

## 基于空间位置关系改进的物体识别上下文模型

李宏伟, 徐晓滨, 文成林

(杭州电子科技大学系统科学与控制工程研究所, 浙江 杭州 310018)

**摘要:**常见的物体识别算法是基于图像局部低维特征的,在图像成像质量较差、分辨率较低情况下存在不确定和歧义性;图像上下文包含了场景信息以及物体之间彼此关联的丰富信息,可作为图像低维特征补充从而有助于提高物体识别率。该文在已有上下文模型基础上,进一步考虑了物体的空间位置关系信息,将图像全局特征、物体同现性关系和空间位置关系信息、局部检测器输出整合到同一个概率框架中,并充分利用树结构图模型高效推理的优势,改进了物体识别性能。最后通过标准图像集进行算法验证测试和对比来说明该文算法的有效性。

**关键词:**上下文模型; 树模型; 同现树; 空间关系; 物体识别

**中图分类号:** TP181

**文献标识码:** A

**文章编号:** 1001-9146(2013)06-0103-04

### 0 引言

常见的单目标检测器仅仅能够识别图像局部的特定物体种类,检测结果可能在图像上下文中形成语义冲突<sup>[1]</sup>。大量关于视觉认知和认知神经的研究表明,人类对实际场景的理解总是建立在一定的上下文信息基础之上<sup>[2]</sup>。针对上下文信息利用的研究已取得很大进展。文献[3]首先利用局部特征来预测每个结点的标签,然后利用同现性关系来调整已预测的物体标签。文献[1]提出了一种基于树模型表示的上下文模型(Tree-Context),本文在其基础上,进一步考虑了物体的空间位置关系信息,将多种信息源整合到同一个概率框架中,利用树结构图模型高效推理的优势来完成图像中的物体识别,所得到的检测结果能够改进物体的识别性能,并给出了具有一致性的场景解释。

### 1 基于空间位置改进的物体识别上下文模型

本文模型在 Tree-Context 模型基础上进一步整合了空间位置关系信息,为区别于文献[1]提出的上下文模型,命名本文所给出的上下文模型为 Relation-Context 模型。主要由上下文模型建模、模型学习和模型推理 3 部分组成。

#### 1.1 上下文模型建模

上下文模型由先验模型和量测模型两部分组成。先验模型是由训练得到的先验知识所建立的,借助于树结构图模型来表示,包含同现性先验模型、空间位置先验模型和空间关系先验模型 3 部分。而量测模型由待检集合学习得到,包括全局图像特征、局部位置和空间关系量测模型。

同现性由二元树模型来表示,结点  $b_i$  代表物体  $i$  是否出现在图像中。选定根结点后,所有二值变量的联合概率分布可通过树结构分解来表示  $p(b) = p(b_{root}) \prod_i p(b_i | b_{pa(i)})$ ,  $pa(i)$  是结点  $i$  的父结点。

收稿日期: 2013-10-08

基金项目: 国家自然科学基金重点资助项目(60934009), 国家自然科学基金资助项目(61004070, 61203094)

作者简介: 李宏伟(1987-), 男, 河南林州人, 在读研究生, 图模型与信息融合。

在对空间位置关系建模时,采用  $L_y$  和  $L_z$  来描述物体的垂直位置和距离观测者的远近,位置变量表示为  $L_i = (L_y, \lg L_z)$ 。  $L_i$  的依赖关系具有和同现二元树一样的依赖结构  $p(L_i | b) = p(L_{root} | b_{root}) \prod_i p(L_i | L_{pa(i)}, b_i, b_{pa(i)})$ 。边势函数  $p(L_i | L_{pa(i)}, b_i, b_{pa(i)})$  表示子结点在其父结点位置条  $L_{pa(i)}$ , 子结点出现变量  $b_i$  以及父结点出现变量  $b_{pa(i)}$  条件下该结点出现的位置,如图 1(a) 所示。

