

论文引用格式: Liu W J, Dong L B and Qu H C. 2021. Small-scale pedestrian detection based on improved R-FCN model. Journal of Image and Graphics, 26(10): 2400-2410 (刘万军, 董利兵, 曲海成. 2021. 改进 R-FCN 模型的小尺度行人检测. 中国图象图形学报, 26(10): 2400-2410) [DOI: 10.11834/jig.200287]

改进 R-FCN 模型的小尺度行人检测

刘万军, 董利兵*, 曲海成

辽宁工程技术大学软件学院, 葫芦岛 125105

摘要: 目的 为了解决传统行人检测算法在分辨率低、行人尺寸较小等情境下检测精度低的问题, 将基于区域全卷积网络(region-based fully convolutional networks, R-FCN)的目标检测算法引入到行人检测中, 提出一种改进 R-FCN 模型的小尺度行人检测算法。方法 为了使特征提取更加准确, 在 ResNet-101 的 conv5 阶段中嵌入可变形卷积层, 扩大特征图的感受野; 为提高小尺寸行人检测精度, 在 ResNet-101 中增加另一条检测路径, 对不同尺寸大小的特征图进行感兴趣区域池化; 为解决小尺寸行人检测中的误检问题, 利用自举策略的非极大值抑制算法代替传统的非极大值抑制算法。结果 在基准数据集 Caltech 上进行评估, 实验表明, 改进的 R-FCN 算法与具有代表性的单阶段检测器(single shot multiBox detector, SSD)算法和两阶段检测器中的 Faster R-CNN(region convolutional neural network)算法相比, 检测精度分别提高了 3.29% 和 2.78%; 在相同 ResNet-101 基础网络下, 检测精度比原始 R-FCN 算法提高了 12.10%。结论 本文提出的改进 R-FCN 模型, 使小尺寸行人检测精度更加准确。相比原始模型, 改进的 R-FCN 模型对行人检测的精确率和召回率有更好的平衡能力, 在保证精确率的同时, 具有更大的召回率。**关键词:** 行人检测; 区域全卷积网络(R-FCN); 可变形卷积; 多路径; 非极大值抑制(NMS); Caltech 数据集

Small-scale pedestrian detection based on improved R-FCN model

Liu Wanjun, Dong Libing*, Qu Haicheng

School of Software, Liaoning Technical University, Huludao 125105, China

Abstract: **Objective** Pedestrian detection is a research hotspot in the field of image processing and computer vision, and it is widely used in fields such as automatic driving, intelligent monitoring, and intelligent robots. The traditional pedestrian detection method based on background modeling and machine learning can obtain a better pedestrian detection rate under certain conditions, but it cannot meet the requirements of practical applications. As deep convolutional neural networks have made great progress in general object detection, more and more scholars have improved the general object detection framework and introduced it to pedestrian detection. Compared with traditional methods, the accuracy and robustness of pedestrian detection based on deep learning methods have been improved significantly, and many breakthroughs have been made. However, the detection effect for small-scale pedestrians is not ideal. This is mainly due to a series of convolution pool operations of the convolutional neural network, which makes the feature map of small-scale pedestrians smaller, have a lower resolution, and lose serious information, leading to detection failure. To effectively solve the problem of low detection accuracy of traditional pedestrian detection algorithms in the context of low resolution and small pedestrian size, an object

收稿日期: 2020-06-23; 修回日期: 2020-08-26; 预印本日期: 2020-09-02

* 通信作者: 董利兵 1558611191@qq.com

基金项目: 国家自然科学基金项目(41701479); 辽宁省自然科学基金项目(20180550529)

Supported by: National Natural Science Foundation of China (41701479)

detection algorithm called region-based fully convolutional network (R-FCN) is introduced into pedestrian detection. This study proposes an improved small-scale pedestrian detection algorithm for R-FCN. **Method** The method in this study inherits the advantage of R-FCN, which employs the region proposal network to generate candidate regions of interest and position-sensitive score maps to classify and locate targets. At the same time, because the new residual network (ResNet-101) has less calculation, few parameters, and good accuracy, this study uses the ResNet-101 network as the basic network. Compared with the original R-FCN, this study mainly has the following improvements: Considering that the pedestrians in the Caltech dataset have multiple scale transformations, all 3×3 conventional convolutional layers of the Conv5 stage of ResNet-101 are first expanded into deformable convolutional layers. Therefore, the effective step size of the convolution block can be reduced from 32 pixels to 16 pixels, the expansion rate can be changed from 1 to 2, the pad is set to 2, and the step size is 1. Deformable convolution can increase the generalization ability of the model, expand the receptive field of the feature map, and improve the accuracy of R-FCN feature extraction. Then, another position-sensitive score map is added in the training phase. Because the feature distinguishing ability of Conv1-3 stages in ResNet-101 is weaker than that of Conv4 stage, a new layer of position-sensitive score map is added after the Conv4 layer to detect multi-scale pedestrians in parallel with the original position-sensitive score map after the Conv5 layer. Finally, the non-maximum suppression (NMS) method often leads to missed detection of neighbor pedestrians in crowded scenes. Therefore, this study improves the traditional NMS algorithm and proposes the NMS algorithm for bootstrap strategy to solve the problem of pedestrian misdetection.

Result The experiment is evaluation on the benchmark dataset Caltech. The experimental results show that the improved R-FCN algorithm improves the detection accuracy by 3.29% and 2.78% compared with the representative single shot multiBox detector (SSD) algorithm of the single-stage detector and the faster region convolutional neural network (Faster R-CNN) algorithm of the two-stage detector, respectively. Under the same ResNet-101 basic network, the detection accuracy is 12.10% higher than the original R-FCN algorithm. Online hard example mining (OHEM) is necessary for Caltech, which has achieved a 7.38% improvement because the Caltech dataset contains a large number of confounding instances in complex backgrounds, allowing the full use of OHEM. In the Conv5 stage of the ResNet-101 network, a deformable convolutional layer is used, which is 0.89% higher than the ordinary convolutional layer. Using the multi-path detection structure can increase the detection accuracy by 2.50%. The bootstrap strategy is used to correct the non-maximum suppression, which is 1.67% better than the traditional NMS algorithm. **Conclusion** The improved R-FCN model proposed in this study makes the detection accuracy of small-sized pedestrians more accurate and improves the phenomenon of pedestrian false detection in the case of low resolution. Compared with the original R-FCN model, the improved R-FCN model has a better ability to balance the accuracy rate and recall rate of pedestrian detection and has a greater recall rate when ensuring the accuracy rate. However, the accuracy of pedestrian detection in complex scenes is slightly low. Thus, future research will focus on improving the accuracy of pedestrian detection in complex scenes.

