

## 基于多层次互补特征的通用目标检测模型

潘泓\* 金立左 夏思宇 夏良正  
(东南大学自动化学院 南京 210096)

**摘要:** 针对实际场景中多类目标检测问题, 该文提出了一种基于多层次特征表示和异质互补描述子的通用目标检测模型。该模型采用基于组件的目标描述思想, 提取目标不同层次的互补特征, 并将其统一到条件随机场(CRF)框架中。目标中单个组件及其局部特征对应 CRF 的一元节点, 组件之间的几何空间结构特征则体现在节点之间的两两连接关系上。通过引入节点支持向量机(SVM)分类器和边缘拓扑结构学习, 极大提高了模型的鉴别能力和灵活性。在 UIUC 多尺度数据集和 PASCAL VOC 2007 数据集上测试结果表明, 该文模型不仅能有效描述多类复杂目标, 还能较好地解决姿态、尺度、光照变化及局部遮挡情况下的目标检测问题。

**关键词:** 通用目标检测; 多层次特征描述; 异构特征; 条件随机场

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2012)07-1531-07

DOI: 10.3724/SP.J.1146.2011.01109

## A Hierarchical and Complementary Feature-based Model for Generic Object Detection

Pan Hong Jin Li-zuo Xia Si-yu Xia Liang-zheng  
(School of Automation, Southeast University, Nanjing 210096, China)

**Abstract:** This paper proposes a novel model based on the hierarchical representation using heterogeneous descriptors for multi-class generic object detection in real-world scenario. Following the idea of part-based object detection, the model extracts complementary features of object class at different levels and represents them with a unified Conditional Random Field (CRF) framework, in which the individual part and its local features correspond to a unary node and the interactions (edges) between pairwise nodes reflect the underlying geometrical structure of the object class. To improve the discrimination and flexibility of the proposed model, Support Vector Machine (SVM) classifier and the learning of edge structure are combined into CRF according to the geometrical topology of object class. Experimental results on UIUC multi-scale dataset and PASCAL VOC 2007 dataset show that the proposed model can not only effectively represent a variety of complex object classes, also successfully detect objects with pose, scale, illumination variations as well as partial occlusions.

**Key words:** Generic object detection; Hierarchical feature representation; Heterogeneous feature; Conditional Random Field (CRF)

### 1 引言

通用目标检测一直是机器视觉和模式识别领域核心和挑战性的研究课题之一, 它要解决的基本问题是, 如何选取合适的特征来描述目标并通过特征学习建立适合多类目标的特征模型, 从而有效检测出场景中的目标和非目标区域, 为后续目标识别算法提供信息。实际应用中, 光照、视角(目标姿态)变化和遮挡等因素导致目标外观和形状发生较大变化, 即使同类目标的不同个体也表现出很大差异,

这些因素都给复杂场景下通用目标检测带来了极大困难。

总体来看, 通用目标检测方法可分为两大类, 即基于整体的目标检测<sup>[1]</sup>及基于局部组件的目标检测<sup>[2-7]</sup>。前者采用判别式模型(discriminative model)提取整个目标区域中具有鉴别力的特征组成高维特征向量进行分类, 这种方法简单快速, 但对目标被局部遮挡或形变较大的柔性目标缺乏鲁棒性。后者则采用产生式模型(generative model)将整个目标分解成各个局部区域一组件, 通过构造组件检测器, 检测是否存在相应组件并结合它们之间的相对位置关系推断场景中是否存在感兴趣的待检测目标。该方法只需提取局部特征和它们的空间位置关系, 能有效解决目标局部遮挡和柔性目标形变较大的问题, 因此对不同特性(刚性、柔性)目标都能取得较

2011-10-26 收到, 2012-02-22 改回

国家自然科学基金(60805002, 90820009), 东南大学优秀青年教师教学科研资助计划(4008001015)和航空科学基金(20115169016)资助课题

\*通信作者: 潘泓 enhpan@seu.edu.cn

好的检测效果。

近期研究<sup>[2-8]</sup>显示,采用异构特征描述子和多层次特征表示方式能有效提高通用目标检测和类别分类能力。文献[2]提取目标的局部主成分分析 SIFT (PCA-SIFT)特征和全局形状上下文(shape context)特征,并采用多层 Adaboost 模型筛选出具有鉴别力的有效特征进行目标识别。文献[3]提取 4 种不同类型的局部描述子组成特征池,通过级联分类器训练目标检测器。文献[4]提取目标边缘成对相邻轮廓段和局部特征描述子,分别构造通用形状码本和特定类别外观码本,融合两种异质互补码本进行多类目标检测和分类。由于异构特征描述子提供了目标数据的互补视图,与基于单一特征类型的算法相比,上述算法都取得了更好的检测和分类结果。另一方面,层次化的特征描述方式<sup>[5-8]</sup>将目标不同尺度下的视图灵活地组织起来,为目标检测和识别提供了丰富信息。

