

8-13-2020

Multi-scale Detection Method for Dense Crowd Target Detection

Jianxin Zhou

1. *National University of Defense Technology, Changsha 410073, China;;*

Jianjun Wu

2. *Hainan Armed Police Force, Hainan 570203, China;;*

Junqiang Xue

2. *Hainan Armed Police Force, Hainan 570203, China;;*

Lin Shuai

1. *National University of Defense Technology, Changsha 410073, China;;*

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Multi-scale Detection Method for Dense Crowd Target Detection

Abstract

Abstract: A multi-scale algorithm was proposed to detect the targets flexibly. In coarse scale, an optimized DPM (Deformable Part Model) method was used to filter out sparse objectives that was represented by whole body. Then the whole scenario was cut into multiple finer regions and the Faster R-CNN (Region-based Convolutional Neural Network) method was trained and utilized to detect dense objects that was indicated by head in fine scale. These two detection results were incorporated via NMS (Non - Maximum Suppression) method, in order to supplement with each other and remove redundancy. The effectiveness of the proposed method has been proved comparing detect accuracy with DPM and R-CNN individually in the final experiment.

Keywords

dense crowd target detection, multi-scale detection, DPM, Faster R-CNN

Authors

Jianxin Zhou, Jianjun Wu, Junqiang Xue, Lin Shuai, Dang Gang, and Zhiquan Cheng

Recommended Citation

Zhou Jianxin, Wu Jianjun, Xue Junqiang, Lin Shuai, Dang Gang, Cheng Zhiquan. Multi-scale Detection Method for Dense Crowd Target Detection[J]. Journal of System Simulation, 2016, 28(10): 2503-2509.

适用于密集人群目标检测的多尺度检测方法

周建新¹, 吴建军², 薛均强², 林帅¹, 党岗¹, 程志全^{3*}(1.国防科学技术大学 计算机学院, 湖南长沙 410073; 2.中国人民解放军海南武警总队, 海南海口 570203;
3.湖南化身科技有限公司, 湖南长沙 410205)

摘要: 针对密集人群场景下的目标检测问题, 提出了一种多尺度的目标检测方法。在粗尺度下, 使用优化的 DPM(Deformable Part Model)检测方法, 将人体全身作为检测对象, 检测整个场景中的稀疏目标; 在细尺度下, 将头部作为检测对象, 使用重新训练的 Faster R-CNN(Region-based Convolutional Neural Network)网络检测稠密人群中的目标。将两种尺度下检测结果通过非极大值抑制(NMS, Non-Maximum Suppression)方法结合在一起, 这样两种方法既互相补充又能去除冗余检测结果。实验结果证明, 相比于单独的 DPM 检测方法和 Faster R-CNN 检测方法, 提出的多尺度检测方法在检测精度上有显著提升。

关键词: 密集人群检测; 多尺度检测; DPM; Faster R-CNN

中图分类号: TP18 文献标识码: A 文章编号: 1004-731X (2016) 10-2503-07

Multi-scale Detection Method for Dense Crowd Target Detection

Zhou Jianxin¹, Wu Jianjun², Xue Junqiang², Lin Shuai¹, Dang Gang¹, Cheng Zhiquan^{3*}(1. National University of Defense Technology, Changsha 410073, China; 2. Hainan Armed Police Force, Hainan 570203, China;
3. Hunan Avatar Science Company, Changsha 410205, China)

Abstract: A multi-scale algorithm was proposed to detect the targets flexibly. In coarse scale, an optimized DPM (Deformable Part Model) method was used to filter out sparse objectives that was represented by whole body. Then the whole scenario was cut into multiple finer regions and the Faster R-CNN (Region-based Convolutional Neural Network) method was trained and utilized to detect dense objects that was indicated by head in fine scale. These two detection results were incorporated via NMS (Non - Maximum Suppression) method, in order to supplement with each other and remove redundancy. The effectiveness of the proposed method has been proved comparing detect accuracy with DPM and R-CNN individually in the final experiment.

