

## 基于样本选择的RGBD图像协同显著目标检测

刘政怡\* 刘俊雷 赵鹏

(安徽大学计算机科学与技术学院 合肥 230601)

**摘要:** 协同显著目标检测的目的是在包含两张及以上相关图像的图像组中检测共同显著的物体。该文提出一种利用机器学习的方法对协同显著目标进行检测。首先, 基于4个评分指标从图像组中选择部分显著目标易于检测的简单图像, 构成简单图像集; 接着, 基于协同一致性的原则, 从简单图像集中提取正负样本, 并用深度学习模型提取的高维语义特征表示正负样本; 再者, 利用正负样本训练的协同显著分类器对图像中的超像素进行分类, 得到协同显著目标区域; 最后, 经过一个平滑融合的操作, 得到最终的协同显著图。在公开数据集上的测试结果表明, 所提算法在检测精度和检测效率上优于目前的主流算法, 并具有较强的鲁棒性。

**关键词:** 目标检测; 协同显著目标; RGBD图像; 深度学习; 分类器

中图分类号: TN911.73; TP391.4

文献标识码: A

文章编号: 1009-5896(2020)09-2277-08

DOI: 10.11999/JEIT190393

## RGBD Image Co-saliency Object Detection Based on Sample Selection

LIU Zhengyi LIU Junlei ZHAO Peng

(School of Computer Science and Technology, Anhui University, Hefei 230601, China)

**Abstract:** Co-saliency object detection aims to discover common and salient objects in an image group which contains two or more relevant images. In this paper, a method of using machine learning is proposed to detect co-saliency objects. Firstly, some simple images are selected to form a simple image set based on four scoring indicators. Secondly, positive and negative samples are extracted from the simple images set based on co-coherence characteristics, and high-dimensional semantic features are extracted by the deep learning model which receives RGBD four-channels input. Thirdly, the co-saliency classifier is trained by positive and negative samples, and co-saliency maps are generated by testing all the superpixels in the images by the co-saliency classifier. Finally, a smooth fusion operation is adopted to generate the final co-saliency map. Experimental results on the public benchmark dataset show that the proposed algorithm is superior to the state-of-the-art methods in terms of accuracy and efficiency, and it is robust.

**Key words:** Object detection; Co-saliency object; RGBD images; Deep learning; Classifier

### 1 引言

视觉注意力机制可以使人类自动捕获到复杂场景中最具有吸引力的那一部分物体。显著目标检测作为计算机视觉领域的一个分支, 致力于使计算机可以自动发现这些显著目标, 并且在过去十多年一直是一个蓬勃发展的研究课题。显著目标检测已应用于分割<sup>[1]</sup>、重定向<sup>[2]</sup>、增强<sup>[3]</sup>、前景标注<sup>[4]</sup>、模糊检测<sup>[5]</sup>等大量视觉任务中。

随着近年来数据量的爆炸式增长, 多个相关的图像需要被人们同时地处理。协同显著目标检测是对传统的单幅图像显著目标检测的拓展, 其目的是在多个相关的图像中发现共同且显著的物体。与单幅图像显著目标检测不同, 协同显著目标不仅需要各自的图像中显著, 并且需要在其他相关图像中也同样显著。

以往对显著目标检测的研究主要集中在RGB图像上, 仅利用颜色信息, 而忽略了深度信息<sup>[6]</sup>, 但是深度信息在许多计算机视觉任务中已被证实是有用的<sup>[7]</sup>。本文将使用深度信息对RGB颜色信息进行补充, 对一组相关的RGBD图像进行协同显著目标检测。

在现有的一些RGBD协同显著目标检测方法中, Song等人<sup>[8]</sup>利用基于bagging的聚类方法检测

收稿日期: 2019-06-03; 改回日期: 2020-03-01; 网络出版: 2020-06-27

\*通信作者: 刘政怡 liuzywen@ahu.edu.cn

基金项目: 安徽省自然科学基金(1908085MF182), 国家自然科学基金(61602004)

Foundation Items: The Provincial Natural Science Foundation of Anhui(1908085MF182), The National Natural Science Foundation of China(61602004)

RGBD图像组的协同显著目标,其中在深度图上提取的平均深度值、深度范围和方向梯度直方图(HOG)被用来表示深度属性。Cong等人<sup>[9]</sup>在超像素和图像级别上探索图像之间的对应关系,在此基础上提出了一种基于多约束特征匹配和交叉标签传播的RGBD图像协同显著目标检测方法。文献[10]提出一种RGBD图像的迭代协同显著目标检测框架,该框架融合了添加方案、删除方案和迭代方案。添加方案旨在通过引入深度先验信息到之前已存在的显著目标检测方法中来生成RGBD图像的显著图,删除方案在一个协同概率函数的基础上专注于捕获图像间的一致性关系,迭代方案通过细化循环作为一个优化过程。

但是上述方法主要利用聚类或者相似度对比的方法获取图像间的协同显著区域,本文提出一种利用机器学习(SVM)的方法来计算协同显著区域的思路,将一组图像中共同显著的区域作为正样本,非显著的区域作为负样本,学习一个协同显著分类器,利用该分类器计算RGBD图像协同显著图,并与初始显著图进行最终的融合。在获得正样本的过程中,并非所有图像的初始显著结果都是最优的,有些复杂图像的初始结果并不尽人意,因此借鉴文献[11],仅从一些简单图像中获取样本,并经过这些样本之间的相似度对比,过滤掉一些可能的噪音,以此训练出了一个更好的协同显著分类器。同时,在特征的表达方面,不同于已有的方法仅提取低水平特征,如RGB, LAB, HSV和LBP等,本文利用已有的深度学习模型提取高维语义特征,也增强了协同显著分类器的性能。