Key words: pedestrian detection; region-based fully convolutional network (R-FCN); deformable convolution; multipath; non-maximum suppression (NMS); Caltech dataset

0 引言

目标检测与识别是计算机视觉中的一个经典难题,研究的内容包括行人、车辆和人脸等。其中,车辆和人脸都是刚性物体,目标的结构稳定,研究起步较早,相应的检测技术也比较成熟,并有相关的产品问世,比如车牌识别和人脸考勤等。而行人具有刚性物体和柔性物体的双重特性,属于目标检测的一类特殊问题,与人脸和车辆相比,外观和姿态的变化

更加多样化,检测难度也更大。此外,行人检测是车辆辅助驾驶、视频监控和人体行为分析等应用领域中基础且至关重要的第一步。因此,继人脸检测之后,行人检测成为图像处理和计算机视觉目标检测领域的另一个研究热点。

基于背景建模和机器学习的传统行人检测方法在特定条件下具有较好的检测精度,但不能满足实际应用的要求。Krizhevsky 等人(2017)将深度学习技术应用到大规模图像分类中,发现基于深度学习表示的特征具有较强层次的表达能力和良好的鲁棒

性,能够较好地解决一些视觉问题。随着两阶段检测器 R-CNN (region convolutional neural network) (Girshick 等, 2014)、Fast R-CNN (Girshick, 2015)、Faster R-CNN (Ren 等, 2017)、R-FCN (region-based fully convolutional network) (Dai 等, 2016) 和单阶段检测器 SSD (single shot multiBox detector) (Liu 等, 2016)、YOLO (you only look once) 系列 (Redmon 等, 2016) 等深度卷积神经网络 (deep convolutional neural networks, DCNNs) 在一般性目标检测方面取得很大进展,越来越多的学者改进通用目标检测框架,将其引入行人检测中。Tian 等人 (2015) 提出一种基于部件的检测方案 DeepParts, 将人体分割成多个部位分别检测,然后将结果组合起来解决行人遮挡问题。Zhang 等人 (2016) 通过分析 Faster R-CNN (region convolutional neural network) 在行人检测中的性能,用 RPN (region proposal network) 提取候选区域,并用随机森林进行分类,取得了当时最佳的检测效果。Mao 等人 (2017) 提出一种改进 Faster R-CNN 的行人检测算法 HyperLearner, 在 VGG (Visual Geometry Group) 网络的基础上,增加了一个分支网络,并与主体网络特征一起送入区域建议网络进行处理,解决了复杂场景下行人检测困难问题。Molchanov 等人 (2017) 提出全卷积 YOLO 神经网络在视频监控中的行人检测方法,通过学习单个神经网络,将检测和分类任务合并到一个使用卷积神经网络的挑战中,在速度上取得了很好效果。Teseima 等人 (2018) 将 FPN (feature pyramid network) 应用于行人检测,利用深度卷积网络固有的多尺度结构,构造了一个在各层次都具有丰富语义的特征金字塔,提高了行人检测算法的性能,对不同尺度的行人具有较强的鲁棒性。Liu 等人 (2019a) 提出一种基于上下文信息改进 SSD 的行人检测方法,通过提取 SSD 卷积神经网络中的浅层特征,将卷积层中的浅层语义信息与深层语义信息融合,进而检测图像中行人的大小,重新设计了不同长宽比预选框,在提高行人检测精度的同时保持了较好的检测速度。

与传统方法相比,神经网络在精度和鲁棒性方面有了显著提高。然而,使用 CNN (convolutional neural networks) 进行行人检测仍然存在两方面问题。1) 如何区分行人与复杂背景。行人出现的区域通常伴随着复杂背景,背景中许多物体可能具有类似行人的明显特征,如广告牌、交通标志和柱子

等。检测器可能无法有效区分行人和背景,增加了网络的误检率。2) 如何有效检测出小尺度行人。因为小尺度行人在图像中占据相对较小的像素点或图像质量不好导致图像更加模糊,高层卷积层提取的行人距离特征不够明显,使得检测器难以识别。Zhao 等人 (2016) 将 Faster R-CNN 算法引入到行人检测中,将候选窗口提取、特征提取和分类器包含在一个框架中,提高了检测速度,但是对小尺度行人检测效果不佳。Xiang 等人 (2018) 提出一种改进 Faster R-CNN 算法,使用最后一层卷积输出的特征图,由于小尺度行人在该特征图中易丢失,导致检测精度明显下降。

大部分行人数据集都存在小尺寸行人图像,严重影响着行人检测器的性能。尤其在 Caltech (Dollár 等, 2009) 数据集中,行人高度在 20 ~ 80 像素的人数超过了总数的 70%,小尺度行人占很大比例。本文针对分辨率低、行人尺寸较小情境下检测精度低的问题,将 R-FCN 目标检测算法引入到行人检测中,提出一种改进 R-FCN 模型的小尺度行人检测算法,通过引进可变形卷积层、增加一条新路径和利用自举策略的非极大值算法,提高了小尺寸行人的检测精度,改善了在分辨率低的场景中行人容易误检现象。

1 R-FCN 算法

1.1 R-FCN 原理

R-FCN 的网络结构主要由残差网络 (residual neural network, ResNet) 和区域建议网络 (region proposal network, RPN) 组成,如图 1 所示。前者用于提取特征,后者用于生成感兴趣区域 (region of interest, RoI)。其中, K 代表锚点数, C 代表物体的类别数。主要过程为:输入的图像先经过一个全卷积网络;在最后一个卷积层后面添加特殊的卷积层生成位置敏感得分图 (position-sensitive score map),同时全卷积网络的最后一个卷积层输出作为 RPN 网络的输入,RPN 网络最后生成 RoIs;通过位置敏感 RoI 池化层 (position-sensitive RoI pooling layer) 将前面的得分图和感兴趣区域作为输入,输出类别信息。

