

A Study of Pedestrian Detection Based on Multi-Channel Template and Convolution Neural Network

Xianyue Pan, Yishan Liu, Ke Tan

Electronic Science and Engineering College, National University of Defense Technology, Changsha Hunan
Email: 597431090@qq.com

Received: Jan. 5th, 2018; accepted: Jan. 23rd, 2018; published: Jan. 30th, 2018

Abstract

In this paper, based on the traditional pedestrian detection framework of RGB channel template and convolution neural network, a pedestrian detection method combining multi-channel template and convolution neural network is proposed. The main contribution is: firstly, the motion detection method is used to detect the interest area, improve pedestrian detection efficiency and reduce the false alarm. Secondly, multi-channel template and maximum fusion strategy are adopted to reduce the possible pedestrian leakage detection based on RGB channel template.

Keywords

Channel Template, Convolutional Neural Network, Pedestrian Detection

结合多通道模板和卷积神经网络的行人检测方法研究

盘先跃, 刘怡杉, 谭 科

国防科技大学电子科学与工程学院, 湖南 长沙
Email: 597431090@qq.com

收稿日期: 2018年1月5日; 录用日期: 2018年1月23日; 发布日期: 2018年1月30日

摘 要

本文在传统的依靠RGB通道模板和卷积神经网络的行人检测框架下, 提出了结合多通道模板和卷积神经

网络的行人检测方法。主要工作是：一是采用运动侦测方法初步检测感兴趣区域，提高行人检测效率和降低虚警。二是采用多通道模板和最大值融合策略，降低单纯依靠RGB通道模板可能存在的行人漏检现象。

关键词

通道模板，卷积神经网络，行人检测

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

对于营区周界警戒应用而言，行人的入侵检测需要考虑复杂天候的影响，譬如，在夜晚照明不足或者采用红外补光照明的情况下，图像的色彩信息丢失，纹理特征显著性降低，此时基于视频分析方式可靠检测行人目标是非常困难的。现有的行人检测方法所使用的训练和测试数据集都是白天的彩色图像，导致现有行人检测方法在夜晚低照度场景、雾雪天气等低对比度场景以及其他类似的复杂天候场景的行人检测正确率偏低，尤其是漏检率偏高[1]。这可能导致采用现有方法研制的营区周界警戒系统的安全性大幅下降。为此，本文针对复杂天候条件，提出一种结合多通道模板和卷积神经网络的行人检测方法，主要工作是融合多通道模板的卷积神经网络计算结果，降低复杂天候条件下依靠单一通道模板可能存在的行人漏检现象。该方法的基本流程如图1所示，主要包括感兴趣区域提取、多通道模板提取、卷积神经网络计算、融合判决四个部分。详细描述如下。

2. 感兴趣区域提取

目前，基于视频分析的行人检测方法中检测正确率较高的方法基本都是采用深度学习算法实现的，这类算法的优点是行人检测正确率高，虚警率低。然而，深度学习算法往往比较耗时，运算效率偏低，这对于营区周界警戒系统来说往往是难以接受的。为了提高行人检测效率，本文首先采用运动侦测技术快速提取视频中的感兴趣区域，仅针对感兴趣区域采用深度学习算法检测行人目标。这样可以在不降低行人检测正确率的基础上大幅提高行人检测效率，对营区周界警戒系统而言很有意义。同时，由于运动侦测可以剔除许多没有运动目标的背景区域，从而也可以降低这些区域引起虚警的可能，进一步降低行人检测的虚警率。

目前，运动侦测方法主要分为三类：时间差分法、背景差分法和光流法。时间差分法是利用图像序列中连续帧(通常为两到三帧)之间的各像素差异来检测运动目标的方法，比较适用于动态环境。但对光照变化非常敏感，虚检现象较为严重。同时，对于低速运动目标容易出现孔洞效应。背景差分法是运动侦测最常用的方法之一，其基本原理是从参考背景图像中减去当前帧图像。而且背景图像在一段时间内实时更新，可以适应动态背景的变化。该方法通常要求摄像机静止，对于动摄像机的运动侦测效果不佳。光流法通过检测由于观察者和参考场景之间的相关运动而产生的光流来检测运动目标，对动摄像机的适应性较好，但运算效率偏低。针对营区周界警戒系统对运算效率和目标完整性要求较高的需求，这里选择背景差分法来进行感兴趣区域的提取。具体是采用高斯混合模型(Gaussian Mixture Model, GMM)来检测运动的感兴趣目标。

