

指导教师： 杨涛

提交时间： 2015-3-29

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 吕青侠

学号： 2012302470

班号： 10011204

自适应偏微分方程学习的视觉显著性检测

(刘日升, 曹俊杰, 林宙辰, 山世光)

摘要

偏微分方程(PDE)已经成功解决了许多低层次的视觉任务。然而, 由于将人类感知和高层次的先验知识合并到一个 PDE 系统中的难度, 使得直接利用 PDE 的视觉显著性检测成为一项非常具有挑战性的任务。本文不是以固定公式和边界条件设计 PDE, 而是提出了一种新的框架, 可以自适应地学习图像的 PDE 系统进行视觉显著性检测。我们假设图像元素的显著性可以从显著性种子(即, 最具代表性的显著元素)的相关性中得到。在这种观点中, 带有狄利克雷边界的一般线性椭圆系统 (LESD)被引入到从种子扩散到其他相关点的模拟中。对于一个给定的图像, 我们首先学习指导图来将人们的先验知识融合到扩散系统中。然后通过优化一个带有这种 LESD 约束的离散子模函数和一个统一的拟阵, 那么对于这张图片显著性种子(即边界条件)就可以被学习, 从而实现了最佳的偏微分方程系统对视觉显著性演变的模拟。对于各种具有挑战性的图片的实验结果表明, 我们所提出的基于学习的 PED 的视觉显著性检测是十分优越的。

1. 介绍

显著性检测作为许多计算机视觉问题(例如, 图像编辑[9], 分割[18], 压缩[12], 目标检测和识别[32])的一个重要组成部分, 在近些年来获得了极大的关注, 而且已经在文献中提出了大量的显著性检测。根据它们所代表的图像显著性的机制, 现有的显著性检测方法大致可以分为两大类: 自下而上和自上而下的方法。自下而上的方法[13, 7, 38, 36, 39, 34, 22, 15]是数据驱动式的, 而且更侧重于图像特征的检测显著性, 例如对比度, 位置和纹理。作为最早的显著性检测的研究者之一的 Itti 等人[13], 考虑了局部对比度和使用由中心向四周变化的图像差异性定义图像显著性。Cheng 等人[7]也研究了全局对比度先验知识。位置对于显著区域建模是另一个重要的先验知识。文献[38]应用了兴趣点的凸壳来估算前景的位置。在文献[39, 36]中将图像边界当做是背景的先验知识。受到最近机器学习技术发展的激发, 压缩传感[34, 22]和运筹学[15]也被运用到图像显

显著性特征的检测之中。文献[34, 22]假设一幅自然图像总是可以分解成鲜明的显著前景和均匀的背景。因此人们可以利用低阶稀疏矩阵分解方法及其扩展做显著性检测。就在最近, Jiang 等人[15]用公式将显著性检测表示成半监督聚类问题, 并且使用充分研究的设施点定位模型, 为显著性区域提取聚类中心。

相比之下, 自上而下的方法[26, 40]往往是任务驱动型的, 将人的感知融入到显著性检测中。例如, 刘等人。[26]提出了一个监督的方法来学习一个显著区域检测图像。杨等人[40]利用字典学习提取区域特征和 CRF 来生成显著性图。

在过去的几十年里, 偏微分方程已经展现了它们解决许多诸如图像恢复、图像平滑、图像修复和多尺度表示[参考文献 5 的简要回顾]等低层次的计算机视觉问题的力量。这主要是因为这些问题在数学物理和生物视觉这些领域的理论研究得到了解决。例如, 尺度空间理论证明了图像的多尺度[23]表示确实是不同时间参数热方程的解。

不幸的是, 现有的 PDE 模型设计方法 (即, 用固定的方程和一般的直观思维作为边界条件来定义 PDE) 不适合处理复杂的诸如视觉显著性检测的视觉任务。这是因为显著性是包含在图像的一种内在信息, 而且它的描述对人的感知依赖非常大。从自下而上的观点看 (即, 图像局部结构), 由于现实世界图像的显著性区域的复杂性, 用固定的公式和边界条件来准确定义一个能描述所有显著性的类型的 PDE 是非常具有挑战的。根据自顶而下的观点, 高层次的人类感知 (例如, 色彩[34], 中心[31], 语义信息[16]) 对显著性检测十分重要。但是将这些先验知识自动融合到传统的 PDE 中去是非常困难的。而且, 现有的大多数 PDE 的边界条件只是用一些对问题 (即, 定性保障[5]和初值[23]) 的一般性理解来定义, 因此不能处理复杂的 (例如, 通过数据和先验所驱动的) 视觉任务。整体来看, 带有固定公式和边界条件的传统 PDE 不能有效的量化和描述复杂的视觉显著性模式, 因此可能解决不了显著性检测问题。

1.1 论文的贡献

在这篇论文中, 我们提出了一个扩散的观点来理解这种机制和调查显著性检测的物理特性。首先, 提出来一个自适应 PDE 系统, 名为带有狄利克雷边界条件的线性椭圆系统 (LESD), 用于描述显著性扩散。接着我们开发了高效技术, 将自底而上和自顶而下的信息融合到显著性扩散中, 并且研究了 LESD 的特定公

式和边界条件。图 1 显示了在事例图像用我们基于 PDE 检测的过程。据我们所知，本文是将学习策略融入到 PDE 中进行显著性检测的第一篇文献。我们总结本文的贡献如下：