1.2 RPN

RPN 网络的基本结构如图 2 所示,左侧分支进行分类,分为前景和背景两类,即候选框中有目标还

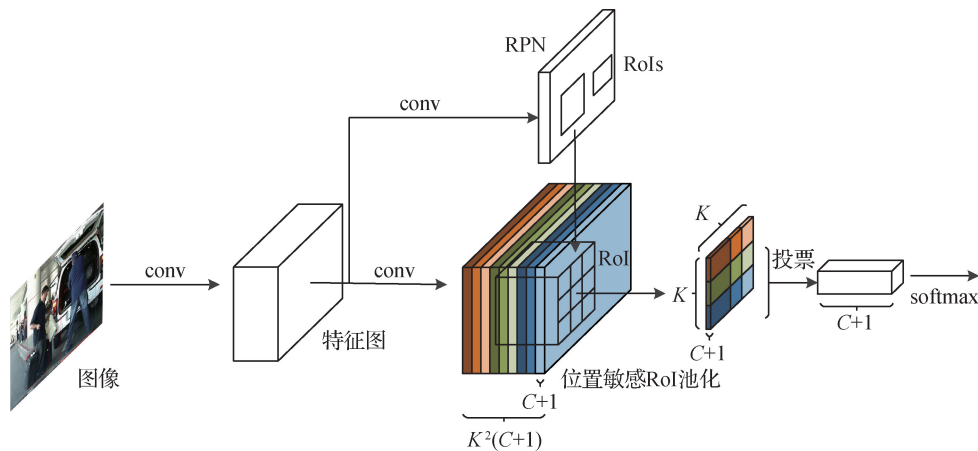


图1 R-FCN 网络结构图

Fig. 1 R-FCN network structure

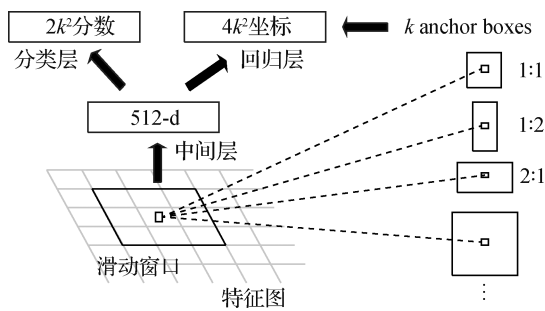


图2 RPN 网络的基本结构

Fig. 2 Basic structure of RPN networks

是无目标;右侧分支进行回归,用于获取边框位置信息的卷积网络。其中 k 代表锚点数, d 代表维度。

RPN 在自身训练的同时,还会生成感兴趣区域。将生成的感兴趣区域和分类的卷积层生成的得分图进行池化,最终得到每个感兴趣区域属于每个类别的概率,将感兴趣区域与回归卷积层的输出进行池化,得到每个 RoI 的 4 个坐标。

1.3 位置敏感得分图

R-FCN 将在共享卷积层的最后一层网络中连接一个卷积层,该卷积层就是位置敏感得分图(position-sensitive score map),该 map 是一层卷积层,其高度 H 和宽度 W 与共享卷积层的一样(即具有同样的感受野),但是通道个数为 $K^2(C+1)$ 。其中 C 表示物体类别种数,再加上 1 个背景类别,所以共有 $C+1$ 类,而每个类别都有 $K \times K$ 个得分图。

为了在每个 RoI 中编码位置信息,通过一个网格将每个 RoI 分成 $K \times K$ 个矩形单元(bins)。对 $W \times H$ 的 RoI 区域,每一个矩形单元的大小 $\approx W/K \times$

H/K 。在 R-FCN 算法中,最后一个卷积层由 K^2 个得分图组成。对于第 i 行第 j 列的矩形单元 bin ($0 \leq i, j \leq K-1$),其得分图计算为

$$r_c(i, j | \Theta) = \sum_{(x, y) \in b(i, j)} z_{i, j, c}(x + x_0, y + y_0 | \Theta) / n \quad (1)$$

式中, $r_c(i, j | \Theta)$ 为第 c 类第 (i, j) 块的池化响应。 $z_{i, j, c}$ 为 $K^2(C+1)$ 维得分图中的输出, (x_0, y_0) 为 RoI 的左上角坐标, n 为图块的像素总数, Θ 表示网络的所有可学的参数。

然后, K^2 位置敏感分数对 RoI 进行投票。在 R-FCN 算法中,只通过对分数求平均进行投票,为每个 RoI 生成 $(C+1)$ 维向量,具体为

$$r_c(\Theta) = \sum_{i, j} r_c(i, j | \Theta) \quad (2)$$

式中, $r_c(\Theta)$ 表示每个类别的平均得分。最后通过 softmax 函数分类出每个类别的概率,并用于训练期间评估交叉熵损失(cross-entropy loss)和推断期间对 RoI 进行排序。具体为

$$s_c(\Theta) = e^{r_c(\Theta)} / \sum_{c'=0}^C e^{r_{c'}(\Theta)} \quad (3)$$

2 改进的 R-FCN 算法

改进的算法继承了 R-FCN 的优点,使用一个区域建议网络(RPN)生成感兴趣候选区域,并使用位置敏感得分图对目标进行分类和定位。He 等人(2016)提出的新的深度残差网络(ResNet-101)具有计算量小、参数少以及精度不降低的特点,所以使用

ResNet-101 作为改进算法的基础网络。与原始 R-FCN 相比,主要有以下改进:1)将 ResNet-101 的 conv5 阶段的所有 3×3 常规卷积层扩展为可变形的卷积层,因为可变形卷积可增加模型的泛化能力,扩大特征图的感受野,可以提高 R-FCN 特征提取的准确度;2)在训练阶段附加另一个位置敏感的得分图,用 conv4 层和在线硬样本挖掘(online hard exam-

ple mining, OHEM)的融合策略检测小尺度行人,与原始 conv5 层后的位置敏感的得分图组合,用于检测不同尺度的行人;3)改进传统非极大值抑制(non-maximum suppression, NMS)算法,提出自举策略的非极大抑制算法(bootstrap non-maximum suppression, B-NMS),可解决行人误检问题。改进的 R-FCN 网络结构如图 3 所示。

