

# 基于几何信息先验分布的似物性推荐方法

刘质彬 赵启阳

(北京航空航天大学计算机学院 北京 100191)

**摘要** 似物性推荐是计算机视觉研究中的热门问题,其目的是用尽可能少的推荐窗口涵盖可能的兴趣目标,以显著地提升目标检测任务的计算效率。从组合几何学角度对该问题进行了分析,一种“完全窗口覆盖”的方法被提出,用少量窗口即可覆盖所有可能目标区域。对于尺寸不大于  $512 \times 512$  的图像,约 19000 个窗口即可覆盖所有尺寸不小于  $16 \times 16$  的目标区域。基于目标矩形的位置、尺寸的先验分布,可以使用贪心策略进一步地缩减窗口数量。为了适应不同图像集在小概率样本上的差异,提出了一种融合了贪心和随机方法的混合机制,其所需的计算量非常小,而且具有很好的泛化能力。在 VOC2007 测试集上,该混合机制可以在 1000 个推荐窗口上取得 94.52% 的召回率,其中在前 10 个热点推荐窗口上的召回率比其他方法平均高出 13.99%~40.29%。

**关键词** 目标检测,似物性推荐,几何信息,完全覆盖集合,混合机制

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.9.060

## Objectness Proposal Based on Prior Distribution of Geometric Characteristics of Object Regions

LIU Zhi-bin ZHAO Qi-yang

(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

**Abstract** Objectness proposal is an emerging problem aiming to improve the efficiency of object detection by reducing candidate windows. The problem was analyzed from the perspective of combinatorial geometric, and a method was proposed to construct full cover sets which cover all possible object rectangles with a rather small amount of windows. For images no larger than  $512 \times 512$ , supposing all object rectangles are not smaller than  $16 \times 16$ , nearly 19000 windows are sufficient to make up a full cover set. By exploiting the prior distribution of locations/sizes of object rectangles, this amount can be reduced further in a greedy mode. In order to address the diversity of low-probability samples of different image sets, a hybrid scheme mixing the greedy and random methods which has good generality was presented. The new scheme recalls 94.52% object rectangles with 1000 proposal windows, and its DRs on the first ten hot proposal windows are 13.99%~40.29% higher than existing methods in average.

**Keywords** Object detection, Objectness proposal, Geometric characteristics, Full cover set, Hybrid scheme

## 1 引言

滑动窗口法是目标检测方法中一种重要的策略<sup>[1-4]</sup>。在通用的目标检测问题中,由于没有目标类别的背景先验知识,因此相关方法必须逐一检查数量巨大的可能的目标矩形区域。近年来提出的通用似物性推荐方法(Genetic Objectness Proposal)可以在预处理阶段大幅减少候选窗口的数量<sup>[5-7]</sup>,已经得到了广泛的关注和探讨<sup>[8-13]</sup>。到目前为止,在似物性推荐问题上,OBN<sup>[6]</sup>、SEL<sup>[10]</sup>、CSVM<sup>[12]</sup>和BING<sup>[13]</sup>等方法的性能最佳。

OBN方法认为,任何物体或者具有清晰的闭合边界,或者视觉特性与周围有明显区分,或者在图像内具有很强的唯一性和显著性,或者兼具多种特点<sup>[6]</sup>。基于这种认识,文献[6]分别构造了几种特征来描述这些特点:窗口内边界附近的边缘密度(Edge Density)和超像素块横跨窗口边缘的程度(Superpixels Straddling)用来描述闭合边界;窗口与其周围区

域的像素颜色差异(Color Contrast)用来描述与环境的外观差异;基于FFT谱残差的多层全局显著性(Multi-scale Saliency)用来评价物体的显著性和唯一性。同时,文献[6]也注意到了物体窗口在位置和尺寸上具有一定的倾向。因此,文献[6]提出了通过朴素贝叶斯结合上述特征对似物性进行预测的OBN方法。

SEL<sup>[10]</sup>则是一种基于图像分割的方法。该方法认为,图像中的物体通过颜色、纹理及相互内嵌的程度与环境区分开来,同时图像内的物体是有层次的。因此其采用了一种自底向上的方法来处理图像。SEL方法在以图像分割方法建立初始区域后,基于相邻区域的相似性,贪心地将小区域组合成大区域。评价相似性时综合了颜色、纹理、相互内嵌的程度以及尺寸。尺寸相似性促使小区域尽早地被组合为大区域。获得所有区域后,该方法基于各区域在图像中的位置对窗口进行排序。

CSVM<sup>[12]</sup>使用级联排序SVM对似物性进行预测。该方

到稿日期:2014-05-18 返修日期:2014-07-17

刘质彬(1990—),男,硕士生,主要研究方向为计算机视觉,E-mail:liuzhibin@nlsde.buaa.edu.cn;赵启阳(1978—),男,博士,讲师,主要研究方向为计算机视觉。

法首先在不同尺度上,以使用4个不同方向的零均值梯度作为特征,用排序SVM训练出尺度相关的模板。这些模板使得在给定图像内,若给定尺寸,重合率高的建议窗口评分更高。而由于该评分只在同一张图像内的同一个尺度上可比较,因此还需要一个修正步骤。通过再次使用排序SVM在每个尺寸训练出一组调整参数,CSVM最终对所有的窗口给出统一的评分。基于该评分对获得的窗口进行排序,即可得到所需的推荐窗口。

