

文章编号:1007-2780(2016)07-0726-07

基于结构标签学习的显著性目标检测

程 黎*, 吴 谨, 朱 磊

(武汉科技大学 信息科学与工程学院, 湖北 武汉 430081)

摘要:提出了一种基于结构标签学习的显著性目标检测算法,将结构化学习方法应用到显著性目标检测中。首先从含有标记的图像中随机采集固定大小的矩形区域,并记录其结构标签;然后使用含结构标签的区域特征构建决策树集合;最后采用监督学习的方法对图像进行优化预测,得到最终的显著图。实验结果表明,本文方法能较准确地检测出图像库中图像的显著性区域,在数据库 MSRA5000 和 BSD300 的 AUC 值分别为 0.891 8、0.705 2,说明本文方法具有较好的显著性检测效果。

关 键 词:显著目标检测;结构标签;决策树;监督学习

中图分类号:TP391.4;TN911.73 **文献标识码:**A **doi:**10.3788/YJYXS20163107.0726

Salient object detection based on structured labels learning

CHENG Li*, WU Jin, ZHU Lei

(College of Information Science and Engineering, Wuhan University of
Science and Technology, Wuhan 430081, China)

Abstract: This paper proposes a salient object detection method based on structured labels learning, applying a structured learning method to salient object detection. Firstly, we get a fixed rectangular region randomly from the local image which includes the labeling, and record the corresponding structured labels. Then, a collection of decision trees is built by using the regional features which includes the structured labels. Finally, the final saliency map is captured by using the supervised learning approach. Experiments show that our method can detect the salient objects accurately, and the AUC scores are 0.891 8 and 0.705 2 on the MSRA5000 and BSD300 datasets, the result shows that our method can achieve good effect in salient object detection.

Key words: salient object detection; structured labels; decision trees; supervised learning

收稿日期:2016-01-25;修订日期:2016-02-29.

基金项目:国家自然科学基金青年项目(No. 61502358)

Supported by National Natural Science Foundation of China(No. 61502358)

* 通信联系人, E-mail: 342362994@qq.com

1 引 言

显著性目标检测可以提取图像中人眼感兴趣的目标区域,在计算机视觉领域有着重要的研究意义,广泛应用于目标识别、图像压缩、图像分割等图像处理技术^[1-2]。

当前显著性目标检测方法主要分为两类:非监督法和监督学习法。非监督法的研究主要集中在如何利用图像的自身属性模拟人类视觉系统进行显著性分析获取显著性区域。最典型的是 Itti 等人提出的基于特征融合的视觉显著性计算模型^[3],采用中心邻域算子分别计算颜色、亮度、方向 3 个显著图,并通过线性加权的方式叠加 3 个显著图形成最终显著图。监督学习法是利用预先设定的分类系统,对有类别标签的图像数据进行学习,对未知类别标签的图像数据进行预测来获取显著性区域。文献^[4]提出了一种基于特征融合的 RGBD 显著目标检测方法,将深度信息引入到区域显著性计算中,利用随机森林预测区域显著值,并通过多尺度融合的方法生成最终显著图。Kim 等人提出了一种基于高维颜色变换的显著性检测方法^[5],利用随机森林获取区域显著值得到初始显著图,然后对初始显著图进行三元(trimap)标记,将低维颜色特征映射为高维颜色空间的特征向量,通过在高维颜色空间寻找一个最优线性组合系数来获取最终的显著图。

传统的监督学习法一般采用逐像素预测显著性的方法,未考虑该像素与附近其他像素的决策结果的空间关系,容易出现误判。结构化的学习目的是挖掘像素与相邻像素之间的空间关系,以求达到更准确地检测效果。Kontschieder 等人利用结构化学习方法对图像进行语义标记^[6]。Dollár 等人将结构化学习的方法运用到边缘检测中^[7],取得了较好的效果。