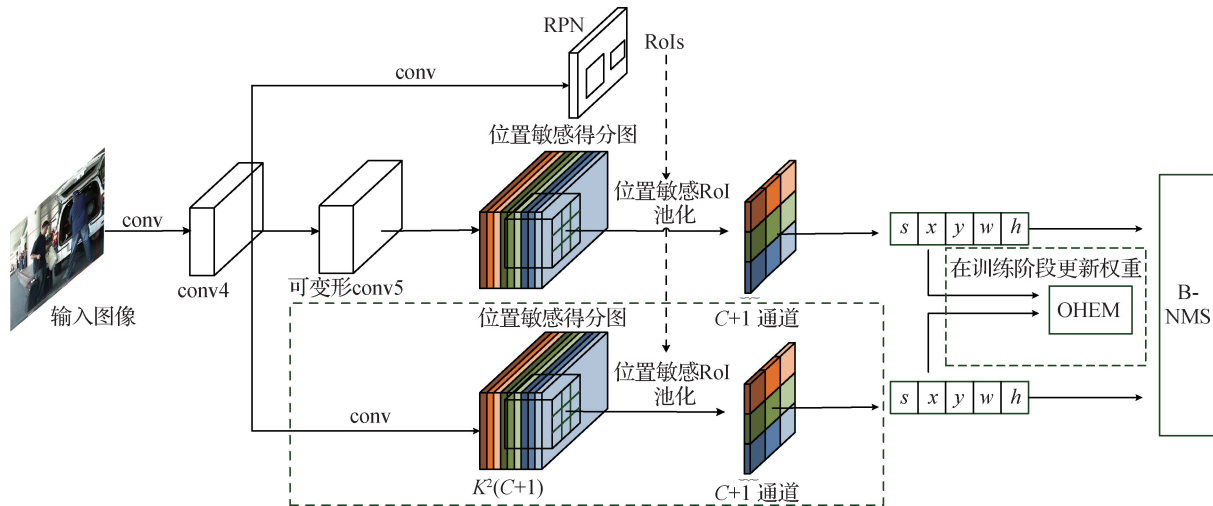


图 3 改进的 R-FCN 网络结构图

Fig. 3 Improved R-FCN network structure diagram

2.1 可变形卷积层

由于姿态、尺度、视点和零件变形等因素引起的几何变化是目标识别和检测的主要问题。Zhu 等人(2019)在可变形卷积网络(deformable convolutional networks, DCNv1)(Dai 等, 2017)的基础上,提出了一种新的变形卷积算法 Deformable ConvNets v2(DCNv2),通过改进可变形结构以及引入偏置(offset),使得特征提取过程能够更加集中于有效信息区域。DCNv2 在此基础上引入权重(modulation),通过分配不同权重给经过的 offset 修正后的区域,实现更加准确的特征提取。

给定 k 个采样点的卷积核,令 w_k 和 p_k 分别表示第 k 个采样点的权重和预先设定的偏移量。例如, $k=9$, $p_k \in \{(-1, -1), (-1, 0), \dots, (1, 1)\}$ 表示感受野的大小,典型的 9 点方格,以 3×3 为主。

传统的卷积输出特征图为

$$y(p_0) = \sum_{p_k \in R} w(p_k) \cdot x(p_0 + p_k) \quad (4)$$

式中,卷积窗口中的每个像素点 p_k 对应权重 w ; p_0 代表窗口输出的每个像素点; x 是输入层像素点

集合。

可变形卷积利用偏移量 Δp_k 和权重 Δm_k 增大规则网格 R 的感受野。设 $x(p_0)$ 和 $y(p_0)$ 分别表示输入特征图 x 和输出特征图 y 中位置 p_0 处的特征。调制后的可变形卷积可表示为

$$y(p_0) = \sum_{k=1}^K w_k \cdot x(p_0 + p_k + \Delta p_k) \cdot \Delta m_k \quad (5)$$

式中, Δp_k 和 Δm_k 分别是学习到的偏置和权重。 Δm_k 位于 $[0, 1]$ 范围内, Δp_k 为实数,取值范围无限制。由于 $p_0 + p_k + \Delta p_k$ 为分数形式,在计算 x 时采用双线性插值。

如图 4 所示,可变形卷积将常规卷积过程分成两条路径,一条路径学习 offset,得到 $H \times W \times 2N$ 的输出(offset), $N = |R|$ 表示网格中像素个数, $2N$ 表示有 x 和 y 两个方向的 offset。另一条路径在得到 offset 以后,对于原始卷积的每一个卷积窗口,都不再是原来规整的滑动窗口(图 4 中绿色框),而是经过平移后的窗口(图 4 中蓝色框),增大了输出特征图的面积。

考虑到 Caltech 数据集中行人有多种尺度变换

情况,将 DCNv2 中的可变形卷积层应用到残差网络中。在 ResNet-101 的 conv5 阶段所有 3×3 常规卷积层 (res5a_branch2b、res5b_branch2b、res5c_branch2c) 中分别引入可变形卷积层、偏移层和权重,因此卷积块的有效步长可以从 32 像素减小到 16 像素,膨胀率从 1 变为 2, pad 设置为 2, 步长为 1, 以增加特征图的分辨率和感受野,实现更加准确的特征提取。

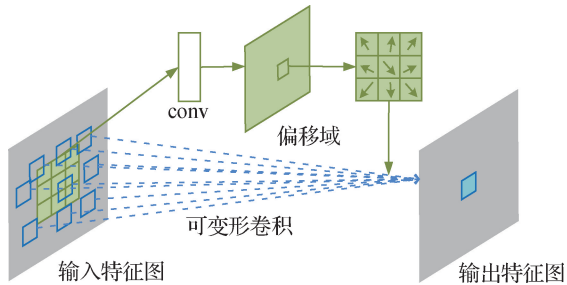


图4 3×3 可变形卷积示例图

Fig. 4 Illustration of 3×3 deformable convolution

2.2 多尺度检测和 OHEM

对残差网络结构分析可知,输入图像与 ResNet-101 中 conv1 - 5 阶段输出的特征图的比分别为 $1:2, 1:4, 1:8, 1:16, 1:32$ 。使用检测子网络在 conv1 - 5 阶段输出的特征图上滑动 1 个像素,相当于分别在原图上滑动 2、4、8、16、32 个像素。由于在 Caltech 数据集中存在大量宽度或高度小于 32 像素的行人,原始 R-FCN 中在 Conv5 阶段后添加位置敏感得分图,会使宽度或高度小于 32 像素的行人在位置敏感得分图中丢失。受 Liu 等人(2016)融合多路径检测结果的启发,在改进网络中加入了另一条检测路径,见图 3 虚线框。由于 ResNet-101 中的 conv1 - 3 阶段特征区分能力弱于 conv4 阶段,因此在 conv4 层后增加一层位置敏感得分图与原始的位置敏感得分图并联来共同检测多尺度行人。