Keywords: dense crowd target detection; multi-scale detection; DPM; Faster R-CNN

引言

面向人群的目标检测是一项基础性的研究课题。目前, 稀疏目标的检测方法研究较多, 而密集

目标的检测由于其本身存在的诸多困难, 给研究人员带来更大的挑战, 如目标互相遮挡、光线、拍摄角度、图像分辨率等因素均制约了密集目标的检测效果。

密集目标检测的方法可以分为三种。第一种是基于特征描述子^[1-8]和支撑向量机(SVM)或深度卷积网络^[9-16]直接检测, 这种方法最为简单, 直接将稀疏目标的检测方法应用到密集目标的检测, 也正



收稿日期: 2016-05-04 修回日期: 2016-08-04;
作者简介: 周建新(1989-), 男, 甘肃平凉, 硕士, 助工, 研究方向为多媒体与虚拟现实技术; 吴建军(1970-), 男, 海南海口, 硕士, 研究方向为视频处理与通信; 薛均强(1982-), 男, 广东深圳, 硕士, 助工, 研究方向为计算机图形学。

<http://www.china-simulation.com>

• 2503 •

是因为简单, 没有充分考虑密集目标的特殊性, 检测效果不好。第二种是间接检测, 利用图像特征归约得到目标密度^[5,17], 然后可以确定任意区域的目标数量, 再结合直接检测的方法提高检测精度, 这种方法对检测精度的提升严重依赖目标密度的准确性, 而归约得到的目标密度往往是不准确的, 对直接检测的精度提升并不高。第三种是多种方法融合^[7-8], 利用各方法之间的差异性进行互充, 取得良好的检测效果, 文献[7-8]的实验结果也证实, 这种方法是进行密集目标检测的有效途径。

由于检测的人群自身的复杂性, 单一的检测方法难以达到理想的状态。在既有的算法中, DPM 检测方法^[2]使用了人体的稳定特征, 在稀疏人群上具有优势; 而 Faster R-CNN^[12]深度神经网络是当前最优的深度学习方法之一。将多种优势算法进行组合, 以达到更佳效果, 是一种直接的想法。

本文采用多尺度融合的思路, 将 DPM 人体检测方法^[2]和 Faster R-CNN^[12]深度神经网络的人头检测方法进行结合, 得到多尺度检测方法。粗尺度下, 目标人群稀疏, 依赖 DPM 稳定地检测人体; 细尺度下, 目标互相遮挡, 依赖 Faster R-CNN 检测人头。图 1 显示了检测效果。蓝绿色框为 DPM 检测方法检测到的目标, 红色框为 Faster R-CNN 后续检测到的目标, 黄色框为二者检测的错误目标。因此, 不同尺度下两种方法的融合可以显著提高目标人群的检测精度。



(a) mall_1 数据集中样本 (b) mall_2 数据集中样本

图 1 多尺度检测方法检测效果

1 检测方法

本文将 DPM 检测方法和 Faster R-CNN 检测方法进行了结合。1.1 部分介绍 DPM 检测方法及其

并行优化, 1.2 部分介绍 Faster R-CNN 检测方法, 最后在 1.3 部分介绍如何将两中检测方法结合在一起。

1.1 基于 DPM 的人体检测

1.1.1 模型

本文采用的 DPM 模型是文献[3]中的星型混合模型, 模型用 PASCAL VOC 2012 数据集进行训练。模型由 m 个组件 (M_1, M_2, \dots, M_m) 构成, 组件内采用图结构 (Pictorial Structures)^[18-19] 表示目标整体与局部之间的位置关系, 局部之间采用弹性连接, 每个部件都代表局部信息。每个组件代表人的一个视角, 由一个 root filter (F_0) 和一组 part filter (F_1, F_2, \dots, F_n) 组成。模型的一个组件可视化后如图 2 所示。

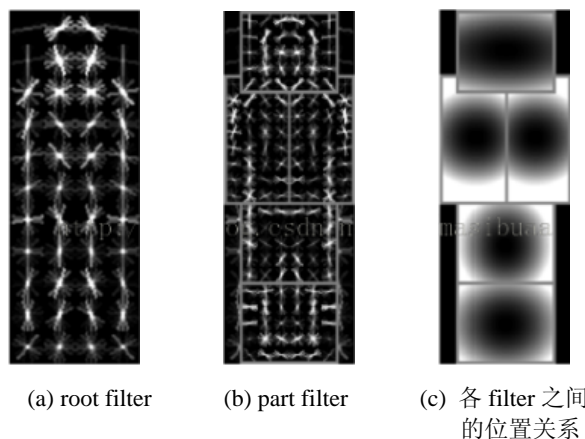


图 2 DPM 模型的描述

1.1.2 特征金字塔

DPM 模型匹配的对象是图像的特征金字塔 H , 图像特征采用 HOG 特征描述子。特征金字塔由一组不同缩放系数图像提取 HOG 特征构成, 如图 3 所示。左边为图像金字塔, 右边为特征金字塔。匹配时, root filter 位于分辨率较低的特征金字塔上层, part filter 位于分辨率较高的特征金字塔下层。由于 DPM 模型各 filter 尺寸是固定的, 特征金字塔确保检测方法可以适用于不同尺寸的目标。本文中 $p = (x, y, l)$ 表示特征金字塔第 l 层的位置 (x, y) ,