- 提供了一个新颖的 PDE 系统，用来描述在显著性扩散中视觉注意力的演变。我们证明了在我们的系统中视觉注意力是一个关于显著性种子的单调子模函数。
- 我们开发了一种有效的方法将自底而上和自顶而下的先验知识融合到 LESD 公式中进行显著性扩散。
- 我们利用 PDE 和拟阵约束得出一个离散优化模型，为 LESD 提取显著性种子。通过进一步证明所提出模型的子模性，使得性能保证得以实现。

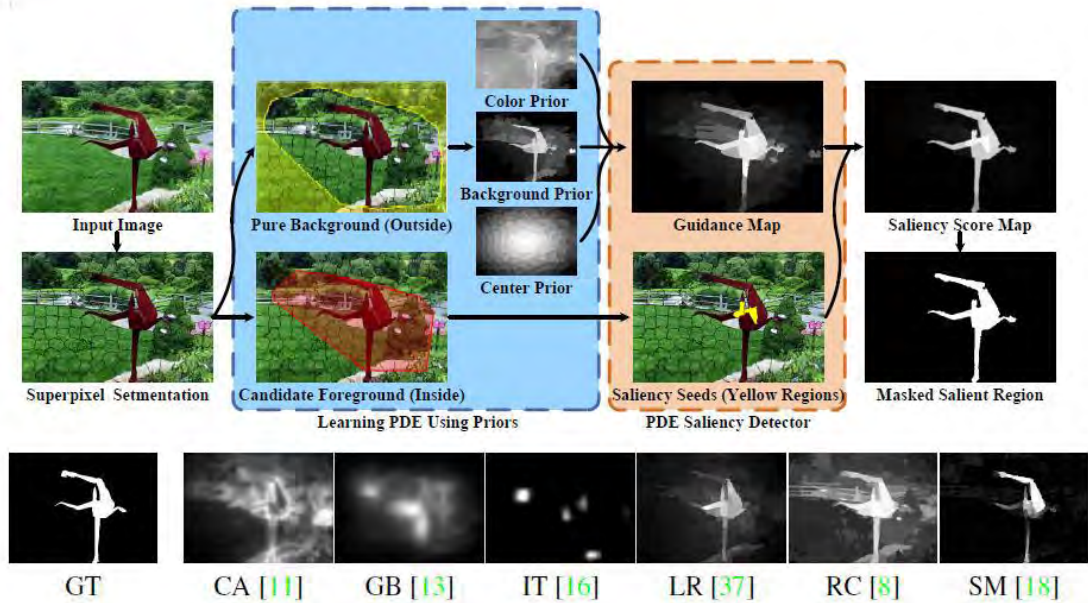


图 1. 在一个示例图片中用基于学习的 LESD 进行显著性检测的流程。橙色区域阐明了我们的 PDE 显著性检测子的核心组成(也就是指导图和显著性种子),这会在第二部分正式介绍。蓝色区域展示了怎样将自底向上和自顶向下的先验知识合并到我们的 PDE 系统中。这个 PDE 学习过程的细节会在第三部分中讨论。图片底部显示的是显著性区域的真实数据(简记为 GT)以及由一些高水准的显著性检测方法计算出的显著性图。

1.2 符号

在此，我们使用小写粗体字母（如 \mathbf{p} ）来表示向量点，花体符号（如 \mathcal{S} ）表示点集。 $|\mathcal{S}|$ 是 \mathcal{S} 的基数。 $\mathbf{1}$ 是全 1 的向量。用 N_p 表示 \mathbf{p} 的邻接点集。 $\|\cdot\|$ 表示 ℓ_2 的范数。假设 f 是在 \mathbf{v} 上的实函数,对于一个给定的点 \mathbf{p} ，其邻接点为 N_p ，我

们用 ∇f 表示 f 的梯度，分解表示为： $\nabla f == [f(p) - f(q_1), \dots, f(p) - f(q_{|N_p|})]$ 。

类似的， v 是 V 上的一个向量，将 v_p 作为在 p 点处的向量。 v 的散度记为 $\text{div}(v)$ ， v 在 p 的可表示为 $\text{div}(V_p) = \frac{1}{2} \sum_{q \in N_p} (V_p(q) - V_q(p))$ ，其中是 $V_p(q)$ 与 q 相一致的向量元素。

2. 使用 PDE 系统进行显著性扩散

这部分提出了一个扩散的观点来理解视觉显著性，以及建立一个 PDE 系统用于模拟图像中的显著性扩散。对于我们的系统同样也做了相应的数值模拟和理论分析的讨论。

2.1 视觉注意的演变

对于一个给定的视觉场景，显著性检测致力于聚焦在那些最可能引起人们注意的区域。本文从扩散点的角度完成这个任务。也就是说，假设我们的注意力最先是被最具有代表性的显著性图像元素（本文称之为显著性种子）所吸引，接着视觉注意力才会转移到所有的显著性区域。

具体来说，记 V 为离散图像域，即：一系列与所有图像元素（例如：像素或斑块）相对应的点。接着，我们定义了一个视觉注意力分数的实值函数 $f(p): V \rightarrow \mathbb{R}$ ，用于估量在 V 上 p 的显著性。假设我们已经知道一系列的显著性种子（记做 S ）及其相应的分数（也就是说，对于 $p \in S$ ，有 $f(p) = S_p$ ），那么我们可将显著性扩散以数学方式表示成一个带有狄利克雷边界条件的改进 PDE：

$$\frac{\partial f(p,t)}{\partial x} = F(f, \nabla f), \quad f(g)=0, \quad f(p) = S_p, \quad p \in S,$$