在行人检测中,误检主要是硬背景实例的混淆造成的。因此,采用 Shrivastava 等人(2016)提出的在线硬样本挖掘 (OHEM) 方法,主要思想是对样本按损失 (loss) 进行排序,选择前面 loss 较小的。该方法主要用来对负样本进行筛选,使得正负样本更加平衡。OHEM 是一种没有计算消耗的硬样本挖掘方法,因为在位置敏感的 RoI 池之前,所有计算都是共享的,所以在每个 RoI 上定义的损失函数是不同检测路径的交叉熵损失和 box 回归损失之和,即

$$L(s, t_{x,y,w,h}) = \sum_{i=1}^2 (L_{cls}^i(s, c^*) + \lambda [c^* > 0] L_{reg}^i(t, t^*)) \quad (6)$$

式中, s 是预测框 $t_{x,y,w,h}$ 的得分, c^* 是 RoI 的真实标签, i 表示第 i 条路径, L_{cls} 和 L_{reg} 分别表示分类的交叉熵损失和使用 smoothL1 的边界框回归损失, λ 用来平衡两者的重要性, $[c^* > 0]$ 表示如果分类正确为 1, 分类错误为 0。对分类错误的样本不进行位置损失排序。

2.3 B-NMS

传统的物体检测流程通常采用多尺度滑动窗口,根据每个物体类别的前景或背景分数对每个窗口计算其特征。但是相邻窗口往往具有相关的分数,这会增加检测结果的假阳性。为了避免出现这样的问题,Neubeck 和 van Gool (2006) 提出非极大值抑制 (NMS) 方法,对检测结果进行后续处理得到最终检测结果。传统的 NMS 算法首先在检测图像中产生一系列的检测框 B 以及对应的分数 S 。当选择最高分数的检测框 M 时,它将从集合 B 中删除并放入最终检测结果集合 D 中。同时,集合 B 中任何与检测框 M 的重叠部分大于重叠阈值 N_t 的检测框也将删除。NMS 算法中的最大问题就是将相邻框的分数均强制为零。在这种情况下,如果一个真实物体出现在重叠区域,则将导致对该物体的检测失败并降低算法的平均精度 (average precision, AP)。在最先进的检测器 (Girshick, 2015; Ren 等, 2017; Dai 等, 2016) 中, NMS 因为极大地减少误报数量而用于获得最终的检测集,但在分辨率低的场景中, NMS 方法往往会导致行人误检。对此, Bodla 等人 (2017) 提出 Soft-NMS 算法,解决了 NMS 算法中存在的问题,用一个稍微小一点的分数替代原有的分数,而并非将相邻框的分数均强制为零,取得了较好效果。但该算法是一种贪心算法,并不能保证找到全局最优的检测框分数重置,仍然存在行人误检现象。

针对传统 NMS 和 Soft-NMS 算法存在的误检问题,本文提出一种改进 NMS 算法的自举策略非极大抑制算法,使用两个框的重叠率 r 和相似性度量 m 来区分这两个框是否属于同一个人,并使用两个阈值 T_r 和 T_m 来挖掘负样本。自举策略的相似性度量 m 定义为

$$m = e^{((1-2\lambda) \times dxy + \lambda \times dr + \lambda \times dh)}$$

$$dxy = \sqrt{\left(\frac{x' - x}{w}\right)^2 + \left(\frac{y' - y}{h}\right)^2}$$

$$\begin{aligned} dw &= \left\| \frac{w'}{w} - 1 \right\| \\ dh &= \left\| \frac{h'}{h} - 1 \right\| \end{aligned} \quad (7)$$

式中, x, y, w, h 分别表示框的中心坐标、宽度和高度, 用于排序队列中的其他框。 x', y', w', h' 用于排序队列 Q 中得分最大的预测框。 $\|\cdot\|$ 表示 L_1 -范数。 λ 表示一个加权参数, 本文设置 $\lambda = 0.3$ 。最后的度量是 3 个偏差的加权和, 表示两个检测框的相似度。

改进的 NMS 算法的具体步骤如下:

输入: 检测框 D , 分值 S , 相似性度量阈值 T_m 。

输出: 结果列表 L 。

1) 选择得分 $S > 0.5$ 的检测框 D , 根据队列 Q 中的得分 S 按降序排序;

2) 根据式(7)计算排序队列 Q 中第 1 个框与其余框的重叠率 r 和相似性度量 m ;

3) 删除排序队列 Q 中重叠率 $r > 0.3$ 的新备份队列 B 中的框;

4) 搜索备份队列 B 中相似性度量 m 大于阈值 T_m 同时重叠率 $r < 0.5$ (即满足条件 $(0.3 < r < 0.5$ 且 $m > T_m$)) 的框;

5) 将 B 中的搜索框追加到已排序队列 Q 的底部, 然后将已排序队列 Q 中的第 1 个框移动到输出列表 L 中;

6) 如果排序队列 Q 为空, 则返回输出列表 L , 否则重复步骤 2) — 5)。

3 实验结果与分析

3.1 实验数据及实验参数配置

实验在深度学习框架 Caffe 上进行。实验平台为 ubuntu14.04 操作系统和 NVIDIA GTX 1080GPU。改进 ResNet-101 在 ImageNet 预先训练的模型, 通过反向传播和随机梯度下降 (stochastic gradient descent, SGD) 进行端到端训练。调整每幅图像的大小, 使较短边缘具有 N 个像素 (在 Caltech 数据集上 $N = 600$ 像素)。本文对 RPN 进行微调, 在训练和测试阶段生成 300 个 proposals, 然后输入到 R-CNN 分类中。采用前一半样本作为小型批处理中的硬示例来更新反向传播中的权值。经过卷积神经网络后, 采用重叠阈值 $T_r = 0.3$ 和相似阈值 $T_m = 0.8$ 的

非极大值抑制方法对预测框进行融合。前 40 k 批次学习率为 0.001, 后 30 k 批次学习率为 0.000 1, 动量为 0.9, 动量衰减为 0.000 5。