两两物体之间的空间相对位置关系对于目标识别十分重要。文献[4]指出,图像中分割区域的关系可以量化为 5 种关系:上、下、内、外、周围。本文在此基础上将物体之间的空间关系改为上(UP)、下(DOWN)、周围(AROUND)、重叠(OVERLAP)4 种。其中物体的内外关系以及重叠程度均可以由重叠属性来表示,物体的空间关系表示为:

$$r_{ij} = \frac{w_{upij}}{UP} + \frac{w_{downij}}{DOWN} + \frac{w_{aroundij}}{AROUND} + \frac{w_{overlapij}}{OVERLAP} \quad (1)$$

这样物体  $i$  和  $j$  的关系可以表示为一个向量:  $r_{ij} = [w_{upij}, w_{downij}, w_{aroundij}, w_{overlapij}]$ 。

量测模型包括图像全局特征和局部检测器输出两部分。概要描述子(gist descriptor)<sup>[5]</sup>是图像的一种低维表示,刻画了图像的粗纹理和空间布局。对于待检图像中的每个候选窗口,检测器都给出了一个检测得分  $s_{ik}$  以及位置变量  $w_{ik}$ 。图 1(b) 给出了物体  $i$  的量测模型,已将 gist 和基线(baseline)检测器的输出整合到了先验模型中,阴影框来表示物体  $i$  的不同候选窗口。

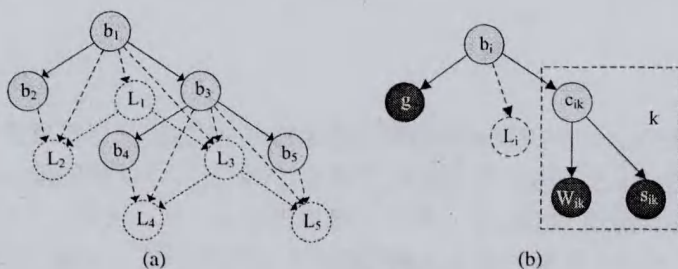


图 1 上下文图模型表示

## 1.2 模型学习

本文采用 Chow-Liu 算法来物体依赖结构,通过选择具有最强依赖关系的物体对来学习树模型,就可以自然形成层级结构模型。空间关系参数可以从训练图像中找到所有出现在同现树种的父子对。对于一幅图像中,若物体对  $(p, s)$  出现多次,则对父物体的所有边界框计算均值,然后再分别对每个子物体计算其空间关系  $r_{prior}(p, s) = \frac{1}{nm} \sum_i^n \sum_j^m r_{ij}(p, s)$  其中  $n$  是训练图像个数,  $m$  是第  $i$  幅图像中物体对  $(p, s)$  出现的次数,  $r_{ij}(p, s)$  是物体对关系向量。

在量测模型中,  $p(g | b_i)$  可以从训练图像的 gist descriptors 训练得到。  $p(c_{ik} | s_{ik})$  可以经由 logistic 回归训练得到,进而可计算出  $p(s_{ik} | c_{ik}) = p(c_{ik} | s_{ik}) p(s_{ik}) / p(c_{ik})$ 。正确检测概率  $p(c_{ik} | b_i)$  可以通过计算训练集合中标签总数和正确检测个数计算得到。

## 1.3 利用树模型进行交替推理

基于上下文的物体识别任务就是在给定概要描述子  $g$ 、基线检测器得到的候选窗口位置变量  $W = \{W_{ik}\}$  和窗口得分  $s = \{s_{ik}\}$  以及空间先验关系  $R = \{R_{ijk}\}$  条件下,推理得到图像中物体是否出现  $b = \{b_i\}$ 、该检测是否正确  $c = \{c_{ik}\}$ 、以及物体的期望位置  $L = \{L_i\}$ 。该问题可以通过以下优化问题解决:

$$\hat{b}, \hat{c}, \hat{L} = \underset{b, c, L}{\operatorname{argmax}} p(b, c, L | g, W, s, R) \quad (2)$$

因为模型中即包含二元变量也有高斯变量,精确推理很困难,需要利用 1.2 节中树模型推理算法来进行迭代推理求解。在变量  $b$  和  $c$  已知条件下,位置变量  $L$  构成高斯树;而在  $L$  和  $R$  条件下,变量  $b$  和



c 共同构成二元树模型。因此,可采用交替推理的方式来得到  $b, c, \mathbf{L}$ 。算法推理步骤如下:

(1) 在第一次迭代过程中,忽略位置信息  $\mathbf{W}$ ,由空间关系  $\mathbf{R}$ ,gist 和窗口评分给出  $b$  和  $c$  的估计  $\hat{b}, \hat{c} = \operatorname{argmax}_{b,c} p(b, c | s, g)$ ;

(2) 计算  $b$  和  $c$  的最大后验估计  $\hat{L} = \operatorname{argmax}_{\mathbf{L}} p(\mathbf{L} | \hat{b}, \hat{c}, \mathbf{W})$ ,以及两两物体的空间关系  $\hat{R} = \operatorname{argmax}_{\mathbf{R}} p(\mathbf{R} | \hat{b}, \hat{c}, \mathbf{W})$ 。然后利用高斯树模型推理可得到物体的期望位置;

(3) 然后在位置估计  $\hat{\mathbf{L}}$  条件下,可以重新估计变量  $b$  和  $c$ :  $\hat{b}, \hat{c} = \operatorname{argmax}_{b,c} p(b, c | s, g, \hat{\mathbf{L}}, \hat{\mathbf{R}}, \mathbf{W})$ ;

(4) 最后一步,每个计算  $b$  和  $c$  联合分布的边缘概率。 $p(b_i = 1 | s, g, \hat{\mathbf{L}}, \hat{\mathbf{R}}, \mathbf{W})$  即为出现预测,而检测变量边缘概率  $p(c_{ik} = 1 | s, g, \hat{\mathbf{L}}, \hat{\mathbf{R}}, \mathbf{W})$  即为检测的正误。

## 2 仿真实验

本文利用 SUN09 数据集来进行算法评估。将 SUN09 数据集分为训练集和测试集,首先用基线检测器对训练集和测试集进行物体检测识别,得到的结果包括检测候选窗口和检测评分。部分实验结果示例如图 2 所示,从左到右依次是基线检测器、Tree-Context 和 Relation-Context 模型的检测结果,通过分析检测结果可看出上下文模型对于物体检测性能的改进。

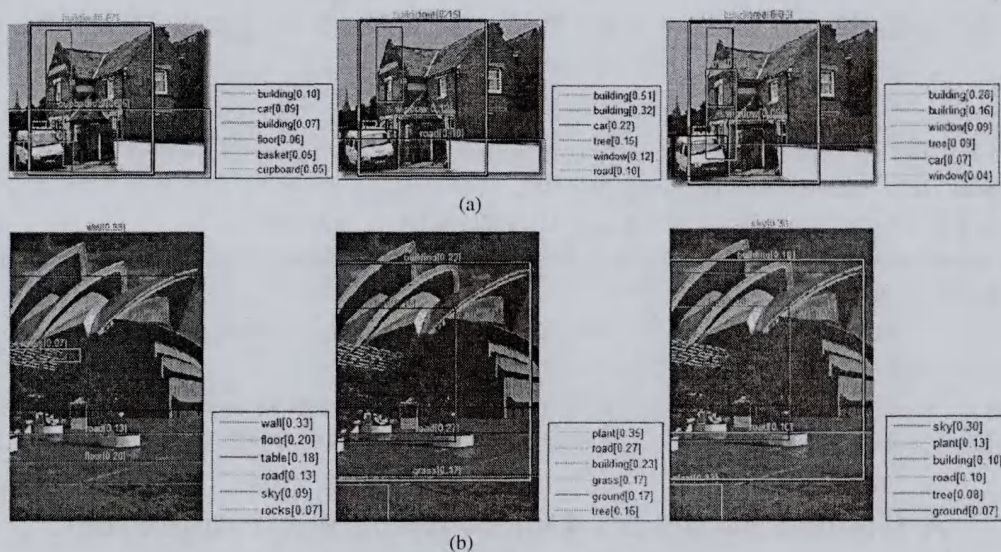


图2 各种上下文模型检测结果

上下文模型的显著优势在于可将上下文语义冲突的检测结果过滤掉,图 2(a) 中两个上下文模型都将检测结果中不符合场景上下文的 cupboard 筛选掉,同样还有图 2(b) 中 rocks 等。此外本文算法由于充分利用了物体的位置空间关系,进一步提高了物体检测正确率。例如图 2(a) 中 Relation-Context 检测结果 6 (window[0.04]),根据空间关系模型“窗户”总是位于“楼房”内部,该结果符合这一关系因而具有较高置信度。