此外,选择和学习合适的目标表示模型也极大地影响目标检测和分类系统的性能。近年来,一些机器学习和数据挖掘模型被成功引入目标检测领域,代表性的工作包括基于 boosting<sup>[9]</sup>的特征选择和模型学习方法,该方法采用 boosting 技术从大量目标特征中选出代表性的鉴别特征,通过聚类形成目标码本,并将码本和新样本特征比对进行目标检测。而文献[10]提出的 constellation 模型和文献[11]提出的 ISM (Implicit Shape Model)模型不仅考虑了目标组件的局部特征,还将各组件间的空间位置关系编码到目标码本中。由于引入了这些空间位置关系,因而极大提高了 constellation 和 ISM 模型的灵活性。但这些模型的问题在于需要考虑目标组件出现的概率分布,当目标构成组件数目较多时,模型就会过于复杂而无法实际使用。考虑到上述问题,近年来,人们逐步采用一些较为复杂的条件随机场(Conditional Random Field, CRF)模型<sup>[6-8,12,13]</sup>和潜在式模型(latent model)<sup>[13,14]</sup>来学习表示通用目标。CRF 和潜在式模型对目标出现概率建模,同时引入隐含变量增强了模型对图像这类结构型数据的处理能力,再考虑到目标组件之间存在着空间邻域约束关系,因此利用 CRF 和潜在式模型对结构型数据的处理能力可以更加自然地对这种邻域关系进行建模。Liu 等人<sup>[12]</sup>采用 CRF 模型学习目标特征,对静止图像和视频序列进行显著性目标检测。Schnitzspan 等人<sup>[13]</sup>将目标分解成多个组件,并用潜在条件随机场(latent CRF)描述目标模型,其中每个组件的标记被当作潜在条件随机场中的隐含变量

节点。由于引入了隐含变量节点,他们提出的潜在 CRF 模型能自动找到符合目标语义解释的组件表示形式。基于类似的思想,Felzenszwalb 等人<sup>[14]</sup>提出了隐条件随机场(hidden CRF)来解决部分组件未标记样本的训练问题。该方法把整个目标表示成各个组件的混和模型,将目标中未标记组件作为目标表示模型中的隐含变量,采用潜在 SVM 分类器预测隐含变量的类别。潜在 SVM 分类器采用迭代方式由已标记样本的组成部件训练得到。

目前的各种目标检测算法或是采用异构特征描述子,或是采用多层次特征表示,尚未将两者有效结合起来。考虑到上述两类目标描述形式的优点和基于组件表示的灵活性,本文将其融合起来,提出了目标的多层次互补特征表示方法,并将其统一到 CRF 框架中,形成了基于多层次互补特征的通用目标检测模型。该模型通过计算图像中不同大小子块的多特征描述子,提取目标局部、区域和全局对应的多层次互补特征。目标中单个组件及其局部互补特征对应 CRF 模型的一元节点(unary node),组件之间的几何空间结构特征则体现在节点之间的两两连接(即边缘组合)关系上。由于采用了基于组件的目标组织方式,又将反映目标不同尺度和不同特性视图的多层次互补特征集中到统一的 CRF 框架下,因此本文提出的通用目标检测模型具有较强的灵活性和鲁棒性,能有效解决目标由于局部遮挡、视角变化和形变造成特征丢失或改变而导致检测失败的问题。我们将本文提出的模型在标准数据集上进行测试,检测结果表明该模型有效提高了通用目标检测算法的性能。

## 2 目标检测模型的描述

给定图像  $I$ ,我们将目标检测转化为两类分类问题,用标记  $L \in \{1, -1\}$  表示图像中是否包含感兴趣的待检测目标( $L=1$ : 包含,  $L=-1$ : 不包含)。对目标识别而言,CRF 可以选择上下文相关特征,解决样本标记不平衡问题,通过对所有特征进行全局归一化,获得全局最优值,因此各种 CRF 的改进模型被广泛应用在图像目标检测和识别领域<sup>[6-8,12,13]</sup>。受这些工作启发,我们也采用 CRF 模型描述上述问题,检测结果通过计算后验概率  $P(L|I)$  得到。假设  $I = \{c_i\}_{i=1, \dots, N}$  是输入观察图像,且  $c_i \cap c_j = \phi$  ( $\forall i \neq j$ ),  $\cup c_i = I$  ( $i=1, \dots, N$ ),则  $I$  中每个组成部分  $c_i$  为局部观察,并存在一个随机变量集  $V = \{1, \dots, N\}$  与  $I = \{c_1, \dots, c_N\}$  对应。假设存在  $m \in V$ ,它邻近组件记为  $n$ ,则随机变量集  $V$  及其相应的邻域关系组成一个

