

融合无监督特征提取层的深度卷积显著性算法

李 想¹⁾, 林 晓¹⁾, 郑晓妹¹⁾, 黄继风¹⁾, 马利庄²⁾

¹⁾ (上海师范大学信息与机电工程学院 上海 200234)

²⁾ (上海交通大学电子信息与电气工程学院 上海 200240)
(lin6008@126.com)

摘 要: 深度神经网络的发展加速了视觉显著性领域的研究。目前, 最有效的基于深度神经网络的显著性检测方法大多数使用端到端训练的深度卷积神经网络。本文采用全卷积神经网络来解决显著性检测问题。首先, 本文提出了一个无监督特征提取层, 来将输入图像预先进行特征提取并将其映射到校直的特征图以作为全卷积网络的输入。其目的是对数据进行降维来减少计算量。接着, 本网络又加入全卷积网络进行端到端学习显著图, 以增强其无监督特征的学习能力。最后, 本文还在训练过程中采用三个损失函数来保证算法稳定收敛和增强算法的输出效果。本文的显著性检测算法在四个公开数据集上进行测试, 与八种先进显著性检测技术在 PR 曲线, F_{γ} -measure, MAE, 运行时间和视觉效果上进行比较。实验表明, 本算法具有较好的检测准确度和较快的运行时间。

关键词: 卷积网络; 无监督特征提取; 显著性检测

中图法分类号: TP391.41 DOI: 10.3724/SP.J.1089.2019.17376

Deep Convolutional Saliency Algorithm Fusing with Unsupervised Feature Extraction Layer

Lin Xiao¹⁾, Li Xiang¹⁾, Zheng Xiaomei¹⁾, Huang Jifeng¹⁾, and Ma Lizhuang²⁾

¹⁾ (The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234)

²⁾ (College of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240)

Abstract: The development of deep neural networks accelerates the study of visual saliency. At present, the most effective deep neural network-based saliency detection methods mostly use end-to-end training deep convolution neural networks. This paper uses a fully convolutional neural network to solve the problem of saliency detection. Firstly, an unsupervised feature extraction layer is proposed to extract the feature vectors on the input image in advance and map it to the calibrated feature map as an input to the supervised layers. The purpose is to reduce the data. Then, the network joins a fully convolutional network to learn the saliency map end-to-end to enhance its learning ability for unsupervised features. Finally, this paper also uses three loss functions in the training process to ensure the stable convergence of the model and enhance the output of the model. The algorithm is evaluated by a large number of experiments. The saliency detection algorithm of this paper is tested on four opened data sets. This paper compares the PR curve, F_{γ} -measure, MAE, run time, and visual effects with eight

收稿日期: 2018-06-29; 修回日期: 2018-11-05. 基金项目: 国家自然科学基金(61872242, 61502220, 61775139). 林 晓(1978—), 女, 博士, 副教授, CCF 会员, 主要研究方向为图像视频处理; 李想(1995—), 男, 硕士研究生, 主要研究方向为图像处理; 郑晓妹(1973—), 女, 博士, 讲师, 主要研究方向为图像视频处理、智能算法; 黄继风(1963—), 男, 博士, 教授, 硕士生导师, 主要研究方向为图像视频处理; 马利庄(1963—), 男, 博士, 教授, 博士生导师, 主要研究方向为图形图像处理、多媒体与虚拟显示。

advanced saliency detection techniques. Experiments show that the algorithm has better detection accuracy and faster running time.

Key words: convolutional neural network; unsupervised feature extraction; saliency detection

显著性检测是计算机视觉、神经科学和视觉感知中的基本问题。在许多研究中,如立体匹配^[1]、视频去模糊^[2]、图像和视频压缩^{[3][4]}、内容感知图像编辑^[5]、目标识别^[6]、语义分割^[7]或视觉跟踪^[8],都使用显著性检测作为预处理过程。随着图像与视频数据在互联网上的快速增长以及对这些媒体实现有效地处理显著目标的需求的不断增加,提出一种准确且高效的显著性检测算法将成为一项很有价值的工作。

到目前为止,显著性检测的方法大致可分为三类。第一类是自下而上的方法^{[10][11]},该分支广泛使用底层线索来计算像素或区域的显著性值。通常,这类方法所使用的线索有:颜色,中心对比^[9],前景先验^[12]和背景先验^[13]。主观上而言,这些特征更能被人类直接感知。然而,并非所有的物体都能用这些底层特征很好的区分。当显著目标的特征不够明显而含有更多语义信息时,检测该目标更需要利用数据的高层信息。第二种是自上而下的方法。此类研究是使用监督方法来学习提取的特征。如, Kim 等人^[14]使用随机森林来学习超像素特征,然后来计算显著图,或者建立有参学习的概率模型^[15]来做显著性检测。在某种程度上,这些方法扩展了显著性检测的泛化能力。然而,它也受到了提取特征的独特性和学习模型的精确度的限制。最后一类是近年来发展的基于深度卷积神经网络(Deep Convolutional Neural Network, DCNN)的显著性检测方法。在这一类研究中,早期是利用深度卷积网络来识别图像局部信息或者是全局加局部信息^{[17][18]}。较新的研究则是以提出一个全卷积的深度神经网络来进行端到端的训练^[16]。基于深度卷积神经网络的方法在深度特征提取和预测方面表现优异并且在显著性检测中取得了非常高的准确性。受这类研究启发,本文设计一个卷积的深度神经网络来解决视觉显著性问题。

与之前的工作不同,本文提出了一个无监督特征图提取层(Unsupervised Feature Extraction Layer, UFEL)用来对图像降维。接着将计算的无监督特征送入一个小尺寸的全卷积网络进行训练,这样可以大大减少计算量而不损失图像信息。训练时,本文在输出端建立多个损失函数,每一个损失函数对应一个尺度的真值图。这样可以让网络充分的学习全局信息,增强算法的鲁棒性。

