



基于移动小车的行人监控系统

邢惠钧¹, 昌硕²

(1. 北京师范大学第二附属中学, 北京 100088; 2. 北京邮电大学, 北京 100876)

摘要: 行人监控属于监控系统中比较重要的一个方面。传统的行人监控手段存在监控区域受限的问题, 即监控设备一旦安装后, 只能监控固定的区域, 目前仍然需要人工判别监控画面中是否有行人。针对这个问题, 设计了基于计算机视觉的移动行人监控系统。通过操控搭载了监控摄像头的移动小车, 实现监控区域的自由切换。另外, 配合成熟的深度卷积神经网络算法和相关滤波器, 实现了监控画面中行人的自主识别、定位。最后对该系统进行了实地测试, 验证了行人监控系统的可行性。

关键词: 计算机视觉; 行人监控; 移动小车; 深度卷积神经网络; 相关滤波器

中图分类号: TP277

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017042

Pedestrian surveillance system based on mobile vehicle

XING Huijun¹, CHANG Shuo²

1. The Second High School Attached to Beijing Normal University, Beijing 100088, China

2. Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: Pedestrian surveillance is one of the most important aspects in the surveillance system. Traditional surveillance equipment can only cover the limit area. Namely, once the surveillance equipment is implemented, it can only monitor a specific region. Besides, the surveillance system can't detect if there is a person in the surveillance picture or not. The professional staffs are needed to determine whether there is a person in the surveillance picture. To solve this problem, a computer vision based on mobile pedestrian surveillance system was designed. By implementing a surveillance camera on a mobile small vehicle and remotely control it, the system could switch the monitoring area. Besides, the system could classify and locate the pedestrian in the picture with the deep convolutional neural network and correlation filters. Finally, the system was tested on the spot, which verified the feasibility of the system.

Key words: computer vision, pedestrian surveillance, mobile vehicle, deep convolutional neural network, correlation filter

1 引言

行人监控属于监控系统比较重要的一个方面, 行人监控的主要目标是实时地识别、定位监控视频序列中的行人。传统的基于摄像头的监控系统仅仅提供了被监控场景

的图像信息, 监控系统并不能告知图像画面中是否有人以及画面中行人的实时位置信息。Dalal 等人^[1]提出了HOG (histogram of oriented gradients) 检测方法, 并利用 SVM (support vector machine, 支持向量机)^[2]对该特征基于人体进行分类。工业界随之将这套人体检测算法用于行人的监

控,以减轻安防人员的工作负担。但该种行人检测方式只适用于背景比较干净,同时图像内的人体无遮挡的情景。对于复杂多变的监控环境,该算法的泛化性并不好,行人检测的准确率下降明显,误检率、漏检率都非常高,未能大规模地普及。

基于深度卷积神经网络的人体识别算法是近几年最热门的研究领域之一^[3-5],基于该种方式的识别算法大大地提高了复杂场景下人体的识别精度。Ouyang 等人^[6]使用深度学习的方法联合训练了特征提取、可变人体模型、遮挡模型以及人体分类器这 4 种以往需要分开考虑的问题。Luo 等人^[7]提出了 SRBM(switchable restricted Boltzmann machine,可切换的受限玻尔兹曼机)去解决复杂背景下不同姿态的人体,该算法联合学习了人体的层级信息和特征图,将人体分为头部、躯干、腿 3 部分分别进行识别。Sermanet 等人^[8]提出了使用离散自编码器的非监督学习去预训练一个针对行人检测的 CNN(convolutional neural network,卷积神经网络)。Li 等人^[9]最近提出了具有尺度感知的基于 fast R-CNN(region with convolutional neural network)的行人检测模型,该模型分别针对大尺度行人和小尺度行人构建了子神经网络用于加权识别,由于针对不同尺度有明确的子模型进行识别,提高了小物体的检测准确率。针对 fast R-CNN 检测过程耗时这个缺点,Ren 等人^[10]通过设计 RPN(region proposal network)来替代 fast R-CNN 中选择性搜索(selective search)过程,有效降低了行人候选区域的生成时间,提高了模型的检测效率。

但是目前开源的 faster R-CNN 模型,更多的是针对多类目标的检测,其中人体数据集只占训练数据集的一部分,而模型的检测效果很大程度上与训练该模型所用的数据集规模有关。针对此种情况,本文基于 faster R-CNN 算法,利用 Caltech Pedestrian Detection Benchmark^[11]数据库微调了该模型,使网络提取的特征对行人具有更好的表征能力,在测试集上相较于传统视频监控所采取的 HOG+

