

报告正文 (国家自然科学基金申请)

(一) 主要学术成绩、创新点及其科学意义

按本年度《国家自然科学基金项目指南》中优秀青年科学基金项目的有关要求，着重阐述所取得的研究成果的创新性和科学价值等（不超过 4000 字）。

1. 引言

人类可以很容易地判断图像中的显著性区域，并注意到图像的重要部分。由于我们可以通过显著性区域来优先分配图像分析与合成所需的计算资源，所以通过计算来检测图像的显著性区域意义重大。提取出的显著性图像可以广泛应用于许多计算机视觉领域的应用，包括对兴趣目标物体图像分割 [10, 14]，目标识别 [21]，自适应压缩 [6]，内容感知图像编辑 [24, 28, 26, 8]，和图像检索 [4] 等。

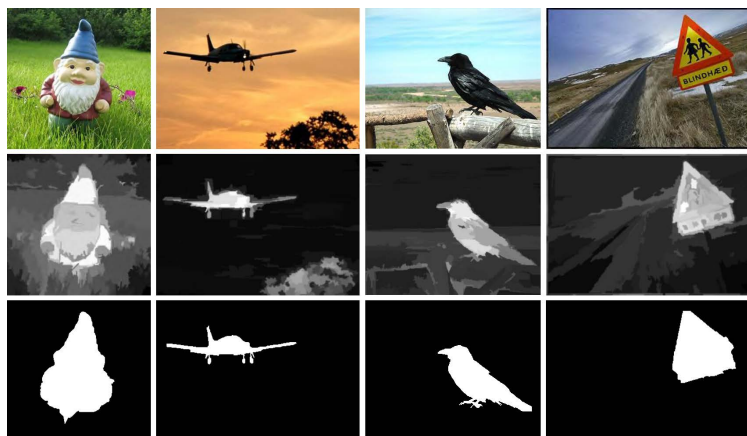


图 1: 给定输入图像 (上), 可以通过全局对比度分析得到高分辨率的视觉显著性图 (中)。这种视觉显著性图可以进一步被用来获取感兴趣物体区域 (下)。

显著性源于视觉的独特性、不可预测性、稀缺性以及奇异性，并且是由颜色、梯度、边缘、边界等图像属性所致。视觉显著性和我们如何感知、处理视觉刺激紧密相关，并且正在被包括认知心理学 [22, 25]，神经生物学 [7, 18] 和计算机视觉 [13, 2] 在内的多个学科进行研究。关于人类视觉注意的理论假设人类视觉系统只详细地处理图像的某个局部，而不是整幅图像。Treisman 和 Gelade [23]，Koch 和 Ullman [15] 的早期工作，以及 Itti, Wolfe 等人的视觉注意理论提议将视觉注意机制分为两个阶段：快速的、下意识的、自底向上的、数据驱动的显著性提取；以及慢速的、任务依赖的、自顶向下的、目标驱动的显著性提取。

我们依据图像的对比度来进行自底向上、数据驱动的显著性检测。人们普遍认为，为了优先响应高对比度刺激，人类的大脑皮质细胞在它们的接受域可能是硬编码的 [19]。基于以下观察，我们提出了提取高分辨率的全局显著性图像的分析方法：

- 基于全局对比度的方法倾向于将大范围的目标和周围环境分离开。这种方法要优于那些通常只在轮廓附近产生较高显著性的局部对比度方法。
- 全局的考虑可以为图像中相似区域分配一个相近的显著性的值，并且可以均匀的突出目标。
- 一个区域的显著性，主要是由它和周围区域的对比度决定，相距很远的区域起的作用较小。
- 为了能够适应大规模图像集处理和高效的图像检索与分类的应用需求，显著图检测算法应该具有简单快速的特点。



图 2: 短题注。

1.1. 我们的方法

我们提出了基于直方图对比度的方法 (HC) 来检测显著性。依据与其它像素的色彩差异来分配像素的显著性值，并用以产生具有全分辨率显著性图像。我们用基于直方图的方法来进行高效处理，并用色彩空间的平滑操作来控制量化的缺陷。我们的算法致力于处理普通的自然景象，对处理具有高纹理的图像场景不一定能达到最佳效果 (见 图??)。

