基于改进全卷积网络的多尺度感知行人检测算法

刘辉**, 彭力, 闻继伟*

江南大学物联网工程学院, 物联网应用技术教育部工程中心, 江苏 无锡 214122

摘要 当前行人检测的一个主要挑战是在复杂的场景中检测出不同尺度的行人,尤其是远尺度行人。由于不同尺度的行人会表现出不同的视觉外观特征,鉴于此提出了一种多尺度感知的行人检测算法。在全卷积网络结构中引进可形变卷积层,扩大特征图的感受野;通过级联区域建议网络提取多尺度行人建议区域,引入多尺度判别策略,定义尺度判别层,判别行人建议区域的尺度类别;构建一个多尺度感知网络,引进软非极大值抑制(Soft-NMS)检测算法,融合每个网络输出的分类值和回归值,获取最终的行人检测结果。实验表明,本文提出的检测算法在基准数据集 Caltech 和 ETH 上的检测误差较低,检测精度优于当前其他检测算法,适用于检测远尺度行人。

关键词 机器视觉; 全卷积网络; 可形变卷积层; 多尺度判别策略; 多尺度感知网络; 行人检测

中图分类号 TP391.4 **文献标识码** A **doi**: 10.3788/LOP55.091504

Multi-Scale Aware Pedestrian Detection Algorithm Based on Improved Full Convolutional Network

Liu Hui**, Peng Li, Wen Jiwei*

Engineering Research Center of Internet of Things Technology Applications of the Ministry of Education, School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, *China*

Abstract A major challenge of pedestrian detection is to detect different-scale pedestrians in complicated scenarios, especially for far-scale pedestrians. Motivated by the experiment that pedestrians with different scales exhibit dramatically different visual features, we propose in this paper a multi-scale aware pedestrian detection algorithm. Firstly, we introduce deformable convolutional layers in full convolutional network structure to expand the receptive field of feature maps. Secondly, we use cascade-region proposal network to extract multi-scale pedestrian proposals and introduce discriminant strategy, and define a multi-scale discriminant layer to distinguish pedestrian proposals category. Finally, we construct a multi-scale aware network, use the soft non-maximum suppression algorithm to fuse the output of classification score and regression offsets by each sensing network to generate final pedestrian detection regions. The experiments show that there is low detection error on the datasets Caltech and ETH, and the proposed algorithm is better than the current detection algorithms in terms of detection accuracy and works particularly well with far-scale pedestrians.

Key words machine vision; full convolutional network; deformable convolution layer; multi-scale discriminative strategy; multi-scale aware network; pedestrian detection

OCIS codes 150.0155; 100.2960; 100.4996; 100.5010

1 引 言

近年来,行人检测一直是机器视觉领域的研究热点,广泛应用于视频监控、智能车辆驾驶、机器人以及人机交互等。通过行人检测研究在图像或视频序列中界定所有行人的边界框,在复杂环境下检测出不同尺度和不同遮挡程度的行人是研究面临的主要挑战。当前诸多学者已经提出了很多有效的行人检测算法,可大致分为两类:基于手工提取特征的算法和基于深度学习提取特征的算法。前期手工提取特征的方法被广泛用于行人检测研究,最具影响力的方法是 Dollar 等[1-3]提出的积分通道特征(ICF)

收稿日期: 2018-03-23; **修回日期**: 2018-04-11; **录用日期**: 2018-04-17

基金项目: 教育部—中国移动科研基金(MCM20182019)

* **E-mail**: wjw8143@aliyun.com; ** **E-mail**: 1391570995@qq.com

算法及由该算法衍生的局部非相关通道特征(LDCF)和聚类通道特征(ACF)算法,结合了3种类型的通道特征,即LUV颜色通道、梯度幅度通道和方向梯度直方图(HOG)[4],参数设置较少,计算复杂度低,尤其在行人检测方面,取得了较好的检测结果。但是其稳健性较差,尤其在复杂环境下对行人远尺度和遮挡处理的检测效果差。文献[5]提出了一种基于在线高斯模型的行人检测候选框的快速生成方法,采用高斯模型拟合行人尺寸分布,获取场景中行人频繁出现的位置以及对应的目标尺度信息,但准确度不高。随着深度学习检测算法[6-8]在行人检测领域的不断应用,文献[6]通过行人属性分析和语义任务来优化行人检测,降低了误检率;文献[7]利用深度学习结合部分部位模型得到一个深度部位模型来解决行人检测中的遮挡问题;文献[8]在加速区域卷积神经网络(Faster-RCNN)[9]的基础上,采用加速区域卷积神经网络框架实现夜间红外图像中的行人检测,用区域建议网络生成候选区域;文献[10]基于边界框回归的方式,通过学习浅层的输入特征,与深层特征结合用于检测不同尺度的目标,但对远距离目标的检测精度太低。