SVM 行人检测方法,显著降低了行人的误识别率。对于本文的监控系统来说,完全依赖于该检测算法是不能满足监控的实时性要求的,系统的摄像头采集帧频率为 10 f/s(帧每秒),而基于 faster R-CNN 的行人检测模型处理一帧图片的平均耗时为 0.26 s。为了解决这个问题,本文在监控系统中引入了基于相关滤波的跟踪算法(处理速度大于摄像头采集帧频率的两倍)^[12],并将其与检测算法融合,通过降低检测算法的检测频率,满足监控系统的实时性要求。

2 基于移动小车的行人监控系统的组成

整个行人监控系统主要由五大部分组成,包括:移动小车数据采集模块、无线数据传输模块、图像处理模块、运动控制模块、客户端显示模块。各个模块之间的协作关系如图 1 所示。

2.1 移动小车数据采集模块

移动小车数据采集模块主要由监控摄像头、摄像头云台、云台控制系统、小车、小车控制系统和供电系统 6 部分组成。摄像头的图像采集帧频率为 10 f/s,采集的图片大小为 640 dpi×360 dpi,图片的位深为 16 位。摄像头云台由两个舵机和放置监控摄像头的平台组合而成,舵机的加入保证了云台具备左右 90°的转向角和上下 90°的俯仰角。云台的作用主要是帮助调整监控摄像头的拍摄方向,进而增大监控范围。小车的驱动方式采用的是直流减速电机,通过控制左右两组电机的转动方向来实现小车的前进、后退以及左右转向,实际操作过程中小车会远程接收控制模块发出的指令并进行移动。移动小车数据采集模块,通过远程遥控可以实现对监控目标物的移动式监控,突破了传统监控手段的位置限制,避免了大量监控摄像头的安装,提高了硬件资源的利用率,有效地克服了传统监控系统监控区域受限、监控不及时、监控资源利用率低 3 方面的问题。整个采集模块的实物如图 2 所示。

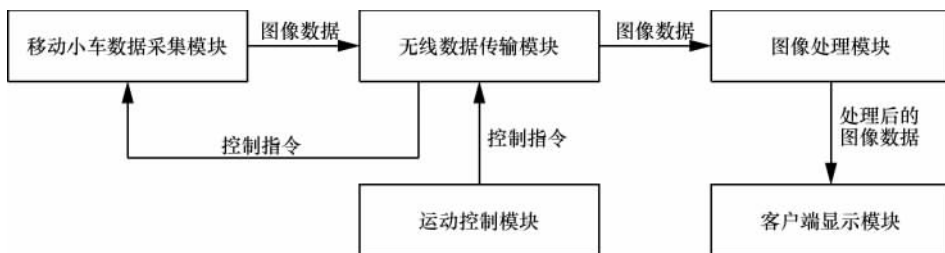


图 1 基于计算机视觉的小型可移动行人监控系统组成

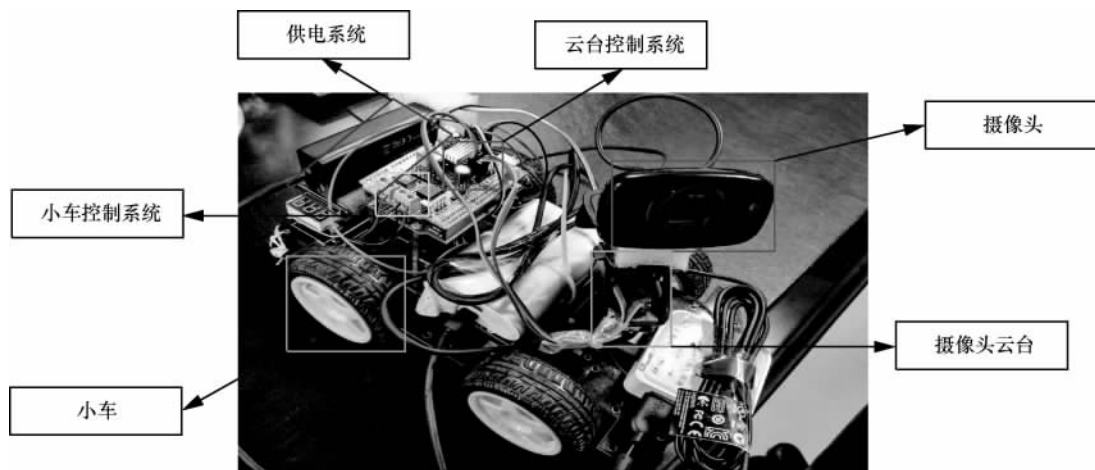


图 2 移动小车数据采集模块实物

2.2 无线数据传输模块

无线数据传输模块采用成熟的 Wi-Fi 传输模块。它主要承担起图像数据传输任务,将实时的图像数据分别传输给图像处理模块和客户端显示模块。

2.3 图像处理模块

图像处理模块为整个系统的核心部分,在接收无线数据传输模块传过来的图像数据后,按照模块的设计逻辑,交替运行检测算法和跟踪算法,并最终达到实时监控的目的。该图像处理模块的运行机制如图 3 所示。

