



## 基于多分辨率滤波通道的多尺度行人检测

张传伟, 曾虹钧, 杨萌月, 李 波, 陈尚瑞

(西安科技大学 机械工程学院, 西安 710054)

**摘 要:** 利用单尺度特征映射进行多尺度行人检测存在准确率和效率较低的问题。为此, 提出一种基于多分辨率滤波通道的多尺度检测方法。应用尺度感知池增强感受域对应性, 通过软决策树实现尺度不变性, 在使用滑动窗口分类策略时, 结合地平线约束和稀疏网格来降低计算成本, 加快行人检测速度。在 Caltech 数据集上的实验结果表明, 该方法的检测准确率达 88.89%, 检测速度达 15.68 frame/s, 其检测准确率优于 VJ、WordChannels 等方法。

**关键词:** 行人检测; 多分辨率滤波通道; 尺度感知池; 软决策树; 滑动窗口

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 张传伟, 曾虹钧, 杨萌月, 等. 基于多分辨率滤波通道的多尺度行人检测[J]. 计算机工程, 2020, 46(2): 235-241.

**英文引用格式:** ZHANG Chuanwei, ZENG Hongjun, YANG Mengyue, et al. Multi-scale pedestrian detection based on multi-resolution filter channels[J]. Computer Engineering, 2020, 46(2): 235-241.

## Multi-Scale Pedestrian Detection Based on Multi-Resolution Filter Channels

ZHANG Chuanwei, ZENG Hongjun, YANG Mengyue, LI Bo, CHEN Shangrui

(College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

**[Abstract]** To address the problem of low accuracy and efficiency of single-scale feature mapping for multi-scale pedestrian detection, a multi-scale detection method based on multi-resolution filter channels is proposed. The scale perception pool is used to enhance the perception field correspondence, and the scale invariance is achieved through a soft decision tree. When using the sliding window classification strategy, the ground plane constraint and sparse grid are combined to reduce the calculation cost and form an acceleration strategy. Experimental results on the Caltech dataset show that the detection accuracy of the method is 88.89%, and the detection rate is 15.68 frame/s, its detection accuracy is higher than VJ, WordChannels and other methods.

**[Key words]** pedestrian detection; Multi-Resolution Filter Channel (MRFC); scale-aware pool; Soft Decision Tree (SDC); sliding window

**DOI:** 10.19678/j.issn.1000-3428.0053684

### 0 概述

行人检测<sup>[1]</sup>的目的是定位图像中的所有行人。在很多实际应用中, 行人检测的速度与准确率同等重要。虽然深度学习的行人检测方法已经实现了较高的准确率, 但其即便使用高端的 GPU, 检测速度仍然很低。而提升决策树(Boosting Decision Tree, BDT)<sup>[2]</sup>方法在该领域具有很强的竞争力, 能够在保证检测准确率的同时, 提高检测速度。

Caltech 行人数据集<sup>[3]</sup>中的行人样本高度范围为 7 像素 ~ 476 像素, 并且经常会表现出大范围的尺

度变化, 在行人检测过程中会构成类内差异。因此, 如何检测不同尺度的行人已成为行人检测的关键问题。

不同尺度的行人在原始图像中具有不同的像素。如果一个特征在不同尺度上保持不变, 那么这一特征可用于多尺度检测。然而, 许多特征都是尺度多变的, 这意味着在大尺度行人中提取的特征与在小尺度行人中提取的特征不同, 即便是被广泛使用的方向梯度直方图(Histogram of Oriented Gradient, HOG)也存在这一问题<sup>[4]</sup>。为了保证尺度不变性, 可构建图像金字塔并计算金字塔每层的特征映射, 然后

**基金项目:** 国家重点研发计划“矿山灾害生命保障救援通道快速安全构建关键技术与装备”(2018YFC0808203); 陕西省组合与智能导航重点实验室开放基金“基于视觉的自主飞行避障与导航技术”(SKLIIN-20180101)。

**作者简介:** 张传伟(1974—), 男, 教授、博士, 主研方向为目标检测、机电系统智能控制; 曾虹钧、杨萌月、李 波、陈尚瑞, 硕士研究生。

**收稿日期:** 2019-01-15 **修回日期:** 2019-03-18 **E-mail:** 1002443877@qq.com

用训练好的行人模型进行滑动窗口检测。由于行人尺度范围较大,金字塔需要包含很多层,因此特征金字塔的构建需要消耗大量时间。

为避免构造特征金字塔,研究者尝试使用单尺度特征映射实现多尺度行人检测,并取得了重要进展<sup>[5]</sup>。本文在 MRFC<sup>[6]</sup>的基础上提出一种基于多分辨率滤波通道的行人检测方法。使用尺度感知池进行采样,并利用差异积分通道(DICs)<sup>[7]</sup>丰富特征获得更好的感受域对应性。同时,使用软决策树在 BDT 级联中构建弱分类器,软决策树的 2 个分支分别用于大尺度和小尺度实例,以解决尺度多变的问题。