实验使用 Caltech 基准数据集。Caltech 数据集由车载摄像机拍摄的约 250 000 幅图像组成, 分辨率为 640×480 像素, 分为 11 部分, 即 set00 – set10。set00 – set05 为训练集, set06 – set11 为测试集。实验选取 54 145 幅至少包含 1 个行人的图像, 按 3 : 1 : 1 的比例, 用 32 487 幅图像作为训练集, 10 829 幅图像作为验证集, 10 829 幅图像作为测试集。这样既能保证训练 CNNs 的数据量充足, 又避免了取所有帧带来的数据冗余。实验将 Caltech 数据集转换成 PASCAL VOC (pattern analysis, statistical modeling and computational learning visual object classes) 数据集 (Everingham 等, 2010) 中定义的方法来评估行人检测模型的结果。

3.2 实验评价指标

实验使用精确率—召回率 (precision-recall, PR) 曲线、平均精度 (mean average precision, mAP) 和损失值作为检测算法的评价指标。

精确率 P 表示正确识别为正样本的正样本占所有识别为正样本的比例。即

$$P = \frac{T_p}{T_p + F_p} \quad (8)$$

式中, T_p (true positives, TP) 表示将正样本正确识别为正样本, 即行人样本识别为行人样本; F_p (false positives, FP) 表示将负样本错误识别为正样本, 即非行人样本识别为行人样本。

召回率 R 表示正确识别为正样本的正样本占所有正样本的比例。即

$$R = \frac{T_p}{T_p + F_N} \quad (9)$$

式中, F_N (false negatives, FN) 表示将正样本错误识别为负样本, 即行人样本识别为非行人样本。

损失值表示预测值与真实值相近程度, 损失值越小, 说明模型的鲁棒性越好。

3.3 实验结果与分析

实验统计改进模型在训练次数每隔 10 000 次的精度, 不同迭代次数的精度值如表 1 所示。可以看出, 随着迭代次数增加, 模型精度值逐步上升, 在 70 000 次时达到顶峰, 在 80 000 次后逐渐下降。本文选取迭代次数为 70 000 次的训练模型进行验证。

R-FCN模型和改进模型的训练损失值图如图5所示。由图5及损失日志文件可知,模型训练开始时,损失值忽高忽低。当训练迭代到40 000次时,学习率衰减10%,损失值波动幅度减少;在70 000次时损失值趋于平稳;当迭代次数达到80 000时,学习率再次衰减10%,损失值开始上升,波动幅度较大;因此,训练迭代次数设置过大,网络可能过度拟合,导致损失值不稳定(Wang等,2015)。

不同基础网络的精度如表2所示。可以看出,ResNet-101更适合本文模型,虽然速度比ResNet-50慢0.024 s,但检测精度比ResNet-50高7.29%。本文算法选取ResNet-101作为基础网络。

表1 不同迭代次数的精度值
Table 1 Precision value of different iterations

迭代次数	精度/%
10 000	49.82
20 000	60.66
30 000	65.33
40 000	72.18
50 000	76.99
60 000	77.54
70 000	77.98
80 000	75.49
90 000	75.13
100 000	74.14
110 000	74.24
120 000	73.32

注:加粗字体为最优结果。

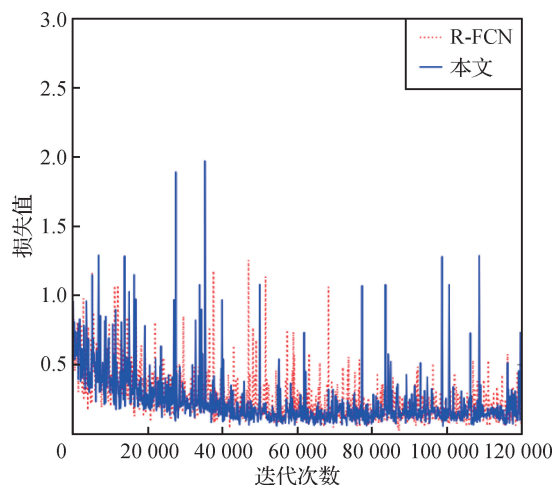


图5 不同方法的损失值

Fig.5 Loss values for different methods

表2 不同基础网络的行人检测精度
Table 2 Pedestrian detection accuracy of different backbone networks

基础网络	精度/%	速率/(s/幅)
ResNet-50	70.69	0.062
ResNet-101	77.98	0.088

注:加粗字体为各列最优结果。

为了验证本文算法的性能,在Caltech数据集(行人高度为20~80像素)中,与两阶段检测器中的Faster R-CNN算法、单阶段检测器中的SSD算法、Kuang等人(2018)的行人检测算法、Liu等人(2019b)的行人检测算法、Yang等人(2020)基于R-FCN改进的检测算法和何姣姣等人(2018)基于R-FCN改进的检测算法进行检测精度对比,结果如表3所示。可以看出,在小尺度行人检测上,改进算法的算法精度较Faster R-CNN、SSD、Kuang等人(2018)和何姣姣等人(2018)的方法分别提高了3.29%、2.78%、5.68%和1.88%。较Liu等人(2019b)和Yang等人(2020)的方法分别降低了1.97%和0.45%。

表3 不同行人检测算法的检测结果
Table 3 Detection results of different pedestrian detection algorithms

算法	精度/%
Faster R-CNN	74.69
SSD	75.20
Kuang等人(2018)	72.30
Liu等人(2019b)	79.67
Yang等人(2020)	78.43
何姣姣等人(2018)	76.10
本文	77.98

注:加粗字体表示最优结果。

不同方法组合的行人检测精度如表4所示。可以看出,1)OHEM对Caltech来说是必要的,实现了7.38%的提升,因为Caltech数据集包含了大量复杂背景下的混淆实例,可以充分利用OHEM。2)在ResNet-101的conv5阶段采用可变形卷积层,比常规卷积层提高了0.89%。3)使用多路径检测结构可以使检测精度增加2.50%。4)利用自举策略对非极大值抑制进行了校正,比传统的NMS算法有了

1.67% 的改善。

表 4 不同组合方法的行人检测精度
Table 4 Pedestrian detection accuracy of different combination methods

方法	精度/%	速率/ (s/幅)
R-FCN	65.88	0.086
R-FCN + OHEM	73.26	0.086
R-FCN + OHEM + DCNv2	73.81	0.087
R-FCN + OHEM + Multi-paths	75.42	0.087
R-FCN + OHEM + DCNv2 + Multi-paths	76.31	0.088
R-FCN + OHEM + DCNv2 + Multi-paths + B-NMS	77.98	0.088