整个图像处理模块依靠开关机制实现检测算法和跟踪算法的相互融合。图像处理模块运行时,第一步会通过检测算法完成整个模块的初始化工作:根据当前环境信息合理配置检测算法和跟踪算法的运行参数;判断监控画面内是否存在行人;一旦存在行人,便通过检测算法记录该行人的位置。第二步检测算法会将记录的行人的位置信息传递给跟踪算法,此时跟踪算法工作并确保行人处于跟踪框内。跟踪算法判断目标物在视频帧中的位置依赖于搜索区域经过相关滤波后输出的 *response* 值,当这个值低于设

定的阈值 *gate* 之后,说明跟踪结果已经不可信了,此时触发检测算法并重新校正跟踪算法。不停地循环第二步中的过程,就完成了对行人的实时监控。

这样做的原因是基于深度卷积神经网络的行人检测模型处理一帧图片平均需要 0.26 s,而追踪算法每秒处理的帧数超过了监控系统摄像头采集的帧频率,达到 64 f/s。检测算法可以以比较大的概率来确定监控画面中是否存在行人以及行人的位置,而追踪算法仅仅利用了视频序列的时间相关性来确定行人的位置,存在跟丢的可能。通过这样一种融合机制,使两种算法实现优势互补,最终确保了监控的实时性和有效性。

2.4 运动控制模块与客户端显示模块

客户端显示模块用来呈现最终的监控画面以及图像处理模块的计算结果。当监控画面中存在行人时,图像处理模块会自动将这个行人框出来。依据客户端显示的画面,操作人员可以远程遥控搭载了摄像头的小车,跟随被监控的行人。而远程遥控的实现是通过运动控制模块来完成的,发出控制指令来指挥小车的运动。对于传统的监控

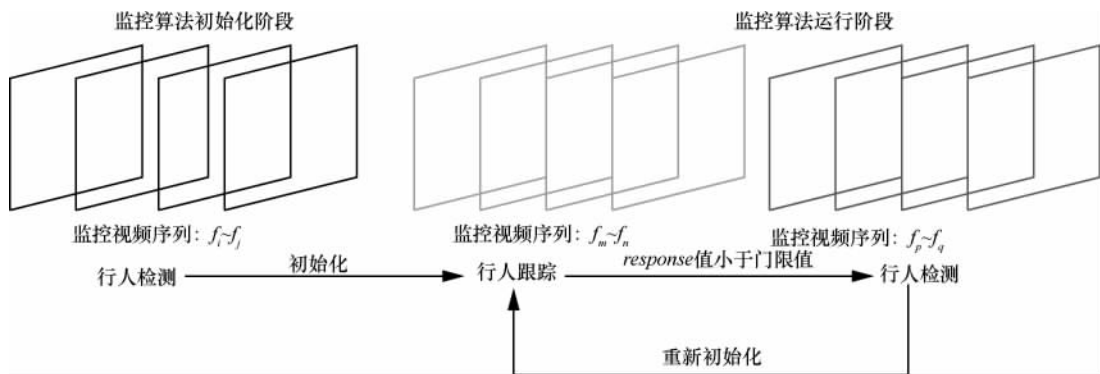


图 3 图像处理模块运行机制

摄像头来说,一旦可疑分子离开摄像头的视野,监控便彻底宣告失败,但基于移动小车的行人监控系统可以有效地避免这种情况的发生。

3 行人监控系统检测算法 faster R-CNN

图像处理模块中行人检测算法是通过 faster R-CNN^[10]来实现的,整个框架本文并未进行改动,只是将该模型在行人数据集上进行了二次训练。整个模型主要包括 3 个部分:特征提取、RPN、R-CNN。图 4 给出了该模型的训练过程。下面会分别讲述每层的实现细节以及后续的模型训练和模型使用过程。

