

多尺度空间金字塔池化 PCANet 的行人检测

夏胡云, 叶学义, 罗宵晗, 王 鹏

(杭州电子科技大学 模式识别与信息安全实验室, 杭州 310018)

摘 要: 针对非理想条件下行人检测的性能和效率问题, 提出多尺度空间金字塔 PCANet。将空间金字塔作为网络的特征池化层, 通过分层池化特征的方式获得图像的显著性特征, 并将底层特征和高层特征级联以获得样本的多尺度特征的向量表示, 输入 SVM 分类器。在 INRIA 和 NICTA 数据库中, 与 HOG、CNN 等算法进行行人检测对比实验, 结果表明, 该算法有更高的正确检测率、更低的漏检率和误检率。

关键词: 行人检测; 深度学习架构; 主成分分析网络; 多尺度特征; 空间金字塔池化; 显著性特征

中文引用格式: 夏胡云, 叶学义, 罗宵晗, 等. 多尺度空间金字塔池化 PCANet 的行人检测[J]. 计算机工程, 2019, 45(2): 270-277.

英文引用格式: XIA Huyun, YE Xueyi, LUO Xiaohan, et al. Pedestrian detection using multi-scale principal component analysis network of spatial pyramid pooling[J]. Computer Engineering, 2019, 45(2): 270-277.

Pedestrian Detection Using Multi-scale Principal Component Analysis Network of Spatial Pyramid Pooling

XIA Huyun, YE Xueyi, LUO Xiaohan, WANG Peng

(Lab of Pattern Recognition and Information Security, Hangzhou Dianzi University, Hangzhou 310018, China)

[Abstract] Pedestrian detection is easily affected by non-ideal factors such as complex background, shooting angle and diversity of human body posture in natural environment. To solve this problem, this paper proposes a Multi-scale Principal Component Analysis Network of Spatial Pyramid Pooling (MS-PCANet-SPP). The feature pooling layer using spatial pyramid pooling method extract the saliency features of the image. The multi-scale features of the input samples can be obtained by cascading the high-level and low-level features, which is input to the SVM classifier. The comparative experiments are performed in the INRIA and NICTA databases. Experimental results show that, compared with HOG, CNN and other algorithms, MS-PCANet-SPP has a higher detection rate, a lower miss rate, and a lower false positive rate.

[Key words] pedestrian detection; deep learning framework; Principal Component Analysis Network (PCANet); multi-scale feature; spatial pyramid pooling; saliency feature

DOI: 10.19678/j.issn.1000-3428.0049831

0 概述

行人检测一直是计算机视觉和模式识别的研究热点和难点。目前行人检测基于统计分类的方法, 通过特征提取和分类与定位^[1], 将行人检测转化成人非人的二分类问题。从分类采用的特征来看, 目前行人检测的研究大致可以分为基于底层特征、基于混合特征、基于机器学习的特征 3 类方法。基于底层特征的方法利用梯度^[2]或者纹理^[3]等特征实现行人检测。该类方法只从单一角度描述行人特征, 容易受复杂背景、遮挡和光照等的影响。基于混

合特征^[4]的方法利用各种不同特征从不同侧面刻画行人特征, 提高检测准确率。该方法会增加特征维度, 导致检测效率降低, 影响检测的实时性。基于机器学习的特征^[5]的方法利用学习网络, 采用统计分析的策略从大量样本中选择出判别能力较强的特征。该方法特征的选择与训练样本密切相关, 若样本集不具有代表性很难选择出好的特征。

上述研究都试图将图像中行人区域的数据信息映射成语义信息, 其难点是对在非理想的自然环境下获得的图像实现稳定的映射。文献[6]提出深度学习的概念, 通过深度结构模拟人脑分层次学习的

基金项目: 国家自然科学基金(60802047, 60702018)。

作者简介: 夏胡云(1994—), 男, 硕士研究生, 主研方向为深度学习、计算机视觉; 叶学义, 副教授、博士; 罗宵晗、王 鹏, 硕士研究生。

收稿日期: 2017-12-25 **修回日期:** 2018-01-30 **E-mail:** xiahuyun043@126.com

机制,构建具有多隐层的机器学习模型来自下而上地学习输入数据的高层语义信息^[7]。相比于传统的机器学习方法,其更能够刻画数据的丰富内在信息。文献[8]提出一种主成分分析网络(Principal Component Analysis Network, PCANet),由主成分分析(Principal Component Analysis, PCA)、哈希运算及局部直方图3个基本操作构成。PCANet已经表现出优异的性能,成为图像分类、识别领域中的一个研究热点^[9-11]。

由于空间金字塔池化对形变目标具有较强鲁棒性^[12]以及多尺度卷积神经网络(Convolutional Neural Network, CNN)^[13]具有较高的分辨率,本文在典型PCANet的基础上,提出一种多尺度空间金字塔池化PCANet(Multi-scale Principal Component Analysis Network of Spatial Pyramid Pooling, MS-PCANet-SPP)。引入空间金字塔池化,运用最大值池化方法逐层构建图像的描述子^[14],在整合并优化原始池化特征的同时,捕获图像的结构化信息。将经空间金字塔池化得到的特征向量级联以提取图像