## 2 本文方法

本文提出的RGBD图像协同显著目标检测的模型如图1所示,主要包括初始显著图计算、简单图像选择、训练样本提取、协同显著分类器训练、协

同显著性计算和平滑性融合等6个步骤。首先利用已有的RGBD图像显著目标检测方法产生初始显著图,接着利用初始显著图计算覆盖得分、紧密度得分、对比度得分和深度置信度得分这4个指标,并依据这4个指标来选择简单图像,构成简单图像集,接着从简单图像集中提取正负样本,训练SVM协同显著分类器,并利用该分类器计算图像的协同显著值,得到协同显著图,最后将协同显著图与初始显著图进行平滑性融合,产生最终显著图。

### 2.1 初始显著图计算

给定一组 $N$ 张RGB图像 $\{I^i\}_{i=1}^N$ 及相应的深度图 $\{D^i\}_{i=1}^N$ 。首先,利用简单的线性迭代聚类算法(SLIC)<sup>[12,13]</sup>将RGBD图像进行超像素划分,形成超像素集 $R^i = \{r_m^i\}_{m=1}^{N_i}$ ,其中 $N_i$ 是第 $i$ 张RGB图像的超像素个数。然后,利用已有的一种RGBD单幅图显著目标检测方法DCMC<sup>[14]</sup>来生成初始显著图 $S_{\text{intra}}$ ,设置 $I^i$ 的超像素 $r_m^i$ 的显著值为 $r_m^i$ 内所有像素的显著值均值,形成超像素级的初始显著图 $S_{\text{intra}}(r_m^i)$ 。

### 2.2 简单图像选择

一组图像中,一部分图像的显著区域和背景区域区别较小,显著区域难以从背景中检测出来。与其相反,另有一些图像的显著区域易于检测。简单图像定义为显著区域容易被检测出来的图像,这些图像的初始显著图的可信度较高,可以利用覆盖得分、紧密度得分、对比度得分和深度置信度得分4个指标来衡量图像的简单性。具体定义如下:

(1) 覆盖得分:在一个好的显著图中,显著区域覆盖整个区域部分不应太多或太少。根据文献[15]的统计结果可知,显著区域占整个区域的百分比服从高斯分布,其均值和方差均为0.2左右。令 $P_{\text{cov}} = |S_f|/|S|$ ,其中 $|S_f|$ , $|S|$ 分别为显著区域和整个区域的像素数。因此覆盖得分 $r_{\text{cov}}$ 被定义为

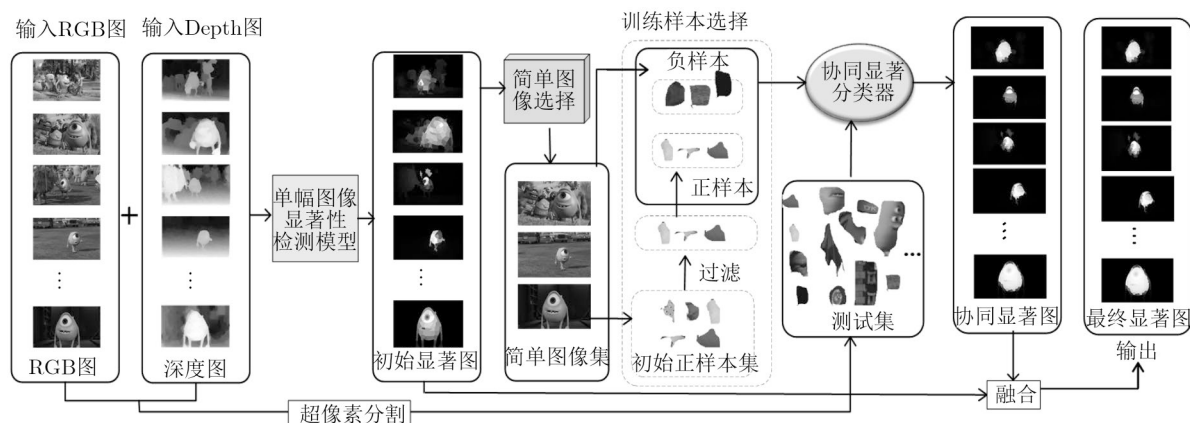


图1 本文提出的RGBD协同显著目标检测方法的框架图

$$r_{\text{cov}} = -\frac{(P_{\text{cov}} - \mu)^2}{\sigma^2} \quad (1)$$

其中,  $\mu = 0.2, \sigma = 0.2$ 。

(2) 紧密度得分：在一个好的显著图中，其显著的像素应该集中在一个紧凑的区域<sup>[16]</sup>。当窗口内显著值之和占整个显著图显著值之和为特定比例时，搜索这个比例下最小的窗口，利用此比例下最小窗口的平均显著值 $S_{\text{com}}^j$ 计算出紧密度得分 $r_{\text{com}}$

$$r_{\text{com}} = \frac{\sum_{j=1}^3 P_{\text{com}}^j \cdot S_{\text{com}}^j}{\sum_{j=1}^3 P_{\text{com}}^j} \quad (2)$$