BING方法<sup>[13]</sup>的主要动机与文献[5,6]相同:图像中大多数物体有着清晰的闭合边界,而且相对于周围的场景,物体展示出一种不同的外观。文献[13]观察到,所有的标定出的物体窗口在缩放到一个固定尺寸上时(文献[13]中为 $8 \times 8$ ),其在梯度上共享一种共同的特征,这个共性由一个用线性SVM从训练图像中学习出的模板 $W$ 来表示。BING方法同时注意到了不同的宽/高配置下的细微区别,因此给出了一个调整权值的模型。总的来说,BING方法包含两个阶段:首先在第一阶段计算出 $W$ ,而后在第二阶段给出调整权值的模型。特别地,为了提高效率,BING方法使用了多个0/1向量来近似 $W$ ,以使用位操作来计算 $W$ 和待选窗口特征的内积。

在似物性推荐方法的性能评价方面,目前通用的指标是“检测率-窗口数量”(下文记作DR-#WIN),其中检测率DR是被推荐窗口所覆盖的真实目标区域在所有目标区域中所占的比例。其中沿用了目标检测中的通用做法,即目标区域与推荐窗口都用矩形来表示。目标区域是否被推荐窗口所覆盖,所采用的评判方法是严格PASCAL-overall标准<sup>[14]</sup>;推荐窗口和目标区域的重合面积与其并集面积之比(下文记作INT-UNION)大于0.5时,即认为目标区域已被推荐窗口所覆盖。可以看到,一个好的推荐窗口将至少有50%属于目标区域,因此该标准看起来相当合理,已经被广泛地用于目标检测方法的性能评价,似物性推荐方法的评价则沿用了这一标准。

在DR-#WIN的评价指标上,近年来提出的似物性推荐方法的性能逐年提高,并无一例外地显著超过了随机猜测方法。其中BING方法<sup>[13]</sup>在PASCAL VOC 2007(下文简称VOC2007)测试集上的性能为最佳:在仅使用1000个推荐窗口时,可以覆盖96.2%的真实目标区域。在Intel i7-3940XM处理器上,该方法的计算效率可以达到处理每张VOC2007图像仅需0.003s,完全可以满足实时处理的要求。在推荐窗口数量增加到5000时,检测率可以进一步达到99.5%。

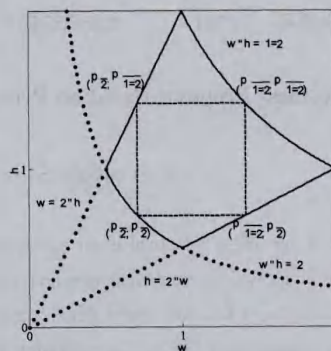
本文从组合几何学的角度对似物性推荐问题进行了分析,并提出了一种构建全覆盖窗口集合的方法,可以使用很少量的窗口覆盖全部可能位置/尺寸的目标区域。基于此结果,本文发现最新的似物性推荐方法更多地在于组合几何学方面有成功的案例,并未显著地使用具体图像的内容信息。这一发现与此前文献的结论是相悖的。此外,在近年来提出的显著物体检测方法<sup>[15-18]</sup>的启发下,通过利用真实目标区域的位置/尺寸的先验分布,可以进一步地缩减全覆盖窗口集合的规模。在此基础上,建立了一种融合了贪心策略和随机策略的混合机制,能够在计算耗时几乎为零的前提下取得接近目前最佳水平的性能。在VOC2007测试集上,该机制可以在1000个推荐窗口上取得94.52%的召回率,而且在前10个推荐窗口上的召回率比其他方法平均高出13.99%~40.29%。

## 2 组合几何学角度的分析

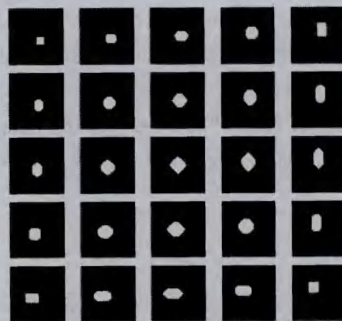
在进行似物性评价时,通常设INT-UNION的下界为0.5。显然该下界值越大,推荐窗口与目标区域的重合度越高。在考量通用性的情况下,本节内的推导将该下界推广到(0,1)的任意值,并用 $\eta$ 表示。

考虑如下问题:在一个图像中,最少需要多少窗口来覆盖所有可能位置、可能尺寸的矩形区域?下文称某个窗口集合 $S$ 为一个完全覆盖集合,当且仅当对于任意一个矩形区域 $\langle x, y, w, h \rangle$ ,至少有一个 $S$ 中的窗口可以覆盖该区域。这里 $x, y, w, h$ 依次为矩形区域左上角的横、纵坐标、矩形宽度和高度。