1 相关工作

视觉显著性的发展经历了两个阶段,第一个阶段是传统显著性检测方法的研究^[20]。在这一时期,显著性检测的方法可分为两大类:自下而上的方法和自上而下的方法^[19]。自下而上的方法是一种面向数据的方法,它通常是根据提取的底层信息建立数学模型来计算显著图。在文献^[21]中,显著值的计算是通过计算超像素的唯一性和空间方差,然后将两个结果获得的显著图结合得到最终的显著图。而另一种方法是主张逐步优化具有特定能量模型的显著图。在文献^[22]中,显著图是利用启发式方法来求得。在图像上建立图模型,然后利用邻接矩阵作为马尔可夫变换矩阵来迭代更新显著图。这类方法的特点是运行高效但不够精确。

而自上而下的方法则是一种面向任务的方法,通常使用监督方法来统计数据的全局特征。例如,文献^[23]利用随机森林来学习手工提取的判别特征并使用学习完成的模型来预测显著图。杨等人^[24]利用字典学习算法并结合条件随机场来优化结果。与自下而上的方法相比,该类方法有更好的鲁棒性。

得益于深度神经网络在计算机视觉邻域的发展,基于深度卷积网络的显著性检测成为最新的研究主流。最近,基于深度卷积神经网络的显著性

检测方法主要有两类方法. 一类方法是将其训练为提取特征器, 然后再对提取出的特征进行分类或是回归来计算其显著性. 李等人^[25]利用 VGG 网络在三个尺度上提取深度特征, 并将其融合在一起进行训练. 另一类方法则是以完全卷积方式端到端地学习检测显著目标. 刘等人^[26]则是训练一个含有多尺度多目标的卷积网络来分割显著目标. 通常, 基于深度卷积的算法在精度上优于传统方法. 然而, 其复杂的结构要经历长时间的训练才能收敛. 而本文则是在深度卷积网络的基础上提出了一个精简的深度网络算法. 其不但运行效率高而且精确也超过了大多数最先进的方法.

2 本文算法

本文提出用一个端到端的深度卷积网络来训练计算显著图. 本文的网络结构如图 1 所示. 图像

被送入网络, 大致经历三个处理过程. 首先图像进入图像无监督特征图提取层, 在该层中, 本文以无监督的方式先将图像分解为大约 256 个超像素并提取对应的设计好 84 维底层特征向量, 然后以线性插值的方式将这些特征向量投影到一个 $(1 \times 84 \times 24 \times 24)$ 的张量(或特征图). 这样以来, 图像经历了第一层网络, 其尺寸降低了超过 100 倍. 对比只降低 2 倍的 VGG 网络, 本文提出的网络大大节省计算时间. 接着, 在得到第一层的结果后, 将其送入一个有监督训练得到的深度卷积网络. 该网络有 8 个模块, 每个模块由 3 层卷积网络组成. 最后三个模块分别设有损失函数, 其目的是使其在训练时能够得到很好的收敛. 经过卷积网络操作后, 再将得到的结果送入组合层. 在该层中, 首先对之前得到的 (24×24) 特征图计算出 (24×24) 的显著图, 再利用反向插值将显著图投影到原图大小的最终效果图.

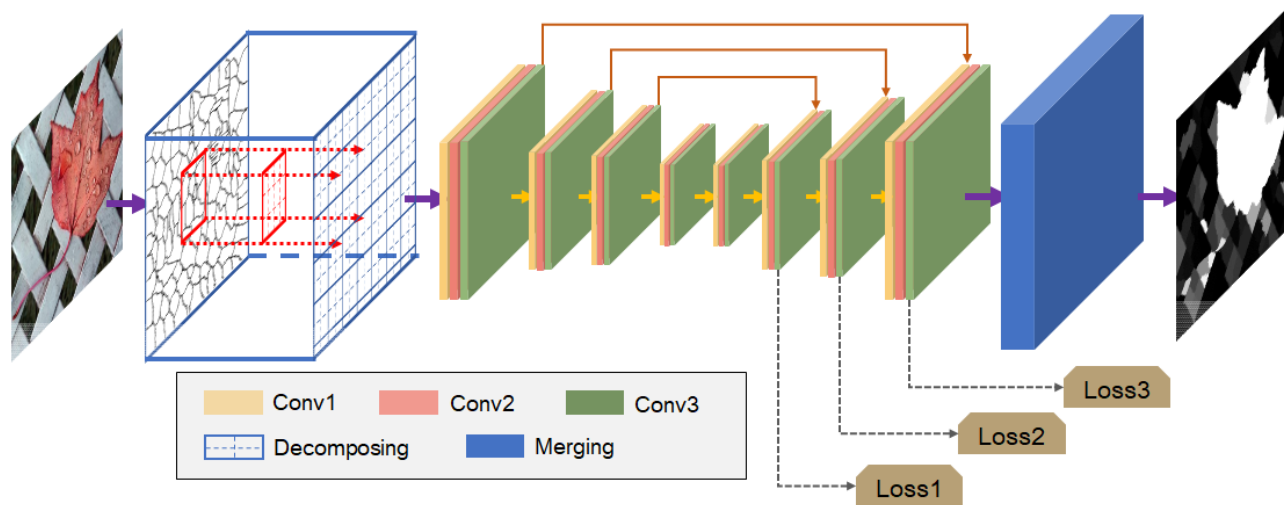


图 1 本文算法流程图

2.1 无监督特征提取层分解图像

本文网络的第一个阶段是以无监督的方式将一张图像映射到一个 $(1 \times 84 \times 24 \times 24)$ 的特征图. 该阶段分为两个过程, 一是特征提取, 二是特征图校直.

