Journal of Computer Applications ISSN 1001-9081 2019-08-10
计算机应用,2019,39(8):2230 - 2234 CODEN JYIIDU http://www.joca.cn

文章编号:1001-9081(2019)08-2230-05 doi:10.11772/j.issn.1001-9081.2018122595

基于 Movidius 神经计算棒的行人检测方法

张洋硕*,苗 壮,王家宝,李 阳
(陆军工程大学 指挥控制工程学院,南京 210007)
(* 通信作者电子邮箱 17625944869@163.com)

摘 要: Movidius 神经计算棒是基于 USB 模式的深度学习推理工具和独立的人工智能加速器,为广泛的移动和嵌入式视觉设备提供专用深度神经网络加速功能。针对深度学习的嵌入式应用,实现了一种基于 Movidius 神经计算棒的近实时行人目标检测方法。首先,通过改进 RefineDet 目标检测网络结构使模型大小和计算适应嵌入式设备的要求;然后,在行人检测数据集上对模型进行重训练,并部署于搭载 Movidius 神经计算棒的树莓派上;最后,在实际环境中对模型进行测试,算法达到了平均每秒 4 帧的处理速度。实验结果表明,基于 Movidius 神经计算棒,在计算资源紧张的树莓派上可完成近实时的行人检测任务。

关键词: 行人检测;深度学习;树莓派;Movidius;嵌入式设备
中图分类号: TP183;TP391.4 **文献标志码:** A

Pedestrian detection method based on Movidius neural computing stick

ZHANG Yangshuo*, MIAO Zhuang, WANG Jiabao, LI Yang
(*College of Command and Control Engineering, Army Engineering University, Nanjing Jiangsu* 210007, *China*)

Abstract: Movidius neural computing stick is a USB-based deep learning inference tool and a stand-alone artificial intelligence accelerator that provides dedicated deep neural network acceleration for a wide range of mobile and embedded vision devices. For the embedded application of deep learning, a near real-time pedestrian target detection method based on Movidius neural computing stick was realized. Firstly, the model size and calculation were adapted to the requirements of the embedded device by improving the RefineDet target detection network structure. Then, the model was retrained on the pedestrian detection dataset and deployed on the Raspberry Pi equipped with Movidius neural computing stick. Finally, the model was tested in the actual environment, and the algorithm achieved an average processing speed of 4 frames per second. Experimental results show that based on Movidius neural computing stick, the near real-time pedestrian detection task can be completed on the Raspberry Pi with limited computing resources.

Key words: pedestrian detection; deep learning; Raspberry Pi; Movidius; embedded device

0 引言

行人检测是目标检测的重要分支,可用于多种不同领域,如视频监控、人员识别和智能汽车驾驶系统。在现实生活中,由于视频或图像中行人姿态、物体遮挡、服装、灯光变化和复杂背景的多样性,行人检测在计算机视觉中仍是一个具有挑战性的任务。近年来,深度学习极大地推动了行人检测技术的发展,在计算机视觉领域引起了广泛的关注;但是,深度学习中的行人检测模型在面向实际应用时还存在着诸多问题亟待解决。

行人检测根据处理过程一般可以分解为生成候选窗口、特征提取和特征分类三个步骤。经典的行人检测方法通常使用基于滑动窗口的方法生成候选窗口,使用梯度方向直方图(Histogram of Oriented Gradients, HOG)[1]或尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)[2]作为特征,并使用支持向量机(Support Vector Machine, SVM)或自适应集成分类器(AdaBoost)作为特征分类方法;但这些方法大多基于手工特征,刻画的是低层次信息,缺乏对行人高层次语义信息的描述。近年来,随着深度学习技术的发展,深度神经网络已被广泛应用于行人检测任务。与传统方法不同的是,深度学习通过深层卷积网络操作抽取图像的高级语义信息来描述行人,具有更好的描述能力。与此同时,伴随计算机硬件性能的不断提高,行人检测算法也在不断优化。郭爱心等[3]在更快、更富有特征层次的卷积神经网络(Faster Regions with Convolutional Neural Network feature, Faster R-CNN)通用目标检测框架[4]的基础上,针对行人特点提出了行人区域建议网络;针对小尺度行人特征信息不足,提出了多层次特征提取和融合的方法;陈光喜等[5]通过设计一个网络与统一的实时监测目标检测 YOLOv2(You Only look Once)网络[6]级联,解决了复杂环境下行人检测不能同时满足高召回率与高效率检测的问题;徐超等[7]提出一种改进的基于卷积神经网络的行人检测方法,使卷积神经网络能选择出更优模型并获得定位更