我们在国际上现有最大的公开基准数据集上广泛地测试了我们的结果，并且与现有最先进的一组显著性图像提取方法 [13, 17, 27, 11, 12, 1, 2, 9] 的结果，以及人工标注的参考数据进行了比较¹。我们的方法和之前已有的方法相比，在正确率和召回率方面都有明显的

¹对 1000 幅测试图得到的结果以及软件原型可以从我们的项目主页上获取: <http://mmcheng.net/salobj/>

时间 (秒)	0.611	0.070	1.614	0.064	0.016	0.109	53.1	0.018	0.019	0.253
代码类型	Matlab	C++	Matlab	Matlab	C++	C++	Matlab	C++	C++	C++

表 1: 计算 Achanta 等人数据集 [2] 中图像的平均用时。该数据集（参见我们主页）中大部分的图像分辨率为 400×300 。这里所示的所有方法的时间是在一个拥有 Dual Core 2.6 GHz CPU, 2GB 内存的机器上测得的。

提高。总地来说，和 HC-map 相比，RC-map 的正确率和召回率更高，但是计算量也更高。令人欣慰的是，利用我们的显著性图像分割出来的区域在大多数情况下符合人工的标注。我们还将提取出的显著性图像应用于图像分割，内容感知的图像缩放，以及非真实感渲染等应用。

2. 相关工作

本文主要关注自底向上的显著性检测相关的相关文献。这类方法或是基于生物学原理的，或是纯计算的，或两者兼顾。这些方法利用亮度、颜色、边缘 [2] 等底层特征属性来决定图像某个区域和它周围的对比度。我们宽泛地把这些算法分类为局部的和全局的两大类。

基于局部对比度的方法利用图像区域相对于局部邻域的稀有度。在 Koch 和 Ullman [15] 提出的非常有影响力的生物启发模型基础上, Itti 等人 [13] 定义了图像的显著性。此定义利用了多尺度图像特征的中心-周围的差异来得到。Ma 和 Zhang [17] 提出了另一种局部对比度分析方法来产生显著性图像，并用模糊增长模型对其进行扩展。Harel 等人 [11] 通过将 Itti 等人的特征图归一化来突出显著部分，并且可以和其它显著图像结合。Liu 等人 [16] 通过将高斯图像金字塔的对比度线性结合，提出了多尺度对比度。最近，Goferman 等人 [9] 同时对局部底层线索、全局考虑、视觉组织规则以及表层特征进行建模来突出显著性物体。这些利用局部对比度的方法倾向于在边缘部分产生高显著性值，而不是均匀地突出整个物体 (见图??)。

基于全局对比度的显著性区域计算方法用一个区域和整个图像的对比度来计算显著性值。Zhai 和 Shah [27] 定义了基于某个像素和其余像素点对比度的像素级显著性。Achanta 等人 [2] 提出了频率调谐方法，此方法用某个像素和整个图像的平均色的色差来直接定义显著性值。但它该方法只考虑了一阶平均颜色，不足以分析复杂多变的自然图像。图 ?? 和图 ?? 展示了这种方法定性和定量的缺点。此外，这些方法忽略了图像各部分间的空间关系，而这个因素对可靠的显著性检测来说可能是至关重要的 (见第3节)。

空间加权区域对比度. 更进一步, 通过在式??中引进空间权值, 我们将空间信息加入进来, 来增加区域的空间影响效果。近邻的区域增大影响, 较远的区域减小影响。特别地, 对任意区域 r_k , 基于空间加权区域对比度的显著性定义为:

$$S(r_k) = \sum_{r_k \neq r_i} \exp(-D_s(r_k, r_i)/\sigma_s^2) w(r_i) D_r(r_k, r_i) \quad (1)$$