在行人尺度检测的研究中,最新的多尺度检测算法[11-13]较之前算法已有明显进步。目前卷积神经网络(CNN)模型在检测远尺度行人方面的挑战主要在于:1)相比于近尺度的行人,通常远尺度的行人保留的信息少得多,而且会包含噪声,导致行人出现模糊的外观和边界;2)检测远尺度行人区域时,只有在视觉特征获得最佳响应的适当范围内才是最有效的,在包含不同尺度行人的复杂场景中更为明显。文献[11]使用基于快速区域全卷积网络(Fast RCNN)[14]的分治策略,该策略使用多个内置子网自适应检测跨越尺度的行人,将小尺度行人与大尺度行人相结合。类似地,文献[12]在不同层作独立预测,确保小目标训练在高分辨率层,大目标训练在低分辨率层,以匹配不同对象的尺度。但这种方法以高级语义特征为代价提供更高的分辨率,会损害网络性能。文献[13]为了提高图像分类正确率,提出一种二分支卷积单元。该卷积单元包含两种类型的滤波器,分别用于提取包含特征通道内与通道间信息的特征,可代替语义特征在图片分类中的辅助功能,减少对网络性能的损耗。

本文提出了一种在检测近尺度和远尺度行人方面性能较为均衡的方法。改进区域全卷积网络(R-FCN)[15]算法学习行人特征,在ResNet-50网络中引入可形变卷积,扩大特征图感受野,提取行人多尺度特征图;通过级联区域建议网络(RPN)提取多尺度行人建议,引进多尺度判别策略,输出远近尺度的行人感兴趣区域(RoI);最后,提出一个远/近尺度检测子网络,以更加均衡的方式实现跨尺度行人的协同检测,与SA-Fast RCNN[11]使用ACF[3]方法提取建议区域相比,本文算法通过改进网络结构的方式来提取行人区域的方法稳健性更好,准确率更高,并且在远尺度行人检测中取得了好的效果。

2 改进的全卷积多尺度感知网络

在R-FCN算法的基础上引进可形变卷积层,扩大特征图的感受野,提取多尺度行人特征图;通过级联RPN,提取多尺度行人RoI,定义一个尺度判别层,使用符号函数对行人RoI进行分类,分别输出到检测网络中;最后根据判别策略构建远尺度和近尺度两个子网络,通过软非极大值抑制(Soft-NMS)算法[16]输出最终的检测结果。网络结构如图1所示。

2.1 改进的全卷积网络

由于输入图片中会包含不同尺度的行人,而且在复杂环境下的远尺度行人很难通过传统方式清楚地判别出来,因此改进ResNet-50网络结构来提取远/近尺度行人的特征图。对于远尺度行人来说,较低的表示层可能有较强的卷积神经元激活,而近尺度的行人则在较高的表示层,有较强的卷积神经元激活。研究发现ResNet-50网络模型的C3~C5层,分别输出不同数量的256、512和1024的尺度特征图,C3~C5层中的空间像素位置分别大致对应于原始图像中32×32、64×64和128×128区域,因此小尺度的行人很容易被忽视。

考虑到Caltech数据集中行人有多种尺度的变化和复杂的遮挡情况,在res5a_branch2b层、res5b_branch2b层和res5c_branch2b层分别引入可形变卷积层和偏移层,卷积核大小为3×3,膨胀大小为2,步长为1,pad为2,形变卷积可增加模型的泛化能力,增大特征图的感受野,如图2所示。一个二维卷积包括两部分:1)在输入特征图 x 中使用规则网格 R 采样;2)通过 w 加权采样总和。传统的3×3、膨胀为1的卷积核结构为长方形,只可以获取固定大小的感受野,其中网格 $R=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\}$ 表示感受野的大小和膨胀,为典型的9点方格。加入偏移区域之后,改变了原卷积长方形的结构,变成9个指向不同方向的偏移点,增大了模型空间采样位置的额

