

一种平衡准确性以及高效性的显著性目标检测深度卷积网络模型

张文明 姚振飞 高雅昆* 李海滨
(燕山大学电气工程学院 秦皇岛 066004)

摘要: 当前的显著性目标检测算法在准确性和高效性两方面不能实现良好的平衡, 针对这一问题, 该文提出了一种新的平衡准确性以及高效性的显著性目标检测深度卷积网络模型。首先, 通过将传统的卷积替换为可分解卷积, 大幅减少计算量, 提高检测效率。其次, 为了更好地利用不同尺度的特征, 采用了稀疏跨层连接结构及多尺度融合结构来提高模型检测精度。广泛的评价表明, 与现有方法相比, 所提的算法在效率和精度上都取得了领先的性能。

关键词: 显著性检测; 深度学习; 分解卷积; 稀疏跨层连接; 多尺度融合

中图分类号: TN911.73; TP391.41

文献标识码: A

文章编号: 1009-5896(2020)05-1201-08

DOI: 10.11999/JEIT190229

A Deep Convolutional Network for Saliency Object Detection with Balanced Accuracy and High Efficiency

ZHANG Wenming YAO Zhenfei GAO Yakun LI Haibin

(School of Electrical Engineering, Yan Shan University, Qinhuangdao 066004, China)

Abstract: It is difficult for current salient object detection algorithms to reach a good balance performance between accuracy and efficiency. To solve this problem, a deep convolutional network for saliency object detection with balanced accuracy and high efficiency is produced. First, through replacing the traditional convolution with the decomposed convolution, the computational complexity is greatly reduced and the detection efficiency of the model is improved. Second, in order to make better use of the characteristics of different scales, sparse cross-layer connection structure and multi-scale fusion structure are adopted to improve the detection precision. A wide range of evaluations show that compared with the existing methods, the proposed algorithm achieves the leading performance in efficiency and accuracy.

Key words: Saliency detection; Deep learning; Decomposed convolution; Sparse cross-layer connection; Multi-scale fusion

1 引言

近年来, 显著性目标检测受到越来越多的关注, 其通过模拟人类视觉注意机制检测识别图像中最重要的目标或者区域。显著性目标检测方法通常是各种计算机视觉应用的第1步, 作为其他视觉任务的预处理步骤, 设计出一种既能够有效学习目标特征保持高检测精度, 又具备较高检测效率的显著性检测网络, 成为了当前研究的关键。

一些早期的深度显著性检测模型利用卷积神经网络(Convolutional Neural Networks, CNN)提取

特征或上下文信息来预测超像素等图像片段的显著性分数。文献[1]提出了两种卷积神经网络, 将局部超像素估计和全局建议搜索结合起来进行显著目标检测。文献[2]提取每个超像素的上下文信息, 并利用CNN特征来计算超像素的显著性值。继CNN之后, 几种利用全卷积网络(Fully Convolutional Networks, FCN)的像素级显著性预测的方法^[3-5]又被提出。文献[3]将底层的空间特征与CNN特征相结合来进行显著性预测。文献[4]提出了一种两级深层次显著性检测网络, 通过从浅层恢复细节来细化粗预测图。文献[5]使用底层线索生成显著先验图, 并利用该先验图以递归方式指导显著性目标预测。

随着深度学习的不断发展, 更多端对端的网络模型相继被提出。文献[6]提出了一种新的空间卷积显著性检测框架网络, 将空间注意机制引入到CNN图层, 提取每个图像样本的全局和局部特征; 全局

收稿日期: 2019-04-08; 改回日期: 2019-08-30; 网络出版: 2020-01-21

*通信作者: 高雅昆 gaoyakun6@163.com

基金项目: 河北省自然科学基金(F2015203212, F2019203195)

Foundation Items: The Nature Science Foundation of Hebei Province (F2015203212, F2019203195)

特征利用区域的内部突出特性,局部特征信息主要用于控制背景噪声。文献[7]提出了一种新的无监督检测方法,网络结构由一个潜在显著性预测模块和一个协同工作并共同优化的噪声建模模块组成。文献[8]利用人为选定的若干特征图构造一个具有3个卷积层的CNN架构,学习它们的深度对比度信息,最终得到改进后的显著性图。文献[9]考虑到生成对抗网络具有较强的特征学习能力,提出了一种新的多尺度对抗性特征学习模型用于图像显著性检测。文献[10]提出了一种新的弱监督框架,用于训练具有多种监督源的显著性检测模型,该方法设计了3个显著性检测网络,分别从类别标签、标题和噪声伪标签中学习得到。

随着方法的不断创新,显著性检测准确度在逐渐提高,但还存在一些问题。首先,大多数基于FCN的显著性检测模型按顺序叠加单尺度、卷积层和最大池化层;由于接收字段有限,学习得到的特征往往上下文信息不够丰富,导致最终的显著性目标预测图在完整性上有所欠缺。其次,早期工作为了提升模型检测精度,采用了复杂、深层的网络体系,模型的检测效率大幅下降。为了解决上述问题,本文提出了一种新的平衡准确性以及高效性的显著性目标检测深度卷积网络模型。本文模型的贡献如下:

(1) 引入卷积核分解,将2维矩阵表分解为1维矩阵的加权组合,在不影响原有网络功能的同时,减少网络参数加快网络的速度;同时,在分解过程中,通过卷积层共享卷积核来提升网络的紧凑度,使网络冗余最小化,进一步减少了网络的参数数量,并通过在1维卷积核之间插入非线性函数,提升了网络的学习能力。