3.1 行人检测模型设计与实现

3.1.1 特征提取

特征提取采用的是 VGGNet^[13],本文采用 16 层的 VGGNet 模型,在使用过程中,去掉了它的全连接层和 softmax 层,因为这两层合起来可以对前面网络的输出进行分类,而这主要针对的是分类任务,并不适合于行人检测任务。最终只留下了全连接层前面的特征提取层,以下统称为特征提取网络。特征提取网络总共包含 5 个卷积层、4 个最大池化层。

(1)卷积层设计

传统的神经网络各个层之间采取全连接的方式,权值参数相对卷积神经网络来说数量非常大。对一张大小为 3 dpi×3 dpi 的灰度图片来说,假设隐藏层神经元为 9,那么全连接的方式需要的权值参数为 36=3×3×4 个,而对于一个卷积核为 3 dpi×3 dpi 的卷积层来说,权值参数只需要 4=

2×2 个。全连接与卷积过程示意如图 5 所示。

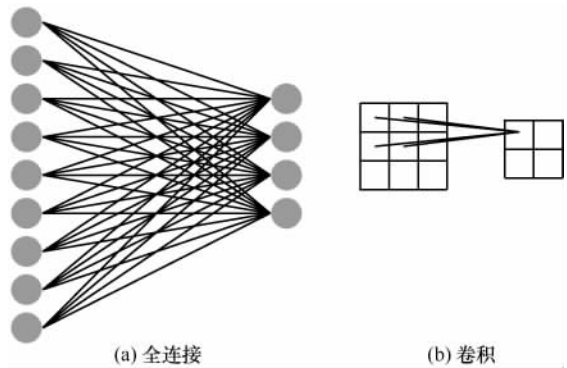


图 5 全连接与卷积过程示意

可以看到,卷积操作能有效地降低参数个数,同时卷积操作说明了这样一种事实,图像的一部分统计特性与其他部分是一样的,这意味着在这一部分学习到的特征也能运用到图像的另一部分上。特征提取网络的所有卷积核大小设置为 3 dpi×3 dpi,卷积核的移动步长设置为 1,另外所有的卷积层对输入的特征图做卷积时,会在前面的特征图四周填充一位 0。对一个输入高为 n 、宽为 m 的特征图做卷积时,输出的特征图大小依旧是 $n \times m$,其计算式如下。

