

基于深度学习的显著区域提取方法

梁 晔^{1,2}, 马 楠², 刘宏哲¹

(1. 北京联合大学信息服务工程重点实验室, 北京, 100101; 2. 北京联合大学机器人学院, 北京, 100044)

摘 要: 目前显著区域提取方法通常会设计多个复杂的网络结构, 导致计算和存储代价较高。深度学习网络本身具有多尺度的特点, 不同的卷积层特征具有不同的空间分辨率, 可以避免复杂网络结构的设计。基于此, 本文设计了一种新颖的基于深度学习的显著性检测网络, 既考虑了特征的多尺度特点, 又考虑了图像中显著区域的大小对显著区域检测结果的影响。实验中以流行的基准数据集作为实验对象, 结果证明了本文方法的优越性能。

关键词: 深度学习; 显著区域; 多尺度

中图分类号: TP391.9 **文献标志码:** A

Deep Learning Based Salient Region Detection

LIANG Ye^{1,2}, MA Nan², LIU Hongzhe¹

(1. Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, 100101, China; 2. College of Robotics, Beijing Union University, Beijing, 100044, China)

Abstract: Several complex networks are usually designed in salient region detection to detect saliency, which inevitably leads to very high computational and storage costs. The deep learning network has the characteristics of multi-scale and different convolution layers have different spatial resolutions, thus the design of complex network structure can be avoided. In this paper, a novel convolution neural network is designed by taking advantage of multi-scale characteristics. Both the multi-scale features and the influence of the size of salient regions are considered to saliency detection. Experiments show the superiority of our method on popular benchmark datasets.

Key words: deep learning; salient region; multi-scale

引 言

伴随着社交媒体网络 and 多媒体技术的迅速发展以及智能拍照设备的普及, 涌现的海量图像和视频数据已经成为人们获取信息和交流的主要载体。相对于多媒体大数据, 现有计算资源相对有限。如何利用脑和认知科学知识找到符合视觉系统信息处理机制的信息选择模型就变得尤为重要。基于视觉机制的选择性注意模型能够让计算机检测出图像中最能引起用户兴趣和体现图像内容的显著区域, 可以为诸多应用提供原始对象, 是目前研究的热点。

基金项目: 北京市自然科学基金(4182022)资助项目; 国家自然科学基金(61871038, 61871039)资助项目; 北京联合大学人才强校优选计划领军计划(BPHR2020AZ02)资助项目。

收稿日期: 2019-09-10; **修订日期:** 2019-12-22

随着显著性提取技术的发展,研究人员已提出多种显著区域检测方法,从基于人工设计特征和浅层模型的检测已发展到目前流行的基于深度学习的检测。传统的显著性检测方法通常使用多种显著性线索或先验信息,但由于使用了低级的人工设计特征和模型,不能识别和理解图像中的语义对象。2006年以来,随着大规模硬件加速设备的出现,特别是计算机图像处理器以及图像大数据的涌现,深度学习技术的数据处理能力不断提升,成为计算机视觉和图像处理领域的研究热点。深度学习模拟了人类视觉系统处理外部信息的方式,通过组合底层特征形成更加抽象的高层特征,具有较强的语义性,在显著区域提取任务中应用越来越广泛。基于深度学习的提取方法较传统方法取得了更好的性能,但仍存在一些问题值得继续研究。

为了获取多尺度特征,基于卷积神经网络的显著性检测方法通常需要训练多个深度网络,尽管取得了较好的性能,但是计算和存储量代价巨大,在实际应用中实用性不强。深度卷积神经网络本身就具有多尺度的层次结构,因此不同空间的卷积层特征图本身就可以看做多尺度特征,不需要通过额外的深度网络构建多尺度特征。基于此,本文提出了一种基于多尺度卷积神经网络的显著区域提取方法,构建了一个新的深度学习网络结构。