无向图  $G(V, E)$ 。其中，节点集合  $V$  中的每个节点  $n$  对应一个二值分类标记  $l_n \in \{1, -1\}$ ，表示该节点对应的局部区域属于目标组件  $l_n = 1$  还是属于背景  $l_n = -1$ 。 $E$  表示无向图的边缘集合，每个边缘  $e$  连接两个独立节点  $(m, n)$ ，整个边缘集合  $E$  描述了节点之间的相互关联关系和 CRF 模型的内在结构。为提高本文模型的特征学习和分类能力，我们将 SVM 分类器融入 CRF 模型，对每个节点训练一个组件 SVM 分类器。此外，我们还引入模型拓扑结构学习来训练组件之间的空间位置关系和整个目标的拓扑结构。组件 SVM 分类器和拓扑结构学习极大提高了本文模型的特征学习和检测器的分类能力。

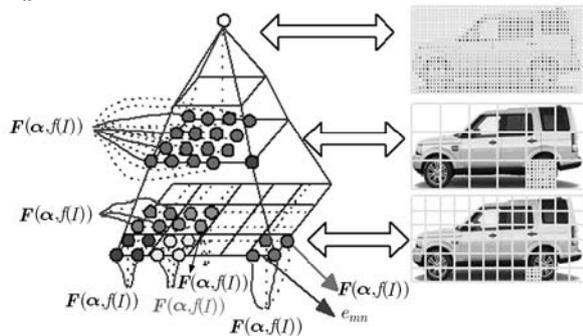
### 2.1 基于 CRF 的目标模型

本文模型建立在概率图模型基础上，通过计算式(1)所示的后验概率函数推断图像  $I$  中是否存在待检测目标。

$$P(L | I; \theta) = \frac{1}{Z(\theta)} \prod_{n \in V} u_n(l_n, I; \theta) \prod_{(m,n) \in E} v_{mn}(l_m, l_n, I; \theta) \quad (1)$$

式中  $Z(\theta)$  是配分函数，它保证模型有适合的归一化概率， $u_n$  和  $v_{mn}$  分别表示一元节点对应的强度函数和相邻节点之间边缘的强度函数， $\theta = \{\alpha, w, e\}$  是 CRF 的模型参数集合，其中， $\alpha$  和  $w$  对应一元节点强度函数的参数，而  $e$  则是边缘强度函数的参数。

图 1 给出了本文提出的基于多层次互补特征的通用目标检测模型。模型中一元节点的强度函数  $u_n$  可由节点所在区域的特定特征描述子  $f_n(I)$  计算得到，而不同尺度层次的一元节点强度函数分别提取了目标局部、区域和全局特征。如图 1 所示，最底层的一元节点强度函数对应目标局部特征。随着层次增加，一元节点个数逐渐减少，每个节点对应区域范围逐渐扩大。直到最顶层，只剩下一个节点，该节点的强度函数对应目标的全局特征。具体来说，对第  $n$  个一元节点，我们将该节点所在区域内的  $M$  个子块分类器响应串接成一个  $M$  维分类器响应矢量  $F(\alpha_n, f_n(I)) = [F(\alpha_{n,1}, f_n(I)), \dots, F(\alpha_{n,M}, f_n(I))]^T$ ，并采用式(2)所示的 softmax 函数定义一元节点强度函数  $u_n$ 。



$$\begin{aligned} \text{混合块分类器} \quad & F(\alpha, f(I)) = F^H(\alpha, f(I)) + F^C(\alpha, f(I)) + F^B(\alpha, f(I)) \\ \text{一元节点分类器} \quad & F(\alpha, f(I)) = [F(\alpha_1, f(I)), \dots, F(\alpha_M, f(I))]^T \end{aligned}$$

图 1 基于多层次互补特征的通用目标检测模型

$$u_n(l_n, I; \alpha, w) = \frac{\exp(l_n \cdot w_n^T F(\alpha_n, f_n(I)))}{\sum_{l \in \{-1, 1\}} \exp(l \cdot w_n^T F(\alpha_n, f_n(I)))} \quad (2)$$

式中  $w_n$  是对应  $M$  个子块分类器响应的加权矢量， $\alpha_n$  是分类器参数。

模型中边缘强度函数反映了节点之间的两两连接关系。对一对节点  $(m, n)$  及其对应节点特征  $f_m(I)$  和  $f_n(I)$ ，将  $f_m(I)$  和  $f_n(I)$  串接成一个特征矢量并仿照一元节点强度函数的 softmax 函数，可得式(3)所示的边缘强度函数  $v_{mn}$ 。

$$v_{mn}(l_m, l_n, I; e) = \frac{\exp([f_m(I), f_n(I)]^T e_{l_m l_n}^{mn})}{\sum_{\{a,b\} \in \{-1, 1\}} \exp([f_m(I), f_n(I)]^T e_{ab}^{mn})} \quad (3)$$