## 1 相关工作

多尺度检测主要利用尺度的不变性。如果特征尺度不变,则可以使用单尺度特征映射来执行检测,再通过调整模型的大小来检测多尺度目标。但是,并非所有用于行人检测的特征都是尺度不变的,且密集图像金字塔会导致计算成本较高。FPDW<sup>[8]</sup>使用指数比例定律计算稀疏特征金字塔和近似中间特征尺度,然而,这种策略通常会导致准确度明显下降。VeryFast<sup>[9]</sup>没有使用特征金字塔,而是根据不同行人的尺度和近似中间分类器尺度,使用指数比例定律来训练一个稀疏分类器金字塔,将测试时间转换为训练时间。FASCF<sup>[10]</sup>使用特征金字塔和分类器金字塔进行多尺度检测。由于 BDT 能够处理类内差异并提供良好的结果,因此一些研究者只使用单尺度特征映射,训练单个行人模型。

为了实现更精确的感受域对应性,本文借鉴空间金字塔池(Spatial Pyramid Pooling, SPP)<sup>[11-12]</sup>的思想,在每个分区中进行特征合并,再对特征映射进行划分,从而将各种尺寸的特征映射转换成固定长度矢量。因为全连接层需要在具有固定长度的矢量中馈送,所以其在基于卷积神经网络(Convolutional Neural Network, CNN)<sup>[13]</sup>的行人检测中非常有用,是 Fast-RCNN 和 Faster-RCNN<sup>[14]</sup>检测器的关键组成部分。不同于 CNN 的最大池化方式,本文使用平均池化来计算积分图。

## 2 尺度感知池

MRFC 是一个基于 CPU 的解决方案,其仅使用单一尺度特征图和单个行人模型进行检测。MRFC 中使用的基本特征映射是 10 个 LUV + HOG 通道。对原始图像计算 10 个通道后,再对 10 个通道连续应用  $3 \times 3$  方框进行 6 次滤波,得到 70 个通道。该过程利用 6 个具有不同标准偏差  $\sigma$  的卷积核对原始

的 10 个通道进行处理,然后将 2 个边缘滤波器(垂直和水平)应用于每个通道中,总共产生 210 个通道。

为了检测多尺度行人,在计算的通道上滑动不同尺度的行人模型,通过网格化方式在通道中进行采样以提取分类特征。网格空间适应于窗口大小,如图 1(a)所示,通过这种方式使不同尺度的行人获得相同数量的特征。在训练时,不同于传统方法将行人尺寸调整到固定尺寸,MRFC 从图像的原始行人尺寸中提取特征。在测试时,无需为多尺度检测计算图像金字塔,而是直接计算这 210 个通道。得到多分辨率滤波通道之后,利用类似于 ACF 检测器的单像素索引进行特征提取。因此,MRFC 检测器的速度非常快。但是,MRFC 方法的最初实现存在一些缺点,感受域对应性问题就是其中之一。如图 1(a)所示,在 MRFC 的实现过程中,特征的感受域不会随着行人的大小而变化。因此,小行人的特征感受域与大行人的特征感受域不对应。在这种情况下,大行人的鼻子特征可能对应于小行人的整个面部,这是不合理的。理想的情况是特征的感受域随行人的尺度而改变,如图 1(b)所示。

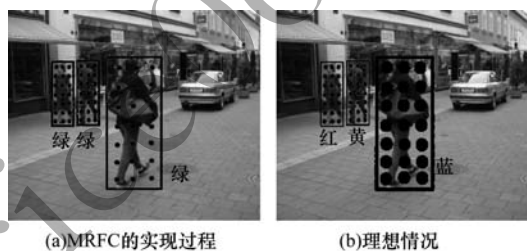


图 1 MRFC 方法中的感受域对应性

Fig. 1 Receptive field correspondence in MRFC method

为了解决上述问题,本文采用类似于 SPP 的方法,将检测窗口划分成  $m \times n$  个单元,并通过在一个或多个单元中的平均池来计算特征,根据检测窗口的大小调整单元的面积,从而使特征感受域更好地对应。本文使用  $23 \times 11$  个单元格,池化区域被约束为不大于  $4 \times 2$  个单元格,从而为每个特征映射出 1 806 个特征。在 MRFC 中,通过不断地卷积形成 7 个不同大小的感受域。图 2 给出本文方法的特征提取过程。其中,图 2(a)表示不同尺度的特征映射被划分为相同数量的单元格,其大小随行人大小而变化,图 2(b)通过由一个或多个单元格组成的不同区域的平均池提取特征,图 2(c)中的上半部分表示特征梯度映射中的池化等效于计算 2 个移位池化区域的差异,其效果与非相邻特征(Nor-Neighboring Features, NNF)类似,下半部分给出 DICs 中的一些判别特征。