的多尺度特征。

1 PCANet

PCANet^[8]是一种深度特征提取网络,由PCA卷积层、非线性处理层和特征池化层组成,如图1所示。其中, I_i 为输入第*i*个样本, $W_l^1(l=1,2,\dots,L_1)$ 为PCA卷积层第1层的第*l*个卷积核, $W_\lambda^2(\lambda=1,2,\dots,L_2)$ 为PCA卷积层第2层的第*λ*个卷积核, L_1 和 L_2 分别为第1层和第2层的卷积核数目, $I_l^1(l=1,2,\dots,L_1)$ 为 I_i 与 W_l^1 卷积的第*l*个输出, $O_l^1(l=1,2,\dots,L_1)$ 为 I_l^1 与所有 W_λ^2 卷积得到的一组输出,每组输出 O_l^1 的数目为 L_2 , $T_l^1(l=1,2,\dots,L_1)$ 为对 O_l^1 中每组输出进行二值化加权得到的一幅“整型图”, f_i 为对所有 T_l^1 进行局部直方图并且级联得到的网络的特征向量。在图1中,符号*表示卷积,红色块表示卷积核与该块进行卷积,得到的卷积结果用黄色块表示(彩色见电子版)。给定*N*个不同的训练样本 $\{I_i\}_{i=1}^N$,其中样本大小为 $m \times n$,下文将对PCANet进行详细描述。

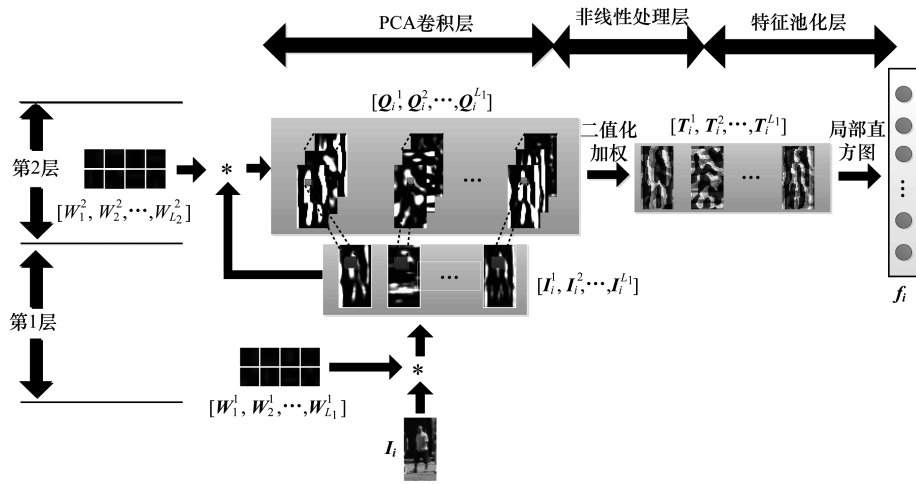


图1 PCANet结构示意图

1.1 PCA 卷积层

PCANet采用2层卷积层提取输入样本的特征。文献[8]表明:当卷积层层数低于2时,分类、识别的效果不好;当卷积层层数高于2时,维度的增加导致计算量急剧增加;卷积层层数等于2是比较合适的。将输入样本 I_i 与第1层卷积核卷积,得到 L_1 个卷积输出:

$$I_l^1 = I_i * W_l^1, l=1,2,\dots,L_1 \quad (1)$$

每一个卷积输出 I_l^1 与第2层的所有卷积核 W_λ^2 卷积得到 L_2 个输出 O_l^1 ,则在第2层, L_1 个输入 I_l^1 将得到 $L_1 L_2$ 个输出:

$$O_l^1 = \{I_l^1 * W_\lambda^2\}_{\lambda=1}^{L_2}, l=1,2,\dots,L_1 \quad (2)$$

1.2 非线性处理层

为加强样本特征表达性^[15],对图1中第2层的每个输出 O_l^1 进行非线性处理,使用赫维赛德阶跃函数 $H(x)$ 作为非线性处理函数二值化输出 O_l^1 ,将 O_l^1 中像素值大于0的置为1,小于0的置为0。对二值化后的输出进行加权处理,加权系数如下:

$$T_l^1 = \sum_{\lambda=1}^{L_2} 2^{l-1} H(I_l^1 * W_\lambda^2), l=1,2,\dots,L_1 \quad (3)$$

通过二值化加权处理使得 O_l^1 中的每 L_2 个输出得到一幅“整型图”。其中,“整型图”表示输出每个像素值的范围是 $[0, 2^{L_2} - 1]$ 之间的整数。因此,每个输入样本 I_i 将得到 L_1 个“整型图”。

1.3 特征池化层

由于上述的非线性处理层得到的像素值为 $[0, 2^{L_2-1}]$ 之间的整数, 而 CNN 中非线性处理层为前一层卷积层进行最大值或均值池化得到的像素值, 因此 CNN 的池化方式不适用于 PCANet^[8]。文献[8]用局部直方图作为 PCANet 的特征池化层操作, 将每一个输出“整型图” T_i^l 分为 B 块, 对每块进行直方图统计, 并将 B 个直方图统计结果按列向量化, 记作 $Bhist(T_i^l)$ 。级联 L_1 个“整型图”所产生的向量得到输入样本 I_i 的特征向量表示:

$$f_i = [Bhist(T_i^1), Bhist(T_i^2), \dots, Bhist(T_i^{L_1})]^T \in \mathbb{R}^{(2^{L_2})L_1B} \quad (4)$$

1.4 预训练卷积核

通过 PCA 预训练的方式获得 PCANet 中卷积核的值, 预训练的步骤为:

步骤 1 对第 i 个输入样本 I_i , 以 $k_1 \times k_2$ 大小的块在样本上滑动截取得到 $c = (m - k_1 + 1) \times (n - k_2 + 1)$ 个块, 将每个块向量化得到 $x_{i,1}, x_{i,2}, \dots, x_{i,c} \in \mathbb{R}^{k_1 k_2}$, 其中, $x_{i,j}$ 表示第 i 个样本 I_i 的第 j 个块向量化后的向量。

步骤 2 对 $x_{i,j}$ 进行去均值级联得到 $\bar{X}_i = [\bar{x}_{i,1}, \bar{x}_{i,2}, \dots, \bar{x}_{i,c}]$, 其中, \bar{X}_i 表示第 i 个输入样本预处理后的结果。

步骤 3 对输入所有 N 个样本进行相同处理并且级联得到:

$$X = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N] \in \mathbb{R}^{k_1 k_2 \times Nc} \quad (5)$$