本文提出了一种基于结构标签学习的显著性目标检测方法,将结构化学习方法应用到显著性检测中。首先从含有标记的图像中随机采集固定大小的矩形区域,并记录其结构标签,然后使用含结构标签的区域特征来构建随机森林,最后采用监督学习的方法进行结构化预测,得到最终的显著图。

2 方法概述

本文方法主要是通过学习特征与不同结构标记之间的关系构造结构森林预测模型,对图像进行结构化预测获取显著图。方法框架图如图 1 所示。

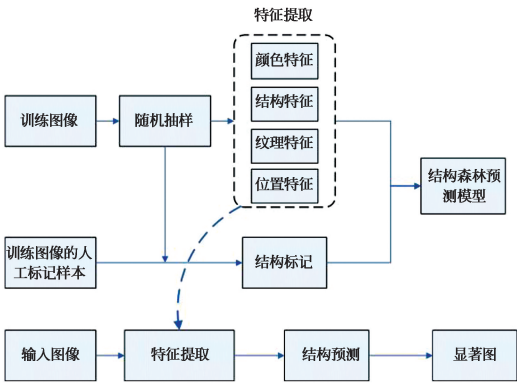


图 1 方法框架图
Fig. 1 Framework of approach

首先以滑动窗口的方式采集固定大小的矩形区域,并记录其结构标签。从采集的矩形区域中随机抽取训练样本及对应结构标记,针对每个样本区域,提取不同类型的特征,如颜色、结构、纹理、位置等对区域进行描述。然后通过学习特征与不同结构标签之间的对应关系构造结构森林预测模型。由于结构标记维数较高且结构复杂,因此本文利用主成分分析法(PCA)对其进行降维处理,转换成传统的随机森林进行训练。训练过程如图 2 所示。

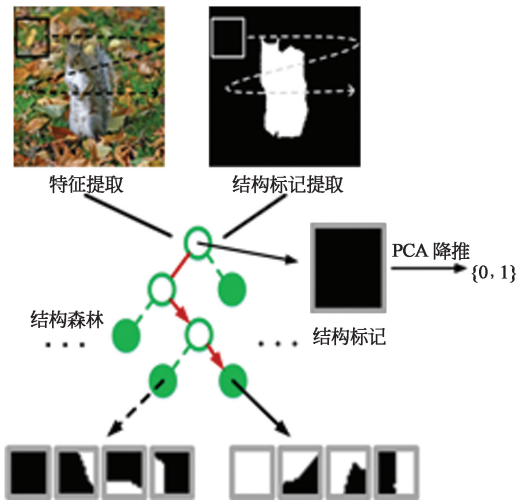


图 2 训练示意图
Fig. 2 Training diagram

结构森林每次对测试图像中一个固定大小的区域中所有像素同时进行预测,如图 3 所示。每一个像素的显著值由所有包含此像素的结构标签共同决策,构造像素级的显著性图。

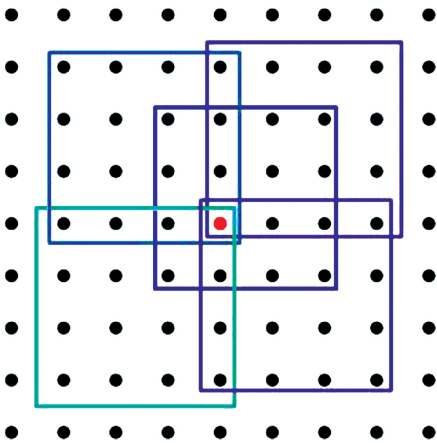


图 3 像素预测示意图

Fig. 3 Pixel prediction diagram

3 特征提取

3.1 区域内部的属性特征

区域内部属性特征主要有 sift 特征、centrist 特征、各种颜色通道特征以及位置特征等。sift 是一种点特征,其本质是提取像素点附近区域内的梯度特征,定义为一种区域梯度的描述子。centrist 是利用图像 census 变换后的统计直方图描述图像特征,定义为一种区域结构的描述子。颜色特征的提取是在某一颜色空间上完成的,分别从 RGB、HSV、LAB 以及 OPP 这 4 个颜色空间中提取均值、方差及统计直方图作为特征向量对区域显著性进行描述;lbp 是一种用来描述图像局部纹理特征的算子,定义为区域纹理特征描述子;位置特征则是用来描述样本区域的中心位置。最后,将这些特征组合起来作为最终的区域特征描述子。具体特征描述如表 1 所示。