收稿日期:2019-01-02;修回日期:2019-04-02;录用日期:2019-04-03。 基金项目:国家自然科学基金资助项目(61806220)。
作者简介:张洋硕(1995—),男,河南三门峡人,硕士研究生,主要研究方向:计算机视觉、目标检测; 苗壮(1976—),男,辽宁辽阳人,副教授,博士,主要研究方向:人工智能; 王家宝(1985—),男,安徽肥西人,讲师,博士,主要研究方向:模式识别、图像检索; 李阳(1984—),男,河北廊坊人,讲师,博士,主要研究方向:机器视觉、机器学习。

准确的检测框; Hosang 等[8] 使用 SquaresChnFtrs 方法生成行人候选窗口并训练 AlexNet[9] 进行行人检测; Zhang 等[10] 使用区域提议网络(Region Proposal Network, RPN) [4] 来计算行人候选区域和级联 Boosted Forest[11] 以执行样本重新加权来对候选区域进行分类; Li 等[12] 训练多个基于 Fast R-CNN[13] 的网络来检测不同尺度的行人,并结合所有网络的结果以产生最终结果。虽然上述工作对行人检测进行很多探索,但是在面对实际应用时,卷积网络模型还存在着模型大小、计算耗时等方面的问题。

本文将行人检测算法放在搭载 Movidius 神经计算棒(Neural Computational Stick, NCS) 的树莓派上,来实现行人的快速检测。树莓派是一种计算资源紧张的设备,因此本文采用深度可分卷积对高精度的 RefineDet 网络[14] 进行改进,构建轻量级卷积神经网络。该网络在选定的嵌入式视觉设备(Raspberry PI 3 板) 上的平均处理速度达到 4 帧/s(Frames Per Second, FPS) 。考虑到有限的计算能力,它符合设计目标。因此,实时行人检测任务可以通过仅执行 4 FPS 的检测算法来完成。

本文的主要工作如下:

1) 针对移动和嵌入式视觉设备的智能监控,对 RefineDet 检测网络进行改进,构建一种轻量级卷积神经网络;

2) 使用该网络在资源受限的搭载 Movidius 神经计算棒的树莓派 3 上运行,以实现嵌入式设备的实时行人检测。

1　改进的行人检测网络

1.1　行人检测网络框架

为了提高嵌入式视觉设备上运行目标检测模型的效率,本文基于高精度的 RefineDet,引入高效的深度可分卷积操作,采用类似 MobileNet[15] 结构来构建骨干网络,改进后的 L-RefineDet网络架构如图 1 所示。