2.1.1 特征提取

无监督特征提取层首先对进入的图像进行超像素分割 SLIC^[23] 成为 R 个同质的区域. 然后从中提取 84 维的特征向量. 其特征向量如表 1 所示,

在每个区域的特征向量包括, 颜色特征, 纹理特征, 空间特征, 和全局特征四个基本的底层特征. 颜色特征是指每个区域的平均颜色值, 主要有 RGB、HSV、Lab 三个颜色空间的平均值来增强特征向量的独特性. 而纹理特征在该层中则是这三种颜色空间对映的颜色直方图. 加入纹理特征是为了弥补仅仅考虑平均颜色值所产生的信息缺失而达到增强区域内部信息的效果. 空间特征指的是超像素的空间属性, 其是用区域的横纵坐标值来表示,

表 1 特征向量

特征	属性	维度	符号
颜色特征	RGB	3	$v_1 \sim v_3$
	HSV	3	$v_4 \sim v_6$
	Lab	3	$v_7 \sim v_9$
纹理特征	rgb 直方图	24	$v_{10} \sim v_{33}$
	hsv 直方图	24	$v_{34} \sim v_{57}$
	lab 直方图	24	$v_{58} \sim v_{81}$
空间特征	区域坐标	2	$v_{82} \sim v_{83}$
全局特征	颜色对比度	1	v_{84}

并且大量实验证明空间信息非常有助于计算显著值。最后是全局特征,其是为了进一步增强特征向量的独特性而提出特征。受众多自底向上的工作启发,本文使用全局颜色对比度来作为最后一维属性。

本文将在实验部分验证,本文所提取出的特征向量计算开销小且非常具有代表性。

2.1.2 特征图校直

由于超像素分割出的图像区域不是规则排列的,为了方便下面的卷积操作,该层的第二步是对提取出的特征向量进行校直,将其投射到规则的(24x24)的阵列上。如图 2 所示,本文校直是采用线性插值方法。具体来说,对校直图上的任意一点 p 其所对应的映射特征向量 v_p 为:

$$v_p = \sum_{r \in N_p} w_{pr} \cdot v_r \quad (1)$$

其中, N_p 代表带 p 的相邻图像区域组成的集合, w_{pr} 为特征向量投影的比例系数,它与点 p 到点 r 的空间距离成正比,并且 $\sum_{r \in N_p} w_{pr} = 1$ 。

总的来说,本文的无监督特征提取层既高效的实现有分辨能力的特征描述子的提取又大大缩小了计算目标的维度,减少了全卷积网络的计算量。并且,该层接收任何尺寸的图像输入,而不必在图像输入前调整其大小。这样可以让网络训练过程当中更充分的学习各个尺寸的数据而不必另加入多尺度结构来完成此目的。

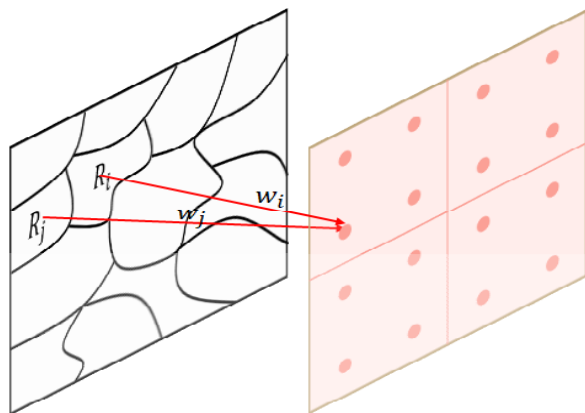


图 2 特征图校直

2.2 有监督的深度卷积网络

2.2.1 卷积网络结构

本文网络的第二个部分为有监督的深度卷积网络层。其任务是将上一层处理得到的特征图进行有监督的训练来实现输出一个对应校直图的显著图结果。如图 1 所示,该网络为全卷积网络,有 8 个卷积子模块,每个模块有三层卷积层。其中前 4 个卷积子模块为下采样模块,输入特征图每经过一层,尺寸都会减小。其目的是为了让网络提取到更高维的特征。然后是上采样模块亦称为反卷积模块,输入特征图每经过该模块,其尺寸都会增大。当特征图增大到与原输入一样大时,其结果就成了对应校直图的显著图输出。并且为了减少在反卷积过程中损失很多深度特征,本文还加入了类似于残差网络的跳跃连接,即每个上采样模块的第一个卷积层除了接受它上一模块的输出以外,还连接的与其大小对应的下采样模块的最后一层卷积层的输出。这样既可保证学习到充分的特征又能逐渐将特征图还原到输入图大小。同时,本网络还抛弃了池化层来更好的保留网络层信息。最后本文大多数模块采用卷积核大小以(1x1)与(3x3)交替排列的方式组合来更好的提炼深度特征信息。

表 2 卷积网层结构

网络层	输入源	输入大小	卷积核	步长	输出大小	通道数	激活函数
Conv_1_1	UFEL	24x24	3x3	1	24x24	128	LReLU
Conv_1_2	Conv_1_1	24x24	1x1	1	24x24	128	LReLU
Conv_1_3	Conv_1_2	24x24	3x3	2	12x12	128	LReLU
Conv_2_1	Conv_1_3	12x12	5x5	1	12x12	256	LReLU
Conv_2_2	Conv_2_1	12x12	1x1	1	12x12	256	LReLU
Conv_2_3	Conv_2_2	12x12	5x5	1	12x12	256	LReLU
Conv_3_1	Conv_2_3	12x12	3x3	1	12x12	512	LReLU
Conv_3_2	Conv_3_1	12x12	3x3	1	12x12	512	LReLU
Conv_3_3	Conv_3_2	12x12	3x3	2	6x6	512	LReLU
Conv_4_1	Conv_3_3	6x6	3x3	1	6x6	1024	LReLU
Conv_4_2	Conv_4_1	6x6	1x1	1	6x6	1024	LReLU
Conv_4_3	Conv_4_2	6x6	3x3	1	6x6	1024	LReLU
Conv_5_1	Conv_4_3	6x6	2x2	1	6x6	1024	LReLU
Conv_5_2	Conv_5_1	6x6	1x1	1	6x6	1024	LReLU
Conv_5_3	Conv_5_2	6x6	3x3	2	12x12	1024	LReLU
Conv_6_1	Conv_5_3, Conv_3_3	12x12	3x3	1	12x12	512	LReLU
Conv_6_2	Conv_6_1	12x12	1x1	1	12x12	512	LReLU
Conv_6_3	Conv_6_2	12x12	3x3	2	12x12	512	LReLU
Conv_7_1	Conv_6_3, Conv_2_3	12x12	3x3	1	12x12	256	LReLU
Conv_7_2	Conv_7_1	12x12	1x1	1	12x12	256	LReLU
Conv_7_3	Conv_7_2	12x12	3x3	2	24x24	256	LReLU
Conv_8_1	Conv_7_3, Conv_1_3	24x24	3x3	1	24x24	128	LReLU
Conv_8_2	Conv_8_1	24x24	3x3	1	24x24	64	LReLU
Conv_8_3	Conv_8_2	24x24	1x1	1	24x24	1	Softmax

本文卷积网络的详细结构如表 2 所示. 其中输入源为该层网络所接受的特征图的输出源, 激活函数 LReLU 为:

$$Y_i = \begin{cases} 0.1 \cdot X_i, & X_i < 0 \\ X_i, & X_i \geq 0 \end{cases} \quad i = 1, \dots, N \quad (2)$$

其中 N 代表输入特征图的大小, X_i 代表输入特征图的第 i 个值, 而 Y_i 为其对应的输出. 另一方面, 激活函数 Softmax 可解为:

$$Y_i = \frac{e^{X_i}}{\sum_{j=1}^N e^{X_j}} \quad (3)$$

其变量符号与公式(2)类似.