1 相关工作

显著性检测方法发展大致经历了3个阶段:基于认知科学和神经科学的检测方法;基于对比度计算的检测方法;基于深度学习的检测方法。最早且最经典的检测模型是Koch和Ullman^[1]提出的基于生物学模型的视觉计算模型。之后Itti等^[2]对Koch和Ullman模型进行改进,并行地提取多尺度、多特征的显著图。由于基于认知科学和神经科学的显著性检测方法计算复杂,基于对比度计算的检测方法逐渐成为研究的重点,该方法可以分为纯计算模型和混合模型^[3-10]。文献[8]采用基于图的流型排序方法进行显著性检测,包括两个阶段的前景和背景对比度排序。文献[9]在显著性检测时采用颜色、位置和大小等多种显著性特征,构建了3个层次的模型进行显著性的计算。文献[10]从重构误差的角度,通过稠密和稀疏特征进行显著图预测。基于人工设计特征的浅层模型虽然取得了一定的成效,但在处理有复杂背景的图像时往往效果不佳。

随着计算机图像处理器的发展以及大规模图像库的出现,深度学习方法越来越多地应用于显著区域提取。文献[11]通过无监督方法学习多个中层滤波器集合,对局部显著区域进行提取,并将局部显著区域结果和卷积神经网络得到的中层显著区域提取结果进行融合。文献[12]采用卷积神经网络计算局部区域块、邻域区域块和整幅图像构成的3个尺度的特征,利用多尺度特征进行显著区域的提取。文献[13]在提取过程中采用2个深度卷积网络:一个用于局部显著图的提取,另一个用于全局显著图的提取,然后将两种显著图进行融合。文献[14]将局部区域信息和全局上下文信息进行融合,进而实现显著区域的提取。这些方法多数采用区域块扫描方式,导致大量的计算冗余。针对此问题,基于全卷积神经网络(Fully convolutional networks, FCN)的提取方法在准确性和速度上都有了较大的改进。文献[15]在多任务学习框架下训练FCN网络;文献[16]在循环FCN网络结构中引入显著性先验取得了更准确的推理结果。文献[17]设计了一个2个层次的网络,首先通过一个网络获得粗粒度的检测结果,然后再通过另一个网络进一步对结果进行细节上的处理。以上文献都证明了多尺度特征对于显著性检测的有效性,然而这些方法共同的问题是网络结构相对复杂,这无疑会对计算量和存储上带来很大的负担。

采用不同空间的采样特征图表示多尺度特征的思想在其他计算机视觉任务中已经证明有效且计算和存储代价小,例如对象识别和分割。受文献[18]启发,本文利用网络结构中的多尺度特征进行显著

性检测。但是本文的网络结构和文献[18]有很大的差别:本文方法是对多个尺度的预测分配不同的权重,权重可以通过有监督的学习得到,并且随着输入图像的不同而不同。

2 显著区域提取方法

2.1 深度学习网络结构

本文提出的卷积神经网络结构由4个模块组成:自底向上的特征提取模块、自上而下的特征连接模块、权重预测模块和显著性预测模块,如图1所示。图中:FC表示全连接(Fully connected);maxout表示激活函数,输出该层的最大值;ReLU表示线性修正单元(Rectified linear unit)。