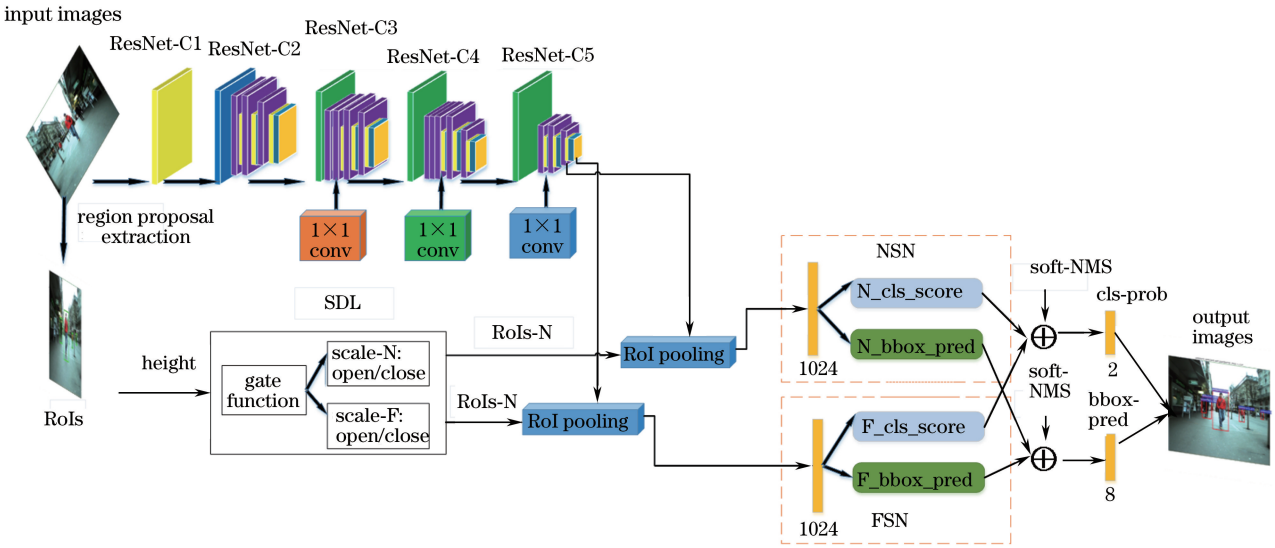

图 1 网络结构

Fig. 1 Network structure

外偏移量。传统的卷积输出特征图 $y(p_0)$ 为

$$y(p_0)=\sum_{p_n\in R}w(p_n)\cdot x(p_0+p_n),\tag{1}$$

式中：卷积窗口中的每个像素点 p_n 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，采样位置为 p_n。可变形卷积利用偏移量 Δp_n（$\{\Delta p_n\mid n=1,2,\cdots,N\}$）来增大规则网格 R 的感受野，其中 $N=|R|$ 表示网格中的像素个数，

$$y(p_0)=\sum_{p_n\in R}w(p_n)\cdot x(p_0+p_n+\Delta p_n),\tag{2}$$

采样偏移位置为 $p_n+\Delta p_n$。由于 Δp_n 为分数，所以(2)式通过双线性插值为

$$x(p)=\sum_{q}G(q,p)\cdot x(q),\tag{3}$$

式中：p 表示任意位置（$p=p_0+p_n+\Delta p_n$）；q 枚举特征图 x 中的所有整体空间位置；$G(q,p)$ 为双线性插值内核，G 函数可以分为两个一维内核的乘积：

$$G(p,q)=g(q_x,p_x)\cdot g(q_y,p_y),\tag{4}$$

式中：$g(x,y)=\max(0,1-|x-y|)$，可以快速计算 $G(q,p)$ 非零时的 q 值。

如图 2 所示，应用可形变卷积把原来的卷积过程分成两路，一个分支学习偏移量，得到 $H\times W\times 2N$ 的偏移输出，其中 $2N$ 表示 x，y 两个方向的偏移。另一个分支在获取偏移之后，在原始卷积的每一个卷积窗口增加偏移，使卷积窗口不再是原来规整的滑动窗口（图 2 中的绿框），而是经过偏移后的窗口（蓝框），增大了输出特征图的面积。可形变卷积引进在 res5 层，因此将该卷积块的有效步长从 32 pixel减少为 16 pixel，所有卷积滤波内核的膨胀从 1 变为 2，以增加特征图的分辨率和感受野。除增加形变卷积层外，还移除了 ResNet-50 模型的平均池化层，在 C3、C4、C5 层的最后增加一个随机初始化的 1×1 卷积，将最后输出通道尺度减少为 1024 维，实现每个卷积层的特征共享。保留了原始输入图像中的空间信息，RoI 池化层将每个 RoI 池化为固定长度的特征向量，该特征向量前向传播到全连接层。