对连接每一组节点对  $(m, n)$  的边缘和这两个节点对应的可能标记组合  $(a, b) \in \{-1, 1\}$  (共 4 种)，我们用分类矢量  $e_{ab}^{mn}$  来描述节点间的空间相关性和不同特征类别的空间分布。

值得注意的是，通过同时考虑单个一元节点对应的局部组件特征和相邻节点之间的相关性(组件间的空间位置关系)，本文提出的模型不仅能提供具有鉴别力的目标多层次互补特征视图，同时还能学习目标内在的拓扑结构关系，这是本文模型区别于其它模型<sup>[10,11,14]</sup>之处。

### 2.2 多层次互补特征表示

我们使用了包括 HOG(Histograms of Oriented Gradients)<sup>[1]</sup>、CS-LBP(Center-Symmetric Local Binary Pattern)<sup>[15]</sup>和 BiCE(Binary Coherent Edge descriptors)<sup>[16]</sup>在内的 3 种异构特征描述子。其中，HOG 属于梯度方向算子，CS-LBP 是纹理算子，而 BiCE 则是基于边缘的描述子，用来描述在特定位置和方向上边缘段。可见，上述 3 种异构特征算子从不同方面描述图像特性，它们的组合提供了图像的互补视图，有助于增强目标的表示能力。

为得到图像的多层次互补特征描述，对每种特征描述子，在图 1 所示的层次结构中，将最底层图像分割成若干个互不重叠的网格(每个网格由  $i \times i$  像素组成)，并将相邻的  $j \times j$  个网格合并成一个子块。对上一层图像，我们将子块的长度和宽度分别加倍。这样随着层次升高，图像中子块个数逐渐减少，直到最顶层，整个图像只对应一个方块。通过计算不同层次上各子块的异构特征，并将相邻子块的特征合并成一个特征矢量，可以得到包括图像局部(最底层)、区域(中间层次)和全局(最顶层)特征表示。对不同特征描述子，我们选取不同的网格和子块设置。根据文献[1,15,16]建议，对 HOG 和 CS-LBP 描述子， $(i, j)$  分别设为(8,2)和(4,1)。对 BiCE 描述子，我们取

$i=8, j=3$ 。为消除特征向量中背景噪声的影响并减少其冗余性,对每种特征描述子分别进行主成分分析(PCA)降维。对每个子块,最终的互补特征由 PCA-HOG(11 维), PCA-CS-LBP(60 维)和 PCA-BiCE(256 维)组成。

为提高模型的鉴别能力,对每个子块的每种描述子分别训练一个 SVM 分类器,并将 3 种 SVM 分类器输出相加得到该子块的混合 SVM 分类器  $F(\alpha, f(I))$  输出。考虑到简化模型的复杂性,根据不同尺度级选择将不同邻域的子块合并成一个一元节点。因此,对每个一元节点  $n$ ,该节点的 SVM 分类器  $F(\alpha_n, f_n(I))$  输出就是该节点包含子块的混合 SVM 分类器输出组成的向量,即  $F(\alpha_n, f_n(I))=[F(\alpha_{n,1}, f_n(I)), \dots, F(\alpha_{n,M}, f_n(I))]^T$ 。

### 3 模型学习

影响本文算法性能的因素包括 CRF 模型参数集合  $\theta=\{\alpha, w, e\}$  和模型的空间结构关系。因此,模型学习就是根据一组给定的  $K$  幅包含目标方框的训练图像  $I=\{I^1, \dots, I^K\}$  及其标记  $L=\{L^1, \dots, L^K\}$ , 训练 CRF 模型参数集合  $\theta$  并学习反映目标类别几何拓扑关系的图模型结构,使模型能尽可能准确地预测新样本是来自目标类别还是其它类别。

#### 3.1 CRF 参数学习

本文使用极大似然法训练 CRF 模型参数,通过使式(4)所示的条件对数似然函数  $\ell(\theta)$  最大化寻找最优参数集合  $\theta=\{\alpha, w, e\}$ 。

$$\ell(\theta) = \sum_{k=1}^K \lg P(L^k | I^k; \theta) + \lg P(\theta) \quad (4)$$