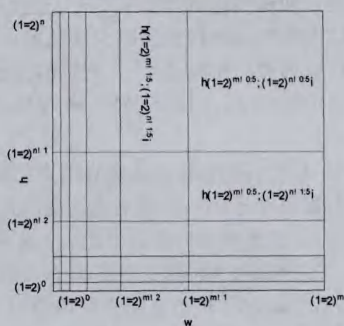
为了构造这样的完全覆盖集合,首先考虑一个在无穷平面上宽和高都为1的单位窗口。对于可被该窗口覆盖的矩形区域的尺寸(下文简称可行尺寸),可以很容易地证明有下述结论。



(a) 可被单位窗口覆盖的 $\langle w, h \rangle$ 的可行范围(由实线包围)



(b) 可被单位窗口覆盖的25种典型的可行位置(白色部分),其中 $h$ 由上到下从 $\sqrt{\eta}$ 增加到 $\sqrt{1/\eta}$ , $w$ 由左到右从 $\sqrt{\eta}$ 增加到 $\sqrt{1/\eta}$



(c) 完全覆盖集合的构造方法,图中的式子为覆盖不同二维区间中所有窗口的推荐窗口的尺寸

图1 完全覆盖集合的构造

引理1 在 $w-h$ 平面上,所有可行尺寸 $\langle w, h \rangle$ 均在被两

条曲线  $w \cdot h = \eta$  和  $w \cdot h = 1/\eta$ , 以及两条直线  $w - \eta \cdot h = 0$  和  $h - \eta \cdot w = 0$  所包围的区域之内。

参见图 1(a) 中被内部实线包围的“梯形”区域。特别需要指出的是, 所有满足  $\sqrt{\eta} \leq w, h \leq \sqrt{1/\eta}$  的尺寸都是可行的, 如图 1(a) 中被虚线正方形包围的区域。由于正方形区域容易拼接, 因此下文将以此为基础构建完全覆盖集合。

对一个可行的尺寸配置  $\langle w, h \rangle$ , 接下来考虑可被单位窗口覆盖的尺寸为  $\langle w, h \rangle$  的矩形区域的可行位置  $\langle x, y \rangle$ 。很容易推出, 当  $w = h = \sqrt{\eta}$  时, 这些矩形的所有可行位置组成一个边长为  $1 - \sqrt{\eta}$  的正方形; 而当  $w = h = \sqrt{1/\eta}$  时, 所有可行位置也组成一个边长为  $\sqrt{1/\eta} - 1$  的正方形。而对任意的  $\langle w, h \rangle$ , 则有:

**引理 2** 对任意的  $\sqrt{\eta} \leq w, h \leq \sqrt{1/\eta}$ , 对尺寸为  $\langle w, h \rangle$  的物体窗口来说, 其所有可行位置所在的范围均包含一个边长为  $1 - \sqrt{\eta}$  的正方形区域; 而且该正方形的边与单位窗口的边平行。

该引理的证明需要针对  $w, h$  的不同组合情形进行讨论, 证明较为直接, 但由于情形较多, 此处略去。图 1(b) 给出了 25 种典型组合下可行的物体位置。根据引理 2, 可以进一步地得到:

**引理 3** 如果  $w, h$  满足  $(1/\eta)^i \leq w \leq (1/\eta)^{i+1}, (1/\eta)^j \leq h \leq (1/\eta)^{j+1}$  (其中  $i, j$  都为非负整数), 则可由宽为  $(1/\eta)^{i+0.5}$ 、高为  $(1/\eta)^{j+0.5}$  的窗口覆盖的物体窗口的所有可行位置能组成一个区域, 该区域包含一个宽为  $(1 - \sqrt{\eta}) \cdot (1/\eta)^i$ 、高为  $(1 - \sqrt{\eta}) \cdot (1/\eta)^j$  的矩形子区域; 而且矩形子区域的边与推荐窗口的边平行。

在一个宽为  $M$ 、高为  $N$  的图像内, 要覆盖所有满足  $(1/\eta)^i \leq w \leq (1/\eta)^{i+1}, (1/\eta)^j \leq h \leq (1/\eta)^{j+1}$  的物体窗口, 需要宽为  $(1/\eta)^{i+0.5}$ 、高为  $(1/\eta)^{j+0.5}$  的窗口的数目, 该数目记为  $s(i, j)$ , 则有:

**引理 4**

$$s(i, j) = \left\lceil \frac{M - (1/\eta)^i}{(1 - \sqrt{\eta}) \cdot (1/\eta)^{i+0.5}} \right\rceil \cdot \left\lceil \frac{N - (1/\eta)^j}{(1 - \sqrt{\eta}) \cdot (1/\eta)^{j+0.5}} \right\rceil$$