其中， g 是一个具有 0 分的环境点（ V 之外）， F 是 F 的一个函数，且其梯度为 ∇f 。

由于上述 PDE 的目的是使视觉注意力从显著性种子转移到其他图像元素，所以我们采用了一个线性离散项 $\text{div}(K_p \nabla f(p))$ ，用它表示分数的函数，其中 K_p 是一个非均匀的度量张量用于控制在 p 处的局部扩散率。为了将我们的感知与高层先验知识合并到扩散处理，我们进一步引入了一个规格化项，它可以用公式表示 $f(p)$ 和指导图 $g(p)$ （将在第三部分讨论它）的差值，形式如下：

$$F(f, \nabla f) = \text{div}(K_p \nabla f(p)) + \lambda(f(p) - g(p)),$$

其中 $\lambda \geq 0$ 是一个平衡参数。

2.2 带有狄利克雷边界的线性椭圆系统

出于显著性检测的目的,我们仅考虑显著性变化稳定时(即没有显著性注意力能进一步被扩散)的情况。这样,我们在符号中省略了时间 t ,且仅对下面 PDE 的求解:

$$F(f, \nabla f) = 0, \quad f(g) = 0, \quad f(p) = \delta_p, \quad p \in \mathcal{S}, \quad (1)$$

这是一个带有狄利克雷边界的线性椭圆系统(LESD)。因此,对于给定一幅图片,显著性检测任务便成为计算用于视觉注意力扩散的 LESD,来实现一种稳定状态的问题。

直到现在,我们已经建立了一个一般的用于显著性扩散的 PDE 系统。图 1 显示了我们的 LESD 带有适当学习的 g 和 S) 可以成功地把图像结构和高层知识合并以模拟显著性扩散,这样可以实现比高水准方法更好的显著性检测结果。因此,对于 LESD 剩下的最重要的问题是开发一个有效的学习框架,将自底向上的图像结构星系与自顶向下的人类先验知识结合起来,使之满足 (1)。在第 3 部分讨论这个问题之前,我们首先为 LESD 提供了必要的数值模拟和理论分析,这大大减小了处理的复杂度。

2.3 离散化

假设 $N_p = \{q_1 \cdots q_{|N_p|-1}, g\}$ 是 p 的集合的相邻集合。在这里第一个 $|N_p| - 1$ 节点在图像域 V 中,并且将在下一部分中指定。环境点 g 与每一个节点相连接[37]。为了检测 p 和其邻接点之间的变化,我们定义了如下面对角矩阵的非均匀的度量张量 K_p :

$$K_p = \text{diag}(k(p, q_1), \cdots, k(p, q_{(|N_p|-1)}), z_g), \quad (2)$$

其中, $k(p, q) = \exp(-\beta \|h(p) - h(q)\|^2)$ 是节点的特征之间的高斯相似度(带有强度参数 β), $h(p)$ 是节点 p 处的一个特征向量, z_g 是一个估量传导耗损的小常量。

然后我们就可以用下面的公式近似 LESD:

$$f(p) = \frac{1}{d_p + \lambda} \left(\sum_{q \in \mathcal{N}(p)} K_p(q) f(q) + \lambda g(p) \right), \quad (3)$$

其中, $K_p(q)$ 是与 q 一致的 K_p 的对角元素, 且满足 $d_p = \sum_{q \in N_p} K_p(q)$ 的对角元素。

基于这个表达式, 我们的 LESD 可以重新定义为一个以

$Af = b$ 的形式表示的线性系统, 这样就可以通过预条件共轭梯度法解决了。

2.4 理论分析

应该强调的是, 视觉注意力分数 f 的确是在 \mathcal{V} 上的一个集合函数, 也就是说, 当 f 是式 (1) 对应于显著性种子集合 S 的解时, $f(S): 2^{\mathcal{V}} \rightarrow \mathbb{R}$ 。这意味着, 我们的 LESD 的解释固有的组合, 因此比在传统的低层图像处理中的 PDE 处理起来会困难得多。这是因为在不知道任何更具体的属性的时候, 一个组合 f 的优化会极其困难 (比如说, 琐碎的最坏情况的指数时间, 甚至是没有近似算法的[21])。幸运的是, 通过证明以下定理, 我们可以利用 LESD 的解的一些好的属性, 比如单调性 (即非递减) 和子模。正如第 3 部分所展示的, 这些结果为我们的显著性检测问题提供了好的保证。

对于图像元素 $p, f(p; S)$ 是视觉注意力分数, 并且起点 $\{s_p \geq 0\}$ 附加在显著性种子集合 S 中, 也就是说, 对于所有的 $p \in S$, 有 $f(p) = s_p$ 。那么 f 是一个关于 $S \subset \mathcal{V}$ 的单调子模函数。

3. 用于显著性检测的学习 LESD

这部分将讨论在给定的图像中, 对于显著性扩散, 怎样自适应地学习一个特定的 LESD。对于一个给定的图像, 我们首先在图像特征空间构造一个无向图, 来模拟图像元素的邻接关系。接着, 我们将人类先验知识和图像结构合并, 来建立扩散函数 (即: 指导图 g)。基于系统的子模块, 我们也为边界条件 (即: 显著性种子 S) 学习提供了一个离散的优化模型。

3.1 特征提取和图像构造

对于一个给定的图像, 我们生成超像素来建立图像元素集 $\mathcal{V} = \{p_1, \dots, p_{|\mathcal{V}|}\}$ 。在这里, 可使用任何边界条件的超像素法, 而且本文采用了 SLIC 算法[3]。接下来, 我们定义了特征向量 $\{h(p), p \in \mathcal{V}\}$ 作为在 CIELAB 颜色空间中超像素的均值。