式(4)中  $P(\theta)=P(w)P(e)$  是防止过训练的正规算子。下面我们分别介绍参数  $\alpha, w, e$  的训练过程。

参数  $\alpha$  是 SVM 分类器参数,可以通过文献[17]中提出的 primal SVM training 算法训练确定。其它参数  $w, e$  则通过共轭梯度法训练得到。为方便训练,我们假设节点参数  $w$  服从标准分布,边缘参数  $e$  服从拉普拉斯分布,即  $P(w) \sim \mathcal{N}(0,1)$ ,  $P(e) \sim \exp(-\|e\|)$ 。求条件对数似然函数  $\ell(\theta)$  关于  $w$  的偏导可得

$$\frac{\partial \ell(\theta)}{\partial w_n} = \sum_{k=1}^K E_{\tilde{P}(L|I^k; \theta)} [F(\alpha_n, f_n(I^k)) \cdot l_n \cdot u_n(l_n, I^k)] - E_{P(L|I)} [F(\alpha_n, f_n(I)) \cdot l_n \cdot u_n(l_n, I)] - w_n \quad (5)$$

式中  $E_{\tilde{P}(L|I)}[\cdot]$  为根据训练数据估计得到的概率分布期望值,  $E_{P(L|I)}[\cdot]$  是本文模型产生的后验概率(见式(1))期望值。同样地,对  $\ell(\theta)$  求其关于  $e$  的偏导可得

$$\frac{\partial \ell(\theta)}{\partial e_{l_n l_n}^{mn}} = \sum_{k=1}^K E_{\tilde{P}(L|I^k; \theta)} [[f_m(I^k), f_n(I^k)]^T \cdot v_{mn}(l_n, l_n, I^k)] - E_{P(L|I)} [[f_m(I), f_n(I)]^T \cdot v_{mn}(l_n, l_n, I)] - \text{sgn}(e_{l_n l_n}^{mn}) \quad (6)$$

由式(5)和式(6)可见,计算偏微分时需要计算条件分布  $P(L_n|I)$  和  $P(l_n, l_n|I)$ 。由于这些概率分布没有闭式的解析表达式,实际计算时我们采用 loopy belief propagation 算法<sup>[18]</sup>求其近似解。

#### 3.2 拓扑结构学习

文献[10,13]的研究结果表明,对柔性和易形变目标,选择一个具有合适拓扑结构的模型来描述目标组件之间的内在空间位置关系,将极大提高这些模型的目标检测准确率。受上述思想启发,我们也希望通过学习待检测目标类别的几何拓扑结构,以提高本文模型的灵活性和鲁棒性。

拓扑结构学习的基本思想就是寻找一组连接相邻节点的边缘集合  $E=\{(m^*, n^*)\}$ , 使这些边缘具有最大对数似然梯度  $\partial \lg P(l=1|I; \theta) / \partial e^{m^* n^*}$ , 因为这样的边缘能最大程度地将目标( $l=1$ )和背景( $l=-1$ )区分开来。为找到满足上述条件的边缘集合  $E=\{(m^*, n^*)\}$ , 我们先清空集合  $E$ , 并采用迭代方式,在现有所有可能的边缘组合  $\{(m, n) \in V \times V - E\}$  中选出使对数似然比最大化的边缘  $(m^*, n^*)$  (如式(7)所示), 将其加入边缘集合  $E=\{(m^*, n^*)\}$  中。为去除误选的边缘,每次迭代结束时,将边缘权值小于阈值  $\tau_1$  或梯度绝对值小于阈值  $\tau_2$  的边缘从集合  $E$  中剔除。

$$(m^*, n^*) = \arg \max_{(m, n) \in V \times V - E} \left\| \frac{\partial \ell(\theta)}{\partial e_{11}^{mn}} - \frac{\partial \ell(\theta)}{\partial e_{-1-1}^{mn}} \right\| \quad (7)$$

### 4 实验结果及讨论

为检验所提模型的目标检测性能,本文先在 UIUC 多尺度汽车数据集<sup>[19]</sup>上进行测试。该数据集包含了各种和训练集中目标大小不同的测试样本。我们随机从训练集中选取 300 幅包含位置信息的汽车图像作为正样本,250 幅不含汽车的图像作为负样本。仿照文献[19]中实验步骤和测试准则,比较了本文模型和其它算法的检测结果。图 2 给出了各种方法的 Precision-Recall(PR)曲线。其中,2AS-SVM<sup>[4]</sup>是融合通用形状码本和特定目标类别外观码本的检测模型;RDM<sup>[20]</sup>目标检测模型是先由产生式模型得到一系列目标位置和尺度信息的假设,再由训练好的判别分类器(如 SVM)检验所有假设,判断目标存在与否并得到其真实位置信息;SPR<sup>[19]</sup>则是基于稀疏组件表示的多尺度汽车检测算法。