步骤 4 使用 PCA 算法对样本集 X 进行计算, PCA 最小化重构误差的过程表达为:

$$\min_{V \in \mathbb{R}^{k_1 k_2 \times L_1}} \|X - VV^T X\|_F^2, \text{ s. t. } VV^T = I_{L_1} \quad (6)$$

其中, I_{L_1} 为 $L_1 \times L_1$ 的单位矩阵, V 为满足 PCA 最小化重构误差的标准正交基。由 PCA 算法可知, 要解出 PCA 卷积层第 1 层卷积核 W_l^1 首先需要计算 X 的协方差矩阵 C :

$$C = \frac{1}{Nc} XX^T \quad (7)$$

$$w_l^1 = q_{L_1}(V) \in \mathbb{R}^{k_1 k_2}, l = 1, 2, \dots, L_1 \quad (8)$$

$$W_l^1 = \text{mat}_{k_1, k_2}(w_l^1) \in \mathbb{R}^{k_1 \times k_2}, l = 1, 2, \dots, L_1 \quad (9)$$

计算协方差矩阵 C 对应的特征向量, 由 PCA 算法可知该特征向量即为所求的标准正交基 V 。式(8)中 $q_{L_1}(V)$ 表示提取 V 的前 L_1 个主成分分量, 将每个主成分分量记作 $w_l^1 (l = 1, 2, \dots, L_1)$, 式(9)将 $w_l^1 \in \mathbb{R}^{k_1 k_2}$ 投影为矩阵 $W_l^1 \in \mathbb{R}^{k_1 \times k_2}$, 得到 L_1 个矩阵 W_l^1 即为 PCA 卷积层第 1 层卷积核。

步骤 5 对 PCA 卷积层第 1 层的所有输出 I_i^l 滑

动截取、向量化、去均值得到 $\bar{Y}_i^l = [\bar{y}_{i,l,1}, \bar{y}_{i,l,2}, \dots, \bar{y}_{i,l,c}] \in \mathbb{R}^{k_1 k_2 \times c}$, 则第 l 个卷积核输出结果去均值化的块表示为 $Y^l = [\bar{Y}_1^l, \bar{Y}_2^l, \dots, \bar{Y}_N^l] \in \mathbb{R}^{k_1 k_2 \times Nc}$, 级联所有卷积核的输出块 $Y^l (l = 1, 2, \dots, L_1)$ 得到:

$$Y = [Y^1, Y^2, \dots, Y^{L_1}] \in \mathbb{R}^{k_1 k_2 \times L_1 Nc} \quad (10)$$

步骤 6 重复步骤 1 ~ 步骤 5, 利用 PCA 算法对 Y 进行计算, 得到 PCA 卷积层第 2 层的卷积核:

$$W_\lambda^2 = \text{mat}_{k_1, k_2}(q_\lambda(YY^T)) \in \mathbb{R}^{k_1 \times k_2}, \lambda = 1, 2, \dots, L_2 \quad (11)$$

文献[8]通过 PCA 卷积层、非线性处理层和特征池化层建立输入数据从底层特征到高层特征的映射, 提取输入样本的高层语义特征。但是, 只提取高层语义特征作为最终特征向量表示的方式忽视了数据的底层特征, 而底层特征包含更多的局部细节信息。文献[13]表明结合底层特征的局部细节信息和高层特征的抽象语义信息的样本多尺度的特征向量表示可有效减少由于多次卷积和池化操作导致的信息损失。同时, 文献[12]通过空间金字塔最大值池化方法对原始冗余特征进行提取, 对提取的特征向量进行聚合, 将不相关信息丢弃后保留图像中重要的信息, 使得网络具有更好的抗噪性和鲁棒性。

2 MS-PCANet-SPP

基于上述分析, 本文提出 MS-PCANet-SPP, 由 PCA 卷积层、非线性处理层和特征池化层组成, 其结构如图 2 所示。类似于 PCANet, 卷积层数为 2, I_i 为输入样本, $W_{l_1}^1$ 和 $W_{l_2}^2$ 分别为 PCA 卷积层第 1 层和第 2 层的卷积核, $O_i^1(l_1)$ 和 $O_i^2(l_2)$ 分别为 PCA 卷积层第 1 层和第 2 层的卷积输出。其中, 黑色块表示对该块进行去均值化, 得到的结果用红色块表示; 在每一层对红色块进行卷积, 卷积的结果用黄色块表示(彩色见电子版)。在非线性处理层, 对 $O_i^1(l_1)$ 进行非线性处理得到的“整型图” $T_i^1(l_1)$, 对 $O_i^2(l_2)$ 进行非线性处理得到的“整型图” $T_i^2(l_2)$ 。在特征池化层引入空间金字塔池化策略逐层构建图像的特征描述子, $f_{i,1}$ 和 $f_{i,2}$ 为每层“整型图”经过空间金字塔池化得到的特征向量, 将经过空间金字塔池化得到的不同尺度的特征向量级联以获得图像的多尺度特征, 即 $f_i = [f_{i,1}; f_{i,2}]$ 为级联 $f_{i,1}$ 和 $f_{i,2}$ 得到的特征向量。其中, $f_{i,1}$ 为 f_i 的低维表示, 因此图 2 对向量进行倒置以表示 $f_{i,1}$ 和 $f_{i,2}$ 的这种级联关系。将 $f_i = [f_{i,1}; f_{i,2}]$ 送入支持向量机(Support Vector Machine, SVM)进行判断: 输入样本图像中是否有行人。