注:加粗字体为各列最优结果。

不同组合方法的行人检测 PR 曲线如图 6 所示。可以看出,1)在 ResNet-101 的 conv5 阶段嵌入可变形卷积层比原始常规卷积层检测的准确率更高;2)采用多路径检测比原始单路径检测小尺寸行人的检测效果有明显提高;3)采用相同组合因素的算法,分别利用传统 NMS 和 B-NMS 进行检测,B-NMS 的检测效果更好;4)改进的 R-FCN 算法对小尺寸行人检测准确率有较好提升,说明了改进方法的有效性。

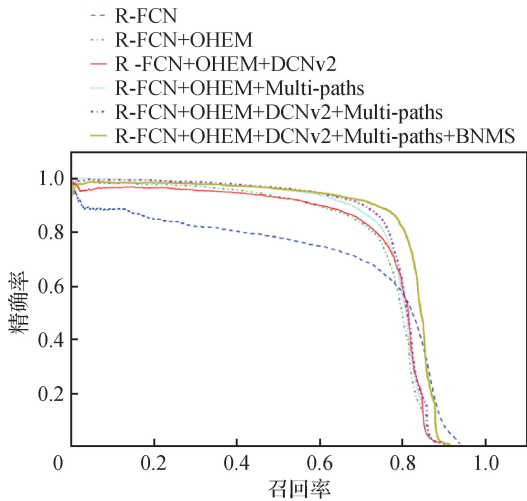


图 6 不同组合方法的行人检测 PR 曲线对比
Fig. 6 Comparison of PR curves of pedestrian detection with different combination methods

检测效果如图 7 所示。通过实验发现,1)从图 7(a)常规卷积与图 7(b)可变形卷积检测效果对比

可以看出,在残差网络的 conv5 阶段嵌入可变形卷积

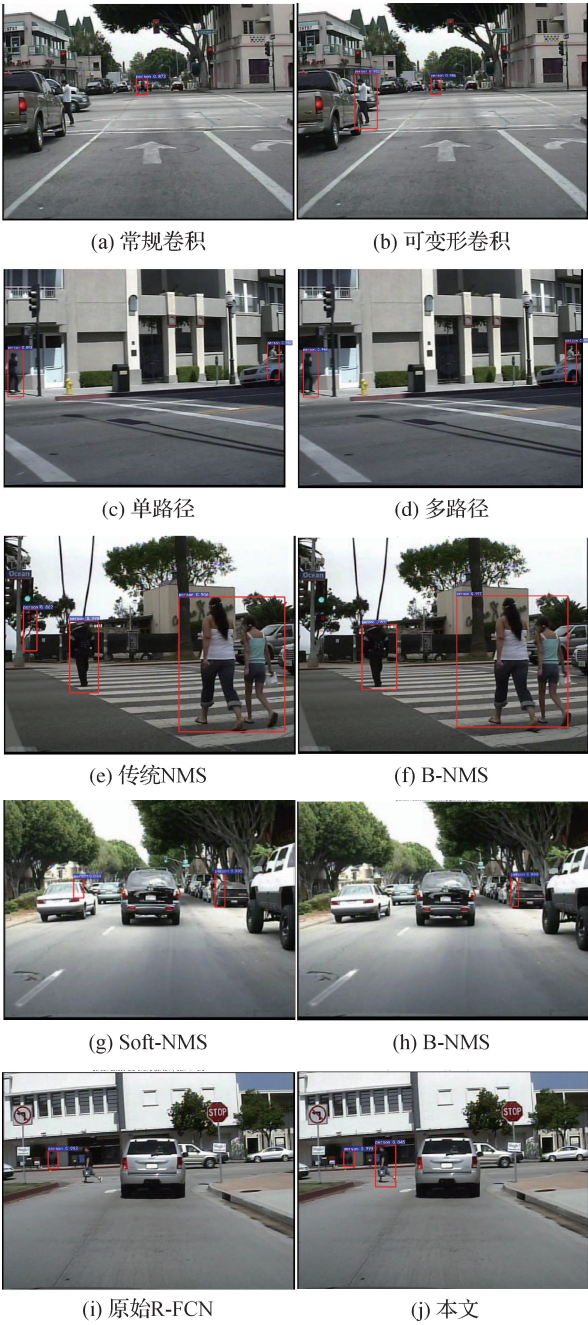


图 7 检测效果对比
Fig. 7 Comparison of different detection effects
((a) conventional convolution; (b) deformable convolution; (c) single-path; (d) multi-path; (e) traditional NMS; (f) B-NMS; (g) Soft-NMS (h) B-NMS; (i) the original R-FCN; (j) ours)

积层,可以使特征提取更加准确。使用常规卷积容易出现漏检现象,而使用可变形卷积层代替常规卷积层,可以明显改善行人漏检现象;2)从图 7(c)单路径与图 7(d)多路径检测效果对比可以看出,在

ResNet-101 的 conv4 层后添加一个新的位置敏感得分图,与 conv5 层后的位置敏感得分图并联,可以提高小尺寸行人的检测精度。在行人被障碍物遮挡的情况下,使用多路径检测方法,可以提高行人检测精度。3)从图 7(e)传统 NMS 与图 7(f)B-NMS 的检测效果对比,及图 7(g)Soft-NMS 与图 7(h)B-NMS 的检测效果对比可以看出,利用自举策略对传统非极大值抑制算法进行校正,明显改善了低分辨率场景的行人误检现象。在复杂场景下使用自举策略的非极大值抑制算法,可以解决行人检测中的误检现象。4)从图 7(i)原始 R-FCN 与图 7(j)本文算法检测效果对比可以看出,本文算法改善了原始算法中对小尺寸行人存在的漏检现象,且检测精度有所提升。

4 结 论

针对行人尺寸较小、外形多变等特点导致的检测率偏低问题,本文在 R-FCN 模型基础上进行改进,提出一种改进 R-FCN 模型的小尺度行人检测算法。通过可变形卷积代替常规卷积,解决了特征提取不准确问题;通过新增一条路径,在 ResNet-101 的 conv4 层后加一层位置敏感得分图与原始位置敏感得分图并联来共同检测多尺度行人,解决了小尺寸行人检测率偏低问题;通过改进传统 NMS 算法,提出自举策略的非极大值抑制算法,解决了行人误检问题。实验在行人数据集 Caltech 上进行评估,结果表明,改进算法相比原始 R-FCN 算法对行人检测精度有很大提升,在低分辨率场景能够有效改善行人误检问题。与具有代表性的 Faster R-CNN、SSD 算法及其他基于 R-FCN 改进的检测算法相比,本文算法能够较精确地检测小尺度行人,检测精度有所上升;同时与最新行人检测算法相比,检测精度不劣于其他最新行人检测算法。