图像结构信息的抽取步骤如下。假设图像域 \mathcal{V} 包含两个部分: 候选前景 \mathcal{F}_c (显著性区域, 也可能包含一些混杂的图像元素)和纯背景 \mathcal{B}_c (非显著性区域)。我们利

用一个漂移凸壳策略来近似估计输入图像的这两个子集。特别地，我们使用 Harris 算子[35]来粗略地检测角点和轮廓点，以及由这些点估计凸壳 C [38]。接着， \mathcal{F}_c 可以通过 C 中的节点来获得。为了进一步识别纯背景节点，我们将通过将邻接节点加到 C 上，定义了一个消耗壳 C' 。然后，通过收集所有 C' 外的节点得到 \mathcal{B}_c 。请看图 2 中 C 和 C' 的一个事例。

现在我们构造了一个无向图 $G = (V, E)$ 来揭示在图像域中的连接关系（即对于每一个 p 的 N_p ），其中 E 是与点集合 V 相应的一组无向边。我们首先定义一个正则图结构以便利用局部空间关系（图 2 (b)）。然后，在 \mathcal{B}_c 中所有的节点被连接到背景的强平滑部分（图 2 (c)）。由于可能存在混杂的图像元素，我们对 \mathcal{F}_c 中的节点不做进一步的连接。最后，所有节点被连接到了一个环境节点 g 。

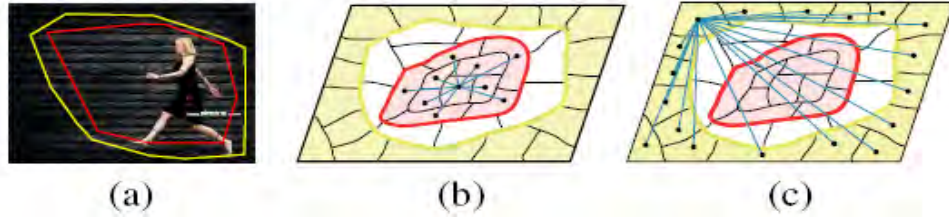


图2 (a) 说明漂移凸壳策略，(b) - (c) 说明连接关系。在 (a) 中红色和黄色的多边形分别表示 C 和 C' ，在 (b) - (c) 中红色和黄色的区域分别代表 \mathcal{F}_c 和 \mathcal{B}_c 。在 (c) 中的线段意味着所有在中节点的相互连接关系。

3.2 使用先验知识学习指导图

本节展示怎样将不同类型的先验知识合并到 PDE 系统中。对于一幅给定的图像，我们首先定义一个背景扩散来评估先验知识的差别。即，我们假设背景分布完全不同于前景，这样，我们执行一个 $\lambda = 0$ 的简化 LESD，实现背景扩散分数，即：

$$\operatorname{div}(K_p \nabla f_b(p)) = 0, \text{ s.t. } f_b(g) = 0, f_b(p) = 1, p \in \mathcal{B}_c$$

这里的边界条件是通过将 \mathcal{B}_c 当作是分数为 1 的背景种子结合，而且，添加了一个分数为 0 的环境点 g 。很容易检测出，背景扩散的解释一个谐波函数，这样， $f_b(p) \in [0, 1]$ 。所以，在 f_b 中的元素可以看作是属于背景的节点的概率。这种观点认为，我们有一个节点属于前景的概率 $f_f(p) = 1 - f_b(p)$ 。通过进一步合并高层

先验知识（如，色彩先验知识图 f_c 和中心先验指示图 f_l ），我们将指导图 $g(p)$ 定义成：

$$g(\mathbf{p}) = f_f(\mathbf{p}) \times f_c(\mathbf{p}) \times f_l(\mathbf{p}), \quad (4)$$

且其值被归一化了。为了给LESD提供良好的边界条件，我们也使用 g 来定义显著性种子的分数，也就是说，对于 $\mathbf{p} \in \mathcal{S}$ ，有 $s_p = g(\mathbf{p})$ 。

3.3 通过子模性质优化显著性种子

由于下面两个原因，我们不能选择 \mathcal{F}_c 中的所有节点作为显著性扩散种子。首先，凸壳可能没有充分抑制 \mathcal{F}_c 中的背景节点（图4（c））；其次，更为重要的是，观察发现，与其邻接点相比，带有极高局部对比度的种子（例如，靠近物体边界的节点、物体上的明节点或暗节点等）也可能导致生成较差的显著性图（图4（d））。因此，在 \mathcal{F}_c 中查找最具代表性的前景节点以定义LESD边界条件是必要的。既然LESD的目标是使视觉注意力分数从种子 \mathcal{S} 转移到整个图像域 V ，那么，我们只要在显著性扩散稳定时，最大化关于 V 中所有图像元素的分数 f 的总和即可，也就是对下面的离散优化问题求解：

$$\begin{aligned} & \max_{\mathcal{S} \in \mathbb{M}^n} L(\mathcal{S}), \\ & \text{s.t.} \begin{cases} f(\mathbf{p}) = \frac{1}{d_{\mathbf{p}} + \lambda} (\sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \mathbf{K}_{\mathbf{p}}(\mathbf{q}) f(\mathbf{q}) + \lambda g(\mathbf{p})), \\ f(\mathbf{g}) = 0, f(\mathbf{p}) = s_{\mathbf{p}}, \mathbf{p} \in \mathcal{S}, \end{cases} \end{aligned} \quad (5)$$