其中特定比例设置为 $P_{\text{com}}^j \in \{0.25, 0.50, 0.75\}, j = 1, 2, 3$ 。

(3) 对比得分：一幅简单的图像，其显著区域和背景区域应有较大的对比度，颜色直方图之间的卡方距离可以用来衡量这种对比度差异。对于一幅图像，将每一个通道的颜色值量化到8个bins上，总共产生 $8^3$ 个bins，依此可以计算出颜色直方图。计算出显著区域和背景区域的颜色直方图 $h_{\text{cf}}$ 和 $h_{\text{cb}}$ ，利用卡方距离计算出对比得分 $r_{\text{cont}}$

$$r_{\text{cont}} = \chi^2(h_{\text{cf}}, h_{\text{cb}}) \quad (3)$$

(4) 置信度得分：一个好的深度图往往具有清晰的层次结构，可以提供准确的显著区域轮廓和边缘信息，有利于从背景中检测到显著区域<sup>[14]</sup>。因此，深度置信度 $r_{\text{conf}}$ 被当作一个指标用来衡量图像的简单与否，定义为

$$r_{\text{conf}} = \exp((1 - m_d) \cdot CV \cdot H) - 1 \quad (4)$$

$$H = -\sum_{i=1}^L (n_i/n_\Sigma) \lg(n_i/n_\Sigma) \quad (5)$$

其中,  $m_d$ 是深度图的平均深度值,  $CV = m_d/\sigma_d$ 是变异系数,  $\sigma_d$ 是图像的标准偏差,  $H$ 是表示深度分布随机性的深度频率熵,  $n_\Sigma$ 是整个深度图的全部像素个数,  $n_i$ 表示第 $i$ 个层次的像素个数,  $L$ 表示深度图的层次数, 整幅图被 $L-1$ 个阈值分割为 $L$ 个层次。 $r_{\text{conf}}$ 越高, 表示深度图的质量越好, 更容易区分前景和背景, 可以被选作简单图像。

最终一幅图像的简单与否的衡量指标 $R$ 定义为:  $R = r_{\text{cov}} \cdot r_{\text{com}} \cdot r_{\text{cont}} \cdot r_{\text{conf}}$ 。 $R$ 值越大, 代表着图像显著目标越容易从背景中被检测出来, 其初始显著图 $S_{\text{intra}}$ 的可靠性越高, 可以作为图像间的协同目标检测提供更准确的依据。输入的 $N$ 张图像的 $R$ 值被计算后, 对 $R$ 值由高到低进行排序, 取靠前的 $1/2$ 的图像作为简单图像集 $\Omega$ ,  $\Omega = \{\Omega^i\}_{i=1}^{N/2}$ 。

## 2.3 训练样本提取

在整个图像组中, 协同显著对象是属于同一类别的, 具有相似的外观, 而且初始显著图提供了有效的单幅图显著描述, 因此协同显著目标基本可以被包含在初始显著图的显著区域内<sup>[17]</sup>。所以在得到简单图像集后, 分别将简单图像集中显著值较高和较低的超像素提取出来作为训练样本的正负样本。

(1) 正样本提取：简单图像集 $\Omega$ 中的图像 $\Omega^i$ 的初始显著值较大的前 $T$ 个超像素被选中作为初始正样本, 它们构成初始正样本集 $\Phi_{\text{init}} = \Phi_{\text{pos}}^1 \cup \Phi_{\text{pos}}^2 \cup \dots \cup \Phi_{\text{pos}}^{N/2}$ 。由于初始显著图是一个粗略的结果, 虽已经过图像排序, 选取了易于显著目标检测的简单图像来提取训练样本集, 但还是不可避免地会在初始正样本集中包含一些背景区域和非协同显著目标区域, 从而降低了初始正样本集的准确度。因此, 一个协同一致性排序方案被用来对初始正样本集进行过滤, 提高构建的正样本集的准确度。

一般而言, 协同显著目标应满足3个约束: (1)类别相同, (2)颜色相同或相似, (3)在深度图中的深度分布相近。基于这3个约束, 使用了一个度量方法来衡量正样本集中各个样本间的协同一致性。首先, 正样本集通过K-means++聚类方法<sup>[18]</sup>被聚类成5类, 并且每一个样本都被分配到你相应的聚类中心 $\{\mathbf{c}_i\}_{i=1}^K$ , 然后协同一致性度量方法定义为

$$\text{mc}(r_m) = \left[ \sum_{\substack{n=1 \\ n \neq m}}^K (1 - \|\mathbf{c}_m - \mathbf{c}_n\|_2) \cdot w_{mn} \right] \quad (6)$$

其中,  $r_m, r_n \in \Phi_{\text{init}}$ ,  $\mathbf{c}_m$ 是正样本 $r_m$ 所对应的聚类中心,  $\mathbf{c}_n$ 是正样本 $r_n$ 所对应的聚类中心,  $K$ 表示正样本集中样本的个数,  $w_{mn}$ 表示正样本 $r_m$ 和正样本 $r_n$ 之间的特征相似度, 定义为

$$w_{mn} = \exp\left(-\frac{E(\mathbf{x}_m, \mathbf{x}_n) + \lambda_{\min} \cdot |d_m - d_n|}{\sigma^2}\right) \quad (7)$$