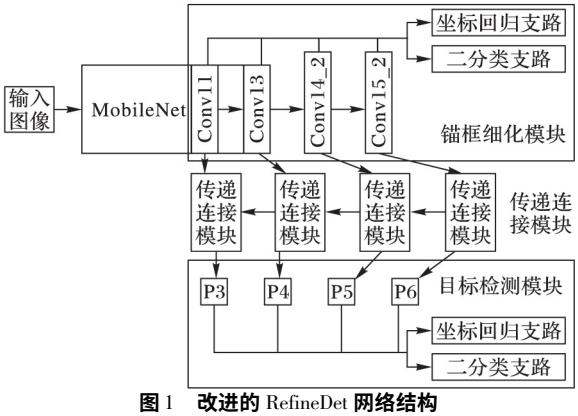


图 1　改进的 RefineDet 网络结构

Fig. 1　Improved RefineDet network structure

该网络架构主要包含三个部分: 锚框细化模块、传递连接模块和目标检测模块。锚框细化模块主要用来得到粗粒度的边界框信息和去除一些负样本,只对物体进行初步的分类和回归,且分类只区分前景和背景; 传递连接模块是用于特征转换操作。对高层次的特征进行反卷积操作,使得特征图之间的尺寸匹配,然后与低层次的特征相加,使锚框细化模块特征转化为目标检测模块特征。目标检测模块用于将输入的多层特征进行进一步的回归和预测多分类标签。

网络架构采用 MobileNet[15] 中的 conv11、conv13 并在后面增加两个卷积层 Conv14_2,Conv15_2,共 4 个特征层作为特征抽取层,以获得不同尺寸的特征。提取特征后进行融合操作,首先是 Conv15_2 特征层经过一个传递连接模块得到对应大小的矩形块(P6) ,接着基于 Conv14_2 的矩形块经过传递连接模块得到对应大小的矩形块(P5) ,此处的传递连接模块相比 P6 增加了反卷积支路,反卷积支路的输入来自于生成 P6 的中间层输出。P4 和 P3 的生成与 P5 同理。

整体来看,一个子模块提取粗粒度的边界框信息,另一个子模块进行细粒度的分类任务,因此能有更高的准确率,而且采用了特征融合,该算法对于小目标物体的检测更有效。

1.2　网络结构

网络的输入图像尺寸为 320 × 320,表 1 描绘了每层的网络结构滤波器大小,其中 Conv0 至 Conv13 与 MobileNet[15] 中结构相同。为了处理不同尺寸的物体,总共提取 4 层特征用作检测,即表 1 中名称为 Conv11、Conv13、Conv14_2、Conv15_2 的特征层,尺寸分别是 20 × 20、10 × 10、5 × 5、3 × 3。每个特征层与不同尺度的锚框相匹配,其中不同层次的锚框采用单发多框目标检测器 (Single Shot multibox Detector, SSD) [16] 三种纵横比设计。

表 1　每层的网络结构滤波器大小

Tab. 1　Network structure filter size of each layer

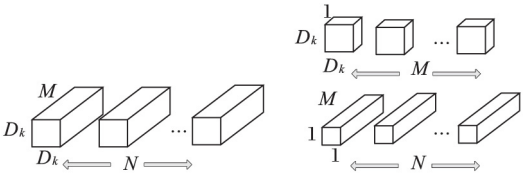
名称	类型/步长	滤波器大小	输入大小
Conv11	Conv dw/s1	3 × 3 × 512 dw	3 × 3 × 512 dw
	Conv/s1	1 × 1 × 512 × 512	1 × 1 × 512 × 512
Conv12	Conv dw/s2	3 × 3 × 512 dw	20 × 20 × 512
	Conv/s1	1 × 1 × 512 × 1 024	10 × 10 × 512
Conv13	Conv dw/s1	3 × 3 × 1 024 dw	10 × 10 × 1 024
	Conv/s1	1 × 1 × 1 024 × 1 024	10 × 10 × 1 024
Conv14_1	Conv/s1	1 × 1 × 1 024 × 256	10 × 10 × 256
Conv14_2	Conv/s2	3 × 3 × 256 × 512	5 × 5 × 256
Conv15_1	Conv/s1	1 × 1 × 512 × 128	5 × 5 × 512
Conv15_2	Conv/s2	3 × 3 × 128 × 256	5 × 5 × 128