由 PR 曲线可见,本文模型的 EER(Equal Error Rate)达到了 95.54%,明显高于其它算法。尽管由

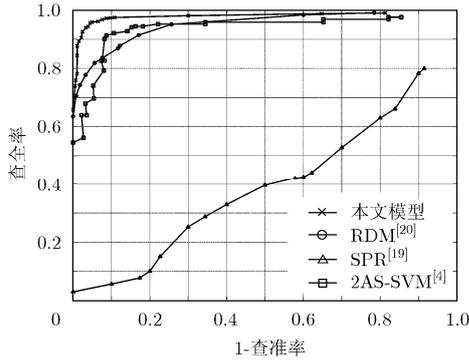


图 2 各种方法在 UIUC 多尺度测试集上的 Precision-Recall 曲线

于融合了形状和外观特征, 2AS-SVM 模型本身的检测性能已经较好(EER = 91.31%)。但 2AS-SVM 模型只提取了目标的单尺度互补特征, 同时只用简单的 ISM 模型描述目标的空间位置关系。而本文模型不仅考虑了多个尺度层次的互补特征, 还采用复杂的 CRF 模型学习和描述目标的内在几何拓扑结构。因此与 2AS-SVM 模型相比, 本文模型 PR 曲线的平均检测精度提高了 5.87%。由于结合了产生式模型和判别式分类器的各自优点, RDM 模型也取得了较好的检测效果(EER = 87.80%)。与上述几种算法相比, 文献[19]中提出的 SPR 算法检测性能明显较差, EER 不到 45%。这是由于, SPR 算法采用像素点灰度值作为目标特征, 并通过聚类方式生成目标码本, 因此, 无论特征的鉴别能力, 还是码本的类别“特殊性”都不强。此外, 与本文模型采用的拓扑结构学习以及 2AS-SVM 和 RDM 算法中采用的 ISM 模型相比, SPR 算法中组件之间位置约束关系太过简单, 无法充分编码目标组件的几何空间模型。

图 3 给出了各种算法在复杂场景下, 目标被局部遮挡(第 1 行)、光照不均匀(第 2 行)和尺度变化(第

3 行)情况下的部分检测结果。其中, 第 1 列到第 4 列分别对应 SPR, RDM, 2AS-SVM 和本文模型的检测结果, 实线框表示正确的检测结果, 虚线框表示检测错误的情况。由结果可见, SPR 模型仅对目标尺度变化具有较好的鲁棒性, 对目标表面光照不均匀和局部被遮挡时, 分别出现漏检和误检, 主要原因在于 SPR 模型中使用的特征和空间模型太过简单。RDM 模型在光照不均匀时, 虽然能检测到目标, 但给出的目标位置不准确(第 2 行第 2 列)。由于 RDM 模型中空间位置约束只考虑了组件的局部空间分布, 在两个目标相距很近时, 通常会把前一物体的一部分和后面物体的一部分误认为是一个完整目标, 因此出现误检(第 1 行第 2 列中的虚线框)。融合了形状和外观特征的 2AS-SVM 较好地克服了上述问题, 在目标被局部遮挡和尺度发生较大变化时, 都能获得较好的检测效果。但 2AS-SVM 只提取单一尺度下目标特征, 同时 ISM 模型也未能充分描述组件空间关系, 对复杂场景中出现的与待测目标相似区域时, 通常会出现误检(第 2 行第 3 列中的虚线框)。而本文模型在多个尺度下提取多种互补异构特征表示目标, 结合 CRF 模型学习并有效描述组件几何拓扑关系, 因此克服了其他方法的缺点, 对上述目标各种变化和局部遮挡都能较准确地提供目标位置信息。

为进一步测试本文模型的灵活性和对多类目标及尺度、姿态、形状变化、局部遮挡和复杂场景的鲁棒性, 我们在 PASCAL VOC 2007 测试集<sup>[21]</sup>上进行实验。该测试集包含 20 种不同的目标类别, 是目前国际上公认的通用目标类别分类、检测和分割的标准测试集。为显示本文模型中各功能模块的有效性, 我们比较了本文模型、只基于多层次 HOG 描述子的模型(hHOG)、只基于最底层尺度级互补特征