其中， $L(\mathcal{S}) = \sum_{\mathbf{p} \in V} f(\mathbf{p}; \mathcal{S})$ 和 $\mathbb{M}^n = \{\mathcal{S} | \mathcal{S} \subset \mathcal{F}_c, |\mathcal{S}| \leq n\}$ 是一个统一拟阵[4]，用于保证 \mathcal{S} 的基数不会超过 n 。由于视觉注意力分数可当做是与种子的节点相关性，那么上面的最大标准自然而然就会趋向于选择带有相对较大的连接子图作为种子（这样才更具有代表性）。因此，在 \mathcal{F}_c 中却具有高的局部对比度（即与其他节点有较少的连接性和路径）的节点，也会被排除在 \mathcal{S} 之外。有人可能会担心，背景节点同样将会有大的 L 值，因为它们可能与 \mathcal{F}_c 外的节点相连。幸运的是，通过学习一个恰当的指导图 g ，我们可以强制让背景区域（如图4（b）中的 g ）具有非

常小的显著性分数（大多数情况接近于0）。这样，在 \mathcal{F}_c 中的背景节点仍然具有相对较小的L值，并且不会被包含在 \mathcal{S} 中。

通常，（5）的效果取决于显著性种子的最大个数 n （图5（a））。在这里，我们提供一个适当的方式来识别 n ，并进一步抑制 \mathcal{F}_c 中的背景节点。首先在 \mathcal{F}_c 上定义一个背景置信函数 $\mathcal{W}(p) = 1 / (1 + g(p)^2)$ ，其中大 $\mathcal{W}(p)$ 的意味着 p 很可能属于背景，而且应该被抑制。因此，我们最大化在（5）中的另一个代价函数 $\hat{L}(\mathcal{S}) = L(\mathcal{S}) - \sum_{p \in \mathcal{S}} \mathcal{W}(p)$ 。基于定理1，可以证明下面关于L和 \hat{L} 的推论。

推论2 $L(\mathcal{S})$ 和 $\hat{L}(\mathcal{S})$ 都是子模函数，而且 $L(\mathcal{S})$ 是关于 \mathcal{S} 的单调函数。L的单调性和子模性以及（5）中的统一拟阵约束意味着，使用贪心算法解决（5）会产生一个 $(1-1/e)$ -逼近[29]。由于非单调性，我们不能保证同样的理论适用于 \hat{L} 。但是，在实际中，通过添加终止条件 $\hat{L}(\mathcal{S} \cup \{p\}) \leq \hat{L}(\mathcal{S})$ ，对于 \hat{L} 的最大化处理可以自动停止，并且可以相应地得到优化的种子集。实验中，我们发现，带有这种终止条件的贪心算法在我们的显著性检测问题中能有效最大化 \hat{L} 。

在这部分的最后，我们在算法1中总结了用于基于学习的LES D框架的算法细节。在图1中也说明了在一个测试图片中我们的显著性监测子的完整流程。

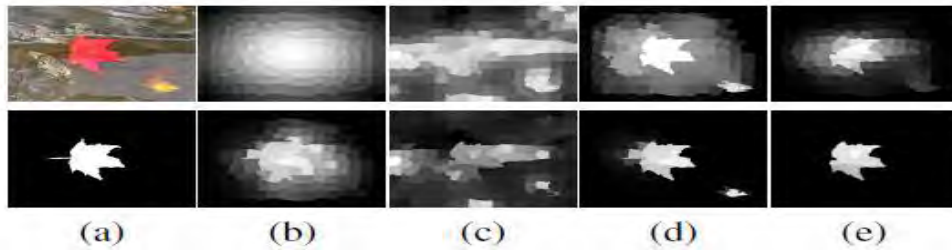


图3 带有不同指导图的显著性扩散。（a）是输入图像及GT显著性区域。（b）-(e)分别为中心先验 f_l ，色彩先验 f_c ，背景扩散先验 f_f ，决定性指导图 g （顶）以及它们对应的显著性图（底）。

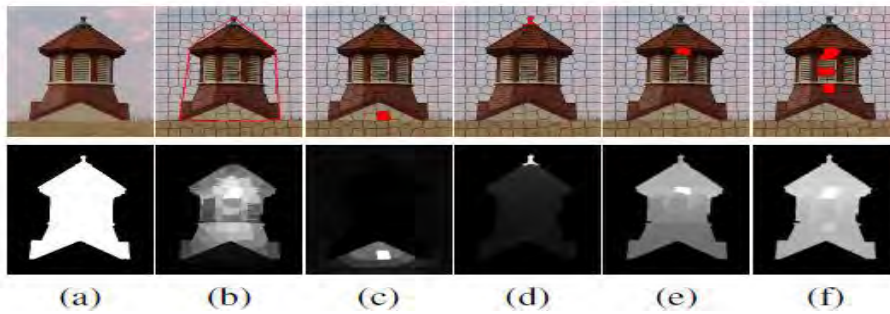


图4 具有不同种子的显著性扩散。（a）为输入图像与GT显著性区域。（b）为 F_c （红色多边形内部）和 g 。(c)-(e)为在 F_c 中使用同一个候选种子的扩散结果：(c)为背景($L=10.6175$)，

(d)为较差的前景(L=1.6818)，(e)较好的前景(L=31.7404)。(f)为最佳种子(L=43.8589)和最终显著图。处于视觉比较的目的，在此，我们采用原始显著图而非规格化后的显著图描述L的值。