该引理可通过将引理 3 中的“覆盖正方形”拼接在一起以覆盖整个图像来证明。构造方法如图 1(c) 所示。接下来, 不失一般性, 假设图像宽为  $M$ , 长为  $N$ , 同时有  $m = \lceil \log_{1/\eta} M \rceil$ ,  $n = \lceil \log_{1/\eta} N \rceil$ 。如果目标矩形的宽和高的最小值均为 1, 则完全覆盖集合中所有窗口的数量为:

$$\begin{aligned} & \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} s(i, j) \\ &= \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left\lceil \frac{M - (1/\eta)^i}{(1 - \sqrt{\eta}) \cdot (1/\eta)^{i+0.5}} \right\rceil \left\lceil \frac{N - (1/\eta)^j}{(1 - \sqrt{\eta}) \cdot (1/\eta)^{j+0.5}} \right\rceil \\ &\leq \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[ \frac{(1/\eta)^m - (1/\eta)^i}{(1 - \sqrt{\eta}) \cdot (1/\eta)^{i+0.5}} \right] \left[ \frac{(1/\eta)^n - (1/\eta)^j}{(1 - \sqrt{\eta}) \cdot (1/\eta)^{j+0.5}} \right] \\ &= \sum_{i=1}^m \sum_{j=1}^n \left\lceil \frac{(1/\eta)^i - 1}{\sqrt{1/\eta} - 1} \right\rceil \cdot \left\lceil \frac{(1/\eta)^j - 1}{\sqrt{1/\eta} - 1} \right\rceil \\ &\leq \sum_{i=1}^m \sum_{j=1}^n \left( \frac{(1/\eta)^i - 1}{\sqrt{1/\eta} - 1} + 1 \right) \cdot \left( \frac{(1/\eta)^j - 1}{\sqrt{1/\eta} - 1} + 1 \right) \\ &\leq (\sqrt{1/\eta} - 1)^{-2} \cdot \sum_{i=1}^m \sum_{j=1}^n (1/\eta)^{i+j} + (\sqrt{1/\eta} - 1)^{-1} \cdot \\ & \quad (n \sum_{i=1}^m (1/\eta)^i + m \sum_{j=1}^n (1/\eta)^j) + mn \end{aligned}$$

$$\begin{aligned} &= \frac{((1/\eta)^m - 1)((1/\eta)^n - 1)}{(\sqrt{1/\eta} - 1)^2 (1 - \eta)^2} + \\ & \quad \frac{n \cdot ((1/\eta)^m - 1) + m \cdot ((1/\eta)^n - 1)}{(1/\eta)^{m+n}} + mn \\ & \leq \frac{(1/\eta)^{m+n}}{(\sqrt{1/\eta} - 1)^2 (1 - \eta)^2} + \frac{n \cdot (1/\eta)^m + m \cdot (1/\eta)^n}{(\sqrt{1/\eta} - 1) \cdot (1 - \eta)} + mn \\ & = O(MN) \end{aligned} \quad (1)$$

特别地, 当  $m = n = 9, \eta = 0.5$ , 即图像宽和高都为 512 时, 该值为 6002500, 远少于所有可能的滑动窗口数量。在实际应用中, 目标区域的尺寸往往具有一定的限制。假设所有目标区域的宽、高至少为  $2^k$ , 则完全覆盖集合中的窗口数量为

$$\begin{aligned} & \sum_{i=k}^{m-1} \sum_{j=k}^{n-1} s(i, j) \\ & \leq \sum_{i=1}^{m-k} \sum_{j=1}^{n-k} \left\lceil \frac{(1/\eta)^i - 1}{\sqrt{1/\eta} - 1} \right\rceil \cdot \left\lceil \frac{(1/\eta)^j - 1}{\sqrt{1/\eta} - 1} \right\rceil \\ & \leq \sum_{i=1}^{m-k} \sum_{j=1}^{n-k} \left( \frac{(1/\eta)^i - 1}{\sqrt{1/\eta} - 1} + 1 \right) \cdot \left( \frac{(1/\eta)^j - 1}{\sqrt{1/\eta} - 1} + 1 \right) \\ & \leq \frac{(1/\eta)^{m+n-2k}}{(\sqrt{1/\eta} - 1)^2 (1 - \eta)^2} + \\ & \quad \frac{(n-k) \cdot (1/\eta)^{m-k} + (m-k) \cdot (1/\eta)^{n-k}}{(\sqrt{1/\eta} - 1) \cdot (1 - \eta)} + (m-k) \\ & \quad (n-k) \\ & = O(\eta^{2k} MN) \end{aligned}$$

特别地, 取  $\eta = 0.5$ , 当所有目标区域的宽、高至少为 16 时, 该数量为 19600; 当至少为 32 时, 则仅仅需要 4225 个窗口。这些数量都远远少于所有可能的滑动窗口数量。尤其需要指出的是, 在完全覆盖集合的构造中, 仅考虑了图像和目标区域的尺寸, 并未考虑图像的内容信息。换句话说, 对不同的图像, 该完全覆盖集合可以不加任何变化地覆盖其中任一图像中的所有目标区域。

最后特别指出, 引理 3 中所获得的数量是覆盖位置、尺寸取任意值的目标区域所需窗口数量的上界, 而这个上界在大多数情况下过于宽松。本文将在第 3 节对其作进一步的改进。

BING 方法<sup>[13]</sup>是目前为止性能最好的似物性推荐方法, 其良好的性能包括了组合几何学方面的因素。在 BING 方法中, 推荐窗口的尺寸呈两倍增长, 这与引理 1—引理 4 中构造完全覆盖集合的方法是一致的; 此外 BING 方法所使用的非最大值抑制步长 0.25 也非常接近引理 2 中所用的步长  $1 - \sqrt{0.5} \approx 0.29$ 。根据前面的引理, 这两项设置使得 BING 可以去除大量的冗余窗口。根据引理 2, 非最大值抑制步长的增加将可能导致部分目标区域无法被覆盖, 而非最大值抑制步长的减小将无谓地增加窗口数量。