图 2 本文方法的特征提取过程

Fig. 2 Feature extraction process of the proposed method

本文使用 10 个基本通道和 20 个梯度通道,总共产生 30 个特征进行映射。与其他类型的卷积特征映射不同的是,本文方法可以使用流式 SIMD 扩展<sup>[15]</sup>计算特征梯度映射。为了快速地进行平均池化,在检测前预先计算积分映射,然后在与区域大小无关的恒定时间内计算区域中的特征值和,即为每一个特征映射(10 个基本特征映射和 20 个梯度特征映射)计算积分映射,并根据卷积换向定律得到式(1)。

$$\Omega(x,y) * G(x,y) * u(x,y) = \Omega(x,y) * u(x,y) * G(x,y) \quad (1)$$

其中, $\Omega$ 是特征映射, $G(x,y)$ 是梯度滤波器, $u(x,y)$ 是阶跃函数。积分可以作为步进函数的卷积,本文只需计算原始的 10 个特征映射的积分映射,然后在此基础上计算 20 个梯度映射。图 3 给出积分图计算的加速策略示意图。

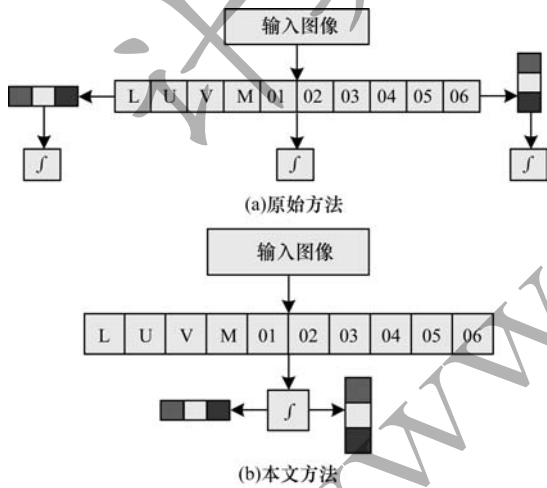


图 3 积分图计算的加速策略

Fig. 3 Acceleration strategy of integral diagram calculation

在 DICs 中提取的特征与 NNF 类似<sup>[16]</sup>,这些特征已被证明对行人检测有效。NNF 与同一水平面中的非相邻矩形区域不同,相较于 NNF,本文的 DICs 的优越性主要表现在以下 3 个方面:

1) NNF 使用固定的模型,其需要图像金字塔,而本文不构造图像金字塔,节省了大量计算时间。

2) NNF 需要对特征计算 2 个区域的平均池,而本文只需要对一个区域计算平均池。

3) NNF 只考虑同一水平面中的 2 个区域,而本文还考虑了在同一垂直面中的 2 个区域,这对于头、肩和脚等部位的检测是非常有用的。

根据定义,本文首先求出区域中的所有特征值,然后除以区域大小来计算平均池特征。为了提高效率,本文不在测试时执行分割操作,而是提前更改每个决策树中单层决策树的阈值。尽管本文只训练一个检测器,但是可将这个检测器切换到  $n$  个检测器中, $n$  为模板尺度的数量。这些检测器有相同的特征指标,但阈值不同,其阈值需要乘以它们所对应的区域大小来获得。

### 3 软决策树

本文使用软决策树在 BDT 中构建弱分类器,其中,树的 2 个分支分别处理大尺度和小尺度的行人。在本文方案中,所有样本都参与单个 BDT 级联的训练。虽然硬决策树中使用的硬决策节点确定性地将样本  $x$  指向其中的一个子节点,但软决策节点会将样本  $x$  指向其左右分支,其概率分别为  $P(L|x)$  和  $P(R|x)$ ,具体计算如式(2)、式(3)所示。

$$P(L|x) = \frac{1}{1 + \exp\left[\frac{\bar{h} - h_x}{\beta}\right]} \quad (2)$$

$$P(R|x) = 1 - P(L|x) \quad (3)$$

其中, $\bar{h}$ 是训练集中行人高度的平均值, $h_x$ 是样本  $x$  的高度, $\beta$ 是控制门函数平滑量的超参数。随着  $h_x$  的增加, $P(L|x)$  变大,样本  $x$  对左分支的权重增加。 $\beta$  反映了不同尺度行人的相似程度,当  $\beta$  较大时,函数  $P(L|x)$  较为平缓,说明不同尺度的实例非常相似,极端情况  $\beta \rightarrow \infty$ ,表示均等地处理所有尺度实例,树会成为平均集合。当  $\beta$  较小时,函数  $P(L|x)$  较为陡峭,说明不同尺度实例的表示存在较大差异,需要以不同方式进行处理,极端情况  $\beta \rightarrow 0$ ,表示节点成为硬决策节点。