用 $\phi(H, p, w, h)$ 表示特征金字塔 H 中 p 位置的宽为 w , 高为 h 的窗口内的特征向量。

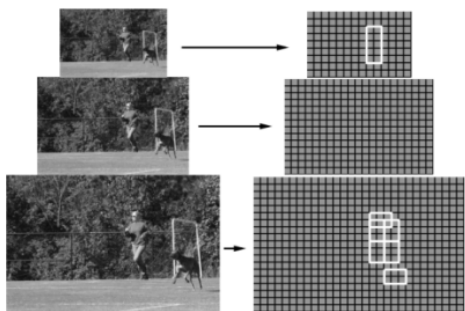


图 3 特征金字塔和人体模型实例

1.1.3 匹配

简而言之, DPM 检测方法的匹配过程就是 root filter 和各 part filter 采用滑窗式的方式在特征金字塔对应区域做卷积操作, 然后得到图像中每个位置的打分 score。最后选取阈值, 对 score 做截断, 再用非极大值抑制方法得到目标位置。

第 k 个组件的打分 $score_k$ 计算公式^[3]如下:

$$score_k(p_0, \dots, p_n) = \sum_{i=0}^n F_i \cdot \phi(H, p_i) - \sum_{i=1}^n d_i \cdot \phi_d(dx_i, dy_i) + b$$

其中,

$$(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i)$$

(x_0, y_0) 表示 root filter 所在的根位置, v_i 是第 i 个部件相对根位置所在的锚位置。因此, (dx_i, dy_i) 表示第 i 个部件相对自身锚位置的形变位移。 d_i 为形变的损耗权重, b 为偏置值,

$$\phi_d(dx_i, dy_i) = (dx_i, dy_i, dx_i^2, dy_i^2)$$

是形变位移的四元偶。每个部件的打分 $score_k$ 是各 filter 的打分减去形变代价, 再加上偏置值。选取 $Max(score_1, \dots, score_m)$ 为 (p_0, \dots, p_n) 处打分。最后选取阈值(实验中默认阈值为 -0.5), 对高于指定阈值的目标包围框做非极大值抑制处理, 得到最终的目标检测结果。

1.1.4 优化加速

针对形成特征金字塔、filter 和特征金字塔匹配过程的卷积操作两处具有并行性的位置, 本文

进行了并行优化, 并对卷积操作调用 AVX 扩展指令集进一步加速, 使 DPM 检测过程的速度大幅提升。

特征金字塔的每一层由对应比例的图像提取 HOG 特征生成。HOG 特征提取过程中, 先对图像每个像素做处理, 包括选取梯度的幅值和方向 (RGB 三通道选取最大值), 用双线性差分方法计算每个像素对临近的四个 cell 的贡献值。由于各像素的操作存在并行性, 可以做并行优化。

DPM 模型的 filter 组和特征金字塔的某一层做卷积操作时, 各个 filter 的卷积操作之间存在并行性, 可以做并行优化。另外, 卷积操作过程中, HOG 特征为 32 维向量, 做 4 次循环展开, 每次展开做 8 维特征卷积操作, 8 维卷积操作调用 AVX 扩展指令集的单指令多数据指令, 一次性处理 8 维向量, 可以进一步优化。

1.2 基于 Faster R-CNN 的人头检测

Faster R-CNN^[12] 用深度卷积网络来实现目标检测。该网络是在 Fast R-CNN^[11] 网络的基础上, 增加 RPN (Regions Proposal Network) 网络, 并且 RPN 网络和 Fast R-CNN 网络共享卷积层。共享卷积层是 Faster R-CNN 网络的重要特点, 它使得提取预选区域 (Region Proposal) 较选择性搜索策略 (Selective Search)^[10] 大幅提速。Faster R-CNN (基于 VGG16 模型) 网络结构如图 4 所示。网络由 Fast R-CNN (绿色框内网络) 和 RPN 网络构成。其中 13 个卷积层由 RPN 和 Fast R-CNN 共享。该网络模型由 ImageNet 预训练模型 VGG16 初始化得到。RPN 网络输出的 proposal_cls_score 是提取的预选区域 (Region proposal) 的打分, proposal_bbox_pred 是预选区域的坐标, proposal_cls_score、proposal_bbox_pred 和 Fast CNN 的 conv5 输出的 Feature Map 一起输入 roi_pool5 空间池化层, 经过两层全连接层, 最后输出检测到的目标概率 cls_prob 和包围框 bbox_pred。