$$h = \frac{n + p \times 2 - stride}{stride} + 1 \tag{1}$$

$$w = \frac{m + p \times 2 - stride}{stride} + 1 \tag{2}$$

其中, p 为填充的 0 的位数, $stride$ 为卷积模板的滑动步长。据此计算可得,输出后的特征图大小与输入的特征图大小一致。

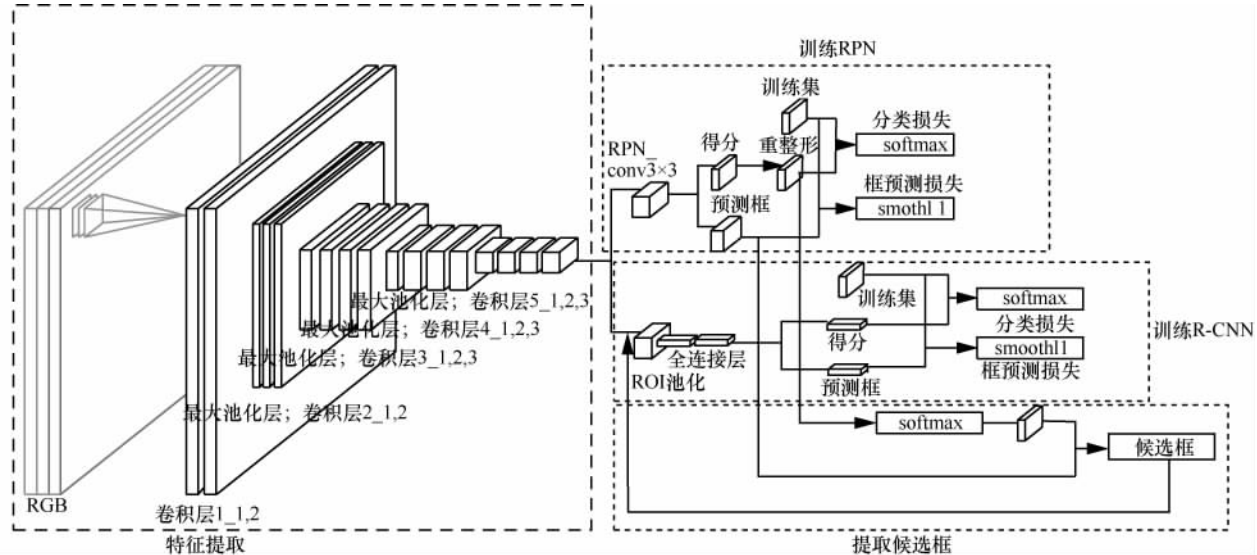


图 4 行人检测模型训练过程



深度卷积神经网络中所使用的激活函数,本文选择的是 ReLu^[14](rectified linear unit)。激活函数的主要作用就是将卷积操作得到的线性叠加值,映射到非线性空间。而传统的 Sigmoid 和 Tanh 函数,在神经网络模型比较“深”的时候,会出现梯度弥散现象,即在网络训练过程中,训练误差形成的梯度在后向传播的过程中,值会变小。导致初始的卷积层参数不更新,或者更新幅度比较小,无法达到训练网络的目的。而 ReLu 激活函数可以解决这个问题。ReLu 的数学表达式如下:

$$F=\max(0,x) \quad (3)$$

(2)池化层设计

为了描述大的图像,一个很自然的想法就是对不同位置的特征进行聚合统计。而这些概要统计特征不仅具有低得多的维度(聚合可以达到降维的目的),同时还会改善结果(不容易过拟合)。对应在本文的模型框架中,即最大池化。其数学表达式如下。

$$y_{ij}=\max(x_{m,n},(m,n) \in region) \quad (4)$$

3.1.2 RPN

任何的行人检测框架都有一个必须实现的重要功能,就是选取待检测的候选区域,本文所使用的 faster R-CNN 模型实现了一个小型卷积网络 RPN 来提取候选区域。在实际应用中,这个小型卷积网络会在特征提取的最后一层特征图 cov5_3 上滑动,在滑动的每一个位置形成 k 个潜在的候选区域,这样 RPN 最后的回归层会有 $4k$ 个输出,隐含了每一个候选区域的 4 个坐标,同时分类层会有 $2k$ 个输出,表示候选区域是物体或者不是物体的概率值。在实际应用中采取 3 种长宽比、3 种面积的候选框,这最终会形成 $k=9$ 个候选区域。对于一个大小为 $W \times H$ 的特征图来说最终会形成 WHk 个候选区域。前面提到,每个 RPN 都有两类输出:回归层的输出和分类层的输出,分类层为传统的 softmax,回归层主要用于调整 RPN 形成的框,尽量使其与实际的框相接近。RPN 最终优化的目标方程为:

$$L=\frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (5)$$

其中, p_i 为 RPN 对于区域 i 的类别预测值, p_i^* 为该区域实际的类别值, L_{cls} 为预测类别值与真实值之间的对数误差; t_i 是 RPN 预测区域 i 所在的位置, t_i^* 是该区域实际的位置, L_{reg} 表示预测位置与真实位置之间的回归误差。最终误差 L 为分类误差与回归误差的权重叠加,相应的调和

参数为 N_{cls} 与 N_{reg} 。优化手段采用 SGD^[15]算法,即调整神经网络的连接参数极小化 RPN 的目标方程。

3.1.3 R-CNN

R-CNN 网络通过利用 RPN 形成的候选框,在每个候选框内依靠 ROI 池化机制形成一个定长的特征向量,然后输入 R-CNN 的分类层与框回归层。注意到 R-CNN 的分类层输出的是具体的类别,不再像 RPN 那样只区分候选框所在的区域是不是物体。需要注意的是,由于每个候选框的大小不一,所覆盖的候选区域也大小不一,这就导致了最终形成的特征向量长度不一,而分类与回归层要求所有的输入向量必须是定长的特征向量。此时,前面提到的 ROI 池化层就派上用场了,它的主要作用就是不管输入的向量尺寸如何,最终都输出一个固定长度的特征向量,方便后续的操作。反映在图 4 的模型训练过程中,即连接 cov5_3 的 ROI 池化层。

3.2 行人检测模型的训练

行人检测模型的训练,是通过对 VGGNet^[13]在 Caltech 数据库上微调生成的。该模型训练的流程主要包括前期的数据预处理和深度卷积神经网络模型的 4 步训练过程。

3.2.1 数据预处理

整个 Caltech Pedestrian Detection Benchmark^[16]数据集分为 10 个视频集,每个视频集包含若干视频序列。为了进行后期的评估,训练数据集的形成过程与主流的检测方法类似,即将前 5 个视频集作为训练数据集,按顺序依次抽取每一个视频集的每一个视频序列,隔 30 帧抽取一张图片作为训练图片,最终形成训练数据集。测试数据集在剩下的 5 个视频集中按抽取训练图片的方式出去测试图片,并最终形成测试数据集。在实践中发现,训练数据集越大,并且越丰富,相同的网络模型经过训练后表现出的性能越优异。为了扩充原有的训练数据集,本文采取数据增强技术,每一张训练图片通过左右对称,扩充为两张图片,以此达到训练数据集加倍的目的。