软决策节点的输出通过其 2 个分支输出的加权和来计算,如图 4 所示。在测试时,每个实例必须通过所有中间节点,直至到达叶节点。该方案仅适用于单个决策树,对于具有数千个树的 BDT,其计算成本较高。因此,本文只使用根节点的软决策节点对其进行简化。在处理多尺度检测问题时,利用左分支对大尺度行人进行分类,而右分支负责小尺度行人分类。

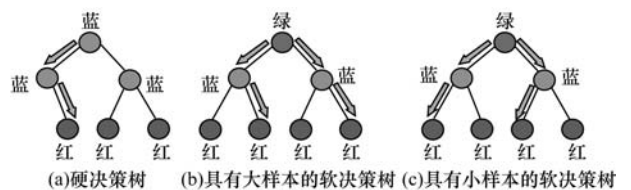


图 4 硬决策树和软决策树的比较

Fig. 4 Comparison of hard decision tree and soft decision trees

图 4 给出硬决策树和软决策树的比较,其中,蓝色、绿色和红色节点分别表示硬决策节点、软决策节点和叶节点,箭头表示样本权重的流动。由图 4(a)可知,硬决策树由硬决策节点和叶节点组成,对于给定的样本,硬决策节点将其所有权重指向其子节点之一。图 4(b)表示软决策树的根节点是一个软决策节点,它根据样本大小将样本权重指向两个子节点,给定一个大样本,软决策节点将更多的权重引导到它的左分支(左分支的箭头比右分支的箭头粗)。图 4(c)表示具有小样本的软决策树,软决策节点将更多的权重指向它的右分支。

在样本  $x$  指向根节点的 2 个分支之后,通过对 2 个分支求值可以得到  $P(y=1|x,L)$  和  $P(y=1|x,R)$ 。其中,  $y \in \{1, -1\}$  是样本标签。整个软决策树给出的  $x$  为正样本的概率为:

$$P(y=1|x) = P(L|x)P(y=1|x,L) + P(R|x)P(y=1|x,R) \quad (4)$$

本文使用 RealSub<sup>[17]</sup> 来训练 BDT,其叶节点不输出概率,而是输出半对数比,如式(5)所示。

$$f(x) = \frac{1}{2} \ln \frac{p(x)}{1-p(x)} \quad (5)$$

其中,  $p(x)$  是叶节点中正样本权重的一部分。为了得到式(4)中的  $P(y=1|x,L)$  和  $P(y=1|x,R)$ ,需要用式(5)的反函数将  $f(x)$  转换成相应的概率,具体如下:

$$p(x) = \frac{e^{2f(x)}}{1 + e^{2f(x)}} \quad (6)$$

在使用式(3)得到  $P(y=1|x)$  后,通过式(5)将其转换为对应的半对数比。在训练过程中,当为每个样本获取半对数比时,将样本权重更新为常用的 RealBoost,并在训练后使用式(6)将所有树节点的半对数比转换为其对应的概率,以提高训练效率。

## 4 加速滑动窗口分类

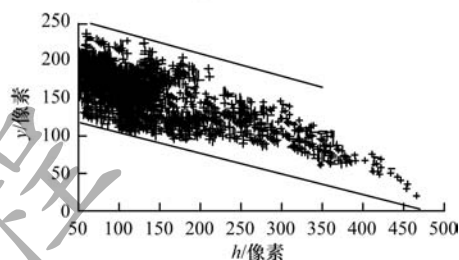
与构造密集金字塔的传统方法相比,使用单尺度特征映射的优点在于,其在特征映射计算阶段节省了大量的计算成本。然而,除了特征映射计算之外,行人检测还存在另一个耗时的过程,即滑动窗口分类。本文采取基于区域平均池的方法,与 ACF 和 LDCF<sup>[18]</sup> 等其他检测器相比,此阶段的计算成本较高。因此,本文的单尺度特征映射必须采取有效的滑动窗口分类策略,否则其优势非常有限。为了验证使用单一尺度特征图的优势,本文将两种简单的加速滑动窗口分类策略进行对比。