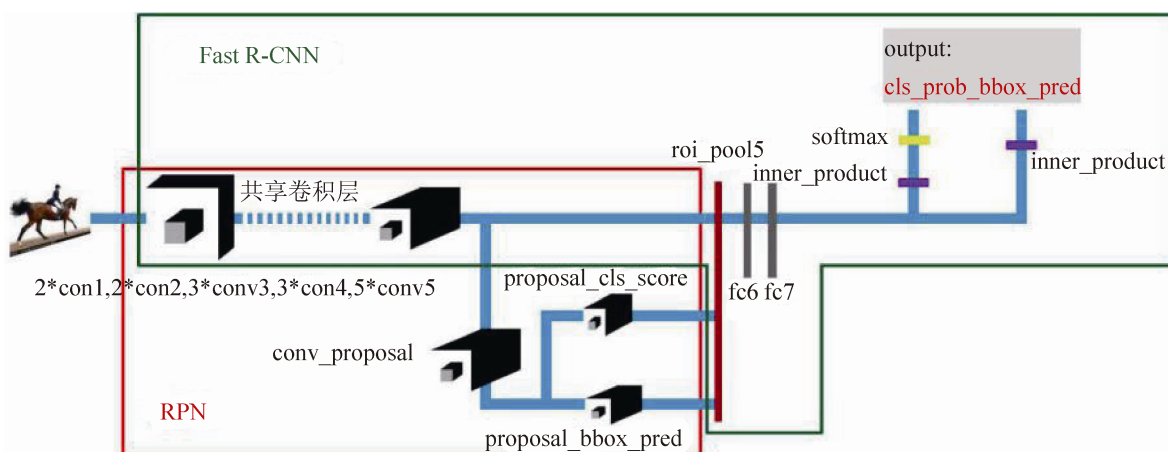


图 4 Faster R-CNN 网络。

1.2.1 Fast R-CNN

Fast R-CNN 网络结构如图 4 中绿框内所示。

Fast R-CNN 进行目标检测的过程如下：

(1) 在原始图片上，提取出预选区域(Region proposal)；

(2) 对整幅图片，通过卷积层获取整个原始图片的 Features Map；

(3) 将预选区域投影到(2)中得到 Features Map。用一个单层的 SPP layer^[20](Spatial Pyramid Pooling layer，如图 4 中 roi_pool5)池化得到预选区域对应的固定维度特征；

(4) 将(3)中得到的特征输入到两个全连接层(图 4 中 fc6 和 fc7)，最后输出两个优化目标。一个是表示目标类型的概率(图 4 中 cls_prob)，另一个是表示目标位置(图 4 中 bbox_pred)。

1.2.2 RPN

RPN 网络结构如图 4 中红色框所示。RPN 网络的输入是图像，然后经过共享卷积层和 RPN 网络最后的一个卷积层(图 4 中 conv_proposal)输出带有打分(图 4 中 proposal_cls_score)的预选区域(图 4 中 proposal_bbox_pred)。每个预选区域是一个 5 维向量(预选区域打分 1 维，预选区域坐标 4 维)，它和共享卷积层输出的 Features Map 共同输入到 Fast R-CNN 网络的池化层(图 4 中的 roi_pool5)。

1.3 结合 DPM 检测方法和 Faster R-CNN 检测方法

DPM 检测方法输出目标全身(root filter)的包围框和各个 part filter 的包围框，我们选取头部对应的 part filter 包围框作为 DPM 检测方法输出的最终包围框。Faster R-CNN 检测方法直接输出头部的包围框。此时，会出现同一个目标被两种方法都检测出来的情况，图像中同一个目标就会有多个包围框。更多的情况是，同一个目标会有不同打分的很多包围框，既可能是由同一方法产生，也可能是由两种方法共同产生。本文采用非极大值抑制的方法去除打分低且重复的包围框。