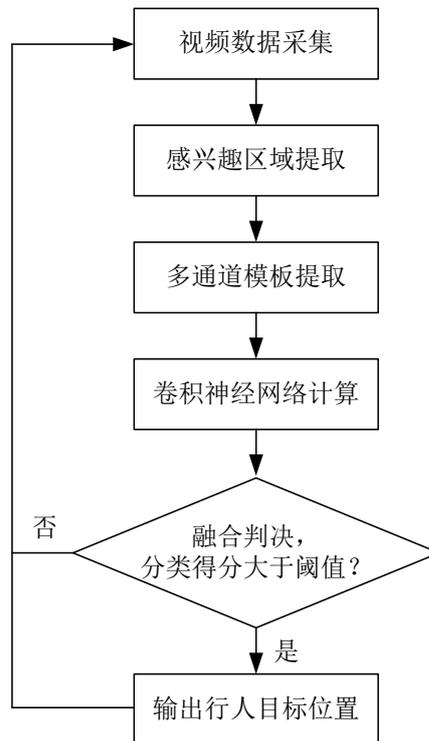


Figure 1. The pedestrian detection process based on multi-channel templates and CNN

图 1. 基于多通道模板和 CNN 的行人检测流程

假设图像中的像素可以用 K 个高斯分布的和来描述, X_t 代表在时间 t 的像素值, 其概率描述为

$$P(X_t) = \sum_{k=0}^K \omega_{k,t} \eta_k(X_t, \mu_{k,t}, \Sigma_{k,t}) \quad (1)$$

其中, K 表示高斯分布的数目, $\omega_{k,t}$ 表示时间 t 的第 k 个高斯分布的权重, 它反映了该分布可以作为背景模型的程度, $\mu_{k,t}$ 是表示时间第 k 个高斯分布的均值, $\Sigma_{k,t}$ 表示时间 t 第 k 个高斯分布的协方差矩阵, 假设 X_t 的分量是独立的, 且方差相同, 即 $\Sigma_{k,t} = \sigma_{k,t} I$, η_k 表示第 k 个高斯分布, 描述如下

$$\eta_k(X_t, \mu_{k,t}, \Sigma_{k,t}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_{k,t}|}} \exp\left(-\frac{1}{2}(X_t - \mu_{k,t})^T \Sigma_{k,t}^{-1} (X_t - \mu_{k,t})\right) \quad (2)$$

在模型更新之后, K 个高斯分布按照 $\frac{\omega}{\sigma}$ 由大到小的顺序重新排序, 选取前 B 个最有可能代表背景的高斯分布, 公式为

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_{k,t} > T_w \right), b = 1, 2, \dots, K \quad (3)$$

其中, T_w 是描述背景在场景中的最小先验概率的阈值。

由于光照等复杂场景的变化, 背景模型也应是一个动态模型, 需要根据场景的变化来自适应更新。模型更新是背景差分算法的重要环节, 对于高斯背景模型而言, 需要更新的是高斯分布的参数, 公式为

$$\omega_{k,t} = (1 - \alpha) \omega_{k,t} + \alpha M_{k,t} \quad (4)$$

其中, α 称为学习率, 取值范围在 0~1 之间。 $M_{k,t}$ 是一个二元值, 当观察像素与高斯分布匹配时, $M_{k,t}=1$, 否则 $M_{k,t}=0$ 。观察像素与高斯分布匹配的条件为

$$|X_t - \mu_{k,t-1}| \leq 2.5\sigma_{k,t-1} \quad (5)$$

如果没有发现与观察像素相匹配的高斯分布, 则将观察像素值作为权重最小的高斯分布的平均值。如果存在与观察像素相匹配的高斯分布, 则相匹配的高斯分布的平均值和方差将被更新, 公式为

$$\begin{cases} \mu_{k,t} = (1-\rho)\mu_{k,t-1} + \rho X_t \\ \sigma_{k,t}^2 = (1-\rho)\sigma_{k,t-1}^2 + \rho(X_t - \mu_{k,t-1})^2 \end{cases} \quad (6)$$