通过将本文提出有监督的深度卷积网络结构与前面所述的无监督特征提取层融合, 本文可完成一个接受并输出相同尺寸图片的快速特征提取的深度卷积网络. 同时本网络在多层之间还加入了跳跃连接用来增强网络的学习能力. 更关键的是本卷积网络是一种端到端的学习算法, 其训

练过程中不需占用额外的存储空间.

2.2.2 多级损失函数

本文在训练时采用多级损失函数来进行训练. 如图 1 所示, 本文分别在模块 6, 7 和 8 的最后一个卷积层之后设有一个损失函数. 在此之前, 本文假设 I 表示输入图像, G 表示真值图, $X^{(l)}, l = 1, 2, \dots, 8$. 为第 l 模块的输入张量. 而变量 $Z^{(l)}, l = 1, 2, \dots, 8$. 为与之对应的输出张量. 在训练中, 本文期望在模块 6, 7 和 8 的输出都可以还原与真值图接近的显著图. 因此第一个损失函数可写为:

$$Loss_1 = \sum_{k=1}^{C^{(6)}} \sum_{i=1}^{W^{(6)}} \sum_{j=1}^{H^{(6)}} \|Z_{i,j,k}^{(6)} - \gamma_1 \cdot \sigma^{(6)}(G)_{i,j}\|_2 \quad (4)$$

在这里 $W^{(6)}, H^{(6)}$, 以及 $C^{(6)}$ 分别代表模块 6 的输出张量宽度, 高度和通道, γ_1 在这里被设为 0.001, $\sigma^{(6)}(G)$ 是将真值图与图像 I 以同样的方式校直后缩

放到 $W^{(6)} \times H^{(6)}$ 的结果, 最后 L2 范数 $\| \cdot \|_2$ 在这里用做距离度量来计算预测结果与真值图之间的差别. $Loss_2$ 函数与 $Loss_1$ 类似, 仅仅是参数 γ_2 被设为 0.02. $Loss_3$ 与前两者不同, 由于其输出层的激活函数为 Softmax 函数. 因此 $Z^{(8)}$ 表示的是一个概率图. 因此对于求 $Loss_3$ 本文使用交叉熵作为度量而不是之前的 L2 范数. 该方程如下:

$$Loss_3 = \sum_{k=1}^{C^{(8)}} \sum_{i=1}^{W^{(8)}} \sum_{j=1}^{H^{(8)}} -\sigma^{(8)}(G)_{i,j} \log(Z_{ijk}^{(8)}) \quad (5)$$

最后本算法将此三个损失函数融合训练, 其最终损失函数为:

$$Loss = \sum_{i=1}^3 t_i Loss_i \quad (6)$$

其中 t_i 为要学习的权重并且 $\sum_{i=1}^3 t_i = 1$.

本算法采用多级损失函数来优化卷积神经网络, 这样可以保证训练的网络模型具有较强的泛化能力以及更好的鲁棒性

2.3 组合层

本算法的最后一个层为组合层, 其作用是把卷积网络的输出根据校正层所使用的投影参数 w_{pr} 进行反操作再投影到层与真值图一样尺寸的最终显著图. 该层与无监督特征提取层一样在网络训练中不需要进行训练. 最终结果如图 3 所示:

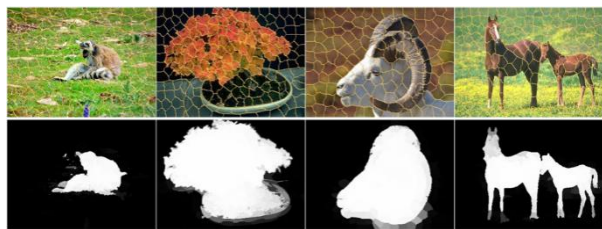


图 3 算法输出结果

从这些样例可以看出本算法可以准确的计算出显著性结果.

3 实验结果

本文的网络模型是由 caffe 框架进行开发的 (网络权值以 caffe 默认的方式初始化), 在 Intel(R) Core i7 3.4 GHz CPU 16 GB RAM Titan X 的设备

上进行网络的训练和测试. 本算法在 MSRA-B^[28]数据集上进行训练, 其含有 5000 张简单目标或背景的图片. 在训练中本网络以梯度下降法训练 80 轮, 每轮批输入 128 张图片. 学习率初始为 $1e-5$, 然后每 10 轮衰减 90%. 动量设为 0.9. 本算法分别与目前的先进算法 DRFI^[32], HDCT^[14], LPS^[33], SMD^[34], wOpt^[13], AMC^[35], LEGS^[17], MDF^[18] 进行比较. 其中, AMC, LPS, SMD 和 wOpt 是传统的自下而上的显著性算法, 而 DRFI 和 HDCT 是结合有监督的自上而下算法, 而 LEGS 和 MDF 是涉及 DCNN 的算法, 即利用卷积神经网络来计算显著图. 本文与该 8 类算法本文分别四个数据集上进行 P-R 曲线性能, F_γ -measure, 和 MAE 等指标的比较. 这些数据集分别为 ECSSD^[27], PASCAL-S^[29], HKU-IS^[30], 和 SOD^[31]. 其中 ECSSD 数据集包含 1000 幅图像样本; PASCAL-S 包含 850 副图像样本; HKU-IS 含有 4446 幅样本; 而 SOD 只有 300 幅图像样本, 却是最具挑战性的数据集. 最后是本算法的训练数据集 MSRA-B, 其含有 5000 张图片, 相对前四个数据集而言, 其目标或背景比较简单.