其中,  $\mathbf{x}_m, \mathbf{x}_n$ 分别为正样本 $r_m$ 和正样本 $r_n$ 的特征, 本文从文献[19,20]提出的RGBD图像显著检测网络中提取其第5层的高维语义特征作为正样本的特征表示, 如正样本 $r_m$ 的特征记为 $\mathbf{x}_m$ ,  $E(\mathbf{x}_m, \mathbf{x}_n)$ 表示正样本 $r_m$ 和正样本 $r_n$ 的欧式距离,  $\lambda_{\min} = \min(r_{\text{conf}}^m, r_{\text{conf}}^n)$ ,  $r_{\text{conf}}^m, r_{\text{conf}}^n$ 分别为使用式(4)计算得到的正样本 $r_m$ 和正样本 $r_n$ 所在图像 $I^m$ 和 $I^n$ 的深度置信度得分,  $d_m, d_n$ 分别为正样本 $r_m$ 和正样本 $r_n$ 的深度值,  $\sigma$ 是用来控制相似度强度的参数, 设置为0.1。

相对于其他的样本, 更大的mc值对应着更高

的协同一致性。也就是说,  $mc$  越大, 样本是协同显著区域的概率越大。协同一致性排序方案就是对初始正样本集  $\Phi_{init}$  中的样本按照  $mc$  的值由大到小排序。将排序后的前 80% 的初始正样本选择为最终的正样本, 记为  $\Phi_{pos}$ 。

(2) 负样本提取: 简单图像集  $\Omega$  中的图像  $\Omega^i$  的初始显著值较小的前  $T$  个超像素被选中作为负样本。选自整个简单图像集中不同的图像的负样本集整合为  $\Phi_{neg} = \Phi_{neg}^1 \cup \Phi_{neg}^2 \cup \dots \Phi_{neg}^{N/2}$ 。

## 2.4 协同显著分类器训练

在提取训练样本后, 为了得到协同显著目标区域, 一个典型的分类器支持向量机(Support Vector Machine, SVM)被用来对输入的图像组进行超像素层次上的分类<sup>[21]</sup>。SVM 是通过寻找最小化训练样本  $(\mathbf{X}, \mathbf{Y})$  的分类误差来学习参数, 以便实现分类效果

$$\left. \begin{aligned} J(\mathbf{X}, \mathbf{Y}) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M \xi_i \\ \text{s.t. } y_i (\mathbf{w}^T \mathbf{x}_i + \mathbf{b}) &\geq 1 - \xi_i; \xi_i \geq 0, \forall i = 1, 2, \dots, M \end{aligned} \right\} \quad (8)$$

其中,  $\mathbf{X}$  是表示样本的高维语义特征,  $\mathbf{Y}$  是训练样本的标签,  $\xi_i$  是松弛变量,  $C$  是惩罚参数。 $\mathbf{w}$  和  $\mathbf{b}$  是需要学习的参数。

## 2.5 协同显著性计算

在训练得到协同显著分类器之后, 使用此模型来计算一组图像的协同显著性值。针对一组  $N$  张 RGB 图像  $\{I^i\}_{i=1}^N$  及相应的深度图  $\{D^i\}_{i=1}^N$ , 首先将 RGBD 图抽象为超像素集  $R^i = \{r_m^i\}_{m=1}^{N_i}$ , 然后同样利用参考文献[19]提出的 RGBD 图像显著检测网络中提取其第 5 层的高维语义特征作为超像素  $r_m^i$  的特征表示  $\mathbf{x}_m^i$ , 超像素  $r_m^i$  的协同显著值计算方法如式(9)

$$S_{\text{cosal}}(r_m^i) = \mathbf{w}^T \mathbf{x}_m^i + \mathbf{b} \quad (9)$$

其中,  $\mathbf{w}, \mathbf{b}$  是在式(8)中被训练得出的参数向量,  $\mathbf{x}_m^i \in \mathbb{R}^d$  是超像素  $r_m^i$  的高维特征。由此, 输入的图像组的协同显著值  $S_{\text{cosal}}$  已被计算得到。

## 2.6 平滑性融合

为了得到一个更好的协同显著图, 一个多层元胞自动机(MCA)<sup>[22]</sup>被用来融合初始显著图  $S_{\text{intra}}$  和协同显著图  $S_{\text{cosal}}$ , 以获得空间上相邻的像素的显著值的一致性。

在多层元胞自动机(MCA)中, 显著图的每一个像素代表着一个元胞, 像素的显著值表示元胞的状态。不同显著图中相同位置的像素互为邻居, 并且在同一个显著图中, 每一个像素的前后左右的像素也互为邻居。如在  $M$  张显著图中, 每个像素共有