其中, ρ 为学习因子, 计算公式为

$$\rho = \alpha \eta_k (X_t | \mu_{k,t-1}, \Sigma_{k,t-1}) \quad (7)$$

其余不匹配的高斯分布保持不变。

在本文中, 依据部分运动侦测实验结果以及高斯背景模型的经验参数, 设置学习率 $\alpha = 0.005$ 、阈值 $T_w = 0.7$ 、高斯分布数量 $K = 6$ 。

3. 多通道模板提取

对于上一小节获取到的每一个感兴趣区域, 本文采用目前检测正确率高的深度学习框架来进行行人检测。相对于原始视频图像, 感兴趣区域的尺寸很小, 行人检测耗时也小。现有基于深度学习的行人检测方法主要是针对 RGB 格式的图像进行的。然而, 营区周界警戒环境的天候复杂度高, 单纯依靠 RGB 格式的图像信息容易造成漏检。为此, 本文提出多通道模板的联合卷积神经网络检测框架, 通过多通道模板增强行人特征的显著性, 通过联合卷积神经网络计算各个通道模板的行人概率, 最后融合各个通道模板的 CNN 计算结果形成最终判决, 可以大幅降低复杂天气条件下的行人漏检问题, 这对提高营区周界警戒系统的安全性具有极大的促进作用。

本文所述的多通道模板包括: RGB 通道模板(记为 RGB_CM)、亮度通道模板(记为 Y_CM)、局部二元模式通道模板(记为 LBP_CM)和梯度通道模板(记为 G_CM)。其中, RGB_CM 也即 RGB 颜色空间的图像模板, Y_CM 也即灰度空间的图像模板, LBP_CM 也即采用局部二元模式对灰度图像进行变换之后得到的图像模板, G_CM 也即对灰度图像求取边缘得到的图像模板。选择这四种通道模板的依据是: 首先, RGB_CM 是目前行人检测常用的通道模板, 该通道模板对于白天场景的行人检测效果很好, 已经在行人检测的各个数据集上得到了验证。其次, 选择 Y_CM 主要原因是摄像机在夜晚模式红外补光条件下图像彩色信息会丢失, 此时采用亮度通道更能体现行人特征的显著性。另外, 选择 LBP_CM 和 G_CM 的主要原因是考虑到营区周界警戒系统需要全天候运行, 环境光照变化比较大, 采用 LBP 通道和梯度通道可以消除光照变化的影响。另外, 在可视条件差的雾雪天气, LBP 通道和梯度通道更能凸显行人的显著特征。这里 LBP 通道主要从纹理的角度来凸显行人特征, 梯度通道主要从边缘的角度凸显行人特征。通过这四个通道模板, 可以增强单纯使用 RGB 通道模板的行人特征的显著性。图 2 展示了四种通道模板的示意图。

下面简要介绍如何依据 RGB_CM 构建 Y_CM、LBP_CM 和 G_CM。

令 $f_{RGB}^{(R)}(x, y)$ 、 $f_{RGB}^{(G)}(x, y)$ 和 $f_{RGB}^{(B)}(x, y)$ 分别表示像素点 (x, y) 在 RGB_CM 上的 R 通道值、G 通道值和 B 通道值, $f_Y(x, y)$ 表示像素点 (x, y) 在 Y_CM 上的像素值, $f_{LBP}(x, y)$ 表示像素点 (x, y) 在 LBP_CM 上的像素值, $f_G(x, y)$ 表示像素点 (x, y) 在 G_CM 上的像素值, 那么有

$$f_Y(x, y) = 0.3f_{RGB}^{(R)}(x, y) + 0.59f_{RGB}^{(G)}(x, y) + 0.11f_{RGB}^{(B)}(x, y) \quad (8)$$