注: 标注 dw 表示为深度卷积层,未标注表示为标准卷积层。

为了增强模型的鲁棒性,模型训练时采用了随机扩展、裁剪和翻转等数据增强策略。在匹配步骤后,大多数锚框是负面的,为了缓解正负样本不平衡问题,网络的正负样本界定的标准基本上和其他目标检测算法类似,和真实的标注的交并比(Intersection-over-Union, IoU) 超过阈值 0.5 的边界框为正样本,负样本是根据边界框的分类损失来选的,对分类损失进行降序排列,按照正负样本 1∶3 的比例选择损失靠前的负样本。

1.3　深度可分卷积

改进网络架构中的关键是深度可分卷积[15] ,因为它极大地降低了算法的复杂度,适用于嵌入式设备的应用。深度可分卷积[2] 通过将每个常规卷积层分成两部分: 深度卷积层(depthwise convolution) 和逐点卷积层(pointwise convolution) ,使计算复杂度更适合于的移动智能设备,其体系结构如图 2 所示。



(a) 常规卷积操作 (b) 深度可分卷积操作

图 2 深度可分卷积与传统卷积

Fig. 2 Depthwise separable convolution and traditional convolution

图 2 比较了深度可分卷积和传统卷积。传统卷积输入 $\boldsymbol{F} \in \mathbf{R}^{D_f \times D_f \times M}$，其中 $D_f \times D_f$ 为图像或特征图大小和 M 输入通道的个数；输出 $\boldsymbol{G} \in \mathbf{R}^{D_g \times D_g \times M}$，其中 $D_g \times D_g$ 为输出特征图大小，N 为输出通道的个数。计算过程涉及到滤波器 $\boldsymbol{K} \in \mathbf{R}^{D_k \times D_k \times M \times N}$，其中 $D_k \times D_k$ 为滤波器核大小，M 和 N 分别对应输入和输出通道的个数。计算过程如式(1)：

$$\boldsymbol{G}_{k,l,n} = \sum_{i,j,m} \boldsymbol{K}_{i,j,m,n} \odot \boldsymbol{F}_{k+i-1,l+j-1,m} \tag{1}$$

式(1)的计算复杂度如式(2)：

$$O_{(\text{conv})} = D_k \times D_k \times M \times N \times D_f \times D_f \tag{2}$$

深度可分卷积由两部分组成：一个是深度卷积层，该层由 M 个滤波器 $\{\boldsymbol{K}^m \in \mathbf{R}^{D_k \times D_k \times 1}, m = 1,2,\cdots,M\}$ 组成，输入 $\boldsymbol{F} \in \mathbf{R}^{D_f \times D_f \times M}$，对每个通道计算输出：

$$\boldsymbol{G}^m \in \mathbf{R}^{D_g \times D_g \times 1};\ m = 1,2,\cdots,M \tag{3}$$

对 M 个输出在第 3 个维度拼接：

$$\boldsymbol{G}' = \text{cat}([\boldsymbol{G}^1,\boldsymbol{G}^2,\cdots,\boldsymbol{G}^M]) \in \mathbf{R}^{D_f \times D_f \times M} \tag{4}$$

另一个是逐点卷积层，该层由滤波器 $\hat{\boldsymbol{Y}} \in \mathbf{R}^{1 \times 1 \times M \times N}$ 组成，输入 $\boldsymbol{G}' \in \mathbf{R}^{D_f \times D_f \times M}$，输出 N 个通道的特征图结果 $\hat{\boldsymbol{G}} \in \mathbf{R}^{D_f \times D_f \times N}$，计算过程如式(5)：

$$\hat{\boldsymbol{G}}_{k,l,n} = \sum_{i,j,m} \hat{\boldsymbol{K}}_{i,j,m,n} \odot \boldsymbol{F}_{k+i-1,l+j-1,m} \tag{5}$$

深度可分卷积的计算复杂度如式(6)：

$$O_{(\text{depth})} = D_k \times D_k \times M \times N \times D_f \times D_f + N \times M \times D_f \times D_f \tag{6}$$