表 1 特征描述

Tab. 1 Feature description

区域内部属性特征				区域的中心-邻域差异对比度及区域与伪背景差异对比度特征			
特征描述	维度	对比度测度	对比度维度	特征描述	维度	对比度测度	对比度维度
sift 特征	128	$\chi^2(h_1, h_2)$	1	RGB 直方图	1 000	$\chi^2(h_1, h_2)$	1
centrist 特征	256	$\chi^2(h_1, h_2)$	1	HSV 直方图	1 000	$\chi^2(h_1, h_2)$	1
RGB 均值	3	$d(V_1, V_2)$	3	LAB 直方图	1 000	$\chi^2(h_1, h_2)$	1
HSV 均值	3	$d(V_1, V_2)$	3	OPP 直方图	1 000	$\chi^2(h_1, h_2)$	1
LAB 均值	3	$d(V_1, V_2)$	3	lbp 直方图	256	$\chi^2(h_1, h_2)$	1
OPP 均值	3	$d(V_1, V_2)$	3				
RGB 方差	3	$d(V_1, V_2)$	3				
HSV 方差	3	$d(V_1, V_2)$	3				
LAB 方差	3	$d(V_1, V_2)$	3				
OPP 方差	3	$d(V_1, V_2)$	3				
空间位置	2						

3.2 区域与邻域对比度特征

区域与邻域对比度特征是将区域内部属性特征与其邻域背景内部属性特征进行对比,对比越强烈则该区域显著性越高。如图 4 所示。

本文选取的样本区域 N_1 是大小为 $d_1 \times d_1$ 的矩形区域,综合考虑计算时间与检测效果,选取

$d_1=16$ 。邻域 N_2 是包含 N_1 的相邻区域,可根据实际情况选择合适的邻域区间,本文选定 N_2 的宽度 $d_2=0.8d_1$ 。分别计算 N_1 与 N_2 的内部属性特征,然后计算它们之间的对比度特征。

颜色空间的均值与方差特征的对比度 $d(V_1, V_2)$ 是通过计算绝对差值获得:

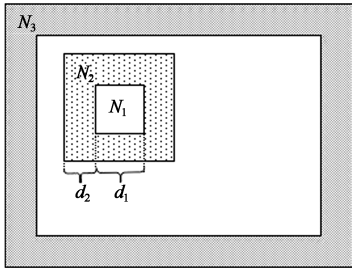


图4 样本区域、邻域以及伪背景区域

Fig. 4 Sample region, surrounding region and background region

$$d(V_1, V_2) = \{ |v_{11} - v_{21}|, |v_{12} - v_{22}|, |v_{13} - v_{23}|, \dots, |v_{1n} - v_{2n}| \}, \quad (1)$$

其中: $V_1 = \{v_{11}, v_{12}, v_{13}, \dots, v_{1n}\}$, $V_2 = \{v_{21}, v_{22}, v_{23}, \dots, v_{2n}\}$, n 表示特征向量的个数。

颜色空间的直方图特征、lbp 特征、centrist 特征及 sift 特征的对比度 $\chi^2(h_1, h_2)$ 是通过计算卡方距离 (Chi-square measure) 获得:

$$\chi^2(h_1, h_2) = \sum_{i=1}^{n'} \frac{2(h_{1i} - h_{2i})^2}{h_{1i} + h_{2i}}, \quad (2)$$