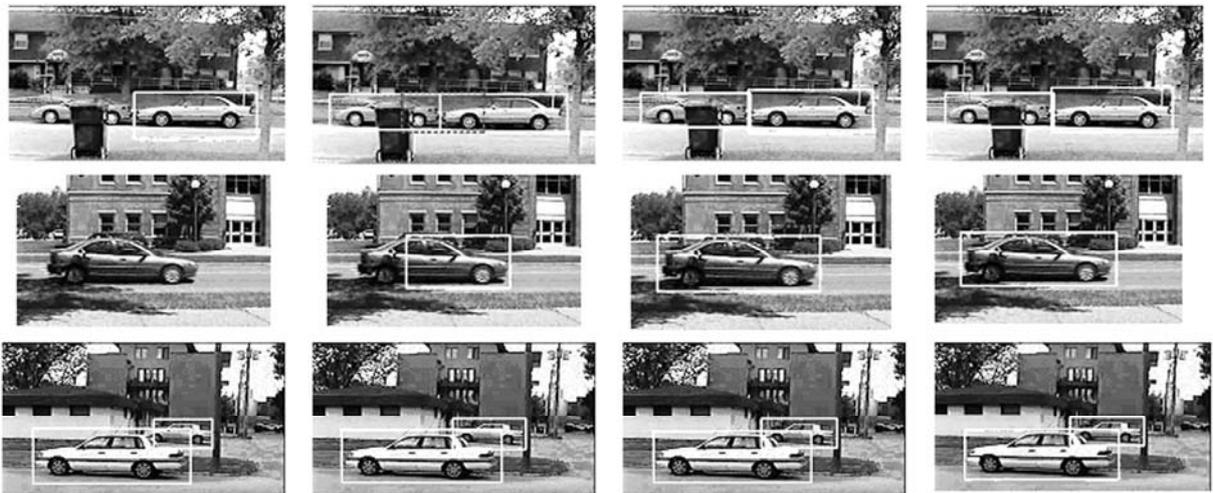


图 3 各种方法在 UIUC 多尺度测试集上的部分检测结果(第 1 列到第 4 列分别对应 SPR, RDM, 2AS-SVM 和本文模型的检测结果)

的模型(complementary)和其它代表性算法的检测性能。表1给出了各种方法对每一类目标的检测平均精度AP(average precision)。在所有检测的20类目标中,本文模型对其中14类目标取得了最好的AP。与文献[13,14]的结果相比,本文模型的AP分别提高了7.26%和5.89%。与只基于多层次HOG描述子和只基于最底层尺度级互补特征的模型相比,本文模型的AP分别提高了4.30%和6.41%。图4给出了本文方法在PASCAL VOC 2007数据集上的部分检测结果。由结果可见,同时使用多层次特征表示和异质互补描述子的本文模型确实可以有效描述多种性质不同(刚性、柔性和易形变)的复杂目标类别,并有助于提高通用目标检测性能。

在2.4 G CPU + 1 G RAM + Windows XP 组成的测试平台上,训练本文模型需要约10 h,检测一整幅640×480的图像大约需要8 s,检测时间主要集中在图像的分层特征计算上。除了对算法进一步优化,也可通过减少尺度级数或降低PCA特征维

数,加快特征计算和比对速度,减少算法检测时间。此外,算法中用到所有异构特征都是基于直方图的特征,因此采用文献[22]提出的积分直方图计算方法可望提高本文算法效率。

## 5 结论

针对实际场景中多类目标检测问题,本文提出了一种基于多层次特征表示和异质互补描述子的通用目标检测模型。该模型采用基于组件的目标描述思想,提取目标不同层次的局部、区域和全局互补特征,并将其融合到统一的CRF框架中。为提高模型对不同目标类别和各种场景变化的鲁棒性,我们将SVM分类器引入CRF框架,并通过模型的拓扑结构训练学习目标类别的内在几何拓扑结构。实验结果表明,本文提出的模型不仅能有效描述刚性、柔性和易形变等多类复杂目标,还能较好地解决了目标在姿态、尺度、光照变化及局部遮挡情况下的检测问题。

表1 各种方法在PASCAL VOC 2007数据集上对20类目标的AP结果比较(%)

目标类别	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow
本文模型	<b>35.4</b>	<b>57.2</b>	<b>9.7</b>	15.3	27.4	<b>45.7</b>	49.0	<b>24.8</b>	15.9	<b>20.5</b>
hHOG	33.5	57.0	7.6	14.8	26.8	45.1	47.5	23.7	13.0	20.2
complementary	32.3	56.6	4.4	13.1	26.6	43.9	47.0	23.1	10.6	20.8
Felzenszwalb <i>et al.</i> <sup>[14]</sup>	32.8	56.8	2.5	<b>16.8</b>	<b>28.5</b>	39.7	<b>51.6</b>	21.3	<b>17.9</b>	18.5
Schnitzspan <i>et al.</i> <sup>[13]</sup>	31.9	57.0	9.1	15.2	26.0	42.7	49.3	14.5	15.2	18.5
目标类别	table	dog	horse	mbik	pers	plant	sheep	sofa	train	tv
本文模型	24.3	<b>12.6</b>	<b>49.5</b>	<b>43.0</b>	36.2	<b>14.9</b>	<b>21.6</b>	<b>26.2</b>	<b>41.4</b>	<b>45.5</b>
hHOG	23.5	11.8	47.9	41.3	35.8	13.7	19.9	25.9	39.3	42.4
complementary	23.2	11.6	47.1	40.9	35.6	13.2	20.4	24.5	39.5	43.6
Felzenszwalb <i>et al.</i> <sup>[14]</sup>	<b>25.9</b>	8.8	49.2	41.2	<b>36.8</b>	14.6	16.2	24.4	39.2	39.1
Schnitzspan <i>et al.</i> <sup>[13]</sup>	24.2	11.8	49.1	41.9	35.7	14.5	18.9	23.3	34.3	41.3