$5M - 1$  个邻居。像素  $i$  的显著值表示为其属于前景  $F$  的概率, 定义为  $P(i \in F) = S$ , 那么  $P(i \in B) = 1 - S_i$  为其属于背景  $B$  的概率。使用阈值  $\gamma$  对显著图进行分割, 高于阈值  $\gamma$  的被认为是前景, 记为  $\eta_i = +1$ , 低于  $\gamma$  被认为是背景, 记为  $\eta_i = -1$ , 其中  $\gamma$  是由 OTSU<sup>[23]</sup> 生成的自适应阈值。像素  $i$  的后验概率定义为  $P(i \in F | \eta_j = +1) = (s_i / (1 - s_i)) \cdot (\lambda / (1 - \lambda))$ , 其中  $\lambda = P(\eta_j = +1 | j \in F) = P(\eta_j = -1 | j \in B)$  是像素  $i$  的邻居  $j$  属于前景  $F$  或背景  $B$  的后验概率。因此, 显著图融合的同步更新规则为

$$\ln(S_m^{(t+1)}) = \ln(S_m^t) + \sum_{j=1}^5 \sum_{k=1}^M \delta(k=m, j>1) \cdot \text{sign}(S_{j,k}^t - \gamma \cdot \mathbf{1}) \cdot \ln\left(\frac{\lambda}{1-\lambda}\right) \quad (10)$$

其中,  $S_m^t$  表示在  $t$  时刻第  $m$  个显著图,  $M$  是不同显著图的个数,  $S_{j,k}^t$  是在  $t$  时刻第  $k$  个显著图中所有邻居中第  $j$  个邻居的显著值的一个向量,  $\delta(k=m, j>1)$  表示在第  $k$  张显著图中像素  $i$  的邻居是其在当前图像中邻接的 4 个邻居还是它在其他图像中相邻的 5 个邻居的指示函数。在经过 TC 次迭代后, 最终的协同显著图被计算为

$$S_{\text{fin}} = \frac{1}{M} \sum_{m=1}^M S_m^{(N)} \quad (11)$$

在本文中,  $M=2$ , 为初始显著图  $S_{\text{intra}}$  和式(8)中计算得出的协同显著图  $S_{\text{cosal}}$ , TC 被设置为 7。

## 3 实验

### 3.1 实验设置

将本文提出的 RGBD 协同显著目标检测方法在两个公开数据集 RGBD CoSal150 数据集和 RGBD CoSeg183 数据集上进行评估, 并且与现有的方法进行了对比。RGBD CoSal150 数据集包含分布在 21 个室内外场景的 150 张 RGBD 图像, 并且每一张 RGBD 图像都提供有像素级的真值图。RGBD CoSeg183 数据集包含 183 张分布在 16 个场景中的 RGBD 图像, 由于背景杂乱和前景复杂, 非常具有挑战性。在本文中, 超像素个数设定为 500, 提取样本时,  $T$  值被设为 20。

在定量评价中, 精度-召回率(P-R)曲线、平均 F-measure (Average F-measure, AF)、平均绝对误差(Mean Absolute Error, MAE)和结构相似性度量 (Structure similarity measure, S-measure) 等 4 个指标被用来对算法进行评估。P-R 曲线通过  $[0, 255]$  的滑动阈值对每个显著图进行二值化, 并和真值进行对比求得准确率  $P$  和召回率  $R$ 。PR 曲线显示了不同

阈值下显著图的查准率与查全率之间的关系。平均F-measure是一个整体的性能测量,被定义为准确率和召回率的加权平均值

$$F_{\beta} = \frac{(1 + \beta^2) P \times R}{\beta^2 \times P + R} \quad (12)$$

其中,  $\beta^2$ 被设置为0.3用来强调准确率<sup>[24]</sup>。平均绝对误差(MAE)被定义为显著图 $S$ 与真值图 $G$ 绝对误差的平均值<sup>[25]</sup>

$$MAE = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |S(i, j) - G(i, j)| \quad (13)$$

其中,  $W$ 和 $H$ 分别为图像的宽和高方向上的像素个数。MAE和F-measure指标都忽略了结构相似性评价,然而,行为视觉研究表明,人类视觉系统对场景中的结构高度敏感。因此,本实验还引入了S-measure<sup>[26]</sup>来进行更全面的评估。S-measure结合了区域感知( $S_r$ )和对象感知( $S_o$ )结构相似性作为它们的最终结构度量

$$S_{\beta} = \alpha \cdot S_o + (1 - \alpha) \cdot S_r \quad (14)$$

其中,  $\alpha \in [0, 1]$ 是平衡参数,被设置为0.5。

### 3.2 定量分析

在本节中,本文所提出的方法与4种先进的方法进行了比较: ESCS<sup>[27]</sup>, CBCS<sup>[28]</sup>, ICFS<sup>[10]</sup>和MCL<sup>[9]</sup>。ESCS和CBCS是对RGB图像进行协同显著目标检测的方法, ICFS和MCL是对RGBD图像进行协同显著目标检测的方法。ESCS和CBCS的实验结果来自于作者提供的代码运行得到, ICFS和MCL的实验结果由作者提供。图2给出了本文提出的方法和其他方法在两个数据集上检测结果示例,从第3行和第6行可以看出,本文方法检测的显著区域更加精准,包含更少的噪点;从第2行和第4行可以看出,本文方法检测的显著区域内部一致性较高。因此,本文方法相较其他方法可以更好地检测到协同显著区域。

图3和表1给出了本文方法与其他方法在两个数据集上的P-R曲线, F-measure, MAE和S-measure等几个评估指标的对比,可以看出本文提出的方法在两个数据集上优于其他的方法。图4和表2给出了本文方法选择简单图像以及采用深度学习获得的高维深度语义特征这两个策略在两个数据集上P-R曲线、F-measure, MAE和S-measure等几个评估指标的对比,可以看出简单图像选择和高维深度语义特征对实验效果是有提升的。图5给出了本文方法中