3.2.2 深度模型框架训练

本文在训练模型的过程中为了让 RPN 和 R-CNN 实现卷积层的权值共享,在训练 faster R-CNN 模型的时候用了 4 阶段^[10]的训练方法,具体如下。

步骤 1 使用 ImageNet^[17]上预训练的模型初始化特征提取网络参数,并微调 RPN。

步骤 2 使用步骤 1 中 RPN 提取 region proposal 训练 R-CNN。

步骤 3 使用步骤 2 的 R-CNN 重新初始化 RPN, 固定卷积层进行微调。

步骤 4 固定步骤 2 R-CNN 卷积层, 使用步骤 3 中 RPN 提取的 region proposal 微调网络。

按照上述 4 个步骤顺序训练, 即可得到最终的训练模型。

3.3 模型检测流程

整个算法的测试流程如图 6 所示, 将监控序列中的某一帧图片输入前面训练好的深度网络模型里面, 最终该模型会输出当前图片是否有行人以及这个行人目前所处的位置。

整个测试流程分为 3 部分, 首先该模型会提取待检测图片的卷积特征, 根据提取的卷积特征通过 RPN 确定可能存在行人的候选区域, 最后将每个候选区域过一遍 R-CNN 最终确定该候选区域是否存在行人 (图 6 中 softmax 输出对应的是目标物为行人的概率) 以及行人的位置 (图 6 中预测框输出对应的是行人在图片中的位置)。

4 行人监控系统跟踪算法的相关滤波器

用于图像追踪的相关滤波器^[18,19]本质是在学习一个判别式分类器, 最终通过搜寻滤波器的最大输出来确定目标物的相对位移距离, 以实现追踪的目的。本文中的具体做法如下, 将输入的图像转换为灰度图像, 用 x 表示。 x 沿着图像上下、左右循环移位构成了训练样本。每一个训练样本表示为: $x_{m,n} \in \{0,1,\dots,M-1\} \times \{0,1,\dots,N-1\}$, 相对应的输入

分类器的高斯标签为 $y(m,n) = e^{-\frac{(m-M/2)^2 + (n-N/2)^2}{2\sigma^2}}$ 。相关滤波器对应的参数 w 同 x 的大小一致, 且它是通过最小化如下计算式来获得的:

$$w^* = \arg \min_w \sum_{m,n} \|wx_{m,n} - y(m,n)\|^2 + \lambda \|w\|_2^2 \quad (6)$$

其中, λ 是正则化参数。式 (6) 的闭式解可以通过快速傅里叶变换来获得。规定大写字母代表相应的傅里叶域信号, 那么最终的参数 w^* 在傅里叶域的形式为式 (7) 所示:

$$W^* = \frac{Y \odot \bar{X}}{X \odot \bar{X} + \lambda} \quad (7)$$

其中, $\bar{X}$ 代表向量 X 的共轭, 符号 $\odot$ 代表的是向量对应元素相乘。基于目标物在当前帧的位置, 划定目标物在下一帧的搜索区域, 将该区域对应的图像块从下一帧剪切出来, 并转换为灰度图像 z 。那么相关滤波器在下一帧的输出值计算如下:

$$r = F^{-1}(W^* \odot \bar{Z}) \quad (8)$$

最终目标物在下一帧的位置相对于当前帧目标物位置的偏移矢量为 r 矩阵中值最大的行与列。将计算出的偏移量与当前帧目标物的位置相加, 便可得到目标物在下一帧的位置, 循环迭代整个过程就可以实现对目标物的追踪。

5 行人监控算法试验结果与分析

本文的主要贡献在于对 faster R-CNN 模型进行了微调, 提高了原有框架的行人检测准确率, 结合跟踪算法将其运用在行人的监控中, 整个监控系统满足了实时性以及有效性, 达到了预期的设计目标。下面将给出单独的基于深度卷积神经网络的检测算法的准确率变化曲线和监控系统的实际测试效果。

5.1 行人检测算法试验结果

本文最终训练得到的行人检测模型在 Caltech 测试集上进行了测试, 测试结果如图 7 所示, 平均漏检率为 23%, 基本上达到了比较高的水平。图 7 为各个检测算法的每张图片的假正率 (false positives per image) 与漏检率变化曲