4. 论述

在本节中，我们会讨论和强调我们提出的基于PDE的显著性检测的一些内容。

4.1 与现有的基于学习的PDE 的比较

最近，刘教授等人[24、25]利用最优控制技术训练PDE 用于图像处理。尽管他们[24、25]和我们的工作目标都在于学习一个PDE 来表述图像，但是我们的学习策略与他们的是不同的。在文献[24、25]中，他们采用了一个非线性的PDE，并学习来自训练图像对（通过手动收集）中的组合系数（即PDE 排列）。而我们的框架考虑的是一个线性椭圆系统，真正地学习PDE 的形式及其边界条件，使自底向上的图像结构与自顶向下的人类感知合并入我们的PDE 系统中。因此，我们能够成功处理更为复杂的显著性检测任务。

算法1 用于显著性检测的学习型LESD

输入：给定的一幅图像I以及必要的参数。

输出：给定图像的显著性图。

- 1: 构造图像在I的超像素上的图G。
- 2: 利用 (4) 式计算指导图g。
- 3: 初始化显著性种子集合 $\mathcal{S} \leftarrow \emptyset$ 。
- 4: **while** $|\mathcal{S}| \leq n$ **do**
- 5: **for** $p \in \mathcal{F}_c/\mathcal{S}$ **do**
- 6: Solve (3) with saliency seeds $\mathcal{S} \cup \{p\}$ for f .
- 7: Obtain the gain $\Delta L(p) = L(\mathcal{S} \cup \{p\}) - L(\mathcal{S})$,
or $\Delta \hat{L}(p) = \hat{L}(\mathcal{S} \cup \{p\}) - \hat{L}(\mathcal{S})$.
- 8: **end for**
- 9: $p^* = \arg \max_{p \in \mathcal{F}_c/\mathcal{S}} \Delta L(p)$ or $\arg \max_{p \in \mathcal{F}_c/\mathcal{S}} \Delta \hat{L}(p)$.
- 10: **if** $\hat{L}(\mathcal{S} \cup \{p^*\}) \leq \hat{L}(\mathcal{S})$ (only for \hat{L}) **then**
- 11: Break.
- 12: **end if**
- 13: $\mathcal{S} \leftarrow \mathcal{S} \cup \{p^*\}$.
- 14: **end while**
- 15: 利用最佳的 g^* 和 \mathcal{S}^* 求解 (3) 实现 f^* 。

4.2之前视觉模型中的子模性

子模性对于离散集合函数来说是一个重要的属性，它在运筹学和机器学习[20]方面有着广泛的应用。就在最近，它也被认同和应用于计算机视觉领域[19, 17, 15]。尽管在文献[15]中提及了在他们的显著性检测子中的子模性，但我们的工作机制与他们的完全不同。特别的，在[15]中的子模优化模型被用作提取聚类中心和图形聚类，而且，在他们的框架中，需要显著性图的计算步骤。相反，我们设计了一个子模优化模型来学习PDE系统的狄利克雷边界条件，通过PDE系统的解直接提取显著图（不需要进一步的后处理）。下面部分的实验结果显示，我们的方法实现起来比[15]具有更精确的显著性区域。

5. 实验结果

实验在三组图片上进行，这些图片由两个数据库生成：MSRA[26]和Berkeley[28]。首先，我们从[2]（MRSA-1000）所提供的1000张图片中选取了MSRA的一个子集。然后，与有5000张图（MRSA-5000）的整个MSRA数据库进行比对。最后，我们对Berkeley图片集中的300多张具有挑战性的图片进行算法测试。我们对所有的测试图片设置超像素数为200。将我们的方法（在比较过程中记做“PDE”）与17个高水准的显著性检测子进行比较，这些检测子包括：IT[13]，AC [1]，CA [9]，CB [14]，FT [2]，GB [10]，GS [36]，LC [41]，LR [34]，MZ [27]，RC [7]，SER[33]，SF [30]，SR [11]，SM [15]，SVO [6] 和XIE [38]等。为了进行量化比较，我们对三组图片的精确度、查全率和特征测量分别进行报道。我们也为比较方法展示了显著图以及真实数据（GT）显著性区域。对于我们的方法，我们在实验中对所有测试图片设置高斯相似 $k(p, q)$ 中的， F 中的 $\beta = 10$ ， F 中的 $\lambda = 0.01$ 。

5.1 定量比较

将我们的方法和其他高水准的方法分别在MSRA-1000、MSRA-5000和Berkeley的图片上进行量化比较。按照[2, 7, 38, 15]中的计算方法，计算平均精确度、查全率和特征测量值。

我们首先比较我们的两个目标函数（即 L 和 \hat{L} ）在MSRA-1000图像集上的性能，并将结果显示在图5（a）中。可以看到， \hat{L} 策略的效果比较好（红色曲线）因为这种非单调性模型可以自适应地确定最佳的 \mathcal{S} 。当我们为 L 定义一个合适的

种子数（这里 $n=10$ ）时，这个单调模型同样能够有较好的效果（黑色曲线）。但是可以看到，基于L的策略的结果取决于显著性种子的数目（蓝色和绿色曲线）。这是因为，小的 n 值可能导致非充分扩散，而过大的 n 值则会将错误的节点引入种子集。基于这样的观察，我们通常在后面的实验中使用 \mathcal{L} 策略。

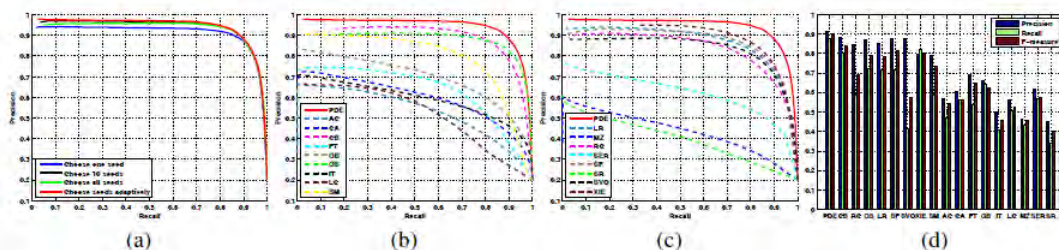


图5 MSRA-1000图片集上的结果。(a)为不同设计选项方法的精确度-查全率曲线。(b) - (c)为所有测试方法的精确度-查全率曲线。(d)为平均精确度、查全率和特征测量值。