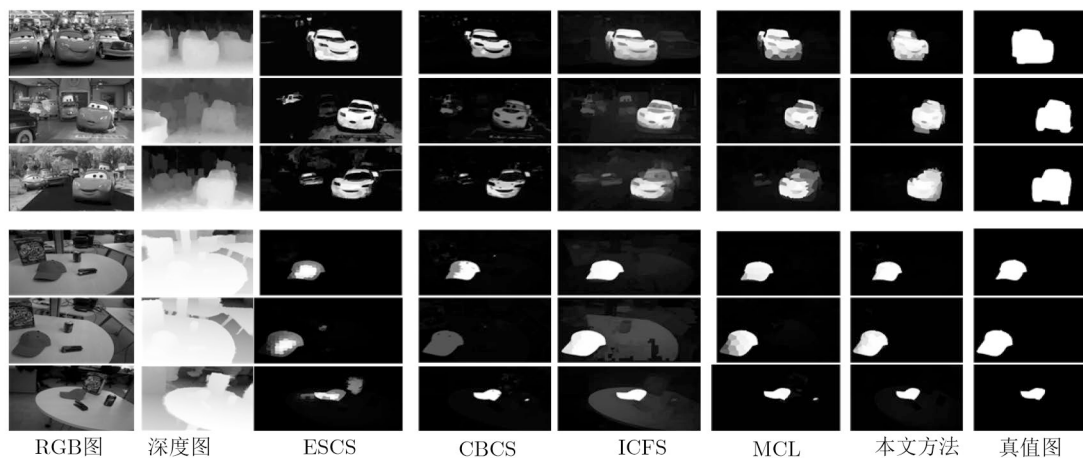


图2 不同方法生成的协同显著图对比

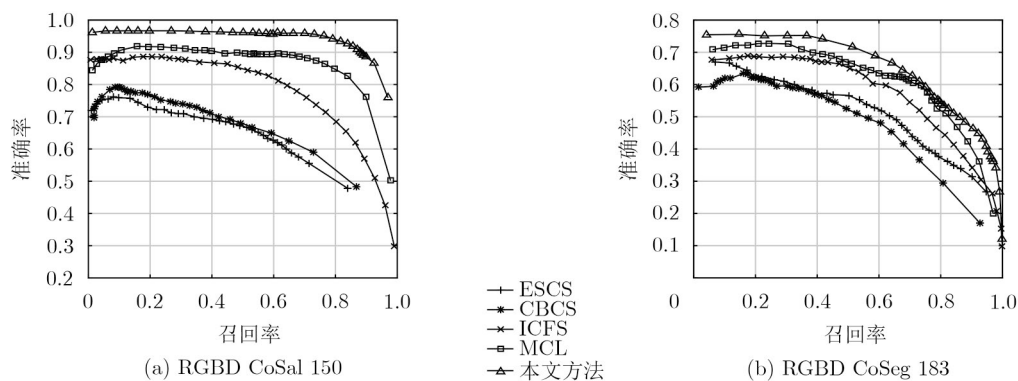


图3 本文算法与其他算法在两个数据集上的P-R曲线对比

影响协同显著目标检测性能的主要参数的对比。由图5(a)可以看出随着所使用的高维语义特征的层数的增加, 方法的性能逐渐提升, 因此选取第5层的语义特征作为样本的特征。由图5(b)可以看出, 当聚类中心个数为5时, 效果最好。实际上, 不同聚类中心个数下的性能是相似的, 表明本文方法对聚类中心这个参数不敏感。

3.3 运行时间分析

表3给出了不同协同显著目标检测方法的运行时间比较。ESCS, CBCS和本文方法运行在MATLAB R2016a平台, 测试环境为4核心3.40 GHz和16 GB RAM的PC上。ICFS和MCL方法的测试环境为4核心3.5 GHz, 16 GB RAM的工作站, 平台为MATLAB R2014b。从中可以看出本文方法虽然比不过ESCS

表 1 不同算法在两个数据集上的测试结果对比

	RGBD CoSal150			RGBD CoSeg183		
	S-measure	F-measure	MAE	S-measure	F-measure	MAE
ESCS	0.625	0.587	0.218	0.636	0.414	0.156
CBCS	0.572	0.582	0.215	0.622	0.365	0.116
ICFS	0.710	0.764	0.179	0.630	0.443	0.163
MCL	0.766	0.810	0.137	0.689	0.488	0.098
本文方法	<b>0.849</b>	<b>0.881</b>	<b>0.089</b>	<b>0.708</b>	<b>0.502</b>	<b>0.081</b>