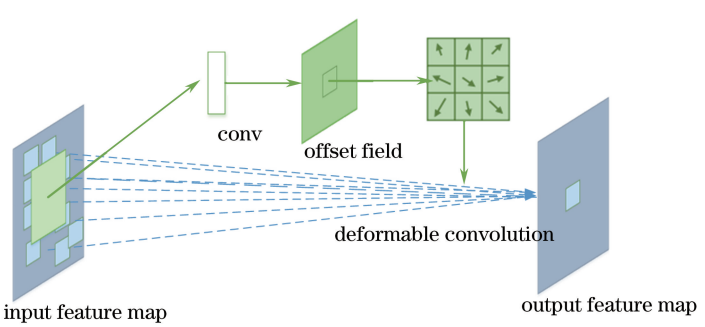

图 2 3×3 形变卷积示例

Fig. 2 Illustration of 3×3 deformable convolution

2.2 尺度判别层

改进 ResNet-50 网络的主要目的是为了提取多尺度行人特征图。考虑到一张图片包含多个不同尺度的行人，引进级联 RPN 提取区域建议，然后输入到尺度判别层(SDL)，判断当前边界框的类别属于近尺度还是远尺度，如图 3 所示。可通过边界框像素的高度或宽度计算行人建议尺度。当行人站在摄像机前时，边界框的像素高度与行人尺度的相关性比像素宽度更高，因此选择行人像素的高度作为划分行人尺度的依据。为了提取不同高度的行人 RoI，定义近尺度 RPN(RPN-N)和远尺度 RPN(RPN-F)，级联后提取不同尺度的 RoI，如图 3 所示。将锚点的尺度改为 32，64，128，256，512，比例为 1∶2，1∶1，2∶1，生成 15 个锚点，用于提取原图片中不同尺度的 RoI。如图 3 所示，通过滑动窗口的方式，RPN-N 获取行人的区域建议(reg-N)和分类得分(cls-N)，然后将 reg-N 作为 RPN-F 的输入，RPN-F 通过 reg-N 的区域坐标值在原图片直接获取行人 RoI，得到最终的得分值(cls-F)和区域建议(reg-F)。对于不同尺度的行人 RoI，通过级联 RPN 的方式将 RPN-N 中输出的 reg-N 作为 RPN-F 的输入，代替了 RPN-F 中锚点的结构，直接计算出最终的区域建议(reg-F)，与滑动窗口相比可以在很大程度上节省计算时间，而且比单个 RPN 提取行人建议的效果好。然后对每个行人 RoI 的高度进行判断，根据级联 RPN 获取的图片输入信息，定义一个尺度判别层，如图 1 所示。

行人建议 j 的边界框高度为 $H_j = y_{j\max} - y_{j\min}$。在当前第 m 层($m=3,4,5$)的行人建议的平均高度 $\bar{H}_m = \sum_{i=1}^{N}(y_{i\max} - y_{i\min})/N$（$N$ 表示当前层训练样本数量，i 表示一个行人边界框，y 表示获取到行人标注框的纵坐标)。定义一个符号函数 $\varepsilon(x)$，尺度判别层工作流程如图 1 所示，

$$S_n = \varepsilon\left(\frac{H_j}{\bar{H}_m} - 1\right), \tag{5}$$

$$S_f = \varepsilon\left(1 - \frac{H_j}{\bar{H}_m}\right), \tag{6}$$

式中：S_n 表示近尺度网络，S_f 表示远尺度网络。符号函数 $\varepsilon(x)$ 表示为

$$\varepsilon(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0 \end{cases}。 \tag{7}$$

如果输入图片中包含的行人 RoI 判定为远尺度行人，则激活 S_f，将 RoI-F 池化后传播到远尺度网络中检测；若判定为近尺度行人，则激活 S_n，将 RoI-N 池化后传播到近尺度网络中训练。由于每张图片会包含多个行人区域，为了减少计算时间，根据当前的判别机制，构建了一个多尺度感知网络(MSAN)，如图 3 所示。

