

基于 HOG 特征和 SVM 分类器的行人检测

温都仁

摘要—行人检测是通过图像局部特征判断图像或视频中是否存在行人并给出精确定位的计算机视觉任务。本次实验以 HOG 特征为局部特征检测因子, 利用 SVM 分类器训练模型完成行人检测任务。该方法对行人特征把控较好, 对不同光照条件和行人朝向具有良好表现, 使用窗口合并的方式提升其在不同尺度图像上的检测表现。

I. 引言

行人检测是利用计算机视觉技术判断图像或视频序列中是否存在行人并给予精确定位, 该技术可应用于智能交通、智能机器人、行人行为分析、车辆辅助驾驶系统等领域。HOG 全称方向梯度直方图, 其核心思想是图像的局部目标的表象和形状能够被梯度或者边缘的方向密度分布很好地描述出来。以这些描述子和人为添加的标签作为训练数据训练得出的模型可用于判断图像或视频序列中是否存在行人。

该任务通过提取大量正负样例的 HOG 特征并人工标注后通过机器训练获得 SVM 分类器, 再将特征应用于目标图像, 以滑窗的形式在目标中进行匹配, 判断其中是否存在行人。考虑到不同尺寸图像中行人所占像素存在差异, 所以可能引起标记框重叠的现象, 参考人脸检测任务中抑制非极大值从而合并窗口的做法朴素合并重叠标记框得到更为准确的结果。

II. 相关工作

A. HOG 特征原理

方向梯度直方图通过计算和统计图像局部区域的梯度方向直方图构成特征, 由于其是在图像的局部方格单元上操作, 所以它对图像集合的光学和形变都能保持良好的不变性。梯度的统计信息则主要存在于边缘, 所以对于行人检测任务而言特征描述子即行人轮廓信息。HOG 特征提取主要有如下几步:

1) 图像灰度化与 Gamma 矫正: Gamma 矫正通过非线性变换使得图像从曝光强度的线性响应变得更接近人眼感受的响应, 是图像预处理中常用的手段, 可

用于调节图像的对比度, 减少图像的光照不均和局部阴影。其计算公式为:

$$f(x) = x^\gamma$$

上式中 γ 为一常数, 本文使用 $\gamma = \frac{1}{2}$ 为矫正因子。

2) 计算梯度直方图: 计算梯度直方图面向图像上的可滑动窗口 (window), 窗口大小选择 64×128 , 计算该窗口中所有像素的梯度以备使用。图像中任意像素点 (x, y) 在 x, y 两个方向上的梯度表示为:

$$\begin{cases} G_x(x, y) = H(x+1, y) - H(x-1, y) \\ G_y(x, y) = H(x, y+1) - H(x, y-1) \end{cases}$$

像素点 (x, y) 处的梯度幅值和梯度方向分别由下面的公式计算而得:

$$\begin{cases} G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \\ \alpha(x, y) = \tan^{-1} \frac{G_y(x, y)}{G_x(x, y)} \end{cases}$$

实际操作中通过梯度算子 $[-1, 0, -1]$ 和 $[1, 0, -1]^T$ 可分别得到像素点 (x, y) 在 x, y 两个方向上的梯度。计算结束后可以得到两个 (64×128) 的梯度矩阵。

3) 8×8 网格计算方向梯度直方图: 在计算得到所有像素点的梯度数值后, 以每 8×8 个像素为单位, 计算其梯度和梯度方向的统计直方图, 该区域被称为细胞单元 (cell)。约定梯度方向每 20° 为一单元, 将 $0-180$ 度分为 9 个单元, 对细胞单元内像素逐个统计。对每个像素而言, 将其梯度按其梯度方向所落入的单元, 并按其权重分配至对应的 1-2 个单元内进行统计。计算结束后得到 16×32 个统计直方图, 每个统计直方图的维数为 (1×9) 。

4) 16×16 网格归一化: 以 16×16 个像素, 即 4 个细胞单元为单位, 对该区域的直方图进行归一化。归一化运算首先对区域内 4 个直方图进行求和, 归一化结果为直方图每一个元素和该值的商, 计算结果为一 36 维向量。