非极大值抑制的原理：(1) 将 DPM 和 Faster R-CNN 两种检测方法的目标打分归一化，分别表示为 $score_1$ 和 $score_2$ 。(2) 将 $score_1$ 和 $score_2$ ，混合，然后从高到低排序得到 $score$ 。(3) 选取包围框重合度阈值(IoU Threshold，实验中取 0.3)，按照算法 1 抑制掉打分低的包围框。

算法 1: 非极大值抑制(NMS)

输入: $score$, $score$ 对应的包围框 $bbox$

输出: 去重之后的包围框 $bbox'$

初始化抑制集合 $supress = \phi$

for $score_i$ **in** $score$ **do**

待处理的包围框对应的打分集合为 $score'$

$score' = score \setminus \{score_i, score_2, \dots, score_n\}$

for $score_j$ **in** $score'$ **do**


```

if ( $bbox_i \cap bbox_j$ ) / ( $bbox_i \cup bbox_j$ )  $\geq$  0.3 then
  suppress = suppress  $\cup$   $bbox_j$ 
end
end
end
 $bbox' = bbox / suppress$ 

```

2 实验结果

2.1 数据集

实验中用到三个数据集(PASCAL VOC 2012, mall_1 和 mall_2), PASCAL VOC 2012 用于训练 DPM 模型。mall_1 和 mall_2 数据集中图片来自于互联网上两个商场监控视频, 用于训练 Faster R-CNN 模型和最后测试。mall_1 数据集有 98 张图片, 尺寸为 1920×1080, 平均每张图片人工标注人头数在 130~148 个之间, 平均 140.86 个, 前 75 张作为训练集, 后 23 张作为测试集。mall_2 数据集有 612 张图片, 尺寸为 640×480, 平均每张图片人工标注人头数在 9~44 个之间, 平均每张 23.8 个, 前 500 张作为训练集, 后 112 张作为测试集。

2.2 实验过程

实验平台基本配置: Windows 7 操作系统, Intel(R) Core(TM) i7-3930K CPU, 16G 内存, NVIDIA GeForce GTX TITAN X 显卡。

2.2.1 对 DPM 检测方法进行并行优化

特征金字塔生成的并行优化。并行优化采用线程级并行, 使用 openMP 技术, 线程数分别设置为 1~6, 每次设置的线程数下测 10 次, 取平均值作为该线程数下的生成特征金字塔时间。表 1 为 mall_1 数据集上实验结果, 当线程数为 4 时, 生成特征金字塔平均时间为 2.83 s, 加速比最高, 为 1.54。

模型匹配过程中卷积操作的并行优化。各 filter 和某层特征金字塔的卷积操作做并行优化时采用 openMP 技术, 32 维 HOG 做 4 次循环展开, 每次展开的 8 维特征采用 AVX 扩展指令进行单指令多数据操作, 一次性处理 8 维特征。并行实验设置线

程数为 1~12, 每个线程数下取 10 次卷积时间。表 2 为 mall_1 数据集上实验结果, 当线程数为 6 时, 平均卷积时间为 3.50 s, 加速比最高, 为 17.01。

表 1 不同线程数下生成特征金字塔的时间和加速比

线程数	平均时间/s	加速比
1	4.37	1
2	3.44	1.27
3	3.23	1.35
4	2.83	1.54
5	3.19	1.37
6	3.24	1.35

表 2 不同线程数下卷积时间和加速比

线程数	平均时间/s	加速比
1*	59.54	1
1	10.44	5.70
4	3.93	15.15
5	3.68	16.18
6	3.50	17.01
7	3.66	16.27
8	3.59	16.59
12	3.72	16.01

注: 1*表示 1 个线程, 未使用循环展开和 AVX 指令, 1 以及其他线程数时都使用循环展开和 AVX 指令。

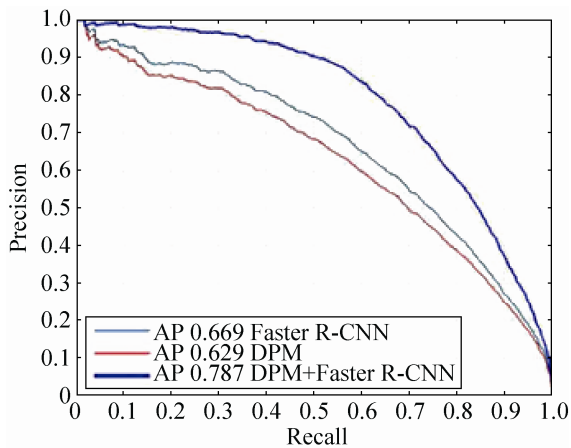
2.2.2 检测结果

DPM 模型训练过程参照文献[2]中训练过程。DPM 模型在 mall_1 测试集上的平均精度(AP, average precision)为 0.629, 在 mall_2 测试集上的平均精度为 0.447, 如图 5 所示。