3.1 评价标准

本文的评价标准采用精确率-召回率曲线 (precision-recall curve, PR 曲线), 准确率-召回率- F_γ -measure 柱状图和平均绝对误差 (mean absolute error, MAE) 来评价显著性检测性能.

3.1.1 PR 曲线

PR 曲线是以精确度作为纵坐标, 召回率作为横坐标所绘制的曲线图. 其中精确度和召回率的计算如下

$$P = \frac{\sum_{i=1}^L \sum_{j=1}^H G(i, j) \cdot S_{bseg}(i, j)}{\sum_{i=1}^L \sum_{j=1}^H S_{bseg}(i, j)}, \quad (7)$$

$$R = \frac{\sum_{i=1}^L \sum_{j=1}^H G(i, j) \cdot S_{bseg}(i, j)}{\sum_{i=1}^L \sum_{j=1}^H G(i, j)}. \quad (8)$$

其中, $G(i, j)$ 和 $S_{\text{bseg}}(i, j)$ 分别是 (i, j) 处真值图和显著图二值分割图对应的值; L 和 H 分别代表图像的宽和高。

3.1.2 F_γ -measure

F_γ -measure 是对精确度和召回率指标的综合, 计算公式为

$$F = \frac{(1 + \gamma^2) \times P \times R}{\gamma^2 \times R + R} \quad (9)$$

其中 $\gamma^2=0.3$ 用来强调准确率。一般地, 该值越大,

则显著性检测算法效果越好。

3.1.3 MAE 值

MAE 是一个更加平衡的评价标准, 其公式为

$$\text{MAE} = \frac{1}{L \times H} \sum_{i=1}^L \sum_{j=1}^H |S(p_{ij}) - G(p_{ij})|. \quad (10)$$

其中 p_{ij} 表示位置 (i, j) 处的像素, $S(p_{ij})$ 和 $G(p_{ij})$ 分别表示在像素 p_{ij} 上的值。MAE 值能反映显著图与真值图之间的差异程度, 其值越小, 说明算法效果优异。

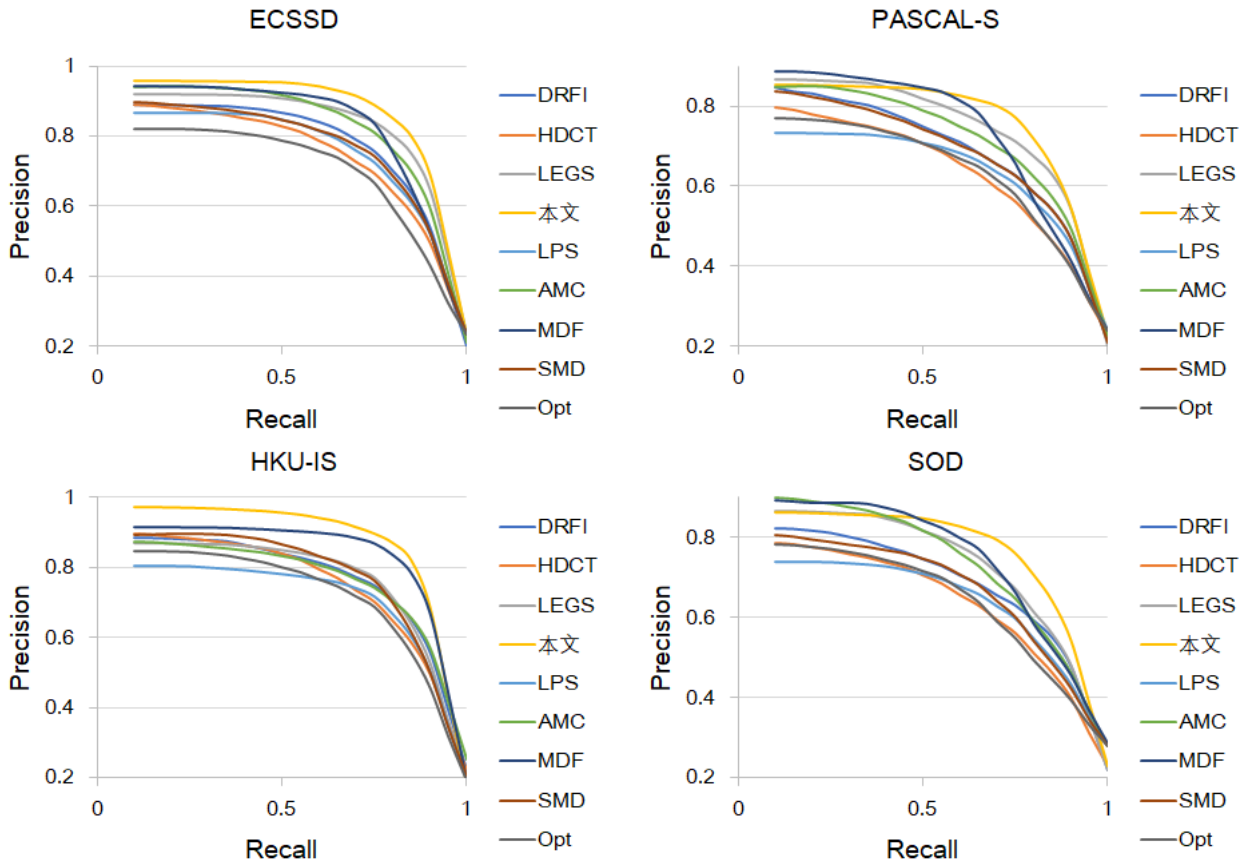


图4 本文与先进算法的 PR 曲线比较

3.2 实验结果比较

本文通过大量的实验来验证本文算法的效果。首先本文的 PR 曲线的对比如图 4 所示, 其中黄色线代表本算法。由这四幅图可看出本算法在四个数据集上远远高于其他算法, 说明本算法的精确度和召回率都非常高并且远远超于其他八种算法。然后是准确率-召回率- F_γ -measure 柱状图的比较

如图 6 所示。该图中的准确率, 召回率 F_γ -measure 是对显著图进行自适应阈值化所到的结果。由此图可以看出, 虽然个别算法如 AMC, MDF 在单一的精确度或召回率的对比下, 有些情况, 其略高于本算法。但若比较精确度和召回率的综合指标可发现, 本算法依然是超过了所有算法。由此可得出本算法不光在综合指标上表现优越, 其在单一指

标上也超过了大部分算法. 最后图 5 展示各算法 MAE 值在四个数据集上的性能比较.