### 3 基于贪心和随机的混合式推荐窗口机制

在第 2 节的完全覆盖集合的构造中, 默认所有目标区域的尺寸和位置的分布是均匀的。然而真实图像应用中的情况并非如此。显著性检测方面的研究指出, 兴趣目标区域倾向于位于图形的中心<sup>[15-18]</sup>, 其尺寸通常来说不会过大或过小。文献<sup>[6]</sup>认为目标区域的位置服从高斯分布, 其宽度和高度较倾向于相等。对 VOC2007 数据集上的目标区域进行统计发现, 图像中目标区域倾向于在图像中心, 如图 2(a) 所示; 高宽

比对数的期望值  $E(\log(h/w)) = E(\log(h)) - E(\log(w))$  在训练集上为 0.1557, 在测试集上为 0.1449, 即图像中的目标区域倾向于高度大于宽度; 同时较大的推荐窗口通常比较小的窗口能覆盖更多的矩形区域, 如图 2(b) 所示; 此外 VOC2007 中的训练及测试集在目标区域的位置和尺寸的分布也是基本一致的。

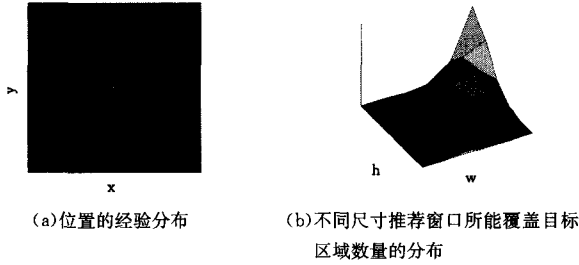


图 2 VOC2007 训练集中目标区域的统计信息

在给出目标区域的尺寸/位置的先验分布后, 仅需要考虑可能被取到的位置和尺寸即可。更进一步, 概率较高的位置和尺寸应当更优先地包含在完全覆盖集合中, 显然这是一种贪心的策略。此外需要指出, 在下面给出的贪心机制中, 本文将所有图像都归一化到一个固定尺寸(实验中选择  $512 \times 512$ ), 其中的目标区域的位置和尺寸也经过同比例缩放变换的处理。

#### 算法 1 生成完全覆盖集合的贪心机制

输入: 图像集  $\{I_n\}$  和目标区域集合  $\{O_i\}$ ; 需要得到的推荐窗口数量  $N$ ;

目标区域的最小宽/高  $2^L$  和最大宽/高  $2^H$

输出:  $N$  个推荐窗口  $\{\langle x_j, y_j, w_j, h_j \rangle\}$  组成的覆盖集

1. 令  $j$  为 1 到  $N$
2. 对  $\{\sqrt{2} \cdot 2^L, \sqrt{2} \cdot 2^{L+1}, \dots, \sqrt{2} \cdot 2^{H-1}\}^2$  中的每个尺寸  $\langle w, h \rangle$
3. 对每个可能的位置  $\langle x, y \rangle$
4. 如果  $\langle x, y, w, h \rangle$  已被选为推荐窗口
5. 跳过本次最内层循环
6. 计算窗口  $\langle x, y, w, h \rangle$  的检测率 DR
7. 将所有被  $\langle x, y, w, h \rangle$  覆盖的物体窗口添加到集合  $C_{x, y, w, h}$  中
8. 将 DR 最大的窗口  $\arg \max DR(x, y, w, h)$  设为第  $j$  组推荐窗口  $\langle x_j, y_j, w_j, h_j \rangle$
9. 从标定物体窗口中删去所有  $C_{x_j, y_j, w_j, h_j}$  中的元素
10. 结束

实验发现, 贪心机制能够成功地缩减完全覆盖集合的规模, 参见第 4 节。然而当尝试以“训练-预测”的模式使用该机制时, 其性能却低于预期水平。其原因在于训练集和测试集上在低概率样本空间中存在着巨大的差异, 因此在训练集上得到的完全覆盖集合并没有很好地考虑测试集中低概率样本的几何属性。

本文使用随机采样来处理小概率样本空间的差异问题。实验发现: 在不同的图像集中, 目标区域尺寸的分布更为相近, 而位置的分布差异相对较大。换句话说, 由训练集得到的尺寸信息较为准确, 而小概率空间中的位置信息则误差较大。因此这里选择仅采用由训练集样本所表征的尺寸的分布, 而简单地假定位置分布是均匀的。虽然这不同于真实情形, 但是通过实验发现, 在无法精确给出真实分布时这种处理方式能够取得较好的性能。

在第 2 节的完全覆盖集合的构造中, 图像尺寸不超过  $2^m \times 2^n$  的图像集所需要的推荐窗口的类型不超过  $mn$  种。因此尺寸分布是在  $mn$  种尺寸类型上的离散分布, 其概率值的估计方法为:

$$p_{w', h'} = \frac{\sum_{\text{所有图像}} \sum_{\text{所有}(x, y)} [\text{窗口}\langle x, y, w', h' \rangle \text{覆盖某个目标区域}]}{\sum_{\text{所有}(w, h)} \sum_{\text{所有图像}} \sum_{\text{所有}(x, y)} [\text{窗口}\langle x, y, w, h \rangle \text{覆盖某个目标区域}]}$$