2.3 多尺度感知网络

SDL 层分别输出不同尺度的 RoI-N 和 RoI-F，为了分别检测不同尺度的 RoI，构建一个多尺度感知网络。MSAN 的网络结构主要由近尺度检测网络(NSN)和远尺度检测网络(FSN)组成，如图 1 所示。每个子网络都有两个输出层，对于每个目标建议 i，第一个输出层输出离散置信值分布 $s_i = (s_0, s_1)$，另一层输出每个边界框回归偏移值 $t_i = (t_x, t_y, t_w, t_h)$。其中，$t_i$ 指定相对于对象提议的原始位置和大小的位置转换和边界框偏移值，t_x，t_y，t_w，t_h 分别表示边界框的坐标和宽度与高度。由于只有行人和背景两类目标，所以最终的分类输出值为边界框为行人的概率和位置信息。引进 Soft-NMS 算法抑制生成框中的冗余信息。传统的 NMS 算法和 Soft-NMS 算法可表示为

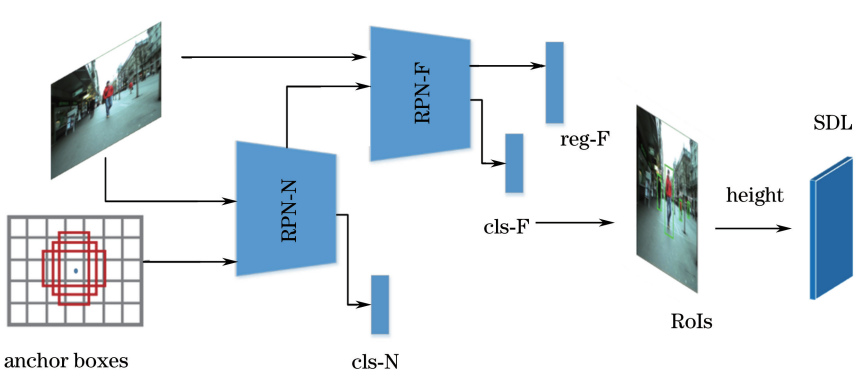

图 3 SDL 流程图

Fig. 3 SDL flow chart

$$s_i=\begin{cases}s_i, & \mathrm{IoU}(M,b_i)<N_t\\ 0, & \mathrm{IoU}(M,b_i)\geqslant N_t\end{cases},\tag{8}$$

$$s_i=\begin{cases}s_i, & \mathrm{IoU}(M,b_i)<N_t\\ s_i\,[1-\mathrm{IoU}(M,b_i)], & \mathrm{IoU}(M,b_i)\geqslant N_t\end{cases}。\tag{9}$$

当重合面积交并比(IoU)小于阈值 N_t 时,检测得分值为 s_i;IoU 大于阈值 N_t 时,得分值为 0。该过程被不断递归地应用于其余检测框。根据算法设计,如果一个物体处于预设的重叠阈值之内,可能会检测不到该待检测物体。Soft-NMS 算法对非最大检测框的检测分数发生衰减而非彻底移除。IoU 大于阈值 N_t 后,得分值为 $s_i\,[1-\mathrm{IoU}(M,b_i)]$。对传统的 NMS 算法进行简单的改动,而且不增加额外参数,可以提高检测精度和检测速度。

每个训练的行人建议都有一个真实的类别 g 和一个真实边界框回归目标值 t^*。多任务在每个目标建议 i 处,损失函数 L 用来联合输出两个检测子网络的参数:

$$L=L_{\mathrm{cls}}(s_i,g)+1[g\geqslant 1]L_{\mathrm{loc}}(t_i^g,t^*),\tag{10}$$

式中:L_{cls} 和 L_{loc} 分别为分类和边界框回归损失函数。L_{cls} 为 Softmax 损失函数,输出分类的准确值;L_{loc} 为 Smooth L_1 损失函数,输出边界框坐标、宽度和高度。$1[g\geqslant 1]$ 表示分类值 $g\geqslant 1$ 时为 1,否则为 0。通过联合两个子检测网络(NSN 和 FSN)的输出,使用 Soft-NMS 算法输出检测结果,本文算法在不同范围的输入尺度下,可以准确输出检测结果。

2.4 算法训练检测步骤

1) 改进 ResNet-50 作为基本的模型,在 res5a_branch2b 层、res5b_branch2b 层和 res5c_branch2b 层分别引进可形变卷积,增加特征图的感受野,提高模型的泛化能力,提取不同尺度的行人特征图。

2) 使用不同锚点的尺度训练级联 RPN。假设锚点分别在 IoU 大于 0.7、0.6、0.5 时为正样本,分别进行训练,实验发现 IoU 大于 0.5 时,70%左右的真实框与级联 RPN 有一个锚点,因此将 IoU 大于 0.5 的锚点设置为正样本。