本算法除了和 MDF 在 ECSSD 数据集上的 MAE 表现相近. 对于其他七种算法本算法的 MAE 值都遥遥领先. 同时本文算法在 PASCAL-S, HKU-IS 和 SOD 数据集上的 MAE 值也远远低于 MDF. 图 7 是不同算法的视觉效果比较. 图中所选的样本有简单目标, 有多目标样本, 也有于背景纹理相似的样本. 本算法都很好的将目标检测出来. 与其他算法效果相比更能说明本算法的优越性. 最后本文比较了本算法与其他几种算法在 100 张 400x300 的图片上的平均运行时间, 如表 3 所示. 对于一张 400x300 的图像来说, 本算法的平均运行时间为 0.06 秒, 其效率远远高于其他算法. 由此, 本文提出了一种即高效又精确的视觉显著性算法.

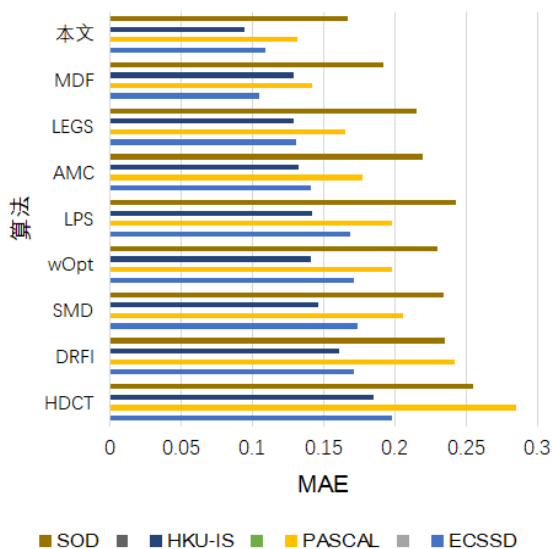


图 5 本文与先进算法的 MAE 比较

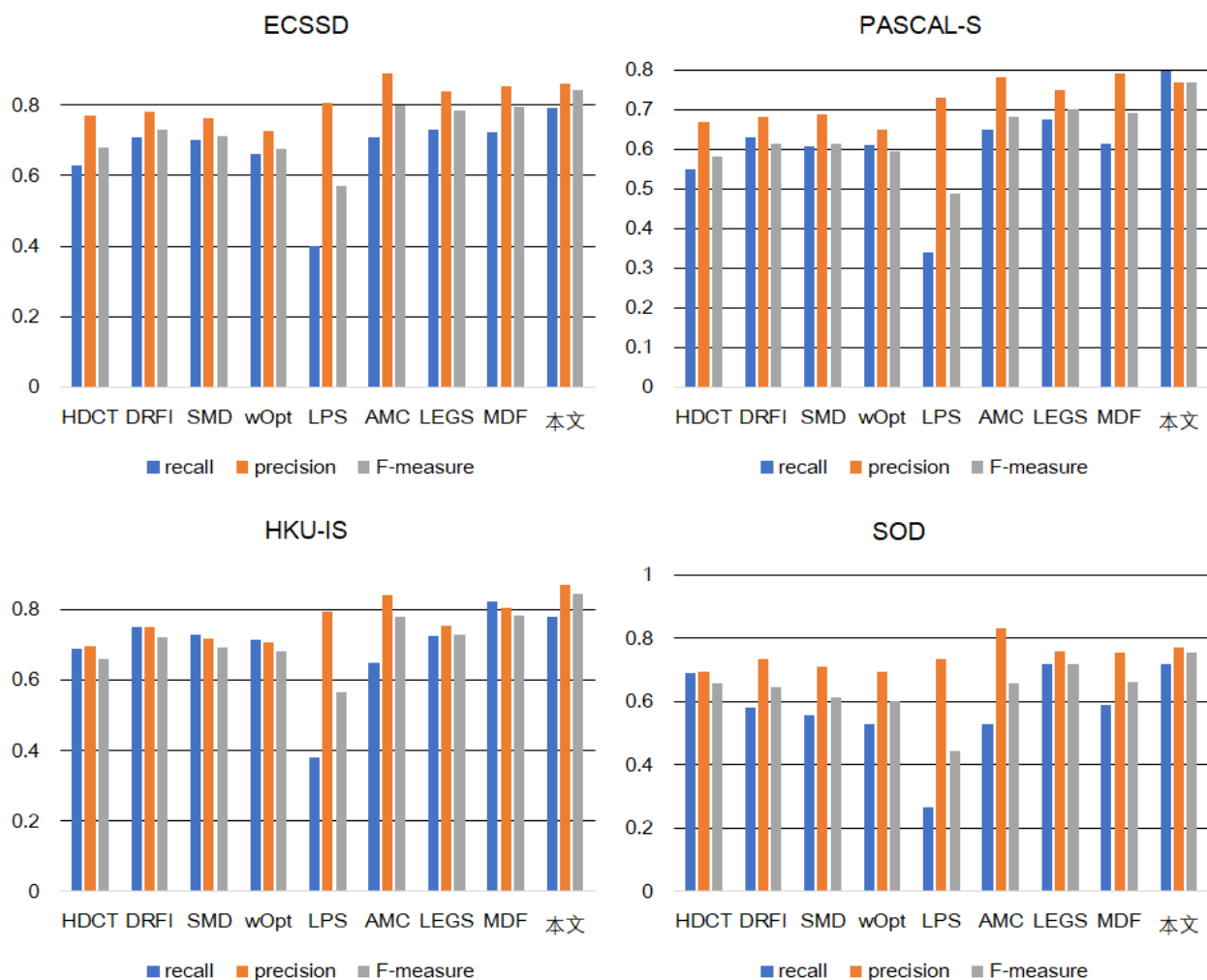


图 6 准确率-召回率-F₁-measure 柱状图的比较

表 3 不同算法的运行效率比较

算法	本文	MDF	LEGS	AMC	LPS	wOpt	SMD	DRFI	HDCT
时间(s)	0.06	7.9	1.85	0.52	0.9	0.23	1.33	5.48	2.18
运行环境	C++(caffe)	Matlab(caffe)	Matlab(caffe)	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab
运行设备	GPU	GPU	GPU	CPU	CPU	CPU	CPU	CPU	CPU