深度可分卷积在每个卷积步骤之后，均接有一个批量标准化层和一个非线性激活的 ReLU 层。

基于式(2)和式(6)，深度可分卷积和传统卷积计算复杂度比较如式(7)：

$$\frac{O_{(\text{conv})}}{O_{(\text{depth})}} = \frac{1}{N} + \frac{1}{D_k^2} \tag{7}$$

深度可分卷积使网络变得更快、更高效，非常适合移动和嵌入式视觉设备。

2 基于神经计算棒的行人检测

2.1 实验装置

2.1.1 树莓派

上述改进模型最终部署于嵌入式视觉设备上，本文选用树莓派 3B + 开发版，该开发板具备 ARM v7 1.2 GHz 处理器和 1 GB RAM，计算能力一般，但其成本低、应用广。

2.1.2 英特尔神经计算棒

树莓派计算能力有限，为了提升所提深度行人检测模型的计算效率，采用 Movidius 神经计算棒加速计算。NCS 是英特尔推出的基于 USB 模式的加速设备，为计算能力不足的移动和嵌入式视觉设备提供深度学习加速功能。NCS 不需要连接到云端，可以直接在本地编译、部署，实现神经网络计算加速。Movidius 神经计算棒内置的 Myriad 2 VPU 提供了强大且高效的性能，支持在 Tensorflow 和 Caffe 框架上直接运行神经网络。

2.2 开发应用流程

本文采用 Caffe 框架来训练模型，部署到搭载 Movidius 神经计算棒的树莓派上。NCS 的环境分为两部分：训练端和测试端。

1) 训练端为一台带 GPU 加速的主机，训练 Caffe 模型，并编译成 NCS 可以执行的 graph；

2) 测试端为一台搭载 Movidius 的树莓派开发板，运行编译好的 graph 格式模型以检测行人。

具体开发应用流程如图 3 所示，主要包括三个步骤。

步骤 1 训练。收集行人数据集，并在带 GPU 运算能力的工作站或者服务器上进行训练，得到 Caffe 格式的行人检测模型。

步骤 2 编译。NCS 的 VPU 无法直接运行步骤 1 中训练的 Caffe 模型，需要将模型转化为 VPU 支持的专用文件格式(graph)。NCS 提供了专门的工具，用于编译、优化 Caffe 模型。

步骤 3 预测。在树莓派 3B + 嵌入式设备上部署 graph 格式行人检测模型，加载 graph 文件到 NCS 执行行人检测。

步骤 1 和 2 通常在运行在训练端，除训练 Caffe 模型外，还安装 NCSDK 编译模块 mvNCCompile，用于将 Caffe 模型转成 NCS 的 graph；步骤 3 运行在测试端，安装 inference 模块，用来支持 graph 的运行。

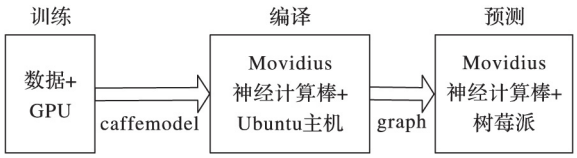


图 3 神经计算棒应用开发流程

Fig. 3 Neural computing stick application development process

2.3 模型训练

标注完善、图像质量高的 PASCAL VOC 数据集是计算机视觉领域中著名的数据集之一，包含日常生活中常见的 20 个类别，person 就是其中一类。本文采用的数据集是基于 PASCAL VOC 2007 和 PASCAL VOC 2012 数据集提取只有 person 类别的目标图片制作而成的 VOC_person 数据集，最终一共包含 8 000 张训练图片和 2 600 张验证图片，并将数据集转化为 LMDB 格式文件以加快读取速度。

本文使用 Caffe 深度学习框架来训练模型。Caffe 是一个高效的深度学习框架，支持命令行、Python 以及 Matlab 程序接口。在训练之前，通过计算每个 RGB 通道的平均值来对数据进行归一化。所提出的模型采用随机梯度下降法(Stochastic Gradient Descent，SGD)进行训练，动量为 0.9，衰减因子为