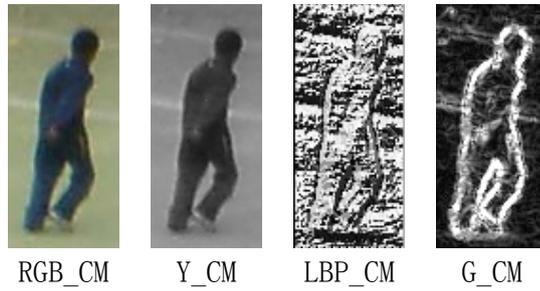


Figure 2. Schematic diagram of four channel templates
图 2. 四种通道模板示意图

$$f_G(x, y) = \sqrt{(f_Y(x+1, y) - f_Y(x-1, y))^2 + (f_Y(x, y+1) - f_Y(x, y-1))^2} \quad (9)$$

$$f_{LBP}(x, y) = \sum_{n=1}^8 f_{LBP}^{(n)}(x, y) 2^{n-1} \quad (10)$$

其中,

$$f_{LBP}^{(n)}(x, y) = \begin{cases} 1, & f_Y^{(n)}(x, y) \geq f_Y(x, y) \\ 0, & \text{其他} \end{cases} \quad (11)$$

这里, $f_Y^{(n)}(x, y)$ 表示像素点 (x, y) 所在 8 邻域的第 n 个像素点的亮度值, 本文将像素点 (x, y) 右侧相邻像素点作为第 1 个邻域像素点, 按照顺时针的方式依次为第 2~第 8 个邻域像素点, 如图 3 所示。

4. 卷积神经网络计算

卷积神经网络(Convolutional Neural Networks, CNN)是一个由多层网络组成的网络结构, 每一层的输出由上一层的输出和当前层的参数决定[2], 可以表示为

$$x^l = f^l(x^{l-1}, \theta^l) \quad (12)$$

其中, x^{l-1} 和 x^l 分别表示第 $l-1$ 层和第 l 层的输出, θ^l 表示第 l 层的参数, 可以表示为

$$\theta^l = [w^l, b^l] \quad (13)$$

这里, w^l 和 b^l 分别表示第 l 层的权重和偏差参数。

f^l 表示第 l 层的网络模型, 也即该层网络所要执行的操作, 主要分为四种类型: 卷积、池化、全连接和回归。

CNN 计算将输入的每一个图像映射到每一个类别的分类得分可以表示为

$$p_c = \text{softmax}(x^{L-1} | \theta^L) \quad (14)$$

其中, L 为 CNN 的网络层数, $p_c \in [0, 1]$, $c = \{0, 1\}$ 。 $c = 0$ 表示类别为非行人, $c = 1$ 表示类别为行人。函数 softmax 的定义为

$$\text{softmax}(x^{L-1}) = \frac{\exp(x_c^{L-1})}{\exp(x_0^{L-1}) + \exp(x_1^{L-1})} \quad (15)$$

卷积的定义是:

$$x_{m', n'}^l = \sum_m \sum_n (w_{m, n, d}^l \cdot x_{m+m', n+n'}^{l-1} + b_d^l) \quad (16)$$