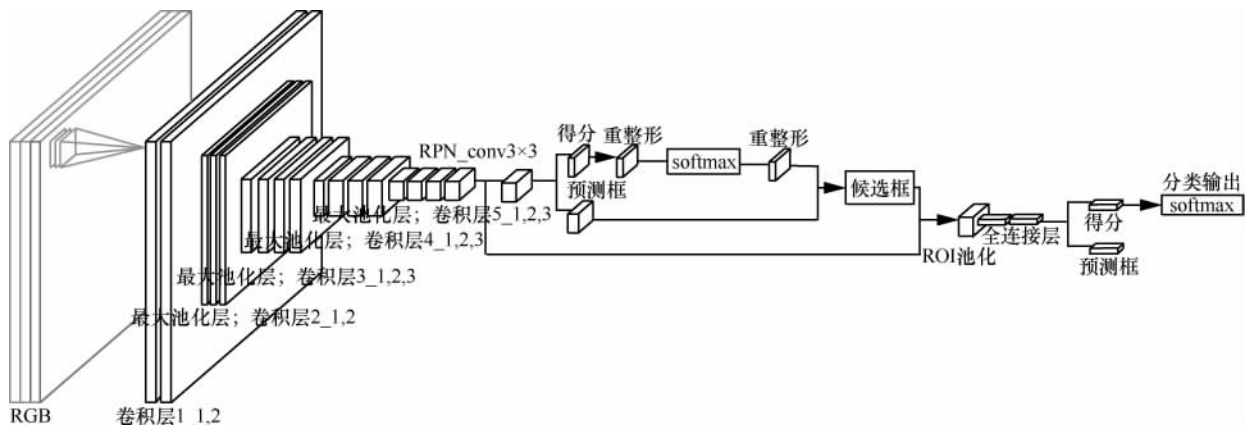


图 6 行人检测算法深度卷积神经网络测试



线图,从图7中可以看出,相比于传统的工业界使用的HOG+SVM模型来说,本文重新训练的行人检测模型,它的平均漏检率降低了45%。虽然训练得到的行人检测模型并没有达到最低的漏检概率,但是通过表1可以发现,相比其他模型来说,本文训练得到的检测模型单张图片的测试时间更低,更有利于实际的应用。由于SAF R-CNN^[24]并没有开源其训练代码,所以表1的最后两行数据直接引用了参考文献[24]中的数据。

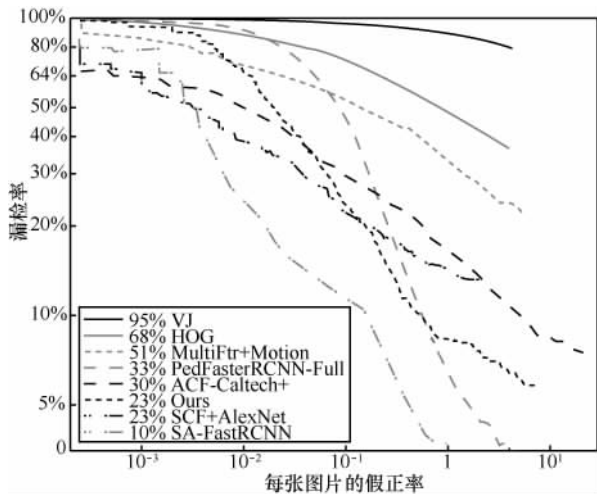


图7 行人检测模型表现曲线^[20,1,21,22,23,24]

5.2 监控算法实际测试效果

本文所设计的基于移动小车的行人监控系统,在实际测试过程中达到了预期的设计目标。下面是实际测试过程



图8 行人监控系统实时监控示例(子图片下的数字代表的是相应的帧序号)



图9 监控失败样例

表1 检测时间与漏检率对比

网络名称	测试时间/s	漏检率
本文提出模型	0.26	23%
SAF R-CNN	0.37	10%
R-CNN	5.31	13%

中的一些示例,如图8所示。

通过这些图片序列可以看到,本文提出的算法可以有效地对出现在摄像头前的行人进行监控。另外在实际测试中,本文设计的监控算法可以达到10 f/s的处理速度,基本满足了行人监控系统的实时性要求。另外在实际测试中发现,该系统对密集人群的检测精度不高,并且存在误识别和漏检行人的情况。图9就是一些监控失败的场景示例图片。