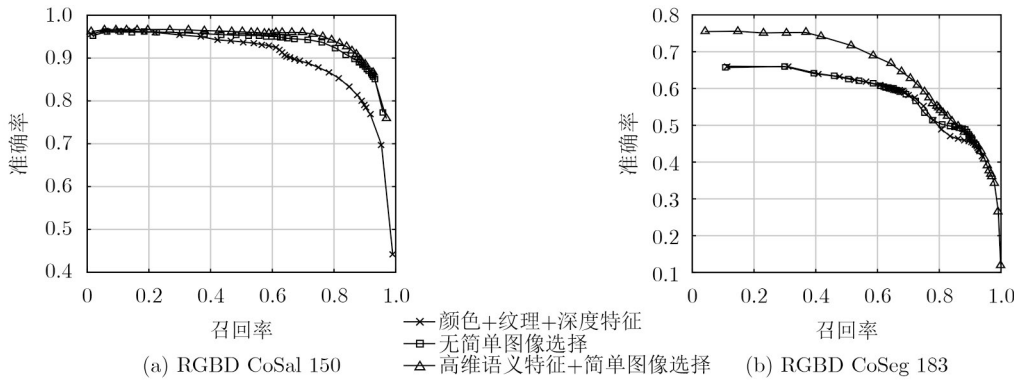


图 4 本文算法两个策略在两个数据集上的P-R曲线对比

表 2 不同模块在两个数据集上的测试结果对比

	RGBD CoSal150			RGBD CoSeg183		
	S-measure	F-measure	MAE	S-measure	F-measure	MAE
颜色+纹理特征	0.816	0.817	0.131	0.661	0.473	0.143
无简单图像选择	0.832	0.837	0.117	0.702	0.477	0.090
高维语义特征+简单图像选择	<b>0.849</b>	<b>0.881</b>	<b>0.089</b>	<b>0.708</b>	<b>0.502</b>	<b>0.081</b>

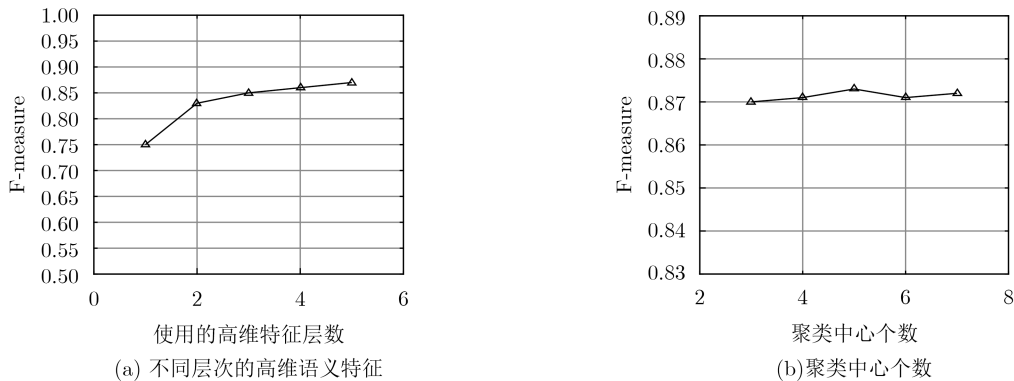


图 5 RGBD CoSal150数据集不同参数的F-measure测量

表3 不同方法平均每副图运行时间比较(s)

方法	ESCS	CBCS	ICFS	MCL	本文方法
时间	2.84	2.43	42.67	41.03	8.76

和CBCS这两种RGB图像协同显著目标检测方法的时间,但是相比于其他两种RGBD图像协同显著目标检测方法的时间,具有较大优势。

#### 4 结论

本文提出了一种协同显著分类器实现RGBD图像协同显著目标检测方法。本方法根据4个评分指标和协同一致性原则,利用简单图像的选择获取了可信度较高的正负样本;同时,在特征的表达方面,采用了深度学习模型从RGBD图像中提取高维语义特征,依此训练出了一个协同分类器,它能够比较准确地分类RGBD图像中的超像素;最后的平滑性融合结合了初始显著图的目标整体性与协同显著图的协同一致性,获得了最优的结果显著图。在公开数据集测试的结果表明,本文算法在检测精度和检测效率上优于目前的主流算法,具有较强的鲁棒性。