3) 将提取到的多尺度行人 RoI 前向传播到尺度判别层进行分类,分别输出两类不同的 RoI-N 和 RoI-F。

4) 将输出的 RoI-N 和 RoI-F 传播到 RoI 池化层,池化为 1024 维的特征向量输入 MSAN 中,引进 Soft-NMS 算法输出最终的检测结果。将阈值设定为 0.5,融合远近尺度检测网络输出的结果。

3 结果与分析

实验平台是 64 位的 Ubuntu16.04 操作系统和 NVIDIA GTX 1070 GPU,采用的软件有 MATLAB 2014a、Python2.7,采用的深度学习框架为 Caffe。在不失一般性的情况下,改进 ResNet-50 在 ImageNet 预先训练的模型,使用随机梯度下降(SGD)进行训练,学习率为 0.001,动量大小为 0.9,权重衰减设置为 0.0005。同时当训练损失达到一个单位时,学习率降低 10%。一张随机选择的图片中,每个 mini-batch 包括 128 个随机采样行人建议,包括 32 个正样本和 96 个负样本。在成绩评估过程中,如果检测框与真实标注框重合面积大于某个阈值,则认为该检测框与真实标注框相匹配。实验表明重合面积大于 0.5 时,行人边界框预测为正值,可检测行人最多,IoU 值小于 0.3 时为负值。IoU 计算公式为

$$\mathrm{IoU}=\frac{\mathrm{area}(B_{\mathrm{dt}}\cap B_{\mathrm{gt}})}{\mathrm{area}(B_{\mathrm{dt}}\cup B_{\mathrm{gt}})}>0.5,\tag{11}$$

式中:B_{dt} 表示最终检测框;B_{gt} 表示真实标注框。若多个 B_{dt} 与 B_{gt} 匹配,则决策得分高的检测框将被选择,没有被匹配的 B_{dt} 记为误检,未被匹配的 B_{gt} 记为漏检。

根据 Dollar 等[1]提出的工具箱为标准对实验结果进行评价,用每幅图像的误检率和漏检率衡量滑动窗口行人检测算法的性能。切割测试图片中包含行人的窗口,从不包含行人的测试集中采集非行人样本,最后将窗口作为测试集来评估算法的性能。每幅图像的误检率和漏检率表达式为

$$f_{\mathrm{fppi}}=\frac{f_{\mathrm{FP}}}{f_{\mathrm{TN}}+f_{\mathrm{FP}}}\times 100\%,\tag{12}$$

$$R_{\mathrm{M}}=\frac{f_{\mathrm{FN}}}{f_{\mathrm{FN}}+f_{\mathrm{TP}}}\times 100\%,\tag{13}$$

式中:f_{TP}、f_{FP}、f_{FN} 和 f_{TN} 分别表示将行人样本分类为行人样本的数量、将非行人样本分类为行人样本的数量、将行人样本分类为非行人样本的数量及将非行人样本分类为非行人样本的数量。R_{M} 值越低,行人检测算法性能越好。

3.1 Caltech 实验结果比较

使用 Caltech 行人数据库进行实验,该数据库采用车载摄像头拍摄,视频总长度大约为 10 h,共包含 2300 个独立的行人。数据集分为 11 个集合:set00~set05 为训练集,set06~set10 为测试集。图片分辨率为 640×480,将边界框高度大于 50 的

行人作为近尺度行人，高度低于 50 pixel 的行人作为远尺度行人。在评估指标方面，采用漏检率总结探测器的性能，选择最新的 CNN 算法 SDN[17]、DeepParts[7]、JointDeep[18]、TA-CNN[6]、RPN＋BF[19]、CCF[20]、MS-CNN[12]、F-DNN[21]、SA-Fast RCNN[11]与本文算法 D-FCN＋MSAN 进行对比。