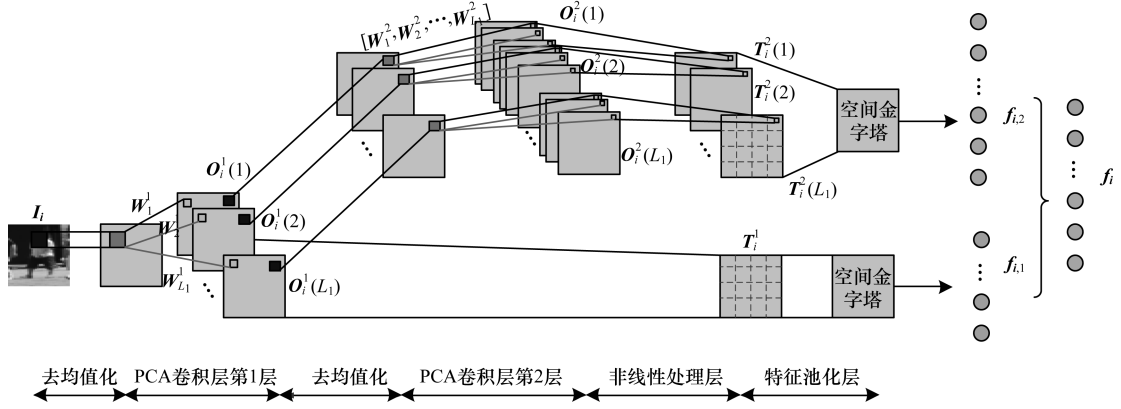


图 2 MS-PCANet-SPP 结构示意图

2.1 多尺度特征及空间金字塔池化

本文描述的多尺度特征表示的是输入样本从底层特征到高层特征的多尺度关系,而空间金字塔表示的是空间像素级的多层次关系。如图 3 所示, U 为局部直方图提取的特征向量集合, $l \in L$ 表示空间第 l 层级的划分, L 是大于等于 0 的整数, 每层会有 2^l 个子集, 即在横坐标以及纵坐标方向都被划分为 2^l 个网格, 得到的图像块个数为 $2^l \times 2^l = 4^l$ 个。分别在 L 层级中, 对特征向量运用池化算法, 将池化的函数表示为:

$$Z = F(U) \quad (12)$$

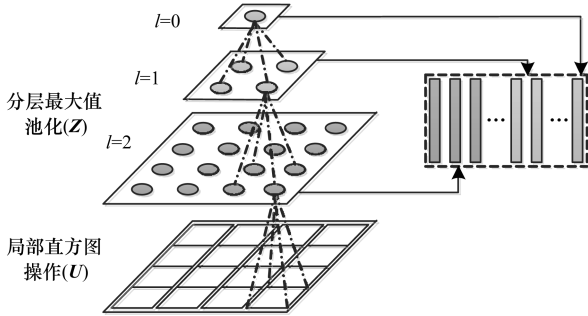


图 3 空间金字塔池化示意图

由于均值池化容易丢失图像中目标的细节信息^[14], 本文采用最大值函数作为空间函数。将从图像提取的特征向量对应的维度最大值表示图像对应的响应程度, 计算局部图像块编码的最大值的公式如下:

$$z_j = \max \{ |u_{1j}|, |u_{2j}|, \dots, |u_{mj}| \} \quad (13)$$

其中, z_j 为 Z 的第 j 个元素, u_{mj} 为矩阵 U 第 m 行第 j 列的元素, m 为特征向量的个数。分别对金字塔每个层级的每个区域的特征使用最大值池化函数, 将所有的层级池化后的向量组合, 完成高层局部特征的表示。在局部直方图基础上引入空间金字塔的优点在于, 空间金字塔采用最大值池化函数对局部噪声有较强鲁棒性, 同时将图像特征向量在池化运算时用空间金字塔的形象表示, 可对图像的特征赋予位置信息, 使分类更加精确。随着划分的层级增加, 区域的密集性也会增加, 相应的向量维数将越来越大, 导致运算复杂度增加。实验结果表明, 当

$L \geq 3$ 时计算复杂度较大且准确率并没有明显提升。因此, 本文使用 $L = 2$ 的空间金字塔模型, 即图像被划分为 3 个层级 ($l = 0, 1, 2$), 将 3 个层级池化后的特征向量拼接作为高层特征的特征向量表示 f_2 。

为引入多尺度策略, 在 PCA 卷积层对第 1 层输出进行二值化加权和空间金字塔池化操作提取输入样本底层特征 f_1 , 将底层特征 f_1 与高层特征 f_2 级联, 使 MS-PCANet-SPP 既能提取到高层语义信息又能获得底层细节信息, 从而获得输入样本的多尺度特征表示。

2.2 算法描述

MS-PCANet-SPP 算法执行步骤如下:

输入 N 张大小为 $m \times n$ 的行人图片样本集

$$\{I_i \in \mathbb{R}^{m \times n} | 1 \leq i \leq N\}$$

输出 特征向量 f_i

步骤 1 对样本集进行去均值化级联得到 $X = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N] \in \mathbb{R}^{k_1 k_2 \times Nc}$ 。

步骤 2 根据式 (9) 计算 PCA 卷积层第 1 层的卷积核 W_1^1 , 将卷积核与输入样本 I_i 进行卷积, 得到第 i 个样本在第 1 层的 L_1 个输出:

$$O_i^1 = \{x_i * W_1^1\}_{l_1=1}^{L_1} \in \mathbb{R}^{m \times n \times L_1} \quad (14)$$

步骤 3 对输出 O_i^1 进行二值化加权处理得到一幅“整型图” T_i^1 , 将 T_i^1 分为 B 块, 提取每块直方图特征 $\{u_i \in \mathbb{R}^{2^{L_1}} | 1 \leq i \leq B\}$, u_i 表示每块直方图特征, 其维数为 2^{L_1} 。如 2.1 节所述, 将空间金字塔设为 3 层, 根据式 (13) 分层池化局部直方图得到池化后每层特征 $z_j = \max \{ |u_{1j}|, |u_{2j}|, \dots, |u_{Bj}| \}$, 级联 3 层特征得到输入样本 I_i 的底层特征向量表示 $f_{i,1} \in \mathbb{R}^{2^{L_1} \times (2^2 \times 2^2 + 2^1 \times 2^1 + 2^0 \times 2^0)}$ 。