### 4.1 地平面约束

对于固定车载相机拍摄的图像,由于地平面约束(Ground Plane Constraint, GPC),一定尺度的行人将不会出现在某些位置,如图 5 所示。其中,图 5(a)表示行人可能被实线框框住,但不可能被虚线框框住。图 5(b)表示在 Caltech 训练集中,行人的可能位置  $(h,y)$  被两条直线限定。



(a)GPC示例



(b)行人位置的分布情况

图 5 GPC 图解

Fig. 5 GPC diagram

目前, GPC 已被广泛用于行人检测,其核心思想是在车载相机的一些有效假设下,行人的投影高度  $h$  与垂直位置  $y$  呈现线性关系。GPC 的一些后处理通过支持向量机(Support Vector Machine, SVM)完成。这种方法旨在提高检测准确率,但其会增加额外的计算成本。本文为了加快滑动窗口分类的速度,在检测阶段只扫描行人可能出现的位置,以节省计算成本。同时,该方法对消除某些误报也有积极的作用。

### 4.2 稀疏网格检测

对于 480 像素  $\times$  640 像素的图像,其有 330 000 多个候选窗口(不同尺度、不同位置)需要被分类,其中大部分属于背景区域。本文真正需要的是峰值得分窗口,即具有局部最高得分的窗口。在非最大抑制(Non-Maximum Suppression, NMS)<sup>[19]</sup> 之后,峰值得分窗口将抑制其邻域中具有较低检测分数的窗口,本文不需要评估所有候选窗口,只需要确保评估所有峰值得分窗口,以降低计算成本。

由于附近位置处的检测器响应是相关的,因此峰值得分窗口的相邻位置通常也具有正响应,说明存在峰值得分窗口的支撑区域(Region of Support, ROS)。对于 BDT 级联,ROS 的大小随着弱分类器的增加而减少。

基于上述分析,本文首先仅评估步长为 3 的稀疏网格  $G_3$ 。如果  $G_3$  中的窗口通过级联的  $k$  个阶段,则触发其  $3 \times 3$  邻域中的每个窗口进行评估,如图 6 所示。在图 6 中,假设窗口  $P$  是峰值得分窗口但不属于  $G_3$ ,则在  $3 \times 3$  邻域中存在窗口  $x_1 \in G_3$ 。窗口  $x_1$  由于 ROS 而具有正分数,  $P$  被  $x_1$  触发。由于  $G_3$  仅占有所有滑动窗口的 1/9,触发窗口的数量较少,因此会大幅降低计算成本。

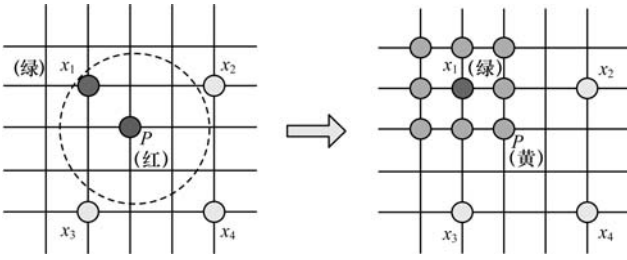


图6 稀疏网格检测策略示意图

Fig.6 Schematic diagram of sparse grid detection strategy

在图6中,首先评估稀疏网格 $(x_1, x_2, x_3, x_4)^{[20]}$ ,假设 $P$ 是峰值得分窗口,其ROS由虚线圆表示。窗口 $x_1$ 位于ROS中,因此它将通过BDT级联的 $k$ 个阶段,并且触发 $3 \times 3$ 邻域中的每个窗口(图6中黄色圆圈)。当 $k$ 增大时,ROS和触发窗口变小,检测速度变高。然而, $k$ 增大容易造成峰值得分窗口丢失,从而降低准确性。在本文实验中,取 $k=20$ 可以实现速度和准确率的平衡。

## 5 实验结果与分析

本文方法既使用了尺度感知和软决策树,又结合了地平面约束和稀疏网格的加速策略,并通过地平面约束和稀疏网格降低滑动窗口分类的计算成本。为了方便描述,下文用SAP、SDT、GPC和SGD分别表示尺度感知池、软决策树、地平面约束和稀疏网格检测算法。在实验中,如果真实目标边界框( $b_g$ )与检测到的边界框( $b_d$ )交集与并集之比(IoU)大于阈值,则将 $b_d$ 视为真值,IoU的计算过程如下:

$$I_{IoU}(b_d, b_g) = \frac{b_d \cap b_g}{b_d \cup b_g} \quad (7)$$

本文将IoU阈值设置为0.5,并与Caltech的结果进行漏检率比较。使用每个图像的假正例(False Positives Per Image, FPPI)曲线作为行人检测的评估指标。计算9个FPPI点处的平均漏检率,并将这些点在 $10^{-2} \sim 10^0$ 的对数空间中进行均匀间隔,最终得到的对数平均漏检率即可作为评价各种检测方法的标准。