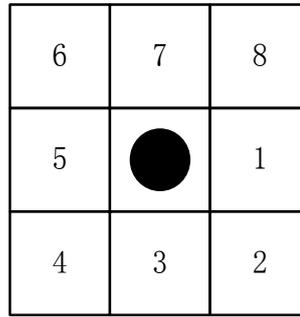


Figure 3. Neighborhood pixel sequence number schematic

图 3. 邻域像素点序号示意图

其中 m 、 n 分别表示图像的高度和宽度， d' 表示滤波器的数量。

卷积之后紧跟激活函数，这里采用 ReLU (Rectified Linear Unit) 激活函数，定义为

$$a(x_{m',n'}^l) = \max(0, x_{m',n'}^l) \tag{17}$$

池化层采用最大值池化，定义为

$$x_{m,n}^l = \max_{m',n' \in L(m,n)} (x_{m',n'}^{l-1}) \tag{18}$$

这里， $L(m,n)$ 表示输入位置。

关于 CNN 架构的选择和初始化，本文采用常用的 VGG 网络[3]进行初始化(具体是将 caffe 平台下已训练好的 prototxt 和 caffemodel 文件加入数据层)，然后使用自己构建的数据进行微调。为了提高行人检测效率，将 VGG 网络的原始输入图像尺寸由 224×224 改为 64×64 。VGG 网络共包含 16 层，即 $L = 16$ 。卷积层的尺寸为 3×3 ，池化层的尺寸为 2×2 。具体采用开源的 Caffe 架构下的 VGG 网络实现。

5. 融合判决

本文使用了四个多通道模板(RGB_CM、Y_CM、LBP_CM 和 G_CM)，每一个通道模板经过 CNN 计算之后都可以由公式(3)~(14)得到一个分类得分，分别记为 $p_c^{(RGB_CM)}$ 、 $p_c^{(Y_CM)}$ 、 $p_c^{(LBP_CM)}$ 和 $p_c^{(G_CM)}$ 。下面考虑如何融合这些分类得分形成最终的判决。

常用的融合方法有四种，包括平均、加权平均、多数投票和加权投票。但是，本文采用多通道模板的目标是降低恶劣天气下行人检测的漏检率。因此，本文采用极大值融合策略，选出各个通道模板对行人类别的分类得分的最大值，表示为

$$p_1 = \max(p_1^{(RGB_CM)}, p_1^{(Y_CM)}, p_1^{(LBP_CM)}, p_1^{(G_CM)}) \tag{19}$$

设定一个得分阈值 T_s ，如果 $p_1 > T_s$ ，则给出检测到行人目标的判决，否则给出未检测到行人目标的判决。在本文中，阈值 $T_s = 0.5$ 。也即，四个多通道模板中只要有一个模板的行人类别分类得分大于设定的得分阈值，就判断检测到行人目标。这样可以大幅降低传统的 RGB 通道模板可能存在的漏检现象。

6. 实验与分析

行人检测是计算机视觉的研究热点之一，目前用于行人检测的公开测试数据集也较多，本节首先介绍行人检测领域的数据集以及针对营区周界警戒应用的自建数据集，然后介绍行人检测的常用评价指标，最后分析和对比本文方法与现有性能较好的行人检测方法的性能指标。

6.1. 行人检测数据集

目前, 行人检测领域的公开测试数据集主要有 6 个, 包括 INRIA、ETH、TUD-Brussels、Daimler、KITTI 和 Caltech。这些数据集依据构建目标的不同都有不同的特点。其中, 常用的是 INRIA 和 Caltech 两个数据集。简要介绍如下:

1) INRIA

INRIA 是最早使用的行人检测数据集, 该数据集包含的图像共有 5264 幅(其中包含行人的图像有 3548 幅)。尽管该数据集规模不大, 但优势在于包含的行人图像是经过裁剪的, 非常适用于行人检测器的训练。

2) Caltech

Caltech 数据集是由穿越美国洛杉矶街头的车辆记录的 2.5 小时 30 Hz 视频构成的, 视频分辨率为 640 × 480。视频中注释了共计 35 万个包围约 2300 个独立行人的包围盒, 也即人工标注的行人窗口数量约为 350,000 个。由于这些数据都是来自真实街区场景的自然行人, 行人存在遮挡现象和姿态变化, 因此检测难度较大。

对于营区周界警戒系统而言, 还需要考虑复杂天候的影响, 譬如夜晚可视程度差的情况。因此, 在现有数据集的基础上, 本文还自建了营区周界警戒数据集, 采用视频录制和互联网收集的方式构建了共有 50 个视频片段, 并标记视频片段中每帧图像中行人出现的位置, 用于训练和测试。

6.2. 常用性能评价指标

行人检测的性能评价指标主要有两个: 真正率(True Positive Rate, TPR)和假正率(False Positive Rate, FPR), 计算公式如下

$$\text{TPR} = \frac{\text{正确检测到的行人数量}}{\text{标记的行人数量}} \times 100\% \quad (20)$$

$$\text{FPR} = \frac{\text{错误检测到的行人数量}}{\text{错误检测到的行人数量} + \text{正确检测到的行人数量}} \times 100\% \quad (21)$$