步骤 4 对步骤 2 中输出 O_i^1 最小化重构误差得到 PCA 卷积层第 2 层的卷积核 $W_1^2 \in \mathbb{R}^{k_1 \times k_2}$, 将 W_1^2 与 O_i^1 卷积得到对应每个 $O_i^1(l_1)$ 的 L_2 个输出 $O_i^2(l_1)$ 为:

$$O_i^2(l_1) = \{O_i^1(l_1) * W_1^2\}_{l_2=1}^{L_2} \in \mathbb{R}^{m \times n \times L_2}, l_1 = 1, 2, \dots, L_1 \quad (15)$$

步骤 5 根据式 (3), 对每个 $O_i^1(l_1)$ 的 L_2 个输出 $O_i^2(l_1)$ 进行计算得到一幅“整型图”, 则对于

第 2 层 L_1 个输入 O_i^1 将得到 L_1 幅“整型图”:

$$T_i^2(l_1) = \sum_{l_2=1}^{L_2} 2^{l_2-1} H(O_i^2(l_1)), l_1 = 1, 2, \dots, L_1 \quad (16)$$

类似于步骤 3 特征池化过程,对每幅“整型图” T_i^2 进行局部直方图,空间金字塔池化得到输入样本 I_i 的高层特征向量表示 $f_{i,2} \in \mathbb{R}^{2^{L_2} L_1 \times (2^2 \times 2^2 + 2^1 \times 2^1 + 2^0 \times 2^0)}$ 。

步骤 6 级联底层特征 $f_{i,1}$ 和高层特征 $f_{i,2}$ 得到输入样本 I_i 的多尺度特征表示 $f_i = [f_{i,1}; f_{i,2}]$ 。

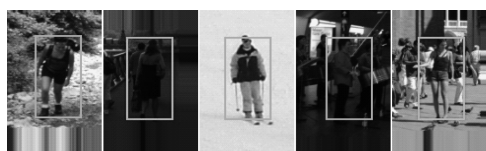
将特征向量 f_i 输入 SVM 完成行人和非行人的判定。SVM 分类函数为:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (17)$$

其中, x 为 SVM 的输入,即特征向量 f_i , θ 为分类的 SVM 的权值矩阵。

3 实验与结果分析

实验计算机配置为 CPU 2.66 GHz,内存 24 GB,采用 Matlab R2014a 平台,使用 Linlinear^[15] 作为 SVM 分类器开发包,选取目前使用最多的行人检测数据库 INRIA^[2] 和规模最大的静态行人数据库 NICTA^[17] 进行行人检测。在 INRIA 数据库中选择 2 416 张行人图片作为训练集正样本,1 126 张行人图片作为测试集正样本,归一化大小为 64×32 ;从 1 218 张高分辨率非行人图片中随机截取 3 654 张大小为 64×32 的图片作为训练集负样本,从 453 张高分辨率非行人图片中随机截取 1 812 张大小为 64×32 的图像作为测试集负样本。NICTA 数据库包含 25 551 张行人样本,随机选取 6 000 张作为训练集正样本,并且从剩余行人样本中选取 6 000 张作为测试集正样本,归一化大小为 80×32 ;从 5 207 张高分辨率非行人图片随机截取 6 000 张大小为 80×32 的图片作为训练集负样本,6 000 张大小为 80×32 的图片作为测试集样本。对图片中单行人样本进行检测时,将除主要行人区域之外的其余行人视为背景考虑,部分实验结果如图 4 所示。



(a)INRIA数据库



(b)NICTA数据库

图 4 单行人样本检测结果

3.1 网络参数选取

MS-PCANet-SPP 的参数包括卷积核、局部直方

图池化大小和空间金字塔池化层数等,其中卷积核的选取对于 PCANet 及 MS-PCANet-SPP 性能有重要影响。卷积核的值采用 1.4 节中 PCA 预训练方式获得;卷积核数目采用文献[8]的结果,即固定每一层卷积核数目 $L_1 = L_2 = 8$;卷积核的大小,通过在 INRIA 和 NICTA 数据库上的实验结果评估其最佳值。卷积核的大小对正确检测率的影响如图 5 所示。

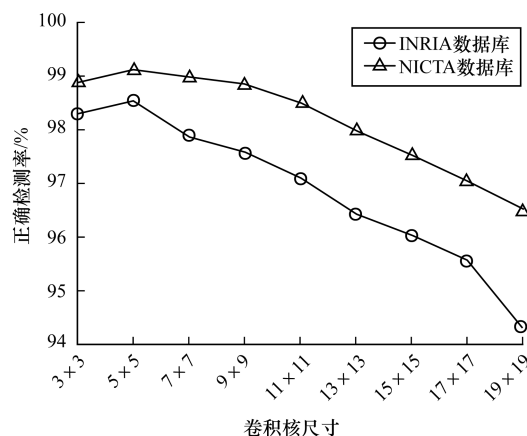


图 5 不同卷积核尺寸对应的网络正确检测率

由图 5 可知,当卷积核大小为 5×5 时,网络对 INRIA 和 NICTA 数据库有最佳的正确检测率。因此,设置 MS-PCANet-SPP 的卷积核尺寸为 5×5 ,固定每一层卷积核数目为 $L_1 = L_2 = 8$ 。并且通过多组实验尝试,兼顾检测时间及检测正确率,本文将 INRIA 数据库局部直方图池化大小设为 16×8 ,池化方式设为非重叠池化,NICTA 数据库局部直方图池化大小设为 20×8 ,池化方式设为非重叠池化。空间金字塔池化参数按文献[12]进行相同设置,将池化层数设为 3 层,从顶层到底层设置池化子区域为 1×1 , 2×2 和 4×4 ,其中空间金字塔池化函数为最大值池化。