模型位置预测精度和出现预测精度的对比如图 3 所示。由图 3 可看出,对于物体出现位置预测两种上下文模型改进效果很接近,这是因为基于空间关系改进主要是关注物体的出现是否符合常规位置,而不对位置做预测改进。而在出现预测精度上,Relation-Context 模型更加充分的利用了上下文信息,相对于 Tree-Context 模型取得了较大改进。

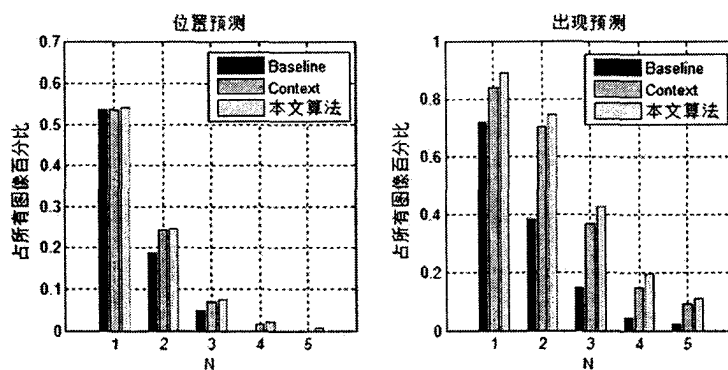


图3 模型位置预测精度和出现预测精度的对比

### 3 结束语

上下文信息包含了场景信息以及物体之间彼此联系的丰富信息,可作为图像低维特征补充来提高目标识别率。本文在原有上下文模型基础上,进一步考虑了物体的空间位置关系信息,将多个信息源整合到同一个概率框架中,并充分发挥树结构图模型高效推理的优势,完成图像中物体的目标识别。此外,本文也存在一些不足和亟待完善的地方,如对于目标空间关系信息的建模利用的是物体在原始图像中的位置,未考虑图像扭曲、图像采集角度等因素,这些问题需要在下一步研究中进行研究与改进。

### 参考文献

- [1] Choi M J, Torralba A, Willsky A S. A tree-based context model for object recognition[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions, 2012, 34(2): 240 - 252.
- [2] 高常鑫. 基于上下文的目标检测与识别方法研究[D]. 武汉: 华中科技大学, 2010.
- [3] Rabinovich A, Vedaldi A, Galleguillos C. Objects in context[C]. Rio de Janeiro: IEEE 11th International Conference on Computer Vision, 2007: 1 - 8.
- [4] Galleguillos C, Rabinovich A, Belongie S. Object categorization using co-occurrence, location and appearance[C]. Anchorage: IEEE Conference, Computer Vision and Pattern Recognition, 2008: 1 - 8.
- [5] Murphy K, Torralba A, Freeman W. Using the forest to see the trees: a graphical model relating features, objects and scenes[J]. Advances in neural information processing systems, 2003, 16: 1 499 - 1 507.

## A Context Model for Object Recognition Improved by Spatial Relationships

LI Hong-wei, XU Xiao-bin, WEN Cheng-lin

(Institute of System Science and Control Engineering, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China)

**Abstract:** Most object recognition algorithms are based on the local low-dimensional characteristics of the images. Images of poor quality or low resolution prompt uncertainty and ambiguity for object recognition. Context information contains scenes and spatial relationships information among objects which can be used as the supplementary for low-dimensional image features to improve the recognition. In this paper, we present an improved context model for object recognition in which we further consider the spatial relationships. We integrate all the information in a unified probabilistic framework taking full advantage of the efficient inference algorithms of tree models. Our model can improve the recognition result and give a consistent scene interpretation of the image. At last, we use a standard dataset to test our algorithm and make comparison to show the effectiveness.

**Key words:** context model; tree model; co-occurrence tree; spatial relationships; object recognition