本文算法虽然比之前算法在检测精度上有较好提升,但在人数较多的复杂场景下行人检测效果略差。下一步将围绕提高模型在人数较多场景下的行人检测精确率展开研究。

参考文献 (References)

Bodla N, Singh B, Chellappa R and Davis L S. 2017. Soft-NMS-

improving object detection with one line of code//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 5561-5569 [DOI: 10.1109/iccv.2017.593]

Dai J F, Li Y, He K M and Sun J. 2016. R-FCN: object detection via region-based fully convolutional networks//Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, USA: ACM: 379-387

Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]

Dollár P, Wojek C, Schiele B and Perona P. 2009. Pedestrian detection: a benchmark//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 304-311 [DOI: 10.1109/CVPR.2009.5206631]

Everingham M, Van Gool L, Williams C K I, Winn J and Zisserman A. 2010. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision, 88(2): 303-338 [DOI: 10.1007/s11263-009-0275-4]

Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and Semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 580-587 [DOI: 10.1109/cvpr.2014.81]

Girshick R. 2015. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]

He J J, Zhang Y P, Yao T Z, Liu K and Xiao J J. 2018. Robust pedestrian detection based on parallel channel cascade network. Pattern Recognition and Artificial Intelligence, 31(12): 1134-1142 (何姣姣, 张永平, 姚拓中, 刘肯, 肖江剑. 2018. 基于并行通道级联网络的鲁棒行人检测. 模式识别与人工智能, 31(12): 1134-1142) [DOI: 10.16451/j.cnki.issn1003-6059.201812009]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90 [DOI: 10.1145/3065386]

Kuang P, Ma T S, Li F and Chen Z W. 2018. Real-time pedestrian detection using convolutional neural networks. International Journal of Pattern Recognition and Artificial Intelligence, 32(11): #1856014 [DOI: 10.1142/s0218001418560141]

Liu S A, Lyu S, Zhang H L and Gong J. 2019a. Pedestrian detection algorithm based on the improved SSD//Proceedings of 2019 Chinese Control and Decision Conference. Nanchang, China: IEEE: 3559-3563 [DOI: 10.1109/CCDC.2019.8832518]

Liu S T, Huang D and Wang Y H. 2019b. Adaptive NMS: refining

- pedestrian detection in a crowd//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 6459-6468 [DOI: 10.1109/cvpr.2019.00662]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands; Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Mao J Y, Xiao T T, Jiang Y N and Cao Z M. 2017. What can help pedestrian detection?//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 3127-3136 [DOI: 10.1109/CVPR.2017.639]
- Molchanov V V, Vishnyakov B V, Vizilter Y V, Vishnyakova O V and Knyaz V A. 2017. Pedestrian detection in video surveillance using fully convolutional YOLO neural network//Proceedings of SPIE 10334, Automated Visual Inspection and Machine Vision II. Munich, Germany; SPIE: 103340Q [DOI: 10.1117/12.2270326]
- Neubeck A and van Gool L. 2006. Efficient non-maximum suppression//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China; IEEE: 850-855 [DOI: 10.1109/ICPR.2006.479]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 779-788 [DOI: 10.1109/cvpr.2016.91]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Shrivastava A, Gupta A and Girshick R. 2016. Training region-based object detectors with online hard example mining//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 761-769 [DOI: 10.1109/cvpr.2016.89]
- Tesema F B, Lin J P, Ou J, Wu H and Zhu W L. 2018. Feature fusing of feature pyramid network for multi-scale pedestrian detection//Proceedings of the 15th International Computer Conference on Wavelet Active Media Technology and Information Processing. Chengdu, China; IEEE: 10-13 [DOI: 10.1109/ICCWAMTIP.2018.8632614]
- Tian Y L, Luo P, Wang X G and Tang X O. 2015. Deep learning strong parts for pedestrian detection//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 1904-1912 [DOI: 10.1109/iccv.2015.221]
- Wang B, Tang S, Zhao R Z, Liu W and Cen Y G. 2015. Pedestrian detection based on region proposal fusion//Proceedings of the 17th IEEE International Workshop on Multimedia Signal Processing. Xiamen, China; IEEE: 1-6 [DOI: 10.1109/mmisp.2015.7340847]
- Xiang X Z, Lyu N, Guo X L, Wang S and El Saddik A. 2018. Engineering vehicles detection based on modified faster R-CNN for power grid surveillance. Sensors, 18(7): #2258 [DOI: 10.3390/s18072258]
- Yang P Y, Zhang G F, Wang L, Xu L S, Deng Q X and Yang M H. 2020. A part-aware multi-scale fully convolutional network for pedestrian detection. IEEE Transactions on Intelligent Transportation Systems, PP(99): 1-13 [DOI: 10.1109/TITS.2019.2963700]
- Zhang L L, Lin L, Liang X D and He K M. 2016. Is faster R-CNN doing well for pedestrian detection?//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands; Springer: 443-457 [DOI: 10.1007/978-3-319-46475-6_28]
- Zhao X T, Li W, Zhang Y F, Gulliver T A, Chang S and Feng Z Y. 2016. A Faster RCNN-based pedestrian detection system//Proceedings of the 84th IEEE Vehicular Technology Conference. Montreal, Canada; IEEE: 1-5 [DOI: 10.1109/VTCFall.2016.7880852]
- Zhu X Z, Hu H, Lin S and Dai J F. 2019. Deformable ConvNets v2: more deformable, better results//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 9308-9316 [DOI: 10.1109/cvpr.2019.00953]

作者简介



刘万军, 1959年生, 男, 教授, 主要研究方向为图像与智能信息处理。

E-mail: liuwanjun@lntu.edu.cn



董利兵, 通信作者, 女, 硕士研究生, 主要研究方向为图像与智能信息处理。

E-mail: 1558611191@qq.com

曲海成, 男, 副教授, 主要研究方向为图像与智能信息处理。

E-mail: quhaicheng@lntu.edu.cn