图4 本文模型在PASCAL VOC 2007数据集上得到的部分检测结果

## 参考文献

- [1] Dalal N and Triggs B. Histograms of oriented gradients for human detection [C]. Proceedings of Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, (1): 886-893.
- [2] Zhang W, Yu B, Zelinsky G J, *et al.*. Object class recognition using multiple layer boosting with heterogeneous features [C]. Proceedings of Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, (2): 323-330.
- [3] Perrotton X, Sturzel M, and Roux M. Mining families of features for efficient object detection [C]. Proceedings of International Conference on Image Processing, Cairo, Egypt, 2009: 857-860.

- [4] Pan H, Zhu Y P, Xia L Z, *et al.*. Combining generic and class-specific codebooks for object categorization and detection [C]. Proceedings of International Conference on Acoustic, Speech and Signal Processing, Prague, Czech, 2011: 2264–2267.
- [5] Epshtein B and Ullman S. Feature hierarchies for object classification [C]. Proceedings of International Conference on Computer Vision, Beijing, China, 2005, (1): 220–227.
- [6] Schnitzspan P, Fritz M, Roth S, *et al.*. Discriminative structure learning of hierarchical representations for object detection [C]. Proceedings of Computer Vision and Pattern Recognition, Miami, FL, USA, 2009: 2238–2245.
- [7] Schnitzspan P, Fritz M, and Schiele B. Hierarchical support vector random fields: joint training to combine local and global features [C]. Proceedings of European Conference on Computer Vision, Marseille, France, 2008, (2): 527–540.
- [8] Ladicky L, Russell C, Kohli P, *et al.*. Associative hierarchical CRFs for object class image segmentation [C]. Proceedings of International Conference on Computer Vision, Kyoto, Japan, 2009: 739–746.
- [9] Freund Y and Schapire R E. Experiments with a new boosting algorithm [C]. Proceedings of International Conference on Machine Learning, Bari, Italy, 1996: 148–156.
- [10] Fergus R, Perona P, and Zisserman A. Object class recognition by unsupervised scale-invariant learning [C]. Proceedings of Computer Vision and Pattern Recognition, Madison, WI, USA, 2003, (2): 264–271.
- [11] Leibe B, Leonardis A, and Schiele B. Combined object categorization and segmentation with an implicit shape model [C]. Proceedings of European Conference on Computer Vision, Prague, Czech, 2004: 17–32.
- [12] Liu T, Yuan Z, Sun J, *et al.*. Learning to detect a salient object [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(2): 353–367.
- [13] Schnitzspan P, Roth S, and Schiele B. Automatic discovery of meaningful object parts with latent CRFs [C]. Proceedings of Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 121–128.
- [14] Felzenszwalb P, Girshick R, McAllester D, *et al.*. Object detection with discriminatively trained part-based models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627–1645.
- [15] Heikkila M, Pietikainen M, and Schmid C. Description of interest regions with local binary patterns [J]. *Pattern Recognition*, 2009, 42(3): 425–436.
- [16] Zitnick C L. Binary coherent edge descriptors [C]. Proceedings of European Conference on Computer Vision, Crete, Greece, 2010, (2): 170–182.
- [17] Chapelle O. Training a support vector machine in the primal [J]. *Neural Computation*, 2007, 19(5): 1155–1178.
- [18] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. 1988.
- [19] Agarwal S, Awan A, and Roth D. Learning to detect objects in image via a sparse, part-based representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(11): 1475–1490.
- [20] Fritz M, Leibe B, Caputo B, *et al.*. Integrating representative and discriminant models for object category detection [C]. Proceedings of International Conference on Computer Vision, Beijing, China, 2005, (2): 1363–1370.
- [21] Everingham M, Van Gool L, Williams C, *et al.*. The PASCAL visual object classes (VOC) challenge [J]. *International Journal of Computer Vision*, 2010, 88(2): 303–338.
- [22] Porikli F. Integral histogram: a fast way to extract histogram in Cartesian spaces [C]. Proceedings of Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, (1): 829–836.
- 潘泓：男，1974年生，博士，副研究员，研究领域为模式识别、目标检测与识别、机器学习、计算机视觉、多媒体信号处理。
- 金立左：男，1972年生，博士，副教授，研究领域为模式识别、目标跟踪、机器学习、计算机视觉、多媒体信号处理。
- 夏思宇：男，1978年生，博士生，研究领域为模式识别、机器学习、计算机视觉。
- 夏良正：男，1942年生，教授，博士生导师，研究领域为图像处理、计算机视觉、数字视频处理、自动目标识别。