3.2 实验对比算法

为了验证 MS-PCANet-SPP 的有效性,将下列算法与 MS-PCANet-SPP 进行对比:

1) HOG: HOG 算法^[2]作为传统特征提取算法,被公认是最具有泛化能力的特征提取方法之一。

2) DL-CLDRC: DL-CLDRC 算法^[18]将深度学习运用于 LDRC 算法,在人脸识别中取得较好的识别效果。

3) CNN-2: 本文设计 2 层的 CNN^[19]对行人数据库进行检测,其卷积核的大小为 5×5 ,第 1 层卷积核数目为 10,第 2 层卷积核数目为 15,选择 Sigmoid 激活函数作为非线性处理函数,池化层池化方式选择均值池化,将通过全连接层得到的特征向量用 Softmax 分类器进行分类输出。CNN-2 采用批量训练方式,设学习率 0.1,将数据集中每 600 张样本设为 1 批,对每批样本进行 200 次迭代。

4) PCANet: PCANet^[8]卷积核尺寸设为 5×5 ,每

一层卷积核数目设为 $L_1 = L_2 = 8$ 。

5) PCANet-1: 在 PCANet 基础上提出仅使用一层 PCA 卷积层的 PCANet 作为训练样本的网络,其卷积核尺寸设为 5×5 ,卷积核数目设为 $L_1 = 8$ 。

6) PCANet-SPP1: 在 PCANet-1 基础上增加空间金字塔池化,其空间金字塔池化层数设为 3 层,从顶层到底层设池化区域为 1×1 、 2×2 和 4×4 ,空间金字塔池化方式为最大值池化。

3.3 实验结果

3.3.1 在 INRIA 和 NICTA 数据库中实验结果

上述算法在 INRIA 和 NICTA 数据库中进行行人检测,其性能对比结果见表 1 和表 2。其中,正确检测率表示行人和非行人样本被正确检测出的百分比,漏检率表示所有的行人样本中未被检测出行人的百分比,误检率表示所有的非行人样本未被检测为非行人的百分比,平均检测速度本文用每秒采样率表示,即每秒能够检测的样本数量。

表 1 各算法在 INRIA 数据库中性能对比结果

| 算法名称 | 正确检测率/% | 漏检率/% | 误检率/% | 平均检测速度/sps |
|-----------------------------|---------|-------|-------|------------|
| HOG + SVM 算法 ^[2] | 97.07 | 2.22 | 3.37 | 2.94 |
| CNN 算法 ^[19] | 95.47 | 8.26 | 2.21 | 4.76 |
| DL-CLDRC 算法 ^[18] | 97.21 | 2.37 | 2.48 | 2.86 |
| PCANet-1 算法 | 96.05 | 3.75 | 4.19 | 4.16 |
| PCANet-SPP1 算法 | 96.29 | 3.55 | 3.75 | 3.70 |
| PCANet 算法 ^[8] | 97.62 | 2.66 | 2.21 | 2.70 |
| 本文算法 | 98.54 | 1.60 | 1.38 | 2.22 |

表 2 各算法在 NICTA 数据库中性能对比结果

| 算法名称 | 正确检测率/% | 漏检率/% | 误检率/% | 平均检测速度/sps |
|-----------------------------|---------|-------|-------|------------|
| HOG + SVM 算法 ^[2] | 98.23 | 1.92 | 1.62 | 4.00 |
| CNN 算法 ^[19] | 97.24 | 2.45 | 3.01 | 6.25 |
| DL-CLDRC 算法 ^[18] | 98.54 | 1.67 | 1.32 | 2.98 |
| PCANet-1 算法 | 97.79 | 2.42 | 2.00 | 4.76 |
| PCANet-SPP-1 算法 | 97.93 | 2.26 | 1.87 | 4.34 |
| PCANet 算法 ^[8] | 98.82 | 1.28 | 1.08 | 2.78 |
| 本文算法 | 99.01 | 1.10 | 0.88 | 2.38 |

根据表 1 和表 2,得出以下结论:

1) HOG 在 2 个数据库中均取得较高的正确检测率,其正确检测率分别为 97.07% 和 98.23%。

2) CNN 在所有算法中正确检测率最低,仅为 95.47% 和 97.24%。这是因为 CNN 使用基于梯度下降的学习方法进行监督训练,导致在网络层数较少时对样本的泛化能力较差。

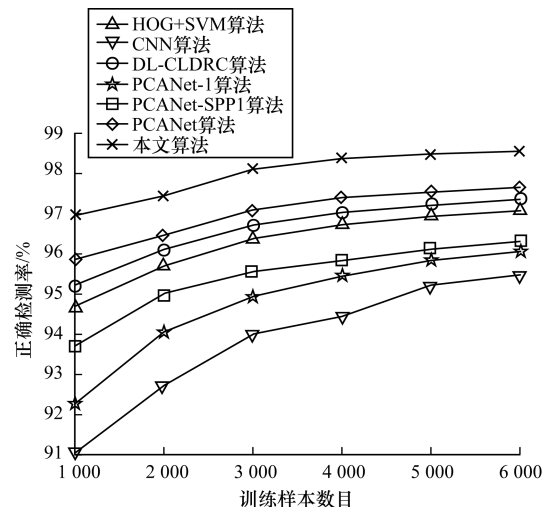
3) PCANet 的正确检测率要高于 PCANet-1,验证了 PCANet 作为一种深度特征提取网络,其网络层数越多特征提取能力越强。