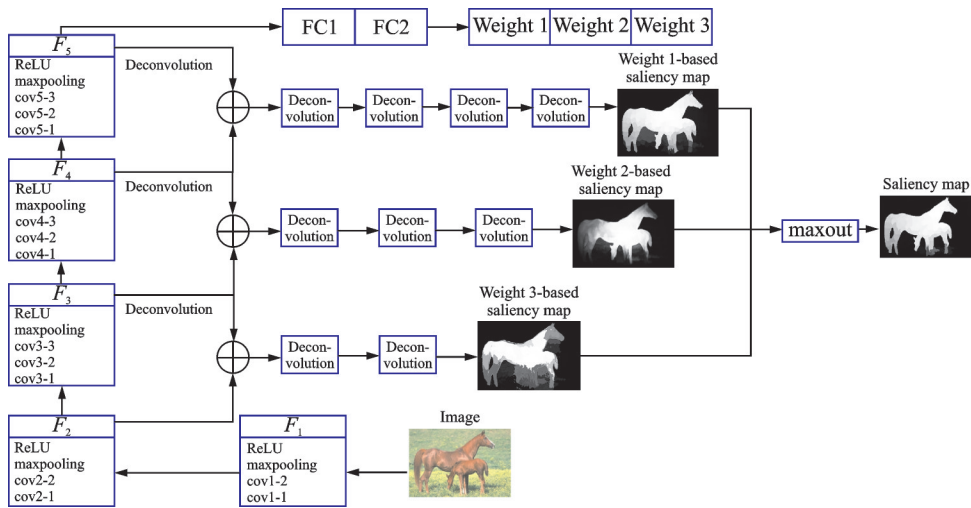


图1 深度学习网络结构

Fig.1 Network structure of deep learning

2.1.1 自底向上的特征提取

将一幅图像输入到网络,特征提取模块进行自底向上进行特征提取形成多尺度特征。采用VGG网络^[19],特征提取模块包括13个卷积层,通过ReLU进行非线性映射和最大池化操作,获得4个尺度的特征,表示为 $\{F_2; F_3; F_4; F_5\}$ 。这样得到4个尺度的特征图,自底向上按着步长为2进行降采样。越靠近下面,分辨率越高,保持较好的细节信息和全局结构,比如 F_2 ;但是特征越低级,包含的语义信息越少。相反,越靠网络顶部,特征越高级,语义性越强,但是也丢失了图像的细节信息。正是由于底层和高层的特征具有各自的优缺点,所以多尺度的特征对于显著性检测都非常重要。

2.1.2 自上而下的特征连接

在网络中,不同层表现出不同的显著性线索,在一定程度上具有互补的特点。为了使多尺度的显著性特征获得更强的表示能力,将底层特征和相邻的高层特征进行连接,表示为

$$\tilde{F}_{k-1} = d_k(F_k) + f_k(F_{k-1}) \quad (1)$$

式中: d_k 为反卷积层操作,包括20个滤波器,卷积核大小为 4×4 ,步长为2; F_k 特征图通过上采样获得和 F_{k-1} 特征图一样的尺寸; $f_k(\cdot)$ 表示进行 1×1 的卷积操作,将 F_{k-1} 作为输入,输出为20个特征通道的特征图。

如图1所示,这个过程开始于网络顶部的特征 F_5 ,迭代直到底部最细分辨率的特征图 F_2 。这样新

的特征 $\{\tilde{F}_4, \tilde{F}_3, \tilde{F}_2\}$ 依次产生,不但包含原始特征的所有信息,并且具有更鲁棒的显著性线索。

2.1.3 显著图预测

有了不同尺度的加强特征 $\{\tilde{F}_4, \tilde{F}_3, \tilde{F}_2\}$,就可以使用这些特征独立地进行显著性预测。3个预测分支对这些特征图进行反卷积操作,分别进行3次、2次和1次。每个卷积核的大小为 4×4 ,步长为2,这样每个分支输出的显著图和输入图像具有相同的尺寸。每个预测分支的中间反卷积层产生20个通道的特征图,通过ReLU进行非线性映射,最后的反卷积层产生1个特征通道的特征图,通过sigmoid归一化到 $[0, 1]$,1表示前景,0代表背景。

由于不同图像的显著区域尺寸不同,因此需要设计最适合的尺度进行显著区域提取。基于此,本文设计的网络能够根据输入的图像预测3个权重,分别分配给3个尺度的显著图。首先,在VGG16的relu5_3的输出与一个全局的平均池化层连接,将 F_5 转化为512维向量。接下来是两个全卷积层,分别有256和3个特征通道,产生3维权重向量 $\hat{\omega} = [\hat{\omega}_2, \hat{\omega}_3, \hat{\omega}_4]$ 。为了防止过拟合,第一个全连接层后面进行dropout操作。输出层通过sigmoid操作将结果归一化到 $[0, 1]$ 。将3个归一化的权重分别赋给3个尺度的显著图,结果为 $\{\tilde{S}_2, \tilde{S}_3, \tilde{S}_4\}$ 。最终的显著图计算如下

$$\hat{S} = f_{\max}(\tilde{S}_2, \tilde{S}_3, \tilde{S}_4) \quad (2)$$

式中 $f_{\max}()$ 为maxout函数,对多尺度特征图的每个位置进行最大池化操作。

2.2 网络训练

假设训练图像集 X 以及对应的基准二值标注集 S ,显著性预测的损失函数为

$$L^S = -S(x, y) \log \hat{S}(x, y) + (1 - S(x, y)) \log (1 - \hat{S}(x, y)) \quad (3)$$