6 结束语

本文基于深度卷积神经网络算法和相关滤波,实现了对目标物的实时监控。通过对faster R-CNN模型在Caltech数据集上二次训练,相较于工业界普遍使用的HOG+SVM模型,该方法显著降低了行人的漏检概率。另外为了满足监控系统的实时性要求,本文引入了基于相关滤波跟踪算法,并通过基于阈值判断的融合机制使检测与跟踪算法交替进行,最终实现对行人的监控,基本满足了监控算法的实时性要求。不过整个监控系统并不能很好地适用密集人群这种场景,同时也存在着行人漏检与误识别的情况,这也是这套系统今后需要改进的地方。

参考文献:

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Computer Vision and Pattern Recognition, June 20–26, 2005, San Diego, CA, USA. New Jersey: IEEE Press, 2005: 886-893.
- [2] SUYKENS J A, VANDEWALLE J. Least squares support vector machine classifiers [J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [3] ZHANG L, LIN L, LIANG X, et al. Is faster R-CNN doing well for pedestrian detection[C]// European Conference on Computer Vision, October 8–16, 2016, Amsterdam, Netherlands. Berlin: Springer, 2016: 443-457.
- [4] SOLANKI D K M S. Pedestrian detection using R-CNN [J]. Group, 2016(12228): 12419.
- [5] OLIVEIRA L, NUNES U, PEIXOTO P. On exploration of classifier ensemble synergism in pedestrian detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(1): 16-27.
- [6] QUYANG W, WANG X. Joint deep learning for pedestrian detection [C]// International Conference on Computer Vision, Dec 1–8, 2013, Sydney, Australia. New Jersey: IEEE Press, 2013: 2056-2063.
- [7] LUO P, TIAN Y, WANG X, et al. Switchable deep network for pedestrian detection[C]//Computer Vision and Pattern Recognition, Jun 23–28, 2014, OH, USA. New Jersey: IEEE Press, 2014: 899-906.
- [8] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]// Computer Vision and Pattern Recognition, Jun 23–28, 2013, ORUSA. New Jersey: IEEE Press, 2013: 3626-3633.
- [9] GIRSHICK R. Fast R-CNN [C]// International Conference on Computer Vision, Jun 7–12, 2015, MA, USA. New Jersey: IEEE Press, 2015: 1440-1448.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems, December 7–12, 2015, Montreal, Canada. New York: Curran Associates, 2015: 91-99.
- [11] DOLLAR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: a benchmark [C]// Computer Vision and Pattern Recognition, June 20–25, 2009, FL, USA. New Jersey: IEEE Press, 2009: 304–311.
- [12] HENRIQUES J.F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computing Research Repository, 2014 abs: 1409-1556.
- [14] NAIR V, HINTON G.E. Rectified linear units improve restricted Boltzmann machines[C]// International Conference on Machine Learning, Dec 12–14, 2010, Washington, DC, USA. New Jersey: IEEE Press, 2010: 807-814.
- [15] SUTSKEVER I, MARTENS J, DAHL G E, et al. On the importance of initialization and momentum in deep learning [J]. ICML, 2013(28): 1139-1147.
- [16] PITOT D, CHRISTIAN W, BERNT S, et al. Pedestrian detection: an evaluation of the state of the art [J]. PAMI, 2012, 34(4): 743-761.
- [17] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [18] BOLME, DAVID S, et al. Visual object tracking using adaptive correlation filters[C]// Computer Vision and Pattern Recognition, Dec 12–14, 2010, Washington, DC, USA. New Jersey: IEEE Press, 2010: 2544-2550.
- [19] HENRIQUES, CASEIRO, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]// European Conference on Computer Vision, Oct 7, 2012, Firenze, Italy. Berlin: Springer, 2012: 702-715.
- [20] VIOLA P, JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137–154.
- [21] WALK S, MAJER N, SCHINDLER K, et al. New features and insights for pedestrian detection [C]// 2010 IEEE Conference on Computer Vision and Pattern Recognition, Dec 12–14, 2010, Washington, DC, USA. New Jersey: IEEE Press, 2010: 1030-1037
- [22] NAM W, DOLLAR P, HAN J H. Local decorrelation for improved pedestrian detection[C]// Advances in Neural Information Processing Systems, Dec 8–13, 2014, Montréal, CANADA. New York: Curran Associates, 2014: 424-432.
- [23] HOSANG J, OMRAN M, BENENSON R, et al. Taking a deeper look at pedestrians[C]// 2015 Computer Vision and Pattern Recognition, Jun 8–12, 2015, Boston, MA, USA. New Jersey: IEEE Press, 2015: 4073-4082.
- [24] LI J, LIANG X, SHEN S M, et al. Scale-aware fast R-CNN for pedestrian detection [J]. arXiv preprint arXiv:1510.08160, 2015.