4) PCANet-SPP1 的正确检测率高于 PCANet-1,验证了本文提出的空间金字塔池化的有效性。

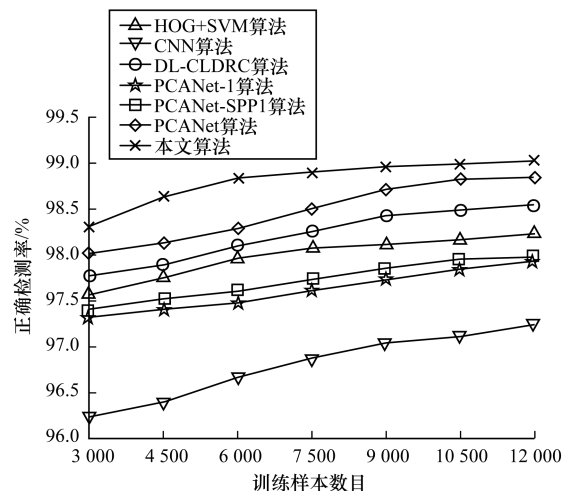
5) MS-PCANet-SPP 在 2 个数据库中的正确检测率最高,并且漏检率和误检率也最低,验证了本文算法的有效性。

6) 由于 MS-PCANet-SPP 相比于其他算法进行了更多特征提取和分层池化,导致网络计算复杂度增加,因此 MS-PCANet-SPP 对于样本的平均检测速度最低。但是从行人检测的安全性考虑,检测率重要性要优于检测速度,并且由于计算机性能的提高,MS-PCANet-SPP 基本可实现实时行人检测。

为验证 MS-PCANet-SPP 的泛化性能和鲁棒性指标,本文在 INRIA 数据库中随机选取 6 组样本作为训练集,其中正样本与负样本比例设为 1:1,样本数分别为 1 000, 2 000, ..., 6 000,测试集数目固定不变。对每组样本进行训练并在测试集进行检测,实验结果如图 6(a) 所示。同样,图 6(b) 为对 NICTA 数据库的检测结果,其训练集的样本数为 3 000, 4 500, ..., 12 000。



(a) 在 INRIA 数据库中检测结果



(b) 在 NICTA 数据库中检测结果

图 6 不同训练样本数目下各算法检测结果

由图 6 可得,在 INRIA 和 NICTA 数据库中随着训练样本数量增加,MS-PCANet-SPP 对测试集的正

确检测率也相应提高,算法的性能仍保持稳定,验证了 MS-PCANet-SPP 具有较好的泛化性能和鲁棒性。

3.3.2 在自建数据集中实验结果

为验证 MS-PCANet-SPP 在非理想的自然条件下行人检测的鲁棒性,从 INRIA、NICTA 和 TUD^[20] 数据库中的测试集选取负责背景、部分遮挡、低分辨率和远距离等的行人样本作为测试集,建立专门针对非理想的自然条件下的行人测试数据集,其中测试数据集正负样本数均为 3 000 张,样本尺寸归一化为 64×32 ,部分测试样本如图 7 所示。



图 7 部分自建数据集样本

表 3 为各算法在自建数据库中的性能比较结果。图 8 为各算法在 NICTA、INRIA 和自建数据库 (横坐标 Our) 中的检测结果。3 个数据库的训练集和测试集,正负样本的比例均相同。由表 3 可知,MS-PCANet-SPP 在复杂背景、部分遮挡、低分辨率等自然条件下正确检测率相比 HOG、CNN、DL-CLDRC 和 PCANet 等算法明显提高,并且漏检率和误检率也更低。由图 8 可得,MS-PCANet-SPP 在自建数据库中相比其他算法具有更低的下降幅度,验证了 MS-PCANet-SPP 对非理想自然条件下的行人检测具有一定的鲁棒性。

表 3 各算法在自建数据集中的性能对比结果 %

| 算法名称 | 正确检测率 | 漏检率 | 误检率 |
|-----------------------------|-------|-------|-------|
| HOG + SVM ^[2] 算法 | 84.52 | 15.12 | 15.81 |
| CNN ^[19] 算法 | 82.74 | 20.87 | 15.01 |
| DL-CLDRC ^[18] 算法 | 89.21 | 11.99 | 10.04 |
| PCANet-1 算法 | 86.14 | 16.43 | 12.25 |
| PCANet-SPP-1 算法 | 87.51 | 14.65 | 11.15 |
| PCANet ^[8] 算法 | 90.23 | 10.66 | 9.21 |
| MS-PCANet-SPP 算法 | 92.34 | 8.88 | 6.90 |

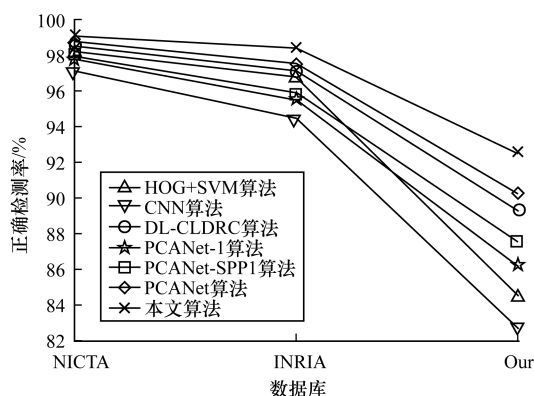


图 8 不同数据库中算法检测结果

4 结束语

行人检测需要解决在非理想条件下的准确性和鲁棒性的问题,本文根据空间金字塔池化对形变目标具有较强鲁棒性,同时受多尺度 CNN 级联不同尺度特征以使得网络既能提取到高层语义信息又能获得底层细节信息的启发,提出一种面向行人检测的深度特征提取网络 MS-PCANet-SPP。通过在原始池化层引入空间金字塔策略,利用空间金字塔最大值池化方法对原始具有冗余的特征进行提取,精简出其中重要的、区分性强的特征,同时将不同尺度的经空间金字塔池化的特征向量级联构成以获得图像的多尺度特征描述,使得 MS-PCANet-SPP 既能提取到高层语义信息又能获得底层细节信息。在 INRIA 和 NICTA 数据库中的对比实验结果表明,MS-PCANet-SPP 能有效提高行人检测的准确率,正确检测率高于 HOG、CNN 和 PCANet 等算法,在漏检率和误检率上也低于其他算法,并且通过进一步实验验证本文提出的 MS-PCANet-SPP 是一个有效的深度特征提取网络,对于复杂背景、拍摄角度以及人体姿态多样等非理想条件下的行人检测具有较强鲁棒性。同时,由于计算机性能的提高,MS-PCANet-SPP 基本能实现实时检测。下一步将在 MS-PCANet-SPP 的基础上对行人检测的显著性检测算法进行研究,以提高算法的检测速度。