0.000 5。初始学习率为 0.005，并使用 multistep 学习率衰减策略。在显存为 11 GB 的 NVIDIA GTX1080Ti 上训练，整个训练过程为 200 000 次迭代。最终训练出的权重文件大小可由 302.6 MB 减小为 62.6 MB。

2.4　模型测试

将训练好的 caffemodel 权重文件通过 mvNCCompile 模块编译成 NCS 可执行的 graph，在测试端上进行测试。将 graph 加载到 NCSDK 中，推断得到目标位置、类别、分数。

3　实验结果与分析

实验使用的检测性能评价指标是平均每幅图像误检率(False Positive Per Image，FPPI)[17]，计算公式如下:

$$FPPI = FP/N_{img} \tag{8}$$

其中: $FPPI$ 表示误检总数; N_{img} 表示测试图片总数。

表 2 是本文算法 L-RefineDet 与其他行人检测方法的测试结果。从实验数据可以看出，本文算法 L-RefineDet 在 VOCperson 数据集上的 FPPI 为 0.27，在搭载 Movidius 的树莓派上平均处理速度达到 4 FPS，与其他行人检测方法相比具有更高的精度和更快的速度。

表 2　L-RefineDet 与其他行人检测方法测试结果比较

Tab. 2　Test result comparison of L-RefineDet and other pedestrian detection methods

测试方法	在 VOCperson 数据集上的 FPPI	在搭载 Movidius 的树莓派上的平均处理速度/FPS
Haar Cascaded	0.41	1.9
HOG + SVM	0.34	0.5
MobileNet-SSD	0.30	5.0
L-RefineDet	0.27	4.0

表 3 是本文算法 L-RefineDet 与其他检测模型大小的比较。与不采用深度可分卷积的 RefineDet 方法相比，本文算法 L-RefineDet 的模型更小，表明深度可分卷积能够极大地降低算法的复杂度。

表 3　L-RefineDet 与其他检测模型大小比较

Tab. 3　Size comparison of L-RefineDet and other detection models

检测模型	模型大小/MB
SSD300	114.6
MobileNet-SSD	22.4
RefineDet	302.6
L-RefineDet	62.6

图 4 为使用 MobileNet-SSD 检测实时场景中行人的效果，图 5 为改进的轻量级网络在搭载 Movidius 神经计算棒的树莓派上处理实时场景中行人的效果，其中包括不同距离和角度的行人检测、遮挡的行人检测。

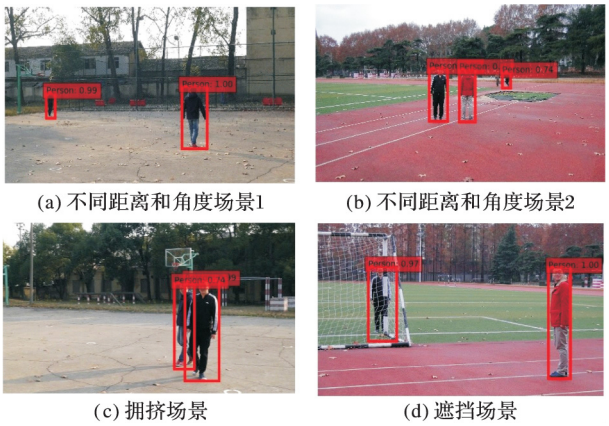
由图 4、5 可以看出，从不同角度和距离捕捉人体对象对检测算法具有很大挑战，当角度和距离不同时，不仅特征有所不同，而且有时人体只是部分可见或以不同的姿势出现。MobileNet-SSD 会对一些特殊角度的行人和小目标行人出现漏检，同时对重叠行人出现漏报，L-RefineDet 对行人检测具有更好的鲁棒性。