在得到尺寸分布模型之后, 即按照该模型采样生成所需数量的窗口。采样过程如下:

#### 算法 2 基于随机采样的推荐窗口生成方法

输入: 各尺寸集合  $\{\langle w_i, h_i \rangle (i=1, \dots, mn)\}$ , 完全覆盖集合中各尺寸窗口的数量  $\{N_{w, h}\}$ , 各尺寸的概率值  $\{p_i = p_{\langle w_i, h_i \rangle}\}$ ; 拟给出的推荐窗口总数  $N_{\text{need}}$

输出: 排序后的  $N_{\text{need}}$  个推荐窗口  $L_{\text{random}}$

1.  $S \leftarrow \emptyset$
2. 对每种尺寸  $\langle w, h \rangle$
3. 在所有位置上以  $\min((p_{w, h} \times N_{\text{need}}) / N_{w, h}, 1)$  的概率采样, 并将其置入窗口集合  $S$
4. 以  $[p_1, p_2, \dots, p_{mn}]$  的概率从  $S$  中不放回地采样, 选出每个排序位置上的窗口, 并记录到列表  $L_{\text{random}}$  中
5. 结束

通过随机采样方法, 尽可能地保证各个位置上的窗口都有机会被选中, 这样就较好地避免了不同图像集在目标区域位置分布上的差异问题。然而, 通过实验发现, 随机采样方法无法保证在推荐窗口数量较少时的性能, 而贪心方法却可以很好地对此予以弥补。为此, 本文进一步给出一种结合了贪心方法和随机采样方法的混合机制。

#### 算法 3 基于贪心和随机采样的混合机制

输入: 贪心方法给出的  $N$  个推荐窗口的列表  $L_{\text{greedy}}$ ; 混合参数  $\alpha$

输出:  $N$  个推荐窗口  $L$

1. 使用算法 2 生成  $N$  个推荐窗口的列表  $L_{\text{random}}$
2. 保留  $L_{\text{greedy}}$  中的前  $N \cdot \alpha$  个窗口, 与  $L_{\text{random}}$  中的窗口列表组成一个大的列表  $L_{\text{total}}$
3. 对  $L_{\text{total}}$  中的每一个窗口  $\langle x, y, w, h \rangle$
4. 删去  $L_{\text{total}}$  中窗口  $\langle x, y, w, h \rangle$  之后, 尺寸为  $\langle w, h \rangle$  且位置与  $\langle x, y \rangle$  的差距不大于  $(0.29w, 0.29h)$  的窗口
5. 结束

其中, 步骤 3 和步骤 4 的作用是根据推荐窗口的排序, 抑制掉在其附近的、相同尺寸的窗口, 目的是在不减少被覆盖的目标区域的数量的前提下, 尽可能地减少推荐窗口的数量。这种非最大值抑制的思路也被用于 BING 和 CSVM。VOC2007 上的实验表明, 参数  $\alpha$  取 0.1 时的混合效果较好。

## 4 实验与分析

实验采用似物性推荐研究常用的数据集 PASCAL VOC2007。该数据集是 PASCAL 测试集的一部分, 常用于目标检测方法的性能验证, 因此也成为了似物性推荐方法的重要测试基准。VOC2007 包括 2501 幅训练图像和 4952 幅测试图像。所有的目标区域都是由人工标定出的、与图像边界平行的矩形, 按照语义内容的不同而分属 20 个类别。实验采用的性能评价标准为常用的“检测率-窗口数量”(DR-#WIN)。

首先使用贪心方法在训练集上生成推荐窗口集合。该集合在取前 1000 个窗口时, 可以达到 99.85% 的召回率, 这证明贪心方法能够很好地缩减完全覆盖集合的规模。在将该集合用于预测测试集中的目标区域时, 在前 15 个窗口上的召回率与训练集上的表现相差无几; 然而随着窗口数量的增加, 召回率缓慢地下降。这说明训练集和测试集在大概率样本空间上的一致性较好, 而在小概率样本空间上的差异较大, 这也符

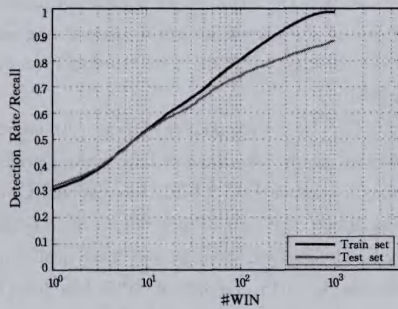
合本文第3节的分析预测。

随后测试混合机制在测试集上的性能表现。在绝大多数窗口数量上,混合机制的性能要高于OBN和CSVM,接近SEL和BING,在1000个推荐窗口上能够召回94.52%的目标区域。需要特别指出的是,混合机制在前10个窗口上的召回率平均要比其他方法高出13.99%~40.29%,尤其是在第一个窗口上的召回率达到了32.7%,显著地超过了其他方法。该部分实验定性的结果参见图3,DR-#WIN性能曲线参见图4(b)。