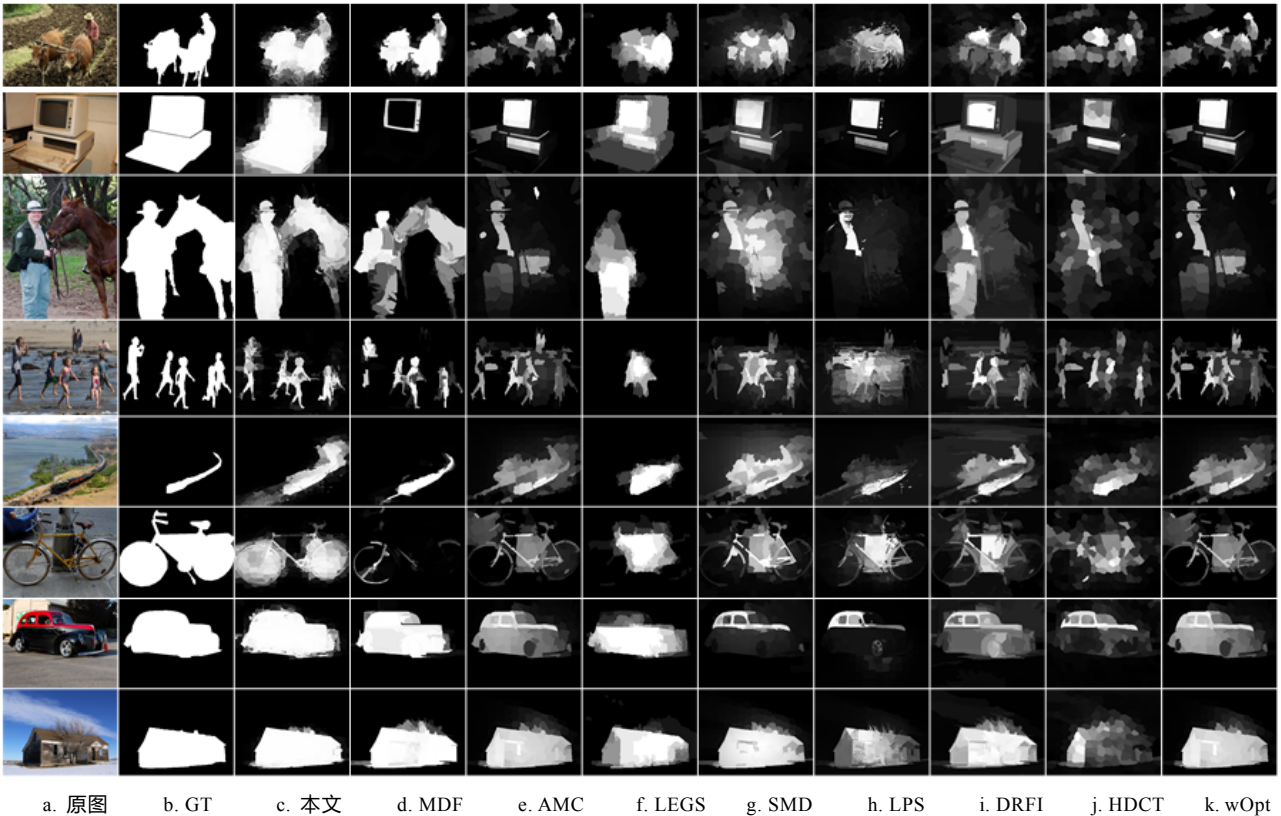


图 7 不同算法显著图的视觉效果对比

3.3 失败样例分析

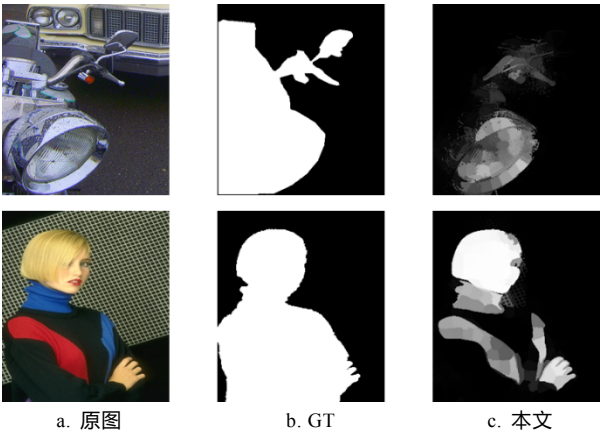


图 8 错误样本

本文算法总体而言具有很好的泛化能力和鲁棒性,但有些情况下,对于前景物体与背景差异不明显或是显著目标的有些区域仅仅在语义上是显著的样本,其表现效果可能会不好.如图 8 所示,其中第一例样本,显著目标与背景的颜色已经面积的差异都很小.最终本算法只检测到车头部分.在第二例中,显著目标是整个人,然而该对象头部以下的区域不具备太多信息非常.但从语义的角度,该区域是整个人物的身体部分.所以其也应为显著区域.但最终本算法只有检测出该目标的头部.究其原因,可能是本文网络所使用的无监督特征提取层在提高提取特征速度的同时却丢失了一些