(a) 不同距离和角度场景1　(b) 不同距离和角度场景2

(c) 拥挤场景　(d) 遮挡场景

图 4　MobileNet-SSD 检测效果图

Fig. 4　MobileNet-SSD detection results



(a) 不同距离和角度场景1　(b) 不同距离和角度场景2

(c) 拥挤场景　(d) 遮挡场景

图 5　L-RefineDet 检测效果图

Fig. 5　L-RefineDet detection results

4　结语

本文在嵌入式视觉设备下进行行人检测的背景下，改进了一种轻量级的人体目标检测算法，并在搭载 Movidius 神经计算棒的树莓派上运行。实验结果表明该算法已经达到了设计目标，取得了令人满意的检测速度和高准确度。但是单根 NCS 一次只能运行一个模型，可以考虑用多根 NCS、多线程做检测，以达到更高的速度。

参考文献(References)

[1] TRIGGS B, DALAL N. Histograms of oriented gradients for human detection [C]// Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2005: 886 – 893.

[2] LOWE D G. Object recognition from local scale-invariant features [C]// Proceedings of the 7th IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 1999: 1150 – 1157.

[3] 郭爱心，殷保群，李运. 基于深度卷积神经网络的小尺度行人检测[J]. 信息技术与网络安全，2018，37(7): 50 – 53, 57. (GUO A X, YIN B Q, LI Y. Small-size pedestrian detection via deep convolutional neural network [J]. Information Technology and Network Security, 2018, 37(7): 50 – 53, 57.)

[4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 – 1149.

[5] 陈光喜,王佳鑫,黄勇,等. 基于级联网络的行人检测方法[J]. 计算机应用,2019,39(1):186-191. (CHEN G X, WANG J X, HUANG Y, et al. Pedestrian detection method based on cascade networks [J]. Journal of Computer Applications, 2019, 39(1): 186-191.)

[6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 6517-6525.

[7] 徐超,闫胜业. 改进的卷积神经网络行人检测方法[J]. 计算机应用,2017,37(6):1708-1715. (XU C, YAN S Y. Improved pedestrian detection method based on convolutional neural network [J]. Journal of Computer Applications, 2017,37(6): 1708-1715.)

[8] BENENSON R, OMRAN M, HOSANG J, et al. Ten years of pedestrian detection, what have we learned? [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8926. Berlin: Springer, 2014: 613-627.

[9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 2012 International Conference on Neural Information Processing Systems. North Miami Beach, FL: Curran Associates Inc, 2012: 1097-1105.

[10] ZHANG L, LIN L, LIANG X, et al. Is Faster R-CNN doing well for pedestrian detection? [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9906. Cham: Springer, 2016: 443-457.

[11] APPEL R, FUCHS T, DOLLÁR P. Quickly boosting decision trees: pruning underachieving features early [C]// Proceedings of the 30th International Conference on Machine Learning. [S. l.]: JMLR, 2013, 28: III-594-III-602.

[12] LI J, LIANG X, SHEN S, et al. Scale-aware Fast R-CNN for pedestrian detection [J]. IEEE Transactions on Multimedia, 2018, 20(4): 985-996.

[13] GIRSHICK. R. Fast R-CNN [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 1440-1448.

[14] ZHANG S, WEN L, BIAN X, et al. Single-shot refinement neural network for object detection [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 4203-4212.

[15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [J]. arXiv E-print, 2017: arXiv:1704.04861.

[16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9905. Cham: Springer, 2016: 21-37.

[17] WOJEK C, DOLLAR P, SCHIELE B, et al. Pedestrian detection: an evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 743-761.

This work is partially supported by the National Natural Science Foundation of China (61806220).

ZHANG Yangshuo, born in 1995, M. S. candidate. His research interests include computer vision, object detection.

MIAO Zhuang, born in 1976, Ph. D., associate professor. His research interests include artificial intelligence.

WANG Jiabao, born in 1985, Ph. D., lecturer. His research interests include pattern recognition, image retrieval.

LI Yang, born in 1984, Ph. D., lecturer. His research interests include machine vision, machine learning.