其中: n' 表示直方图的维度。

3.3 区域与伪背景区域对比度特征

通常一幅图像的显著性区域不会出现在图像的边缘,因此有必要考虑区域与背景边缘的差异对比。如图4,伪背景区域 N_3 是样本图像四周的背景边框。区域与伪背景区域的对比度特征计算方法与3.2节所述方式相同,区别在于对比对象不同。

4 预测模型

本文在随机森林模型的基础上,通过学习特征与不同结构标记之间的对应关系构造结构森林模型对图像进行预测。

4.1 随机森林

随机森林指的是利用多棵树对样本进行训练和预测的一种分类器组合,每棵树都是一个分类器,随机森林的预测结果是由这些树综合投票所得。

单棵决策树是一个树形结构的分类器,假设输入特征 $x \in X$, 其对应的输出标记是 $y \in \{0, 1\}$, θ 是独立同分布的随机向量,那么在节点 j 处,定义一个分裂函数为:

$$h(x, \theta_j) \in \{0, 1\}, \quad (3)$$

通过这个分裂函数将当前节点存储的训练数据分给其子节点。若 $h(x, \theta_j) = 0$, 那么将训练数据分配给左节点; 否则, 分配给右节点, 直到分裂至叶子节点为止。

每棵树采用递归的方式单独训练。通过寻找参数 θ_j 获取最优分裂函数 $h(x, \theta_j)$ 。定义信息增益准则如下:

$$I_j = I(S_j, S_j^L, S_j^R), \quad (4)$$

其中: S_j 为节点 j 处的数据集, $S_j^L = \{(x, y) \in S_j | h(x, \theta_j) = 0\}$ 表示将训练数据分配给左节点, $S_j^R = S_j \setminus S_j^L$ 表示将训练数据分配给右节点。选择参数 θ_j 最大化信息增益 I_j , 用数据集 S_j^L 与 S_j^R 分别对左节点和右节点进行递归训练, 当到达树的最大深度值或者当训练数据低于固定阈值时停止。

那么对于一个多级分类器, 标准的信息增益定义是:

$$I_j = H(S_j) - \sum_{k \in \{L, R\}} \frac{|S_j^k|}{|S_j|} H(S_j^k), \quad (5)$$

其中: $H(\cdot)$ 表示香农熵或者基尼杂质。

4.2 结构森林

对于传统随机森林模型, 其输入样本对应为一个二值标记。结构森林模型对应的输出标记不局限于一个数值或者一个标记, 充分利用输入样本中像素与像素间的空间关系, 将输出标记扩展为一个结构化的输出标记。

首先, 在学习结构森林分类器时, 将输入特征的对应标记 $y \in \{0, 1\}$ 扩展为一个结构标记 $y \in Y = \{0, 1\}^{\omega \times \omega}$, ω 为输入样本宽度。结构标记 Y 包含样本所有像素点对应的标记, 如图5中的结构标记 Y 所示。由于传统随机森林分裂函数是二值分裂函数, 而结构标记 Y 是 $\omega \times \omega$ 维, 每个结构标记可能有 $2^{\omega \times \omega}$ 种状态。为了降低节点分裂的代价, 减少信息增益的计算量, 需要对高维的结构标记进行降维处理, 本文选用 PCA 将高维结构标记 Y 映射到 1 维, 结构标记的信息仍存储在当前节点中, 如图5所示。