这里, 正确检测到行人的含义是: 检测到的行人窗口与人工标记的行人窗口的重合度大于 50%。错误检测到行人的含义是: 检测到的行人窗口与人工标记的行人窗口的重合度不大于 50%。重合度定义为

$$\text{IoU}^{(r, \tilde{r})} = \frac{\hat{S}^{(r, \tilde{r})}}{S^{(r)} + S^{(\tilde{r})} - \hat{S}^{(r, \tilde{r})}} \times 100\% \quad (22)$$

其中, r 和 \tilde{r} 分别表示标记的行人窗口和检测到的行人窗口, $S^{(r)}$ 和 $S^{(\tilde{r})}$ 分别表示 r 和 \tilde{r} 的面积, $\hat{S}^{(r, \tilde{r})}$ 是指 r 与 \tilde{r} 之间交叉区域的面积。

另外, 行人检测效率也是行人检测性能的重要评价指标之一, 本文采用平均检测耗时来描述行人检测效率, 该值越小说明行人检测效率越高。平均检测耗时是指平均对每一帧视频图像进行行人检测所耗费的时间, 不包含视频解码、显示以及分类器训练耗时。由于耗时测算与处理器和处理环境性能关联大, 因此实验时需要在相同平台上进行耗时测试, 本文使用的计算性平台参数如表 1 所示, 后续不再赘述。

6.3. 不同方法的性能对比分析

为了评价本文提出的结合多通道模板和卷积神经网络的行人检测方法(简记为 MC_CNN 方法)的性能, 将该方法与目前行人检测领域性能较好的方法进行对比实验。这里选取了三种方法进行对比实验。其中, 文献[4]是一种基于特征的行人检测方法, 采用 HOG 特征和 SVM 分类器实现行人的检测, 简记为 HOG_SVM 方法。文献[5]是一种基于深度学习的行人检测方法, 采用深度卷积神经网络(Deep Convolutional

Table 1. Computer platform parameters**表 1.** 计算机平台参数

内容	参数
CPU	Intel I5 3.2 GHZ
内存	8 G
操作系统	Windows 7 64 位
软件平台	Matlab 2012
深度学习库	Caffe
计算机视觉库	OpenCV 2.4.11

Neural Networks)进行行人目标检测,简记为 DCNN 方法。文献[6]是一种结合运动检测和特征分类的行人检测方法,采用光流法检测运动区域,然后结合 HOG 特征和 SVM 分类器进行行人的检测,简记为 OF_HOG_SVM 方法。这三种方法可以从不同的侧面来印证本文方法的优越性。需要说明的是,四种方法都使用 INRIA 数据集、Caltech 数据集的 set00~set05 子集以及自建数据集的前 20 个视频片段进行训练。测试时,首先在 Caltech 数据集的 set06~set10 子集上进行测试,然后再在剩余的自建数据集上进行测试,分别统计各个数据集上的行人检测性能指标。其中,图 4 展示了 Caltech 数据集上四种方法的性能指标。图 5 展示了自建数据集上四种方法的性能指标。表 2 给出了四种方法的平均检测耗时对比结果。下面结合实验结果进行分析。