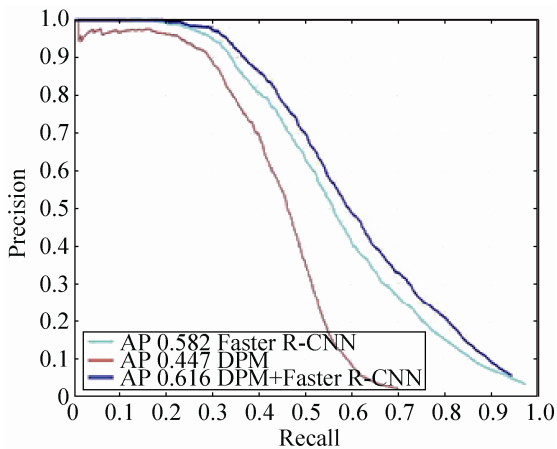
Faster R-CNN 模型训练过程参照文献[8]中四步训练法。实验过程中, 最后共享卷积层输出为 $W \times H$ (宽为 W , 长为 H , mall_1: 62×35, mall_2: 47×35), 每个位置放置 4 个不同大小的正方形 anchor(边长为 16×[0.5,0.9,1.2,1.5]), 在长和宽两个方向上步长均为 4 个像素, 所以每张图片总共有 $4 \times W \times 4 \times H \times 4$ 个 anchor。gt_pred_IoU 表示样本的真实包围框和预测的包围框的面积交集和并集的比, 正样本取 gt_pred_IoU \geq 0.7, 负样本取 IoU \leq 0.3。由于样本中负样本占绝大多数, 为

避免训练参数偏向负样本,在优化过程中正负样本比例取 4:3。Faster R-CNN 方法在 mall_1 上训练的模型在测试集上的平均精度(AP)为 0.669,在 mall_2 上训练的模型在测试集上的平均精度(AP)为 0.582,如图 5 所示。

多尺度检测方法在 mall_1 数据集上的平均精度(AP)为 0.787,较 Faster R-CNN 检测方法平均精度提高 17.64%,较 DPM 检测方法平均精度提高 25.1%。在 mall_2 数据集上,多尺度检测方法的平均精度为 0.616,较 Faster R-CNN 检测方法平均精度提高 5.82%,较 DPM 检测方法平均精度提高 29.04%。可见,DPM 检测方法和 Faster R-CNN 检测方法之间具有较大差异性,可以互补,共同了提高检测精度。



(a) 三种方法在 mall_1 数据集上的检测精度曲线



(b) 三种方法在 mall_2 数据集上的检测精度曲线

图 5 三种方法检测精度比较

2.2.3 非极大值抑制(NMS)阈值对多尺度检测方法平均检测精度的影响

多尺度检测方法通过非极大值抑制(NMS)方法将 DPM 检测方法和 Faster R-CNN 检测方法结合在一起。非极大值抑制的阈值(IoU Threshold)选取对平均精度(AP)有很大影响。表 3 和图 6 分别表示选取不同阈值时,剩下的检测目标数量和 mall_1 数据集上的平均精度(AP)。从表 3 可以看出,阈值越低,抑制掉的候选目标越多,这其中既包括冗余的候选目标,也包括真实存在的互相遮挡的目标。从图 6 可以看出,较低的阈值导致相对较低的召回率(recall)和相对较高的精度(precision)。反之,较高的阈值导致相对较高的召回率和相对较低的精度。当阈值取 0.30 时,平均精度最高,为 0.787。

表 3 不同 IoU Threshold 对检测目标数量的影响

数据集	NMS 前	IoU Threshold	NMS 后
mall_1	5 765 557	0	511 903
		0.20	656 480
		0.30	818 223
		0.40	1 006 635
		0.50	1 274 556
		0.60	1 751 762
mall_2	41 983 530	0	381 581
		0.20	491 255
		0.30	611 537
		0.40	753 542
		0.50	955 968
		0.60	1 308 553