将本文方法与VJ、HOG+SVM、HOG-LBP、Multifu+Motion、WordChannels、ACF-Caltech、SCF+AlexNet和MRFC+Semantic等其他具有代表性的方法进行比较。其中,WordChannels和MRFC+Semantic与本文方法相同,均基于单尺度特征映射,所以在对比时更具说服力。

### 5.1 实验配置

本文实验的软硬件配置如下:图像视频处理平台为华硕台式机,CPU为Inter Core i7,频率为2.5 GHz,GPU为NVIDIA GTX1080,120 GB固态硬盘,

内存大小为4 GB,使用OpenCV框架,并参考该开源项目上一些层次结构的具体实现方式。

### 5.2 结果分析

Caltech行人数据集是目前规模最大,使用范围最广的行人检测数据集。它由250 000个640像素×480像素(大约137 min)的视频组成,分为11个样本,前6个样本用于训练,后5个样本用于测试。标准评估在测试集每个视频的30帧处执行,总共产生4 024个图像。

从每4个连续帧中抽取一个图像作为训练图像。由于性能评估仅需要检测高于50像素的行人,因此将最小的模型尺寸设置为64像素×32像素。为了提高效率,将通道降采样2倍,则最小模型的特征映射大小为32像素×16像素。因为训练数据较多,所以本文在每次自举循环中使用两倍数量的弱分类器。

图7给出漏检率随FPPI变化的曲线,对于每个FPPI点,漏检率越低越好。可以看出,在深度学习方法中,SCF+AlexNet的漏检率为23.09%,而在非深度学习方法中,基于单一尺度特征映射的WordChannels和MRFC+Semantic的漏检率分别为42.30%和16.84%,而本文方法的漏检率为10.11%,得到整个FPPI范围内的最低漏检率,优于其他所有的对比方法。

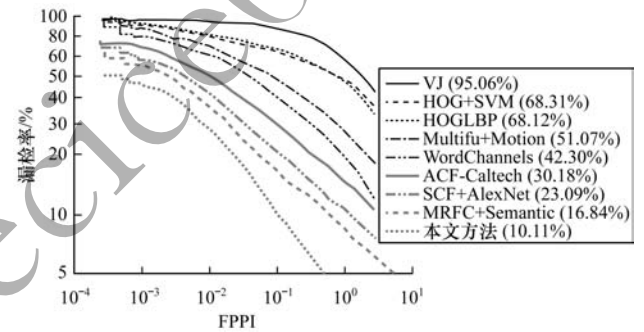


图7 漏检率随FPPI变化的曲线

Fig.7 Curve of the missed detection rate varying with FPPI

此外,本文还在小尺度、不规则宽高比和部分遮挡的条件下进行实验。将小尺度行人高度设置在55像素~85像素,结果如图8所示。可以看出,本文方法的漏检率低至10.17%。将宽高比定义为 $W/H=0.43 \pm 0.1$ ( $W$ 和 $H$ 分别是行人宽度和高度), $W/H$ 的分布集中在0.43,其检测结果如图9所示,可以看出,本文方法的漏检率为14.35%,低于WordChannels和MRFC+Semantic方法。部分遮挡表示行人被遮挡的面积不超过40%,其检测结果如图10所示。在部分遮挡条件下,本文方法的漏检率为17.01%。在上述3个条件下,遮挡对于多尺度行人检测的影响较大,但是从整体来看,本文方法的检测准确率在所有对比实验中最高。