1) FPR 指标对比分析

由图 4 和图 5 可见,本文方法和文献[5]所述方法的 FPR 值明显低于其他两种方法,究其原因,主要是因为本文方法和文献[5]所述方法都是采用深度学习框架实现行人目标的检测,而文献[4]和文献[6]所述方法都是采用结合特征与分类器的方法实现行人目标的检测。这说明,深度学习框架在行人目标检测应用中的虚检率低,也即把非行人目标检测为行人目标的概率低。另外,本文方法的 FPR 值低于文献[5]所述方法,文献[6]所述方法的 FPR 值也低于文献[4]所述方法,这主要是由于本文方法和文献[6]所述方法采用运动侦测滤除了背景区域的干扰,降低了在这些区域将非行人目标检测为行人目标的概率,从而降低了 FPR 的值。对比图 4 和图 5 中各方法的 FPR 值可以发现,文献[4]所述方法在自建数据集下的 FPR 值增大明显,这是因为自建数据集下的干扰目标较多,容易造成虚检。而文献[6]所述方法通过运动侦测过滤了大量干扰目标,因此 FPR 值增大不显著。文献[5]所述方法采用深度学习框架进行行人目标的检测,虚检现象不显著,故 FPR 增加也不明显。本文方法通过感兴趣区域提取环节过滤了大量干扰目标,又结合卷积神经网络的深度学习框架实现了可靠的行人检测,因此 FPR 值也没有明显增加。

2) TPR 指标对比分析

由图 4 和图 5 可见,本文方法的 TPR 值高于文献[5]所述方法,这是因为本文采用多通道模板增强了行人特征的显著性,在特征融合时采用最大值融合策略降低了行人目标的漏检率。本文方法的 TPR 值也高于文献[4]和文献[6]所述方法,主要原因是后两种方法在提取行人目标特征时受行人姿态和遮挡情况的影响较大,如果训练数据集中的行人姿态和遮挡情况不完备,那么在测试时难免存在行人漏检的现象。对比图 4 和图 5 中各方法的 TPR 值可以发现,文献[4]、文献[5]和文献[6]所述方法在自建数据集下的 TPR 值都有明显的下降,这是因为自建数据集中行人目标的光照、对比度等条件不理想,导致这三种方法存在不同程度的漏检。而本文方法融合了多通道模板,在 RGB_CM、Y_CM、LBP_CM 和 G_CM 四个通道模板上分别检测行人目标,避免某通道行人特征不显著造成的漏检现象,因此 TPR 值没有明显下降。