然后根据随机森林中决策树的节点分裂函数策略, 遍历整个结构标记空间将样本进行分类。最后在每个叶子节点得到最终的结构标记。

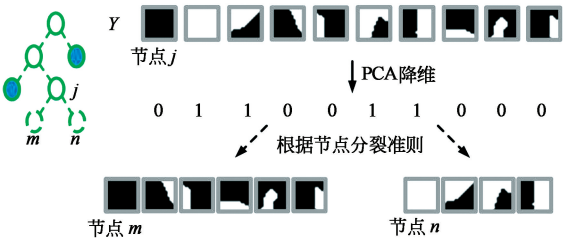


图 5 节点分裂示意图

Fig. 5 Node splitting diagram

5 仿真实验结果与分析

本文在公用数据库 MSRA5000^[8] 和 BSD300^[9] 上进行实验。

本文从 MSRA5000 数据库中随机选取 2 000 张图像及其对应的人工标记作为训练样本。包含有显著像素的为正样本,反之为负样本。每张图像

上分别选取 350 个正、负样本。在样本选取时可能存在一张图像无法选够 350 个正样本的情况,那么选取该图像正样本数量的最大值。正样本分为两类:强正样本与弱正样本。当显著区域的面积超过样本面积的一半时为强正样本,反之则为弱正样本。如果弱正样本的比例过大,那么会影响最终检测效果,因此本文选取强正样本数为正样本总数的 2/3,弱正样本数为正样本总数的 1/3。

测试集合包括 2 个数据库,一个是 MSRA5000 中剩余的 3 000 张图像,另外一个完整的 BSD300 中的 300 张图像。本文采用上述训练过程产生的预测模型对测试集合中的图像进行检测,并与其他 10 种方法进行比较,包括:AMC^[10]、BL^[11]、BSCA^[12]、CHM^[13]、HCT^[5]、HSD^[14]、RBD^[15]、RC^[16]、RCoDiBP^[17]、ST^[18]。部分检测结果如图 6 所示。

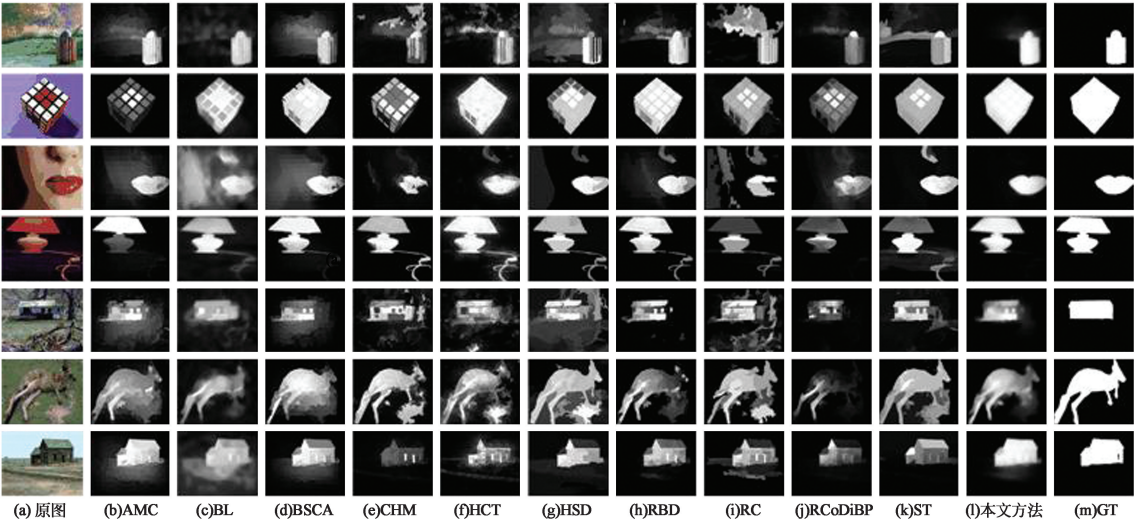


图 6 显著图对比图

Fig. 6 Visual comparison of saliency maps

由图 6 可知,本文的方法可以较准确地获取图像的显著性区域,相比其他方法更接近人工标记的结果。本文方法由于考虑到像素与相邻像素的空间关系,能够有效避免相邻的同属于前景/背景的像素点被判决为不同的显著类型,检测结果更加准确。如第 2 张魔方图,最显著的目标是魔方,本文方法能更准确地检测出图像的魔方区域。

精度-召回率是衡量显著性检测效果指标之一。对于样本显著图 I,用固定在灰度范围[0, 255]之间的阈值对其进行分割,得到 256 个二值图。通过这些二值图与人工标记的 ground-truth