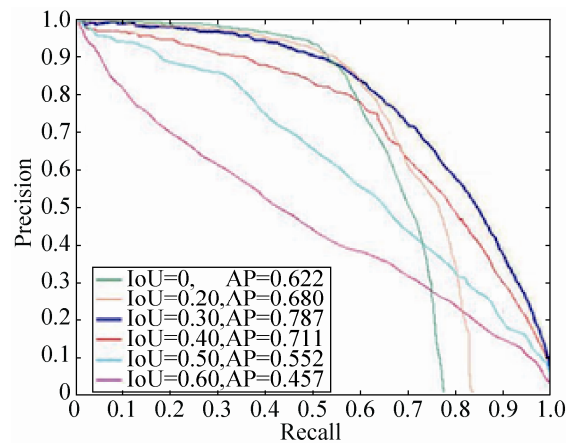


图 6 不同的 IoU 阈值对应的不同的平均检测精度

3 结论

本文提出了一种多尺度的目标检测方法, 针对密集人群场景中的目标检测问题, 将 DPM 模型匹配方法和 Faster R-CNN 检测方法互相结合, 相比单纯的 DPM 检测方法和 Faster R-CNN 检测方法提高了平均检测精度。另外, 我们对 DPM 检测方法两处具有并行性的地方做了优化操作, 大幅度提升了人群的检测速度。

下一步, 我们会扩充数据集, 进一步训练模型, 并对本文提出的多尺度检测方法进一步优化, 提高本方法的鲁棒性和实用价值。

参考文献:

- [1] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005, IEEE Computer Society Conference on. USA: IEEE, 2005, 1: 886-893.
- [2] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (S0162-8828), 2010, 32(9): 1627-1645.
- [3] Felzenszwalb P F, Girshick R B, McAllester D. Cascade object detection with deformable part models [C]//Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. USA:IEEE, 2010: 2241-2248.
- [4] Wang X, Yang M, Zhu S, et al. Regionlets for generic object detection [C]// Computer Vision (ICCV), 2013 IEEE International Conference on. USA: IEEE, 2013: 17-24.
- [5] Lempitsky V, Zisserman A. Learning to count objects in images [C]//Advances in Neural Information Processing Systems. Canada:NIPS,2010: 1324-1332.
- [6] Ren X, Ramanan D. Histograms of sparse codes for object detection [C]// Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. USA: IEEE, 2013: 3246-3253.
- [7] Haroon Idrees, Imran Saleemi, Cody Seibert, Mubarak Shah. Muti-Source Multi-Scale Counting in Extremely Dense Crowd Images [C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013. USA: IEEE, 2013: 2547-2554.
- [8] Bansal A, Venkatesh K S. People Counting in High Density Crowds from Still Images [J]. arXiv preprint arXiv (S2331-8422), 1507.08445, 2015. <https://arxiv.org>.
- [9] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. Spain:NIPS , 2012: 1097-1105.
- [10] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. USA:IEEE,2014: 580-587.
- [11] Girshick R. Fast r-cnn [C]//Proceedings of the IEEE International Conference on Computer Vision. USA: IEEE, 2015: 1440-1448.
- [12] 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.Canada:NIPS, 2015: 91-99.
- [13] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [J]. arXiv preprint arXiv (S2331-8422), 1506.02640,2015. <https://arxiv.org>.
- [14] Wang C, Zhang H, Yang L, et al. Deep People Counting in Extremely Dense Crowds [C]//Proceedings of the 23rd ACM international conference on Multimedia. Australia,ACM, 2015: 1299-1302.
- [15] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks [M]// Computer Vision – ECCV 2014. Springer International Publishing, 2013:818-833.
- [16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv (S2331-8422), 1409.1556, 2014. <https://arxiv.org>.
- [17] Rodriguez M, Laptev I, Sivic J, et al. Density-aware person detection and tracking in crowds [C]//2011 International Conference on Computer Vision. Spain:IEEE, 2011: 2423-2430.
- [18] Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition [J]. International Journal of Computer Vision (S0920-5691), 2005, 61(1): 55-79.
- [19] Fischler M A, Elschlager R A. The representation and matching of pictorial structures [J]. IEEE Transactions on computers (S0018-9340), 1973, 22(1): 67-92.
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. USA: IEEE, 2015: 1904-1916.