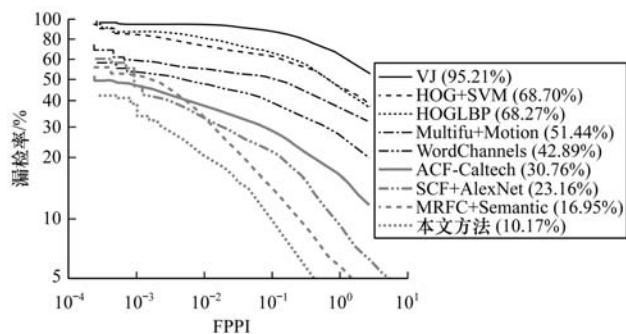


图 8 小尺度条件下的检测结果

Fig. 8 Detection results under small scale conditions

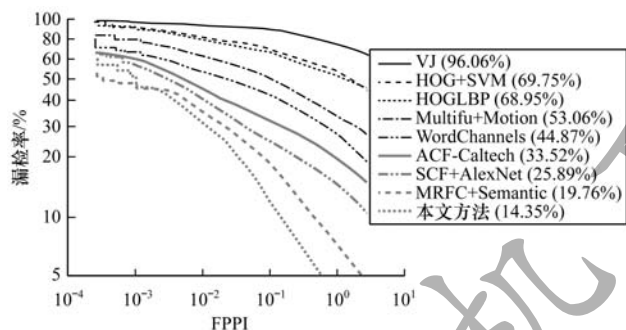


图 9 不规则宽高比条件下的检测结果

Fig. 9 Detection results under the condition of irregular aspect ratio

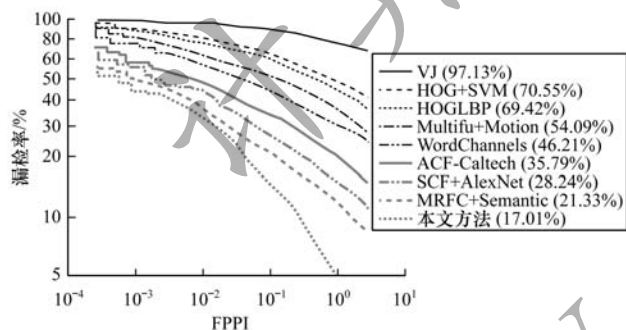


图 10 部分遮挡条件下的检测结果

Fig. 10 Detection results under partial occlusion condition

为了进一步评估本文方法的检测效率和准确率,分别对 4 个组件进行不同的组合,结果如表 1 所示。从表 1 可以看出,单使用 SAP 所达到的准确率和检测速度最低。将 SAP 作为基准,分别观察其他 3 个组件对 Caltech 数据集所产生的效果可以发现, SAP + SDT 组合的准确率比 SAP 高 4.91%,说明尺度感知和软决策树的结合可以有效提高检测准确率。而当结合 SAP + SDT + GPC + SGD 4 个组件时,检测速度也得到提高,可以达到 15.68 frame/s。由此说明,本文方法不仅能够实现对多尺度行人的快速检测,也可以达到较高的准确率。

表 1 在不同组合下的准确率和检测速度比较

Table 1 Comparison of accuracy and detection speed in different combinations

组合	准确率/%	速度/(frame · s <sup>-1</sup> )
SAP	81.06	5.89
SAP + SDT	85.97	10.07
SAP + SDT + GPC	87.95	13.72
SAP + SDT + GPC + SGD	88.89	15.68

### 5.3 真实场景下的行人检测

为了测试本文方法的实时性,使用高清摄像头拍摄的校园内自然真实场景进行实验,场景主要是教学楼前的走道。在训练得到的模型中输入一段视频进行实验,其中部分帧的检测结果如图 11 所示。可以看出,所有帧中的行人目标都能够完整地检测出来,且概率较大,目标边界恰好框出目标。



图 11 部分帧的检测结果

Fig. 11 Detection results of some frames

然而,从图 11(a)可以看出,采用 SAP + SDT + GPC + SGD 相结合的算法也存在一些误检测,这是因为非深度学习在提取特征时,会将一个亮度很大的圆形区域视为人头,并将其下连接的区域视为身体和腿,因此出现了误检测的情况。由图 11(b) ~ 图 11(d)可知,如果行人处于较远的位置或者骑车经过时,检测窗口未能全部准确识别出来,这些需要在后期研究中进行改进。

## 6 结束语

为了提高多尺度行人检测的准确性,本文受 MRFC 的启发,利用单尺度特征映射对多尺度行人检测问题进行研究。对 MRFC 的缺点进行分析,并分别通过尺度感知池和软决策树改善感受域对应性和尺度不变性,使用地平线约束和稀疏网格来降低计算成本。在 Caltech

数据集上的实验结果表明,该方法可实现实时有效的行人检测,与其他先进的检测方法相比,采取 SAP + SDT + GPC + SGD 检测器不仅可以实现快速检测并且检测准确率较高。下一步将在, SAP + SDT + GPC + SGD 检测器的基础上对多尺度行人的实时性检测问题进行研究,以提高算法的实时性能。