参考文献

- [1] 苏志松,李绍滋,陈淑媛,等. 行人检测技术综述[J]. 电子学报,2012,40(4):814-820.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2005:886-893.
- [3] MU Y, YAN S, LIU Y, et al. Discriminative local binary patterns for pedestrian detection i personal album [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2008:1-8.
- [4] TUZEL O, PORIKLI F, MEER P. Pedestrian detection via classification on Riemannian manifolds [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2008,30(10):1713-1727.
- [5] VIOLA P, JONES M. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [6] HINTON G, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks [J]. Science,2006,313(5786):304-304.

- [7] 余凯,贾磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,50(9):1799-1804.
- [8] CHAN T H, JIA K, GAO S, et al. PCANet: a simple deep learning baseline for image classification? [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5017-5032.
- [9] GAO F, DONG J, LI B, et al. Automatic change detection in synthetic aperture radar images based on PCANet [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1792-1796.
- [10] WANG Z, BI Z, WANG C, et al. Traffic lights recognition based on PCANet [C]//Proceedings of Chinese Automation Congress. Washington D. C., USA: IEEE Press, 2016: 559-564.
- [11] WANG S, CHEN L, ZHOU Z, et al. Human fall detection in surveillance video based on PCANet [J]. Multimedia Tools and Applications, 2016, 75(19): 11603-11613.
- [12] YANG J, YU K, GONG Y, et al. Linear spatial pyramid matching using sparse coding for image classification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2009: 1794-1801.
- [13] SERMANET P, LECUN Y. Traffic sign recognition with multi-scale convolutional networks [C]//Proceedings of International Joint Conference on Neural Networks. Washington D. C., USA: IEEE Press, 2011: 2809-2813.
- [14] HOAI M. Regularized max pooling for image categorization [J]. Journal of the British Institution of Radio Engineers, 2014, 14(3): 94-100.
- [15] 顾凌云,吕志文,杨勇,等. 基于 PCANet 和 SVM 的谎言测试研究 [J]. 电子学报, 2016, 44(8): 1969-1973.
- [16] CHANG C C, LIN C J. LIBSVM: A library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27.
- [17] OVERETT G, PETERSSON L. Boosting with multiple classifier families [C]//Proceedings of IEEE Intelligent Vehicles Symposium. Washington D. C., USA: IEEE Press, 2007: 1039-1044.
- [18] SHAILAJA K, ANURADHA B. Effective face recognition using deep learning based linear discriminant classification [C]//Proceedings of IEEE International Conference on Computational Intelligence and Computing Research. Washington D. C., USA: IEEE Press, 2016: 1-6.
- [19] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [C]//Proceedings of NIPS'12. Nevada, USA: Curran Associates Inc., 2012: 1097-1105.
- [20] WOJEK C, WALK S, SCHIELE B. Multi-cue onboard pedestrian detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2009: 794-801.

编辑 刘盛龄

(上接第269页)

- [7] 徐勇军,高梅国,毛二可. 智能运输系统中的雷达车辆检测器 [J]. 北京理工大学学报, 2001, 21(2): 256-259.
- [8] 陈振学,汪国有,刘成云. 基于计算机视觉的汽车流量检测统计 [J]. 华中科技大学学报(自然科学版), 2006, 34(5): 46-49.
- [9] 王洪建,李志敏. 基于视频图像的车辆流量实时检测系统 [J]. 光学精密工程, 2005, 13(增刊): 222-226.
- [10] 檀甲甲,张建秋. 实时采集道路车流量信息的视频新方法 [J]. 仪器仪表学报, 2008, 29(1): 158-166.
- [11] 张桂梅,孙晓旭,陈彬彬,等. 结合分数阶微分和 Canny 算子的边缘检测 [J]. 中国图象图形学报, 2016, 21(8): 1028-1038.
- [12] 张鹏,黄毅,阮雅端,等. 基于稀疏特征的交通流视频检测算法 [J]. 南京大学学报(自然科学版), 2015, 51(2): 264-270.
- [13] 陈银,任侃,顾国华,等. 基于改进的单高斯背景模型运动目标检测算法 [J]. 中国激光, 2014, 41(11): 245-253.
- [14] 张立国,杨瑾,李晶,等. 基于小波包和数学形态学结合的图像特征提取方法 [J]. 仪器仪表学报, 2010, 31(10): 2285-2290.
- [15] 常志国,李晶,胡云鹭,等. 基于视频的车流量统计算法 [J]. 计算机系统应用, 2016, 25(7): 187-191.
- [16] 齐美彬,鲜柯,蒋建国,等. 一种基于车辆遮挡模型的车流量统计算法 [J]. 仪器仪表学报, 2010, 31(6): 1335-1341.
- [17] FELZENSZWALB P F, SCHWARTZ J D. Hierarchical matching of deformable shapes [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2007: 1-8.
- [18] 晏建洋,吴建星. 基于 LabVIEW 和 MATLAB 的矿山微震信号小波分析与研究 [J]. 安全与环境工程, 2016, 23(3): 125-128.
- [19] 徐峰,刘婷薇,李平,等. 基于 LabVIEW 与 MATLAB 混合编程的手势识别系统 [J]. 电子设计工程, 2017, 25(8): 32-36.

编辑 刘盛龄