式中 (x, y) 代表输入图像的坐标; $\hat{S}(x, y)$ 为 (x, y) 位置的显著值。

本文采用有监督学习的方式进行权重训练。设置图像的权重向量 $\omega = [\omega_1, \omega_2, \omega_3]$ 。通过Moore-Neighbor追踪方法找到基准二值标注集 S 中所有非连接的显著区域,并将这些区域表示为 $O = \{O^{(i)}\}_{i=1}^N$, N 代表检测到的显著区域的个数。对于每个显著区域 $O^{(i)} \in O$,计算它的面积和整幅图像的比例 $r^{(i)}$ 。计算完毕后,得到 $\tilde{r} = \{r^{(i)}\}_{i=1}^N$,向量 ω^i 的每个分量设置为

$$\omega_k^i = \begin{cases} 1 & \exists r^{(i)} \in \tilde{r}, t_k^{\min} < r^{(i)} \leq t_k^{\max} \\ 0 & \text{其他} \end{cases} \quad (4)$$

式中 t_k^{\min} 和 t_k^{\max} 分别代表显著区域面积占整幅图像比例的最小和最大阈值,阈值的设置如表1所示。

根据显著区域大小的标注结果 ω ,权重预测的损失函数定义为

$$L^\omega = - \sum_{i=1}^N \sum_{k=2}^4 (\omega_k^i (1 - \hat{\omega}_k^i) + (1 - \hat{\omega}_k^i) \log (1 - \hat{\omega}_k^i)) \quad (5)$$

式中 $\hat{\omega}_k^i (k \in \{2, 3, 4\})$ 为预测值。

将 L^S 和 L^ω 整合在一起作为整个深度网络的损失函数进行网络的训练,表示为

$$L = L^S + \alpha L^\omega \quad (6)$$

表1 显著区域面积占整幅图像比例的阈值设定
Table 1 Threshold setting of the proportion of salient region to the whole image

k	4	3	2
t_k^{\min}	0.50	0.25	0
t_k^{\max}	1.00	0.50	0.25

式中 α 参数用于平衡两种损失值。

3 实验验证

3.1 实验设置

本文通过 THUS10K 数据集^[20]进行训练,并通过水平翻转和多尺度变换等数据增强策略来缓解过拟合,扩充后的数据集大小为 15 KB,训练和测试的数量比为 2:3。采用随机梯度下降方法来训练网络,每次参与迭代的样本数量(Batch)为 256;冲量值(Momentum)为 0.9;正则化项的权重为 0.000 5;学习率初始值为 0.01,当损失稳定时学习率以 0.1 的速度下降;对每层的输出采用比率为 0.5 的 drop-out 操作来防止过拟合;训练迭代次数为 80 次。式(6)中的参数 α 经验设置为 0.1。使用 Caffe 框架进行训练和测试,计算机具有 i7-4790 CPU 和 TITAN-X GPU。

实验中采用 4 个流行的数据集 ECSSD^[9], DUT-OMRON^[8], PASCAL-S^[21] 和 HKU-IS^[12]。其中, ECSSD 包含 1 000 幅图像,图像来自于网络,具有复杂的背景;DUT-OMRON 包含 5 168 幅图像,图像从 140 000 幅图像中手工挑选。PASCAL-S 数据集包含 850 幅图像,图像具有杂乱的背景。HKS-IS 包含 4 447 幅图像,图像集具有一定的复杂度。

在量化的性能评价中,采用当前流行的性能评价指标:(1) F -measure 值;(2)AUC 值(ROC 曲线下面的面积);(3)平均绝对误差(Mean absolute error, MAE),反映显著图和基准二值标注之间的误差。

F -measure 值是由查准率和查全率加权计算出来的数值,计算公式为

$$F_{\beta} = \frac{(1 + \beta^2) \cdot p \cdot r}{\beta^2 \cdot p + r} \quad (7)$$

式中: p 表示查准率; r 表示查全率;权重 β^2 代表 F -measure 值侧重于查准率还是查全率,一般取为 0.3,表示查准率的重要性较大;也有文献将 β^2 设置为 0.5,表示查全率和查准率重要性一样。本文实验中 β^2 设置为 0.3。