图 4(a)所示为近尺度检测结果，本文算法 D-FCN＋MSAN 与最新的 F-DNN 算法的漏检率相同，为 7.10%，比 SA-Fast RCNN 算法低 0.6%。对于远距离行人[图 4(b)]，本文算法达到了最低漏检率 45.61%，比 SA-Fast RCNN 算法的 79.80%和 F-DNN 算法的 55.40%降低了 34.39%和 9.79%。还测试了 R-FCN 算法和加入形变卷积之后的 R-FCN (D-FCN)和 SA-Fast RCNN 算法的检测精度和速度，并与本文算法性能作对比，结果如表 1 所示。可以看出 R-FCN 算法的多尺度检测效果比本文算法好，引进形变卷积后精度提升了 6%左右，检测速度变慢。本文算法 D-FCN＋MSAN 在训练集上迭代 20 万次生成最终的检测模型，在测试集上检测了 4024 幅图像，共耗时 12 min，检测每幅图像平均耗时 0.18 s，比 D-FCN 算法慢 0.05 s，主要原因在于加入 MSAN 网络后，网络结构更为复杂，导致训练和检测速度下降，但远尺度检测精度提高了 10%左右，性能远优于 SA-Fast RCNN 算法。综合考虑，本文算法 D-FCN＋MSAN 在 Caltech 数据集上的检测准确率优于现有其他算法。

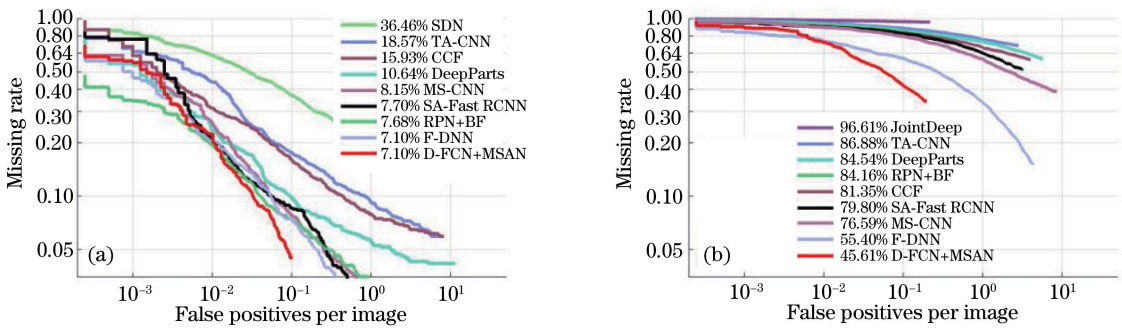

图 4 Caltech 数据集的结果比较。(a)近尺度；(b)远尺度

Fig. 4 Comparison of results on the Caltech benchmark. (a) Near-scale; (b) far-scale

表 1 漏检率与检测速度

Table 1 Missing rate and detection rate

Algorithm	R-FCN	D-FCN	D-FCN＋MSAN	SA-Fast RCNN
R_M(N-scale) /%	18.09	12.46	7.10	7.70
R_M(F-scale) /%	60.33	54.36	5.61	79.80
Time consumption per image /s	0.11	0.13	0.18	0.32

3.2 ETH 数据实验比较

为了验证本文算法的适用性，用 ETH 数据集进行检测。ETH 数据集中包含行人的环境比较复杂，而且互相遮挡严重。选择几个经典的算法用于对比，包括 ACF[3]、LDCF[2]、SDN[17]、RPN＋BF[21]、TA-CNN[6]。在图 5(a)所示近尺度检测中，D-FCN＋MSAN 算法与最新的 CNN 算法 TA-CNN 相比漏检率下降了 2.36%；在图 5(b)所示远尺度检测中，与 TA-CNN 算法相比漏检率下降了 9.14%。

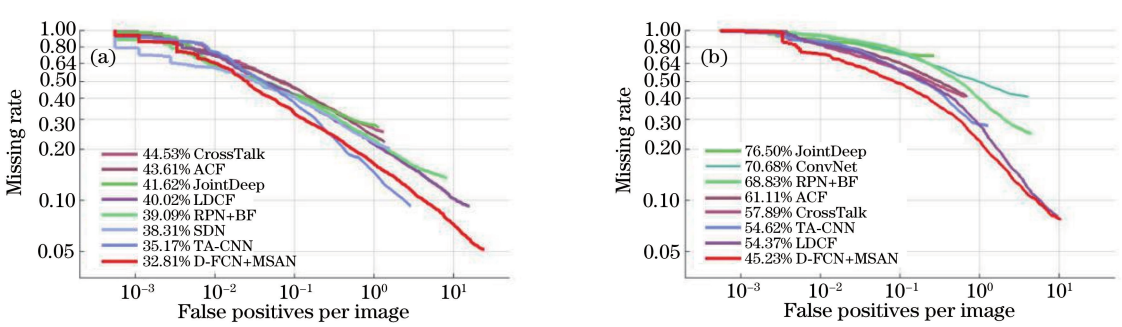

图 5 ETH 数据集的结果比较。(a)近尺度；(b)远尺度

Fig. 5 Comparison of results on the ETH benchmark. (a) Near-scale; (b) far-scale

4 结 论

提出的多尺度感知行人检测算法主要是为了解决辅助车辆驾驶系统中,复杂场景下的远尺度行人检测精度较低的问题。使用级联 RPN 提取多尺度行人建议,引进多尺度判别策略,将不同尺度的行人建议进行分类处理。本文算法在两个广泛应用的数据集 Caltech 和 ETH 上取得了很好的检测精度,但检测速度有待提高,如何提高检测速度将是下一步研究的主要方向。