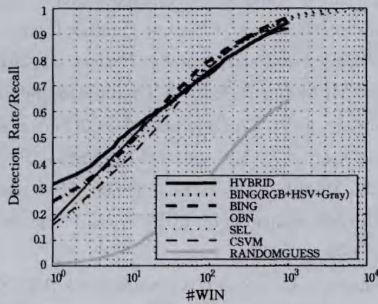


按照从上到下的顺序,各行给出了取1、10、100和1000个推荐窗口时的结果。目标区域被标记为黑色,而最佳推荐窗口被标记为白色。

图3 推荐窗口数量逐渐增加时,混合机制在VOC2007上的实验结果



(a) 训练图像上的覆盖集的性能



(b) 混合机制性能与其他方法的比较

图4 贪心机制和混合机制的性能

在混合机制的计算效率方面,所有的计算量都在随机采样和非最大值抑制方面。其中随机采样需要遍历不超过 $O(MN)$ 个不同的窗口,而非最大值抑制需要执行 $N_{need}$ 次抑制,因此整体的计算复杂性为 $O(MN)$ ,亦即与像素数量呈线性关系。VOC2007测试集中的图像在i5 3317U上的平均运行时间大约为0.002s。而在同样的测试环境下,BING的平均运行时间约为0.012s。另外,由于混合机制实际上并未使用图像内的信息,因此实际应用中可以预先生成归一化的推荐窗口,在应用于图像时,对窗口进行相应的缩放,这样对每

张图像仅需进行 $4N_{need}$ 次乘法运算。

表1中列出了混合参数 $\alpha$ 对检测率的影响。据此,在实验中取 $\alpha$ 为0.1。

表1 1000个推荐窗口时不同 $\alpha$ 下平均检测率

$\alpha$	0	0.05	0.1	0.3	0.5	1
DR(%)	80.39	89.01	94.52	93.70	92.22	88.64

**结束语** 本文从组合几何学的角度对似物性推荐问题进行了分析,并且给出了完全覆盖集合的一种构造方法以及最小完全覆盖集合的一个上界。在此基础上,结合了贪心方法和随机采样方法的混合机制能够在通用的测试集上取得接近目前最佳水平的性能表现,而其计算耗时则完全能够满足实时处理的要求。该方法能够有效地缩减推荐窗口的数量,因此可以在几乎不降低召回率的前提下,大幅地提升有关目标检测方法的计算效率,对计算机视觉的相关应用也能提供较好的工具支持。

### 参考文献

- [1] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2005. San Diego, United States; IEEE, 2005: 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, 2010, 32(9): 1627-1645
- [3] Lampert C H, Blaschko M B, Hofmann T. Beyond sliding windows: Object localization by efficient subwindow search[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, United States; IEEE, 2008: 1-8
- [4] 杨扬, 李善平. 分割位置提示的可变形部件模型快速目标检测[J]. 自动化学报, 2012, 38(4): 540-548  
Yang Yang, Li Shan-ping. Fast object detection based on deformable part-based model cued by segmentation position [J]. Acta Automatica Sinica, 2012, 38(4): 540-548
- [5] Alexe B, Deselaers T, Ferrari V. What is an object? [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2010. San Francisco, United States; IEEE, 2010: 73-80
- [6] Alexe B, Deselaers T, Ferrari V. Measuring the objectness of image windows [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2189-2202
- [7] 时健, 郭延文, 杜振龙, 等. 一种基于网格参数化的图像适应方法[J]. 软件学报, 2008, 19(1): 19-30  
Shi Jian, Guo Yan-wen, Du Zhen-long, et al. A Mesh Parameterization-Based Image Retargeting Method [J]. Journal of Software, 2008, 19(1): 19-30
- [8] Endres I, Hoiem D. Category independent object proposals[C]//European Conference on Computer Vision, 2010. Crete, Greece: Springer Berlin Heidelberg, 2010: 575-588
- [9] Endres I, Hoiem D. Category-independent object proposals with diverse ranking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(2): 222-234
- [10] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154-171
- [11] van de Sande K E A, Uijlings J R R, Gevers T, et al. Segmentation as selective search for object recognition[C]//2011 IEEE

International Conference on Computer Vision, Barcelona, Spain; IEEE, 2011; 1879-1886

[12] Zhang Z, Warrell J, Torr P H S. Proposal generation for object detection using cascaded ranking SVMs[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2011; 1497-1504

[13] Cheng M M, Zhang Z M, Lin W Y, et al. BING: Binarized normed gradients for objectness estimation at 300fps[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2014 . Columbus, United States; IEEE, 2014; 3286-3293

[14] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338

[15] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259

[16] Gao D, Mahadevan V, Vasconcelos N. On the plausibility of the discriminant center-surround hypothesis for visual saliency [J]. Journal of Vision, 2008, 8(7): 1-18

[17] 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

[18] Mahadevan V, Vasconcelos N. Biologically inspired object tracking using center-surround saliency mechanisms [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 541-554

(上接第 284 页)

随机选取 6 个场景对应的失真图像对作为训练样本, 剩下的 2 个场景对应的失真图像作为测试样本, 以保证测试图像和训练图像内容上的完全独立。表 3 和表 4 分别列出了本文方法和相关文献中的方法在 LIVE3D\_II 库上的实验结果。