3.2 方法比较

在比较流行方法时,或者使用作者公布的代码,或者使用作者提供的显著图结果。浅层模型有 4 个:MR^[8], WCtr^[22], MBS^[23]和 BSCA^[24];深层模型有 5 个:MDF^[12], MCDL^[25], SELD^[26], DCL^[27]和 DS^[15]。本文方法简称为 MultiS。

每种方法在数据集 ECSSD, DUT-OMRON, PASCAL-S 和 HKU-IS 中的检测结果和融合后结果如表 2—5 所示。表 2 显示在 ECSSD 数据集上本文融合方法的 AUC 和 F -measure 值最高,MAE 次低。表 3 显示在 DUT-OMRON 数据集上本文融合方法的 AUC 值最高, F -measure 值次高,MAE 次低。表 4 显示在 PASCAL-S 数据集上本文融合方法的 AUC 值和 F -measure 值最高,MAE 最低。表 5 显示在 HKU-IS 数据集上本文融合方法的 AUC 值最高, F -measure 值次高,MAE 次低。整体来说,本文方法的性能最好。

在数据集 PASCAL-S, ECSSD, HKU-IS 和 DUT-OMRON 挑选典型的显著图,结果如图 2—5 所示,每一组图片出现的顺序为原始图片,以及分别用标准二值标注法、本文方法, BSCA, DCL, DS, MBS, MCDL, MDF, MR, SELD 和 WCtr 方法的实验结果。从图中可以看出本文方法的检测结果具有更高的完整性和更清楚的边界。

表 2 不同方法在 ECSSD 数据集上的结果
Table 2 Results on the dataset ECSSD by different methods

方法	<i>F</i> -measure	AUC	MAE
MR	0.637 2	0.834 4	0.236 6
BSCA	0.657 9	0.845 2	0.233 2
WCtr	0.611 9	0.822 5	0.225 7
MBS	0.634 1	0.826 1	0.230 9
SELD	0.769 8	0.865 6	0.156 1
DCL	0.699 1	0.889 8	0.161 5
DS	0.716 5	0.893 6	0.189 2
MDF	0.728 8	0.858 7	0.173 5
MCDL	0.735 2	0.863 1	0.170 2
MultiS	0.773 8	0.908 5	0.160 3

表 4 不同方法在 PASCAL-S 数据集上的结果
Table 4 Results on the dataset PASCAL-S by different methods

方法	<i>F</i> -measure	AUC	MAE
MR	0.406 4	0.727 3	0.246 1
BSCA	0.574 4	0.865 1	0.191 7
WCtr	0.415 0	0.664 6	0.323 1
MBS	0.461 4	0.724 9	0.286 6
SELD	0.574 4	0.738 8	0.266 1
DCL	0.523 1	0.747 6	0.261 2
DS	0.542 3	0.751 7	0.279 1
MDF	0.383 4	0.624 8	0.316 4
MCDL	0.527 3	0.736 6	0.273 9
MultiS	0.721 2	0.900 5	0.146 3

表 3 不同方法在 DUT-OMRON 数据集上的结果
Table 3 Results on the dataset DUT-OMRON by different methods

方法	<i>F</i> -measure	AUC	MAE
MR	0.406 4	0.727 3	0.246 1
BSCA	0.574 4	0.865 1	0.191 7
WCtr	0.588 2	0.871 8	0.143 8
MBS	0.575 7	0.868 4	0.156 7
SELD	0.676 7	0.913 9	0.092 4
DCL	0.629 3	0.921 6	0.097 2
DS	0.674 9	0.940 1	0.120 4
MDF	0.663 2	0.897 5	0.091 6
MCDL	0.679 2	0.914 4	0.089 1
MultiS	0.678 9	0.931 7	0.089 5

表 5 不同方法在 HKU-IS 数据集上的结果
Table 5 Results on the dataset HKU-IS by different methods

方法	<i>F</i> -measure	AUC	MAE
MR	0.644 8	0.867 1	0.187 9
BSCA	0.660 9	0.909 9	0.174 8
WCtr	0.662 1	0.908 1	0.142 4
MBS	0.680 6	0.919 6	0.174 3
SELD	0.799 1	0.958 1	0.073 6
DCL	0.747 5	0.977 3	0.071 9
DS	0.811 3	0.980 8	0.079 6
MDF	0.769 8	0.968 7	0.129 1
MCDL	0.742 9	0.932 9	0.092 9
MultiS	0.800 8	0.987 2	0.073 2