#### 参 考 文 献

- [1] WANG Wenguan, SHEN Jianbing, LI Xuelong, *et al.* Robust video object cosegmentation[J]. *IEEE Transactions on Image Processing*, 2015, 24(10): 3137–3148. doi: 10.1109/TIP.2015.2438550.
- [2] LEI Jianjun, WU Min, ZHANG Changqing, *et al.* Depth-preserving stereo image retargeting based on pixel fusion[J]. *IEEE Transactions on Multimedia*, 2017, 19(7): 1442–1453. doi: 10.1109/TMM.2017.2660440.
- [3] LI Chongyi, GUO Jichang, CONG Runmin, *et al.* Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior[J]. *IEEE Transactions on Image Processing*, 2016, 25(12): 5664–5677. doi: 10.1109/TIP.2016.2612882.
- [4] CAO Xiaochun, ZHANG Changqing, FU Huazhu, *et al.* Saliency-aware nonparametric foreground annotation based on weakly labeled data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(6): 1253–1265. doi: 10.1109/TNNLS.2015.2488637.
- [5] PANG Yanwei, ZHU Hailong, LI Xuelong, *et al.* Motion blur detection with an indicator function for surveillance machines[J]. *IEEE Transactions on Industrial Electronics*, 2016, 63(9): 5592–5601. doi: 10.1109/TIE.2016.2564938.
- [6] LEI Jianjun, LIU Jianying, ZHANG Hailong, *et al.* Motion and structure information based adaptive weighted depth video estimation[J]. *IEEE Transactions on Broadcasting*, 2015, 61(3): 416–424. doi: 10.1109/TBC.2015.2437197.
- [7] YANG Jingyu, GAN Ziqiao, LI Kun, *et al.* Graph-based segmentation for RGB-D data using 3-D geometry enhanced superpixels[J]. *IEEE Transactions on Cybernetics*, 2015, 45(5): 927–940. doi: 10.1109/TCYB.2014.2340032.
- [8] SONG Hangke, LIU Zhi, XIE Yufeng, *et al.* RGBD co-saliency detection via bagging-based clustering[J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1722–1726. doi: 10.1109/LSP.2016.2615293.
- [9] CONG Runmin, LEI Jianjun, FU Huazhu, *et al.* Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation[J]. *IEEE Transactions on Image Processing*, 2018, 27(2): 568–579. doi: 10.1109/TIP.2017.2763819.
- [10] CONG Runmin, LEI Jianjun, FU Huazhu, *et al.* An iterative co-saliency framework for RGBD images[J]. *IEEE Transactions on Cybernetics*, 2019, 49(1): 233–246. doi: 10.1109/tcyb.2017.2771488.
- [11] CHEN M, VELASCO-FORERO S, TSANG I, *et al.* Objects co-segmentation: Propagated from simpler images[C]. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia, 2015: 1682–1686. doi: 10.1109/ICASSP.2015.7178257.
- [12] ACHANTA R, SHAJI A, SMITH K, *et al.* SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274–2282. doi: 10.1109/TPAMI.2012.120.
- [13] GUO Jingfan, REN Tongwei, and BEI Jia. Salient object detection for RGB-D image via saliency evolution[C]. 2016 IEEE International Conference on Multimedia and Expo, Seattle, USA, 2016: 1–6.
- [14] CONG Runmin, LEI Jianjun, ZHANG Changqing, *et al.* Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion[J]. *IEEE Signal Processing Letters*, 2016, 23(6): 819–823. doi: 10.1109/lsp.2016.2557347.
- [15] MAI Long and LIU Feng. Comparing salient object detection results without ground truth[C]. The 13th European Conference on Computer Vision, Zurich, Switzerland, 2014: 76–91. doi: 10.1007/978-3-319-10578-9\_6.
- [16] LI Lina, LIU Zhi, and ZHANG Jian. Unsupervised image co-segmentation via guidance of simple images[J]. *Neurocomputing*, 2018, 275: 1650–1661. doi: 10.1016/j.neucom.2017.10.002.
- [17] CONG Runmin, LEI Jianjun, FU Huazhu, *et al.* HSCS: Hierarchical Sparsity based co-saliency detection for RGBD images[J]. *IEEE Transactions on Multimedia*, 2019, 21(7): 1660–1771. doi: 10.1109/TMM.2018.2884481.
- [18] ARTHUR D and VASSILVITSKII S. k-means++: The advantages of careful seeding[C]. The Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, USA,

- 2007: 1027–1035.
- [19] HUANG Posheng, SHEN C H, and HSIAO H F. RGBD salient object detection using spatially coherent deep learning framework[C]. The 23rd IEEE International Conference on Digital Signal Processing, Shanghai, China, 2018: 1–5.
- [20] LIU Zhengyi, SHI Song, DUAN Quntao, *et al.* Salient object detection for RGB-D image by single stream recurrent convolution neural network[J]. *Neurocomputing*, 2019, 363: 46–57. doi: 10.1016/j.neucom.2019.07.012.
- [21] HAN Junwei, CHENG Gong, LI Zhenpeng, *et al.* A unified metric learning-based framework for co-saliency detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 28(10): 2473–2483. doi: 10.1109/tcsvt.2017.2706264.
- [22] QIN Yao, FENG Mengyang, LU Huchuan, *et al.* Hierarchical cellular automata for visual saliency[J]. *International Journal of Computer Vision*, 2018, 126(7): 751–770. doi: 10.1007/s11263-017-1062-2.
- [23] OTSU N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1): 62–66. doi: 10.1109/TSMC.1979.4310076.
- [24] BORJI A, CHENG Mingming, JIANG Huaizu, *et al.* Salient object detection: A benchmark[J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5706–5722. doi: 10.1109/TIP.2015.2487833.
- [25] WANG Wenguan, SHEN Jianbing, and SHAO Ling. Consistent video saliency using local gradient flow optimization and global refinement[J]. *IEEE Transactions on Image Processing*, 2015, 24(11): 4185–4196. doi: 10.1109/TIP.2015.2460013.
- [26] FAN Dengping, CHENG Mingming, LIU Yun, *et al.* Structure-measure: A new way to evaluate foreground maps[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 4558–4567.
- [27] LI Yijun, FU Keren, LIU Zhi, *et al.* Efficient saliency-model-guided visual co-saliency detection[J]. *IEEE Signal Processing Letters*, 2015, 22(5): 588–592. doi: 10.1109/LSP.2014.2364896.
- [28] FU Huazhu, CAO Xiaochun, and TU Zhuowen. Cluster-based co-saliency detection[J]. *IEEE Transactions on Image Processing*, 2013, 22(10): 3766–3778. doi: 10.1109/TIP.2013.2260166.
- 刘政怡: 女, 1978年生, 副教授, 研究方向为计算机视觉、深度学习.
- 刘俊雷: 男, 1995年生, 硕士生, 研究方向为计算机视觉.
- 赵 鹏: 女, 1976年生, 副教授, 研究方向为智能信息处理、机器学习.
- 责任编辑: 陈 倩