所有17种方法在MSRA-1000上的精确度-查全率曲线如图5(b)和(c)。使用一个自适应阈值[2]的平均精确度、查全率和特征测量值展示在图5(d)中。我们同样对MSRA数据库中的5000张图片进行了实验。为了实现更为合理的比较结果，这里我们使用了精确的人为标签标记，而非之前文献中的包围盒，来评估显著性检测结果。实验结果如图6所示。Berkeley图片集比MSRA图片集更具有挑战性，因为其中的许多图片包含多个具有不同大小和位置的前景目标。我们将比较结果表述在图7中。

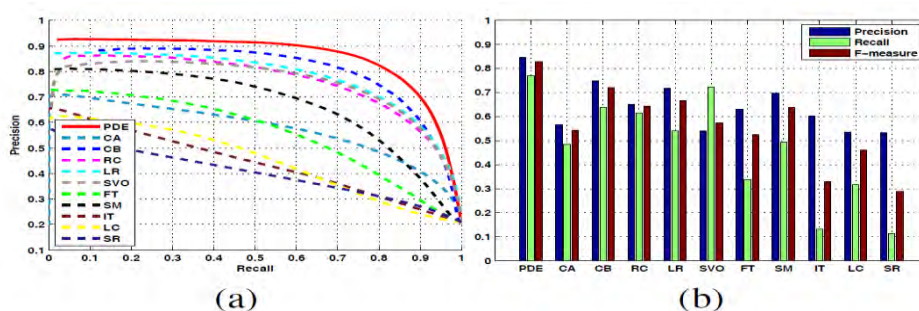


图6 MSRA-5000图片集上的结果。(a)为精确度-查全率曲线。(b)为平均精确度、查全率和特征测量值。

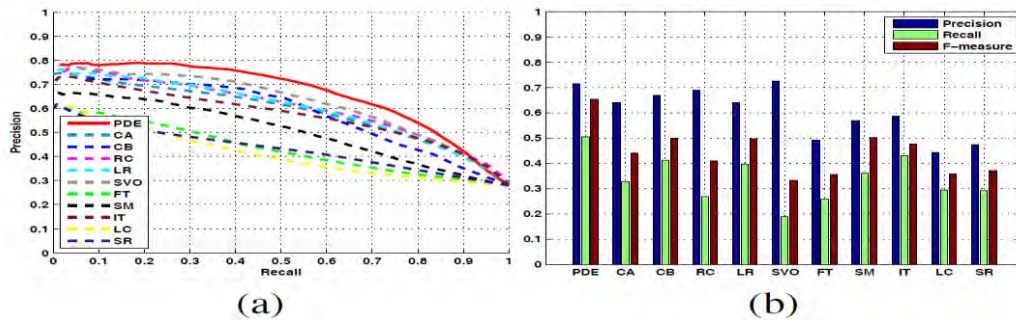


图7 Berkeley图片集上的结果。(a)为精确度-查全率曲线。(b)为平均精确度、查全率和特征测量值。

基于中心-周围对比度的方法，例如IT [13]，GB [10]和CA [9]，只能检测显著目标的边界部分。使用最近的超像素方法，如CB [14]和RC [7]，能够检测显著性目标。但是它们无法抑制背景区域，而且同样会导致较低的精确度-查全率曲线。在图5 (b) 中，我们观察到GS[36]在查全率超过0.96 的时候，与我们的精确度相近。然而，在那种方法中，当背景颜色与边界的颜色完全不同时，边界策略的侧地距离趋向于将背景部分当做是显著性区域。因此，在大多数情况下，在同样的查全率时，他们的精确度远比我们的低。可以看到，我们的PDE 显著性检测子在所有的三个具有挑战性的图像集中，都有全局最佳的效果。这些实验结果也证明了，学习策略可以成功地将自底向上和自顶向下的信息合并到显著性扩散中。

5.2 定性比较

在图8中，我们展示了通过一些典型的显著性检测子计算得到的显著图。作为一种基于眼睛定位预测的方法，IT [13]只能辨别中心-四周的差异，而忽略了大部分目标信息。在LR [34]中的简单低秩假设在图像包含复杂结构时可能会无效。RC [7]利用超像素突出目标更多的一致性，但是对于复杂背景的情况，这种方法[9, 10, 7]通常就有些力不从心了。在SM [15]中，如果一个显著性目标中包含有和背景相近的颜色的区域，那么，这些区域将被指定为背景区域。结果，它们就会和背景区域具有相同的显著性值。相反，我们提出的方法可以成功强调显著性区域，并且保持物体的边界，这样，产生的结果就会更接近真实结果。