图对比,计算 I 的精度与召回率。最后,通过计算数据库中所有图像的精度与召回率,获得综合的评价标准即精度-召回率曲线。假设 ground-truth 图的显著区域像素点个数为 S_g ,通过阈值分割所得二值图的显著性区域像素点个数为 S_a ,定义两者交集:

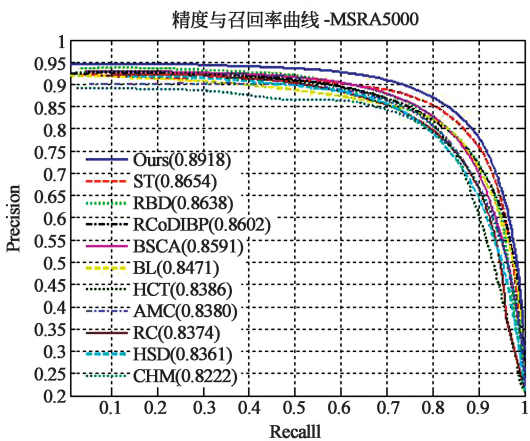
$$S = S_g \cap S_a, \tag{6}$$

那么精度(Precision)和召回率(Recall)的计算公式为:

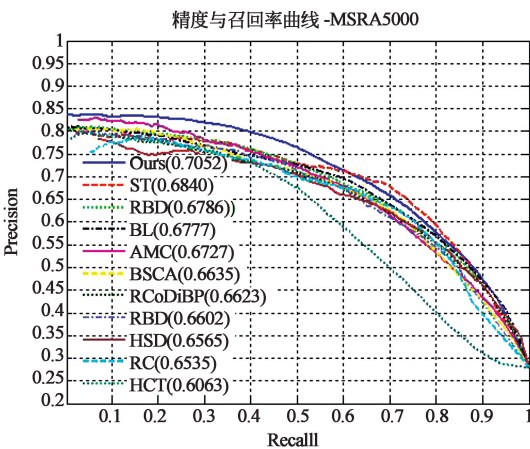
$$\text{precision} = S / S_a, \tag{7}$$

$$\text{recall} = S / S_g. \tag{8}$$

本文方法与以上 10 种方法分别在数据库 MSRA5000 和 BSD300 中获取显著图,通过对比显著图与 ground-truth 图计算精度和召回率,所得的精度-召回率曲线如图 7 所示。



(a)精度与召回率曲线-MSRA5000
(a)Precision-Recall curves on MSRA5000 dataset



(b)精度与召回率曲线-BSD300
(b)Precision-Recall curves on BSD300 dataset

图 7 精度-召回率曲线图
Fig. 7 Precision-Recall curves

由图 7 可看出,在数据库 MSRA5000 中,本文方法的精度-召回率曲线高于上述 10 种方法,在数据库 BSD300 中,本文方法的精度-召回率曲线大部分高于其他方法。在数据库 MRSA5000 与 BSD

300 中的 AUC 值分别为 0.891 8、0.705 2。说明本文方法在显著性检测中有较好的检测效果。

为了进一步分析本文方法的检测效果,将本文方法的检测速度与其他方法进行对比,测试结果如表 2 所示。

表 2 检测速度的对比	
Tab. 2 Comparison of detecting speed	
方法名称	每幅图的平均时间/s
AMC	0.08
BL	8.70
BSCA	0.86
CHM	8.90
HCT	2.47
HSD	0.28
RBD	2.44
RC	0.25
RCoDiBP	0.85
ST	103.92
本文方法	3.45

由表 2 可知,本文方法的检测速度介于其他 10 种方法中间。其中,监督学习的方法由于提取的特征通常比较复杂,检测时间相对较长,如 BL、CHM、HCT 以及本文方法。而本文方法与 BL、CHM、HCT 这 3 种方法对比,检测速度相对较快。