5) 计算 *HOG* 特征向量: *HOG* 特征向量是输入图像的特征描述, 即一个窗口, 一个窗口由 24×128 的像素区域组成, 在网格滑动步长为 8 的情况下, 一个窗口由 105 个 16×16 的区域组成, 每个 16×16 的区域由一个 36 维的向量描述, 故一个窗口的 *HOG* 特征向量的维数为

$$36 \times 105 = 3780.$$

B. SVM 分类器

SVM(Support Vector Machines), 支持向量机是定义在特征空间上间隔最大的线性分类器, 是一种二分类模型。SVM 的训练目标是极大化支持向量(离分割超平面最近的点)与分割超平面的距离(图 1)。在本次实验中输入支持向量机的进行分类的样本是图像的 *HOG* 特征向量, 实验并未手工实现 SVM, 所以对其原理和步骤不再进行赘述。

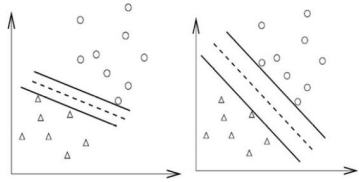


图 1. SVM 分类器训练目标示意

C. 检测流程和使用方法

图像的 *HOG* 特征提取和训练/检测流程如图 2 所示, 实验中 *HOG* 特征提取调用了 `opencv.HOGDescriptor`, 源码文件 `HOG.py` 是自行实现的 *HOG* 特征检测算法, SVM 分类器调用了 `sklearn` 下的 `SVC` 方法, 模型的保存和加载使用了 `python pickle` 库。

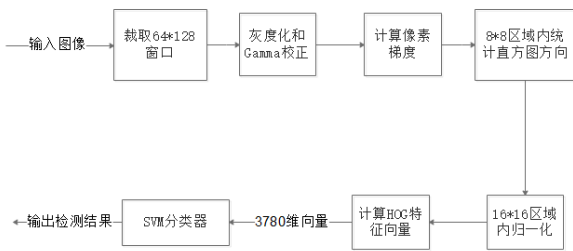


图 2. 行人检测流程

III. 数据集与方法

A. 数据集

本次实验使用的是 INRIA 数据集, 数据集包含正例 2416 项, 分别为 1239 幅包含行人的图像和其左右镜像后的图像, 负例 1218 项(裁减后 12180 项)。其中各抽取 20% 的数据作为测试集, 故最终的数据规模是训练样本共 13136 (其中正例 1932, 负例 9744), 测试样本共 1698 (其中正例 386, 负例 2436 幅)。

B. 实验方法

本次实验通过提取 1932 个正样例和 9744 幅负样例的 *HOG* 特征(3780 维), 将其特征向量和标签投入 SVM 分类器进行训练后对测试数据进行分类测试, 对比了同等条件下不同模型的真阳率($TPR = \frac{TP}{P}$)和假阳率($FPR = \frac{FP}{N}$)。并观察了若干张尺寸未定的图像做了不同尺度图片上的行人检测表现。

IV. 实验结果

A. 结果概述

默认使用 *HOG* 特征 SVM 线性核函数进行训练的模型对行人检测的效果良好, 在测试数据上的准确率约有 0.97, 假阳率达到 0.007。但是对于部分尺寸较大的图像而言其结果不尽人意, 存在较为严重的误检现象, 主要存在于树木等杆状物。

B. 不同尺度图片上的 *HOG* 特征检测与局部优化

由于模型的 *HOG* 窗口大小为 64×128 , 而现实中检测目标不可能都是固定大小的窗口, 目标像素的多少从某种程度上取决于图像大小, 所以对于不同尺寸的图像而言如果仅做固定窗口的检测会出现窗口重叠的现象。针对该类问题提出一点优化: 对检测结果中的重叠窗口做简单合并处理, 结果对比如图 3 所示。进行合并操作时对滑窗检测结果中重叠的部分, 将其对应标记框范围扩大到包含两个小矩形标记框所包围像素之并集的最小矩形像素集。

C. 不同 SVM 核函数实验结果分析

SVM 核函数种类有 `rbf` (高斯核函数)、`Linear` (线性核函数)、`Poly` (多项式核)和 `Sigmoid` (双曲正切函数 `tanh` 核), 其中除了 `linear` 线性核函数, 其他核函数可用于解决非线性可分问题。不同核函数训练得到的模型在同一个数据集上的检测结果差异较大, 表 1 展示