表 3 LIVE3D\_II 库 SROCC 对比(后 4 种为无参考方法)

	WN	JP2K	JPEG	Blur	FF	ALL
You	0.909	0.894	0.795	0.813	0.891	0.786
CyclopeanMS-SSIM	0.940	0.814	0.843	0.908	0.884	0.889
Benoit	0.923	0.751	0.867	0.455	0.773	0.728
BRISQUE <sup>[12]</sup>	0.846	0.593	0.769	0.862	0.935	0.769
文献[13]	0.950	0.867	0.867	0.900	0.933	0.880
本文方法	0.931	0.913	0.731	0.919	0.944	0.890

表 4 LIVE3D\_II 库 PLCC 对比(后 4 种为无参考方法)

	WN	JP2K	JPEG	Blur	FF	ALL
You	0.912	0.905	0.830	0.784	0.915	0.800
CyclopeanMS-SSIM	0.957	0.834	0.862	0.963	0.901	0.900
Benoit	0.926	0.784	0.853	0.535	0.807	0.748
BRISQUE <sup>[12]</sup>	0.844	0.599	0.787	0.927	0.925	0.780
文献[13]	0.947	0.899	0.901	0.941	0.932	0.895
本文方法	0.950	0.918	0.717	0.976	0.938	0.905

从表 3 和表 4 中可以看出, 本文方法在 LIVE3D\_II 库中非对称失真图像上也得到了不错的效果, 在所有图像上的主观感知一致性要优于文献[12, 14]的结果, 同时在单个的 JP2K 和 Blur 失真类型上也体现了优势。这进一步说明了本文方法的有效性和可靠性。

**结束语** 本文利用立体图像对及合成图像对, 使用小波变换系数特征, 提出了一种基于支持向量回归的无参考立体图像质量评价算法。实验结果表明该算法与人类主观感知有着很好的一致性。但该方法在单个的 JPEG 失真类型上效果不尽如人意, 这也是接下来需要提高的地方。

## 参考文献

[1] Wang Z, Simoncelli E P, Bovik A C. Multi-scale structural similarity for image quality assessment[C]//IEEE Asilomar Conference on Signals, Systems and Computers. Pacific Grove, CA, 2003; 1398-1402

[2] Sheikh H R, Bovik A C. Image information and visual quality [J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444

[3] You J, Xing L, Perkis A, et al. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis[C]//International workshop on video processing and quality metrics. Scottsdale, Arizona, USA, 2010

[4] Xing L, You J, Ebrahimi T, et al. A perceptual quality metric for stereoscopic crosstalk perception[C]//IEEE International Conference on Image Processing. Hong Kong, China, 2010; 4033-4036

[5] Rittermann M. A proposal for the quality assessment of 3D video objects[C]//International Workshop on Image Analysis for Multimedia Interactive Services. Lisboa, Portugal, 2004

[6] Akhter R, Sazzad Z M P, Horita Y, et al. No-reference stereoscopic image quality assessment[C]//Proceedings of SPIE. San Jose, California, USA, 2010; 17-21

[7] Shao F, Lin W S, Gu S B, et al. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics[J]. IEEE Trans. Image Process, 2013, 22(5): 1940-1953

[8] Perera G C V, De S V, Kondoz A M, Dogan An improved model of binocular energy calculation for full-reference stereoscopic image quality assessment[C]//IEEE International Conference on Acoustic, Speech and Signal Processing, 2014; 594-598

[9] Gorley P, Holliman N. Stereoscopic image quality metrics and compression[C]//Proceedings of SPIE. San Jose, California, USA, 2008; 1-11

[10] Gu K, Zhai G, Yang X. No-Reference Stereoscopic IQA Approach: From Nonlinear Effect to Parallax Compensation[J]. Journal of Electrical and Computer Engineering, 2012

[11] Sazzad Z M, Akhter R, Baltes J. Objective no-reference stereoscopic image quality prediction based on 2D image features and relative disparity[J]. Advances in Multimedia, 2012, 2012; 1-16

[12] Mittal A, Moorthy A K, Bovik A C. Blind/referenceless image spatial quality evaluator[J]. IEEE Trans. Image Process, 2012, 19(2): 75-78

[13] Chen Ming-jun, Lawrence K, Cormack Bovik A C. No-Reference Quality Assessment of Natural Stereopairs[J]. IEEE Transactions on Image Processing, 2013, 22(9): 3379-3391

[14] Chen Ming-jun, Su D K, Kwon C C, et al. Full-reference quality assessment of stereopairs accounting for rivalry[J]. Signal Processing; Image Communication, 2013, 28(9): 1143-1155

[15] 金波, 李朝锋, 吴小俊. 结合 NSS 和小波变换的无参考图像质量评价[J]. 中国图象图形学报, 2012, 17(1): 33-39

Jin Bo, Li Chao-feng, Wu Xiao-jun. No-reference image quality assessment based on natural scene statistics and wavelet [J]. Journal of Image and Graphics, 2012, 17(1): 33-39

[16] Moorthy A K, Su Che-chun, Mittal A, et al. Subjective evaluation of stereoscopic image quality [J]. Signal Processing; Image Communication, 2013, 28(8): 870-883