参 考 文 献

[1] Dollár P, Tu Z, Perona P, *et al*. Integral channel features[C] // British Machine Vision Conference, 2009: 7-10.

[2] Nam W, Dollár P, Han J H. Local decorrelation for improved pedestrian detection[C] // Advances in Neural Information Processing Systems, 2014: 424-432.

[3] Dollár P, Appel R, Belongie S, *et al*. Fast feature pyramids for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1532-1545.

[4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2005: 886-893.

[5] Qin J, Wang M H. Fast pedestrian proposal generation algorithm using online Gaussian model [J]. Acta Optica Sinica, 2016, 36(11): 1115001.
覃剑, 王美华. 采用在线高斯模型的行人检测候选框快速生成方法[J]. 光学学报, 2016, 36(11): 1115001.

[6] Tian Y, Luo P, Wang X, *et al*. Pedestrian detection aided by deep learning semantic tasks[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5079-5087.

[7] Tian Y, Luo P, Wang X, *et al*. Deep learning strong parts for pedestrian detection[C] // IEEE International Conference on Computer Vision, 2015: 1904-1912.

[8] Ye G L, Sun S Y, Gao K J, *et al*. Nighttime pedestrian detection based on faster region convolution neural network[J]. Laser & Optoelectronics Progress, 2017, 54(8): 081003.
叶国林, 孙韶媛, 高凯珺, 等. 基于加速区域卷积神经网络的夜间行人检测研究[J]. 激光与光电子学进展, 2017, 54(8): 081003.

[9] 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, 2015: 91-99.

[10] Liu W, Anguelov D, Erhan D, *et al*. SSD: single shot multibox detector[C] // European Conference on Computer Vision, 2016: 21-37.

[11] Li J, Liang X, Shen S M, *et al*. Scale-aware fast R-CNN for pedestrian detection[J]. IEEE Transactions on Multimedia, 2017, 20(4): 985-996.

[12] Cai Z, Fan Q, Feris R S, *et al*. A unified multi-scale deep convolutional neural network for fast object detection[C] // European Conference on Computer Vision, 2016: 354-370.

[13] Hou C C, He Y Q, Jiang X H, *et al*. Deep convolutional neural network based on two-stream convolutional unit[J]. Laser & Optoelectronics Progress, 2018, 55(2): 021005.
侯聪聪, 何宇清, 姜晓恒, 等. 基于二分支卷积单元的深度卷积神经网络[J]. 激光与光电子学进展, 2018, 55(2): 021005.

[14] Girshick R. Fast R-CNN[C] // IEEE International Conference on Computer Vision, 2015: 1440-1448.

[15] Dai J, Li Y, He K, *et al*. R-FCN: object detection via region-based fully convolutional networks[C] // Advances in Neural Information Processing Systems, 2016: 379-387.

[16] Bodla N, Singh B, Chellappa R, *et al*. Soft-NMS-improving object detection with one line of code[C] // IEEE International Conference on Computer Vision, 2017.

[17] Luo P, Tian Y, Wang X, *et al*. Switchable deep network for pedestrian detection[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 899-906.

[18] Ouyang W, Wang X. Joint deep learning for pedestrian detection[C] // IEEE International Conference on Computer Vision, 2014: 2056-2063.

[19] Zhang L, Lin L, Liang X, *et al*. Is faster R-CNN doing well for pedestrian detection?[C] // European Conference on Computer Vision, 2016: 443-457.

[20] Yang B, Yan J, Lei Z, *et al*. Convolutional channel features[C] // IEEE International Conference on Computer Vision, 2015: 82-90.

[21] Du X, El-Khamy M, Lee J, *et al*. Fused DNN: a deep neural network fusion approach to fast and robust pedestrian detection[C] // IEEE Winter Conference on Applications of Computer Vision, 2017: 953-961.