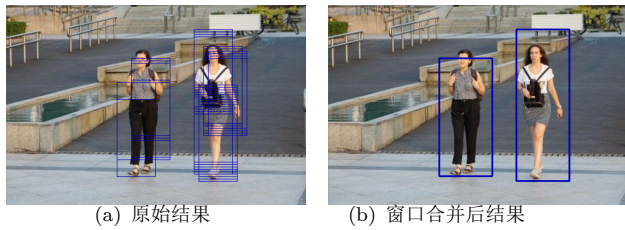


图 3. 不同尺度图片行人检测结果处理

了几种核函数训练结果, 其中 rbf 和 poly 两种核函数在当前测试集上的效果较为显著。

表 I
不同 SVM 核函数训练结果测试对比

	linear	rbf	poly	sigmoid
TPr	0.9479	0.9635	0.9562	0.2458
FPr	0.0073	0.0057	0.0033	0.1371

对不同尺度图片的检测结果也不同, 高斯核函数训练数模型容易漏检, 多项式核函数训练模型容易误检, 线性核函数结果相对稳定, 但对于大图也容易产生误检的现象, 而 sigmoid 核函数模型预测结果和数据都表明该方法不适合本任务。

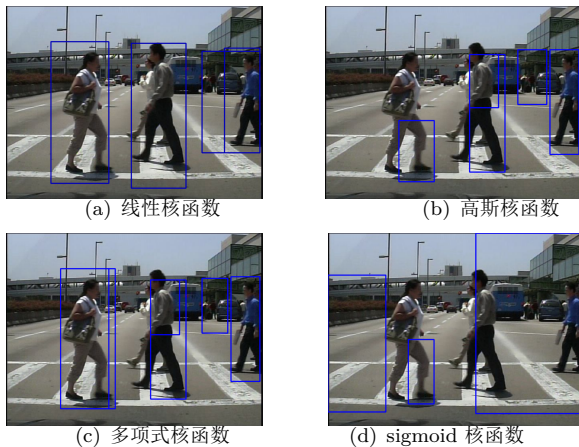


图 4. 不同核函数 SVM 分类器行人检测结果对比

V. 存在问题

本次实验实现的是 HOG+SVM 最基本的组合使用, 实验结果在同源数据集上表现良好, 但是对于外部数据集, 即一些网络上的图片检测情况不尽人意, 主要是以下两方面:

1) 误检问题: 采用滑窗检测图像局部特征的过程中对图像中的树木、邮筒、栏杆等与行人较为接近的物体时容易产生误检现象。这是因为 HOG 特征描述的还是以边缘为主的特征, 所以当图像中存在与行人轮廓接近的物体时容易产生误检。

2) 多尺度检测问题: 由于采用了 sklearn 中的 SVC 所以并没有多尺度检测的库函数, 所以导致检测只在一个尺度上进行。虽然采用窗口合并的方式解决了部分窗口重叠问题可以标记大尺寸图片中的行人位置, 但是对于小于 64×128 的行人像素而言无法做到精确化, 只是标记其所在的范围。

VI. 总结

经历本次实验, 主要有以下几点结论:

1) HOG 特征描述局部像素区域的特征: 局部特征描述子 HOG 特征和 SIFT/SURF 局部特征的最大区别在于 HOG 描述的是局部区域的特征, 而 SIFT 和 SURF 描述的则是特定像素点的特征, 其一个特征向量包含的信息丰富于 SIFT 和 SURF 特征, 因此 HOG 可以用来实现区域化的任务。与此同时 HOG 特征描述的是一个窗口的特征而不是整个图像的梯度信息, 其包含的信息非常有限, 但也正是得益于此, HOG 特征描述的图像局部特征信息对光照, 形变等具有良好的鲁棒性, 从而可以检测不同光照条件和形状 (主要指行人朝向和细微动作) 的行人。

2) 视觉任务有多尺度检测的必要性: 多尺度图像检测问题表明图像金字塔不仅是只作为 SIFT 特征检测中的一种数据结构, 也是图像检测任务中不可或缺的一步, 通过多尺度图像检测和坐标变换可以得到更为准确的检测结果。

3) 视觉任务中有数据预处理的必要性: 通过这次的实践认识到了更深一层的图像处理, 对图像也表示也从像素级过渡到了向量级, 由处理一个图像变到了处理大量图像。在实验伊始耗费了大量的时间精力对 Caltech 数据集进行预处理, 后由于图像裁减问题更换了数据集, 在这个过程中对数据预处理的重要性有了深刻的认识。

参考文献

- [1] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 886-893 vol. 1.