其中 $D_s(r_k, r_i)$ 为区域 r_k 和 r_i 的空间距离, σ_s 控制空间权值强度。 σ_s 越大, 空间权值的影响越小, 导致较远区域的对比度会对当前区域显著性值做出较大的贡献。两个区域的空间距离定义为两个区域重心的欧氏距离。在我们的试验中, $\sigma_s^2 = 0.4$, 像素坐标归一化到 $[0, 1]$ 区间。

3. 实验比较

我们在 Achanta 等人提供的公开测试集 [2] 上测试了我们的方法。据我们所知, 此测试集是此类数据最大的测试集, 并且由人工精确标注了显著性区域。我们将基于全局对比度的方法与其它 8 个当今最好的方法进行了比较。

在我们的方法中, 我们迭代应用 GrabCut [20] 来改善二值化显著性图像后得到的分割结果 (见图??)。传统的 GrabCut 方法是由人工选中矩形区域来进行初始化操作, 而我们用一个固定阈值二值化后的显著性图来得到显著性分割, 并用这个显著性分割来自动地进行 GrabCut 初始化。这个阈值的我们经验性的选择固定阈值实验中与 95% 召回率对应的阈值。

初始化之后, 我们迭代运行 GrabCut 来改进显著性分割结果 (在我们的实验中最多迭代 4 次)。在每一次迭代后, 我们用膨胀和腐蚀操作来得到新的 Trimap 以进行下一次迭代。如图??所示, 膨胀后仍然落在外面的区域设置成背景, 在腐蚀区域内的设置成前景, 其余的区域为 Trimap 中的未知。Grabcut 本身是用高斯混合模型和 Grapcut 进行迭代, 来改善每一步的区域分割效果, 靠近初始显著性物体区域的部分成为显著性物体的几率更大。因此, 我们的新的初始化方法可以使 GrabCut 包含显著性区域附近的显著性区域, 并根据颜色特征的差异排除非显著性区域。在算法实现中, 我们设置了狭窄的图像边界区域 (15 像素宽) 作为背景来提高边界区域的收敛速度。

图??展示了我们显著性分割算法的两个实例。在旗子的例子中, 不需要的区域被正确地排除; 在花朵的例子中, 我们的方法扩展了初始显著性区域并收敛得到了精确的分割结果。为了保持一致性, 我们用召回率 95% 的阈值对显著性图像进行二值化 (见图??)。我们

在基准数据集 [2] 上比较了平均正确率, 召回率以及 F -测量, 其中 F -测量定义为:

$$F_{\beta} = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 \times Precision + Recall}. \quad (2)$$

基于内容感知的图像缩放. 在内容敏感的图像缩放中, 显著性图像经常用来指定图像的相对重要区域 (见 [3]). 我们用提取出的显著性图像进行了图像缩放实验. 实验中采用了 Zhang 等人提出的 [28] 内容敏感图像缩放方法².

参考文献

- [1] R. Achanta, F. Estrada, P. Wils, and S. Ssstrunk. Salient region detection and segmentation. In ICVS, pages 66–75. Springer, 2008.
- [2] R. Achanta, S. Hemami, F. Estrada, and S. Ssstrunk. Frequency-tuned salient region detection. In CVPR, pages 1597–1604, 2009.
- [3] R. Achanta and S. Susstrunk. Saliency Detection for Content-aware Image Resizing. In ICIP, 2009.
- [4] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. Sketch2photo: Internet image montage. ACM Trans. Graph., 28(5):124:1–10, 2009.
- [5] M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Repfinder: Finding approximately repeated scene elements for image editing. ACM Trans. Graph., 29(4):83:1–8, 2010.
- [6] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 still image coding system: an overview. IEEE Trans. Consumer Elec., 46(4):1103–1127, 2002.
- [7] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. Annual review of neuroscience, 18(1):193–222, 1995.
- [8] M. Ding and R.-F. Tong. Content-aware copying and pasting in images. The Visual Computer, 26:721–729, 2010.
- [9] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In CVPR, 2010.
- [10] J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. IEEE TCSV, 16(1):141–145, 2006.
- [11] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In NIPS, pages 545–552, 2006.
- [12] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, pages 1–8, 2007.