信息。尽管如此, 本算法对于大多数样本都可以准确的检测到显著目标。其表现性能正如图 4, 图 5, 以及图 6 所示。

4 结 语

本文提出了一个端到端的深度卷积神经网络来解决显著性检测问题。相比之前的基于深度卷积网络的视觉显著性的工作, 本算法有三点创新。首先是本文设计了一个无监督特征提取层将每个图像划分为区域, 并从每个区域提取全面的特征向量, 通过校正层将其映射到可被卷积网络接受的特征图。这样即快速减低维度加快训练和推断速度又提取出重要且有代表性的特征信息。然后是本文加入深度卷积网络与无监督特征提取层融化进行端到端的学习。最后是在训练时, 该网络加入了三级损失函数来控制网络训练过程, 这样不仅保证网络稳定收敛还能大大提高网络的精确度。本文还在四个公开数据集上来对本算法的 P-R 曲线, MAE, 和 F_{γ} -measure 进行评估。实验证明本算法表现优异, 优于八种最先进的算法。

参考文献(References):

- [1] Ji Yuhang, Ma Lizhuang. Stability-based tree for disparity refinement[J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(12): 2159-2167(in Chinese)
(季雨航, 马利庄. 基于稳定树的立体匹配视差优化算法[J]. 计算机辅助设计与图形学学报, 2016, 28(12): 2159-2167)
- [2] Lin Xiao, Shen Yang, Ma Lizhuang, *et al.* Image resizing based on shape and structure-preserving of salient object[J]. Computer Science, 2014, 41(12): 288-292(in Chinese)
(林 晓, 沈 洋, 马利庄, 等. 显著物体形状结构保持的图像缩放方法[J]. 计算机科学, 2014, 41(12): 288-292)
- [3] Qiao Congbin, Sheng Bin, Wu Wen, *et al.* Video deblurring by segmentation from motion[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(11): 2108-2115(in Chinese)
(谯从彬, 盛 斌, 吴 雯, 等. 基于运动分割的视频去模糊[J]. 计算机辅助设计与图形学学报, 2015, 27(11): 2108-2115)
- [4] Laurent I. Automatic foveation for video compression using a neurobiological model of visual attention[J]. IEEE Trans Image Process, 2004, 13(10):1304-1318.
- [5] A Shape-Preserving Approach to Image Resizing[J]. Computer Graphics Forum, 2009, 28(7):1897-1906.
- [6] Rutishauser U, Walther D, Koch C, *et al.* Is bottom-up attention useful for object recognition?[C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004.
- [7] Wei Y, Liang X, Chen Y, *et al.* STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(11):2314-2320.
- [8] Wu H, Li G, Luo X. Weighted attentional blocks for probabilistic object tracking[J]. Visual Computer, 2014, 30(2):229-243.
- [9] Cheng M M, Zhang G X, Mitra N J, *et al.* Global Contrast Based Salient Region Detection[C]// Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.
- [10] Bruce N D B, Tsotsos J K. Saliency, attention, and visual search: An information theoretic approach[J]. Journal of Vision, 2009, 9(3):5-5.
- [11] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(11): 1254-1259
- [12] Jiang P, Ling H, Yu J, *et al.* Salient Region Detection by UFO: Uniqueness, Focusness and Objectness[C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. IEEE, 2013.
- [13] Zhu W, Liang S, Wei Y, *et al.* [IEEE 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Columbus, OH, USA (2014.6.23-2014.6.28)] 2014 IEEE Conference on Computer Vision and Pattern Recognition - Saliency Optimization from Robust Background Detection[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014:2814-2821.
- [14] Kim J, Han D, Tai Y W, *et al.* Salient Region Detection via High-Dimensional Color Transform[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014.
- [15] Yang J, Yang M H. Top-Down Visual Saliency via Joint CRF and Dictionary Learning[C]// Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
- [16] Hou Q, Cheng M M, Hu X, *et al.* Deeply Supervised Salient Object Detection with Short Connections[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018:1-1.
- [17] Wang L, Lu H, Ruan X, *et al.* Deep networks for saliency detection via local estimation and global search[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
- [18] Li G, Yu Y. Visual Saliency Detection Based on Multiscale Deep CNN Features[M]. IEEE Press, 2016.
- [19] Borji A, Sihite D N, Itti L. Salient Object Detection: A Benchmark[M]// Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012.
- [20] Borji A, Cheng M M, Hou Q, *et al.* Salient Object Detection: A Survey[J]. Eprint Arxiv, 2014, 16(7):3118.
- [21] Hornung A, Pritch Y, Krahenbuhl P, *et al.* Saliency filters: Contrast based filtering for salient region detection[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012.
- [22] Schölkopf B, Platt J, Hofmann T. Graph-Based Visual Saliency[J]. Proc of Neural Information Processing Systems, 2006, 19:545-552.
- [23] Jiang P, Ling H, Yu J, *et al.* Salient Region Detection by UFO: Uniqueness, Focusness and Objectness[C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. IEEE,

- 2013.
- [24] Yang J, Yang M H. Top-Down Visual Saliency via Joint CRF and Dictionary Learning[C]// Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
 - [25] Li G, Yu Y. Visual Saliency Based on Multiscale Deep Features[J]. 2015.
 - [26] Liu N, Han J. DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
 - [27] Yan Q, Li X, Shi J, et al. Hierarchical Saliency Detection[C]// 2013.
 - [28] Liu T, Sun J, Zheng N N, et al. Learning to detect a salient object[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2007.
 - [29] Li Y, Hou X, Koch C, et al. The Secrets of Salient Object Segmentation[J]. 2014.
 - [30] Li G, Yu Y. Visual Saliency Based on Multiscale Deep Features[J]. 2015.
 - [31] Movahedi V, Elder J H. Design and perceptual validation of performance measures for salient object segmentation[C]// Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010.
 - [32] Jiang H, Wang J, Yuan Z, et al. Salient Object Detection: A Discriminative Regional Feature Integration Approach[C]// Computer Vision and Pattern Recognition. 2013.
 - [33] Li H, Lu H, Lin Z, et al. Inner and Inter Label Propagation: Salient Object Detection in the Wild[J]. IEEE Transactions on Image Processing, 2015, 24(10):3176-3186.
 - [34] Peng H, Li B, Ling H, et al. Salient Object Detection via Structured Matrix Decomposition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4):818-832.
 - [35] Zhang L, Ai J, Jiang B, et al. Saliency Detection via Absorbing Markov Chain with Learnt Transition Probability[J]. IEEE Transactions on Image Processing, 2017:1-1.