图 2 PASCAL-S 数据集结果举例

Fig.2 Some results on the dataset PASCAL-S



图3 ECSSD数据集结果举例

Fig.3 Some results on the dataset ECSSD

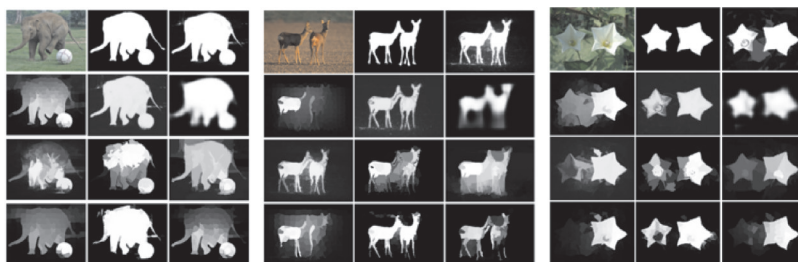


图4 HKU-IS数据集结果举例

Fig.4 Some results on the dataset HKU-IS

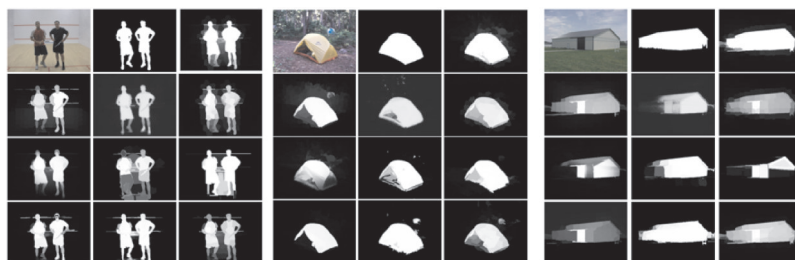


图5 DUT-OMRON数据集结果举例

Fig.5 Some results on the dataset DUT-OMRON

3.3 实验分析

在本文的网络结构中将底层特征和相邻的高层特征进行连接,获得了加强的特征 $\{\tilde{F}_4, \tilde{F}_3, \tilde{F}_2\}$,具有更鲁棒的显著性线索。假如只使用 \tilde{F}_4 或 \tilde{F}_3 或 \tilde{F}_2 去进行显著性检测,则本质是基于单尺度特征的检测方法,性能必定没有本文所提方法的性能好。

一些传统的显著图融合方法多是对多幅显著图进行简单地加和平均或是简单相乘取平均,这些显著图融合方式将各种显著图同等对待,把各种显著性检测的权值设为同一数值,这在实际中是不合理的。因为对于一幅图片甚至是每一个像素点,各种显著性检测方法的检测效果都不同,为此单尺度提取的结果也理应进行不同的对待。由于每幅图像显著区域的大小存在差异,本文方法将显著区域大小的影响纳入到深度网络结构的设计中,能够根据图像中显著区域的尺度自适应地进行融合,性能较好。

4 结束语

为了更好地利用图像的多尺度特征,目前很多基于深度学习的显著区域提取方法会设计各种各样复杂的网络结构,导致了高的计算量和存储代价。针对此问题,本文设计了一种新颖的深度学习显著性检测网络,把不同卷积层的特征看成多尺度特征,避免了复杂的网络结构,并且充分考虑了不同尺寸的显著区域对显著性检测的影响,实验结果表明了本文方法的有效性。

参考文献:

- [1] KOCH C, ULLMAN S. Shifts in selective visual attention: Towards the underlying neural circuitry[J]. *Human Neurobiology*, 1985, 4: 219-227.
- [2] 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.
- [3] WANG Jingdong, JIANG Huaizu, YUAN Zejian, et al. Salient object detection: A discriminative regional feature integration approach[J]. *International Journal of Computer Vision*, 2017, 123(2): 251-268.
- [4] SHEN Xiaohui, WU Ying. A unified approach to salient object detection via low rank matrix recovery[C]//*Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Providence, USA: IEEE, 2012: 853-860.
- [5] CHENG Mingming, WARRELL J, LIN Wen-Yan, et al. Efficient salient region detection with soft image abstraction[C]//*Proceedings of IEEE International Conference on Computer Vision*. Sydney, NSW, Australia: IEEE, 2013: 1529-1536.
- [6] LI Xi, LI Yao, SHEN Chunhua, et al. Contextual hypergraph modelling for salient object detection[C]//*Proceedings of IEEE International Conference on Computer Vision*. Sydney, NSW, Australia: IEEE, 2013: 3328-3335.
- [7] MARGOLIN R, TAL A, ZELNIK-MANOR L. What makes a patch distinct[C]//*Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 1139-1146.
- [8] YANG Chuan, ZHANG Lihe, LU Huchuan, et al. Saliency detection via graph-based manifold ranking[C]//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 3166-3173.
- [9] YAN Qiong, XU Li, SHI Jianping, et al. Hierarchical saliency detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Portland, Oregon, USA: IEEE, 2013: 1155-1162.
- [10] LI Xiaohui, LU Huchuan, ZHANG Lihe, et al. Saliency detection via dense and sparse reconstruction[C]//*Proceedings of IEEE International Conference on Computer Vision*. Sydney, Australia: IEEE, 2013: 2976-2983.
- [11] LIN Yuetan, KONG Shu, WANG Donghui, et al. Saliency detection within a deep convolutional architecture[C]//*Proceedings of the Association for the Advancement of Artificial Workshop*. Québec, Canada: AAAI Press, 2014: 839-848.
- [12] LI Guanbin, YU Yizhou. Visual saliency based on multiscale deep features[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015: 5455-5463.
- [13] WANG Lijun, LU Huchuan, RUAN Xiang, et al. Deep networks for saliency detection via local estimation and global search [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015: 3183-3192.
- [14] ZHAO Rui, OUYANG Wanli, LI Hongsheng, et al. Saliency detection by multi-context deep learning[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015: 1265-1274.
- [15] LI Xi, ZHAO Liming, WEI Lina, et al. Deep saliency: Multi-task deep neural network model for salient object detection[J]. *IEEE Transactions on Image Processing*, 2016, 25(8): 3919-3930.
- [16] WANG Linzhao, WANG Lijun, LU Huchuan, et al. Saliency detection with recurrent fully convolutional networks[C]//*Proceedings of European Conference on Computer Vision*. [S.l.]: Springer, 2016: 825-841.
- [17] LIU Nian, HAN Junwei. DHSNet: Deep hierarchical saliency network for salient object detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 678-686.
- [18] LIN TY, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid networks for object detection[C]//*Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition. Jishou, China: IEEE, 2017: 2117-2125.

- [19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2019-09-08]. <https://arXiv.org/abs/1409.1556>.
- [20] CHENG Mingming, ZHANG Guoxin, MITRA N J, et al. Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569-582.
- [21] LI Yin, HOU Xiaodi, KOCH C, et al. The secrets of salient object segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 280-287.
- [22] ZHU Wangjiang, LIANG Shuang, WEI Yichen, et al. Saliency optimization from robust background detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 2814-2821.
- [23] ZHANG Jianming, SCLAROFF S, LIN Zhe, et al. Minimum barrier salient object detection at 80 FPS[C]//Proceedings of IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1404-1412.
- [24] QIN Yao, LU Huchuan, XU Yiqun, et al. Saliency detection via cellular automata[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 110-119.
- [25] ZHAO Rui, OUYANG Wanli, LI Hongsheng, et al. Saliency detection by multi context deep learning[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 1265-1274.
- [26] LEE G, TAI YW, KIM J. Deep saliency with encoded low level distance map and high level features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 660-668.
- [27] LI Guanbin, YU Yizhou. Deep contrast learning for salient object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 478-487.

作者简介:



梁晔(1978-),女,博士,讲师,研究方向:图像分析与处理,E-mail:liangye@buaa.edu.cn。



马楠(1978-),女,博士,教授,研究方向:图像处理和数据挖掘,E-mail:xxtman-an@buaa.edu.cn。



刘宏哲(1971-),女,博士,教授,研究方向:图像处理和语义计算,E-mail:liuhongzhe@buaa.edu.cn。

(编辑:张黄群)