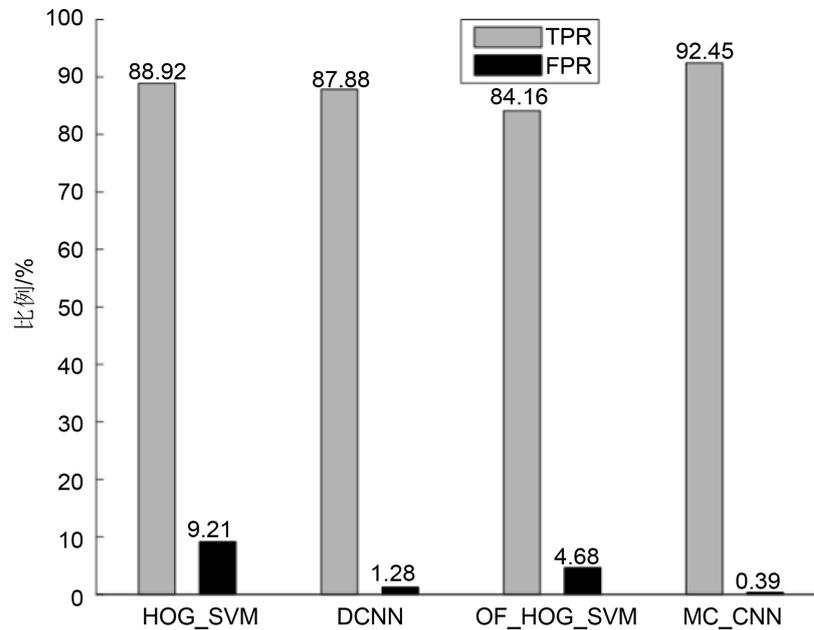


Figure 4. Caltech data set test results

图 4. Caltech 数据集测试结果

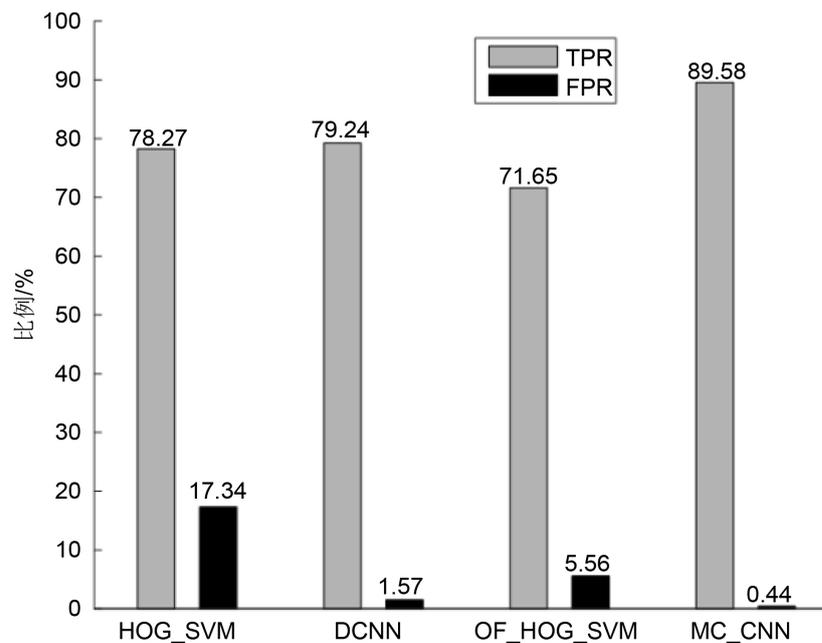


Figure 5. Self-built data set test results

图 5. 自建数据集测试结果

3) 运算效率对比分析

由表 2 可见, 文献[5]所述方法的平均检测耗时最多, 其次是文献[4]所述方法, 本文方法和文献[6]所述方法的平均检测耗时相当, 明显小于其他两种方法, 说明本文方法和文献[6]所述方法的运算效率高大于其他两种方法。分析其原因, 主要是因为本文方法和文献[6]所述方法都使用了运动检测方法过滤了大量背景区域, 从而大幅降低了检测耗时。

Table 2. Average time-consuming comparison of four methods under two data sets
表 2. 两个数据集下四种方法的平均检测耗时对比

方法	平均检测耗时/ms
HOG_SVM	753
DCNN	1791
OF_HOG_SVM	79
MC_CNN	82

7. 结束语

本文在传统的依靠 RGB 通道模板和卷积神经网络的行人检测框架下,提出了结合多通道模板和卷积神经网络的行人检测方法。主要贡献在于:一是采用运动侦测方法初步检测感兴趣区域,提高行人检测效率和降低虚警。二是采用多通道模板和最大值融合策略,降低单纯依靠 RGB 通道模板可能存在的行人漏检现象。实验结果表明,本文方法的真正率高、假正率低,检测耗时少。后续研究重点是进一步提高检测效率,以期达到实时处理的效果。

参考文献 (References)

- [1] 李权. 面向安全监控的异常声音识别的研究[D]: [硕士学位论文]. 长沙: 湖南师范大学, 2015.
- [2] 张永良, 张智勤, 吴鸿韬, 等. 基于改进卷积神经网络的周界入侵检测方法[J]. 计算机科学, 2017, 44(3): 182-186.
- [3] Ribeiro, D., Nascimento, J.C., Bernardino, A., *et al.* (2016) Improving the Performance of Pedestrian Detectors Using Convolutional Learning. *Pattern Recognition*, **61**, 611-619.
- [4] Zhou, Z. and Xu, L. (2014) Pedestrian Detection Based on HOG and Weak-Label Structural SVM. *Journal of Computational Information Systems*, **10**, 367-374.
- [5] Tomè, D., Monti, F., Baroffio, L., *et al.* (2016) Deep Convolutional Neural Networks for Pedestrian Detection. *Image Communication*, **47**, 482-489. <https://doi.org/10.1016/j.image.2016.05.007>
- [6] Zhang, S., Bauckhage, C., Klein, D.A., *et al.* (2016) Moving Pedestrian Detection Based on Motion Segmentation. *Multimedia Tools & Applications*, **75**, 6263-6282. <https://doi.org/10.1007/s11042-015-2571-z>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
 左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
 期刊邮箱: csa@hanspub.org