### 参考文献

- [1] JIN Peifei, ZHOU Li, LIU Jian, et al. Pedestrian detection based on ROI extraction [J]. Computer Engineering and Design, 2016, 37(11): 3035-3039. (in Chinese)  
靳培飞,周莉,刘健,等.基于感兴趣区域提取的行人检测[J].计算机工程与设计,2016,37(11): 3035-3039.
- [2] SON J, JUNG I, PARK K, et al. Tracking-by-segmentation with online gradient boosting decision tree[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 3056-3064.
- [3] WANG Taiqing, GONG Shaogang, ZHU Xiatian, et al. Person re-identification by video ranking [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 688-703.
- [4] LE T H N, ZHENG Y, ZHU C, et al. Multiple scale Faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Computer Society, 2016: 46-53.
- [5] ZHANG Wei, MA Bingpeng, LIU Kan, et al. Video-based pedestrian re-identification by adaptive spatiotemporal appearance model [J]. IEEE Transactions on Image Processing, 2017, 26(4): 2041-2054.
- [6] GAO Changxin, WANG Jin, LIU Leyuan, et al. Temporally aligned pooling representation for video-based person re-identification[C]//Proceedings of IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2016: 4284-4288.
- [7] LÜ Rui, SHAO Zhenfeng. A pedestrian detection algorithm based on sparse multi-scale image segmentation and cascade deformable part model [J]. Geomatics and Information Science of Wuhan University, 2016, 41(11): 1544-1549. (in Chinese)  
吕瑞,邵振峰.基于稀疏多尺度分割和级联形变模型的行人检测算法[J].武汉大学学报(信息科学版), 2016, 41(11): 1544-1549.
- [8] MUKASA T, XU J, STENGER B. 3D scene mesh from CNN depth predictions and sparse monocular SLAM [C]//Proceedings of IEEE International Conference on Computer Vision Workshop. Washington D. C., USA: IEEE Press, 2018: 912-919.
- [9] ZHOU Jiqiang. Research on multi-class target detection and multi-target tracking algorithms in surveillance video [D]. Harbin: Harbin Institute of Technology, 2017. (in Chinese)  
周纪强.监控视频中多类目标检测与多目标跟踪算法研究[D].哈尔滨:哈尔滨工业大学, 2017.
- [10] BAO Gang, LAI Jun. Radar cross section reduction of a cavity in the ground plane: TE polarization [J]. Communications in Computational Physics, 2014, 15(4): 895-910.
- [11] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [12] SHI Zenglin, YE Yangdong, WU Yunpeng, et al. Crowd counting using rank-based spatial pyramid pooling network [J]. Acta Automatica Sinica, 2016, 42(6): 866-874. (in Chinese)  
时增林,叶阳东,吴云鹏,等.基于序的空间金字塔池化网络的人群计数方法[J].自动化学报,2016,42(6): 866-874.
- [13] FU Xinchuan, YU Rui, ZHANG Weinan, et al. Delving deep into multiscale pedestrian detection via single scale feature maps [J]. Sensors, 2018, 18(4): 1063.
- [14] ZHANG Chenglong, XIA Xiaojun, BAI Song, et al. Target track recording system based on kernelized correlation filters tracking algorithm [J]. Computer Systems and Applications, 2017, 26(5): 113-118. (in Chinese)  
张乘龙,夏筱筠,柏松,等.基于KCF跟踪算法的目标轨迹记录系统[J].计算机系统应用,2017,26(5): 113-118.
- [15] HENRIQUES JF, RUI C, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [16] WANG Liqing, HUANG Songjie. Image retrieval algorithm based on multiscale convolutional neural network [J]. Software Guide, 2016, 15(2): 38-40. (in Chinese)  
王利卿,黄松杰.基于多尺度卷积神经网络的图像检索算法[J].软件导刊,2016,15(2): 38-40.
- [17] IZUMI I. Image registration with subpixel accuracy of DCT-sign phase correlation with real subpixel shifted images [C]//Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis. Washington D. C., USA: IEEE Press, 2017: 210-215.
- [18] LI Qi. Research and application of deep learning in pedestrian detection [D]. Beijing: Beijing University of Posts and Telecommunications, 2017. (in Chinese)  
李蹊.深度学习在行人检测中的研究与应用[D].北京:北京邮电大学, 2017.
- [19] DENG Wei, LIU Bingham. Deep learning-based pedestrian detection combined with semantics [J]. Computer Systems and Applications, 2018, 27(6): 167-172. (in Chinese)  
邓炜,刘秉瀚.联合语义的深度学习行人检测[J].计算机系统应用,2018,27(6): 167-172.
- [20] ZHOU Ying, ZHAO Xiaozhe, LU Chao. Uncertain target detection based on multi-scale feature clustering algorithm [J]. Fire Control and Command Control, 2019, 44(4): 166-170. (in Chinese)  
周颖,赵晓哲,逯超.基于多尺度特征聚类算法的不确定目标检测[J].火力与指挥控制,2019,44(4): 166-170.