6 结 论

本文方法将结构化学习的方法应用到显著性目标检测中,通过学习特征与不同结构标签之间的对应关系构造结构森林模型,最后采用监督学习的方法进行优化预测,得到最终的显著图。实验结果表明,本文方法在数据库 MRSA5000 与 BSD300 中的 AUC 值分别为 0.891 8、0.705 2,说明本文方法具有较好的显著性检测效果。

参 考 文 献:

[1] 魏昱. 图像显著性区域检测方法及应用研究[D]. 济南:山东大学,2012.
WEI Y. Research on image salient regiondetection methods and applications [D]. Ji'nan: Shandong University, 2012. (in Chinese)

[2] 柯洪昌,孙宏彬. 图像序列的显著性目标区域检测方法[J]. 中国光学,2015,8(5):768-774.

- KE H C, SUN H B. A saliency target area detection method of image sequence [J]. *Chinese Optics*, 2015, 8(5): 768-774. (in Chinese)
- [3] 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.
- [4] 杜杰, 吴谨, 朱磊. 基于区域特征融合的 RGBD 显著目标检测[J]. *液晶与显示*, 2016, 31(1): 117-123.
- DU J, WU J, ZHU L. RGBD salient object detection based on regional feature integration [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(1): 117-123. (in Chinese)
- [5] KIM J, HAN D, TAI Y W, *et al.* Salient region detection via high-dimensional color transform [C]. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH: IEEE, 2014: 883-890.
- [6] KONTSCIEDER P, ROTA BULÒ S, PELILLO M, *et al.* Structured labels in random forests for semantic labeling and object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(10): 2104-2116.
- [7] DOLLÁR P, ZITNICK C L. Fast edge detection using structured forests [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(8): 1558-1570.
- [8] LIU T, YUAN Z J, SUN J, *et al.* Learning to detect a salient object [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(2): 353-367.
- [9] MOVAHEDI V, ELDER J H. Design and perceptual validation of performance measures for salient object segmentation [C]. *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops (CVPRW)*, San Francisco, CA: IEEE, 2010: 49-56.
- [10] JIANG B W, ZHANG L H, LU H C, *et al.* Saliency detection via absorbing Markov chain [C]. *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV)*, Sydney, NSW: IEEE, 2013: 1665-1672.
- [11] TONG N, LU H C, RUAN X, *et al.* Salient object detection via bootstrap learning [C]. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA: IEEE, 2015: 1884-1892.
- [12] QIN Y, LU H C, XU Y Q, *et al.* Saliency detection via Cellular Automata [C]. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA: IEEE, 2015: 110-119.
- [13] LI X, LI Y, SHEN C H, *et al.* Contextual hypergraph modeling for salient object detection [C]. *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV)*, Sydney, NSW: IEEE, 2013: 3328-3335.
- [14] YAN Q, XU L, SHI J P, *et al.* Hierarchical saliency detection [C]. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR: IEEE, 2013: 1155-1162.
- [15] ZHU W J, LIANG S, WEI Y C, *et al.* Saliency optimization from robust background detection [C]. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH: IEEE, 2014: 2814-2821.
- [16] CHENG M M, MITRA N J, HUANG X L, *et al.* Global contrast based salient region detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 569-582.
- [17] ZHU L, KLEIN D A, FRINTROP S, *et al.* A multisizesuperpixel approach for salient object detection based on multivariate normal distribution estimation [J]. *IEEE Transactions on Image Processing*, 2014, 23(12): 5094-5107.
- [18] LIU Z, ZOU W B, LE MEUR O. Saliency tree: a novel saliency detection framework [J]. *IEEE Transactions on Image Processing*, 2014, 23(5): 1937-1952.

作者简介:程藜(1990—),女,湖北武汉人,硕士,主要从事图像处理与显著目标检测方面的研究。E-mail: 342362994@qq.com

吴谨(1967—),女,安徽芜湖人,教授,博士生导师,主要从事图像处理与模式识别方面的研究。E-mail: 471854385@qq.com

朱磊(1982—),男,湖北武汉人,博士,讲师,主要从事图像处理与机器学习方面的研究。E-mail: 179901529@qq.com