²我们采用作者公开的代码

- [13] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, 20(11):1254–1259, 1998.
- [14] B. Ko and J. Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. *J Opt Soc Am*, 23(10):2462, 2006.
- [15] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurbiology*, 4:219–227, 1985.
- [16] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, T. X., and S. H.Y. Learning to detect a salient object. *IEEE TPAMI*, 33(2):353–367, 2011.
- [17] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, pages 374–381, 2003.
- [18] S. K. Mannan, C. Kennard, and M. Husain. The role of visual salience in directing eye movements in visual object agnosia. *Current biology*, 19(6):247–248, 2009.
- [19] J. Reynolds and R. Desimone. Interacting roles of attention and visual salience in v4. *Neuron*, 37(5):853–863, 2003.
- [20] C. Rother, V. Kolmogorov, and A. Blake. “Grabcut”– Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.
- [21] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? In *CVPR*, pages 37–44, 2004.
- [22] H. Teuber. Physiological psychology. *Annual Review of Psychology*, 6(1):267–296, 1955.
- [23] A. M. Triesman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.
- [24] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.*, 27(5):118:1–8, 2008.
- [25] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, pages 5:1–7, 2004.
- [26] H. Wu, Y.-S. Wang, K.-C. Feng, T.-T. Wong, T.-Y. Lee, and P.-A. Heng. Resizing by symmetry-summarization. *ACM Trans. Graph.*, 29(6):159:1–9, 2010.
- [27] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, pages 815–824, 2006.
- [28] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin. A shape-preserving approach to image resizing. *Comput. Graph. Forum*, 28(7):1897–1906, 2009.

(二) 拟开展的研究工作

着重阐述拟开展的研究工作的科学意义和创新性,技术路线、研究方案等的可行性(不超过 4000 字)。

1. 总结与展望

我们提出了基于全局对比度的显著性计算方法,即基于直方图对比度(HC)和基于空间信息增强的区域对比度(RC)方法。HC方法速度快,并且产生细节精确的结果,RC方法可以产生空间增强的高质量显著性图像,但与此同时具有相对较低的计算效率。我们在国际上现有最大的公开数据集上测试了我们的方法,并与之前已有八种最好的其它方法进行了比较。实验结果表明,我们提出的方法在正确率和召回率上都明显优于其它方法,并且简单而高效。

在未来的工作中,我们计划研究包含空间关系且保留详细细节的显著性图像的高效计算算法,并且希望研究能够处理具有复杂纹理的背景图像的显著性检测算法,以克服我们现有算法在处理这类情况中存在的缺陷。最后,我们还希望显著性图像的检测过程中进一步考虑人脸、对称性等高级因素。我们相信显著性图像可以应用于高效物体检测[10],可靠图像分类,鲁棒的图像景物分析[5],并提高图像检索效果[4]。

致谢. 本项目受到了国家 973 计划(2011CB302205),国家 863 计划(2009AA01Z327),国家自然科学基金(U0735001),以及国家核高基计划(2011ZX01042-001-002)的支持. 程明的工作受到了 Google PhD fellowship, IBM PhD fellowship, 以及教育部博士研究生学术新人奖的资助。

(三) 其他需要说明的问题

1. 申请人同年申请不同类型的国家自然科学基金项目情况(列明同年申请的其他项目的项目类型、项目名称信息,并说明与本项目之间的区别与联系)。

申请人同年参与 XXX 项目,有助于促进本项目的实施。

2. 具有高级专业技术职务(职称)的申请人是否存在同年申请或者参与申请国家自然科学基金项目的单位不一致的情况(如存在上述情况,列明所涉及人员的姓名,申请或参与申请的其他项目的项目类型、项目名称、单位名称、上述人员在该项目中是申请人还是

参与者，并说明单位不一致原因)。

不存在上述情况。

3. 具有高级专业技术职务（职称）的申请人是否存在与正在承担的国家自然科学基金项目的单位不一致的情况（如存在上述情况，列明所涉及人员的姓名，正在承担项目的批准号、项目类型、项目名称、单位名称、起止年月，并说明单位不一致原因）。

不存在上述情况。

4. 其他。