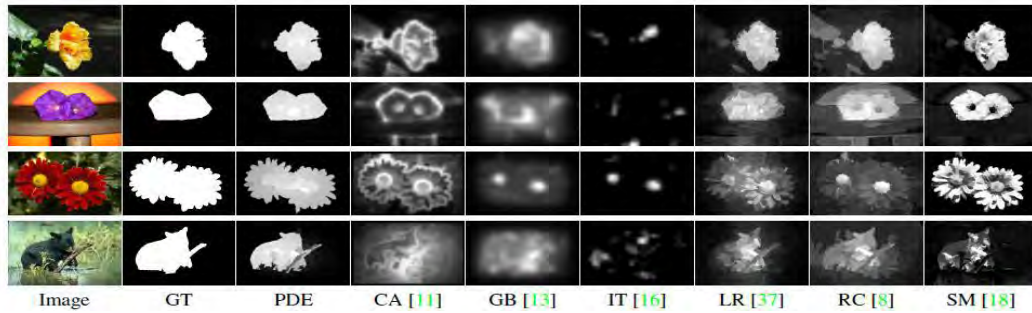


图8不同方法的定性比较。上面三行诗在MSRA中的例子，底部那行是在Berkeley中的例子。

6. 结论

本文开发了一个PDE 系统用于显著性检测。我们定义了一个带有狄利克雷边界的线性椭圆系统 (LESD)，以便模拟图像中显著性的扩散，并证明了该解的子模性。接着，我们求解出一个子模块最大化模型来优化边界条件，使高层先验知识合并到学习的PDE 公式中。在各种具有挑战性的图片集上，我们将我们的PDE 与许多高水准的技术进行了评估测试，结果显示我们的PDE 对于显著性检测具有很大的优越性。今后，我们计划扩展子模PDE 的学习技术，使更为复杂的人类感知和在计算机视觉中其他具有挑战性问题的先验知识合并起来。

致谢

刘日升对Kim和Zhong的讨论交流表示感谢。刘日升是由国家自然科学基金和中国博士后科学基金资助。曹俊杰、林宙辰和山世光是由国家自然科学基金资助。

参考文献

- [1] R. Achanta, F. Estrada, P. Wils, and S. Sussstrunk. Salient region detection and segmentation. In ICVS, 2008.
- [2] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In CVPR, 2009.
- [3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE T. PAMI, 34(11):2274–2282, 2012.
- [4] G. Calinescu, C. Chekuri, M. Pal, and J. Vondrak. Maximizing a monotone submodular function subject to a matroid constraint. SIAM J. Computing, 40(6):1740–1766, 2011.
- [5] T. Chan and J. Shen. Image processing and analysis: variational, PDE, wavelet, and stochastic methods. SIAM, 2005.
- [6] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai. Fusing generic objectness and visual saliency for salient object detection. In ICCV, 2011.
- [7] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global

- contrast based salient region detection. In CVPR, 2011.
- [8] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008.
- [9] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE T. PAMI*, 34(10):1915–1926, 2012.
- [10] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In NIPS, pages 545–552, 2006.
- [11] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, 2007.
- [12] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE T. IP*, 13(10):1304–1318, 2004.
- [13] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE T. PAMI*, 20(11):1254–1259, 1998.
- [14] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. Automatic salient object segmentation based on context and shape prior. In BMVC, 2011.
- [15] Z. Jiang and L. S. Davis. Submodular salient region detection. In CVPR, 2013.
- [16] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In ICCV, 2009.
- [17] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade. Distributed cosegmentation via submodular optimization on anisotropic diffusion. In ICCV, 2011.
- [18] B. C. Ko and J.-Y. Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. *JOSA A*, 23(10):2462–2470, 2006.
- [19] V. Kolmogorov and R. Zabini. What energy functions can be minimized via graph cuts? *IEEE T. PAMI*, 26(2):147–159, 2004.
- [20] A. Krause and D. Golovin. Submodular function maximization. *Tractability: Practical Approaches to Hard Problems*, 3, 2012.
- [21] A. Krause and C. Guestrin. Beyond convexity: Submodularity in machine learning. In *ICML Tutorials*, 2008.
- [22] C. Lang, G. Liu, J. Yu, and S. Yan. Saliency detection by multitask sparsity pursuit. *IEEE T. IP*, 21(3):1327–1338, 2012.
- [23] T. Lindeberg. *Scale-space theory in computer vision*. Springer, 1993.
- [24] R. Liu, Z. Lin, W. Zhang, and Z. Su. Learning PDEs for image restoration via optimal control. In ECCV, 2010.
- [25] R. Liu, Z. Lin, W. Zhang, K. Tang, and Z. Su. Toward designing intelligent PDEs for computer vision: An optimal control approach. *Image and Vision Computing*, 31(1):43–56, 2013.
- [26] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *IEEE T. PAMI*, 33(2):353–367, 2011.
- [27] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, 2003.
- [28] V. Movahedi and J. H. Elder. Design and perceptual validation of performance measures for salient object segmentation. In *CVPR Workshops*, 2010.
- [29] G. L. Nemhauser and L. A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*,

3(3):177–188, 1978.

- [30] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In CVPR, 2012.
- [31] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In SIGGRAPH, 2004.
- [32] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? In CVPR, 2004.
- [33] H. J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of vision*, 9(12), 2009.
- [34] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In CVPR, 2012.
- [35] J. Van De Weijer, T. Gevers, and A. D. Bagdanov. Boosting color saliency in image feature detection. *IEEE T. PAMI*, 28(1):150–156, 2006.
- [36] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In ECCV. 2012.
- [37] J. Weickert. *Anisotropic diffusion in image processing*, volume 1. Teubner Stuttgart, 1998.
- [38] Y. Xie, H. Lu, and M.-H. Yang. Bayesian saliency via low and mid level cues. *IEEE T. IP*, 22(5):1689–1698, 2013.
- [39] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In CVPR, 2013.
- [40] J. Yang and M.-H. Yang. Top-down visual saliency via joint CRF and dictionary learning. In CVPR, 2012.
- [41] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, 2006.