(2) 在显著性目标检测中,深层次信息与浅层次信息有着互补的关系,因此本文在网络设计中加入了跨层连接以及多尺度融合,充分利用了深层次信息与浅层次信息互补的特性,进一步提升了网络检测目标的完整性,从而提升了网络精度。本文主要结构如下:第2节中对本文的整体网络架构进行详细阐述;第3节为实验验证及对比;第4节为结论。

2 网络模型

为了使模型达到实时性的要求,本文提出的模型以结构较为简单的VGG-16^[11]网络作为参考网络进行设计。在设计中去掉了VGG-16网络的全连接层,保留卷积层得到的空间信息以保证网络的计算效率。修改后的网络提供了5种不同尺度的特征信息,利用跨层连接结构,可以使得每个卷积层学习利用多层次的上下文信息。最后,通过一个多尺度融合模块融合各层特征信息,并通过上采样恢复特征图尺寸,得到最终的显著性目标预测图。整体网络框架图如图1所示,网络框架主要由卷积分解模块、跨尺度连接模块和多尺度融合模块等部分组成。图1框架图中实线箭头线为网络直连主干结构,虚线箭头线所指部分为跨尺度连接结构,带十字圆块为多尺度融合结构,其中卷积块(convblock)结构由下采样及卷积分解模块实现,3×3卷积结构(conv3)利用空洞卷积模块实现,rate代表空洞卷积采样率,最后两层为上采样层(upblock)。

2.1 卷积结构分解

为了提高网络的检测效率,本文在框架中采用了一种新的分解卷积结构。分解卷积结构如图2所示,其中的3×3的2维卷积分解为两个1维卷积。由文献[12]可以得到,任何2维卷积可以由1维卷积的组合表示。由 $W \in R^{C \times d^h \times d^v \times F}$ 表示一个典型的2维卷积层的权重,其中 C 是输入特征图的数目, F 是输出特征图的数目, $d^h \times d^v$ 表示每个特征映射的内核大小。

为了简单起见,考虑相同的内核大小 d ,很容易看到,通过分解将任意2维卷积核的权重 $W_{2D} \in R^{C \times d \times d \times F}$ 降为1对 $W_{1D} \in R^{C \times d \times 1 \times F}$,从而得到具有 $2 \times (C \times d \times F)$ 个参数的1对卷积核。因此,这种因式分解可用于原网络中的3×3卷积分解为2个3×1的卷积,参数量从9变成6,减少了33%,从而提高了计算效率。1维组合通过最小化冗余来改善模型的紧凑性,并通过在1维卷积之间插入非线性性来从理论上提高学习能力。2维卷积计算公式为

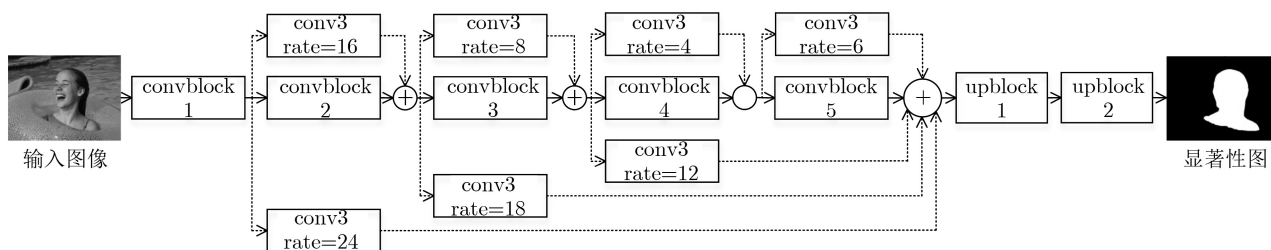


图1 整体框架图

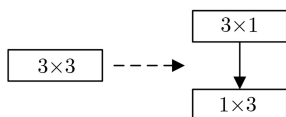


图2 卷积分解示意图

$$B(i, j) = \sum_{m=0}^d \sum_{n=0}^d A(m, n) \otimes I(i - m, j - n) \quad (1)$$

其中 I 为输入图像， A 为卷积模板， m 和 n 为卷积核大小。其分解后的1维卷积计算公式为

$$B(i, j) = \sum_{m=0}^d \sum_{n=0}^d A_2(n) \otimes A_1(m) \otimes I(i - m, j - n) \quad (2)$$

其中 A_1, A_2 为分解后的1维卷积模板。

不同卷积结构的数据对比如表1所示。其中2维卷积代表采用原始2维卷积结构的网络，分解卷积为采用分解卷积结构的网络；参数量为不同卷积结构下网络中的参数量估计值，参数量会直接影响网络的检测效率；准确率为在当前结构下训练的网络进行显著性目标检测后的结果图与真值图对比后计算的值；使用时间为当前结构下网络对单幅图像进行检测时所需的时间。由表1中数据可知，分解卷

结构	参数量(10^6)	准确率(%)	使用时间(s)
2维卷积	5.16	89.3	0.026
分解卷积	3.75	89.7	0.017

积结构能够在保证检测精度的同时有效降低参数量，提升计算效率。

2.2 稀疏跨层连接结构

为了解决网络退化问题，同时充分利用各级特征，提高网络的检测精度。本文参考了稠密网络中的跨层结构。为了使得所设计的网络可以充分利用上下文信息，解决由于特征信息不充足造成非显著性目标误检为显著性目标的问题。本文以VGG-16^[11]直连网络为基础网络，并在网络中引入跨层连接结构。稠密网络中的跨层结构为稠密结构，每一层都会与后面所有层进行连接，这样往往造成层级信息冗余，不同于稠密连接结构，本文的跨层连接结构只将每一层与其下一层及最后一层间进行连接，这样网络中除第一个卷积块外，后面的每一个卷积块都可以得到前面一个卷积块的特征图输出作为当前卷积块的输入。本文在模型中采用了如图3(b)所示的稀疏跨层连接结构。

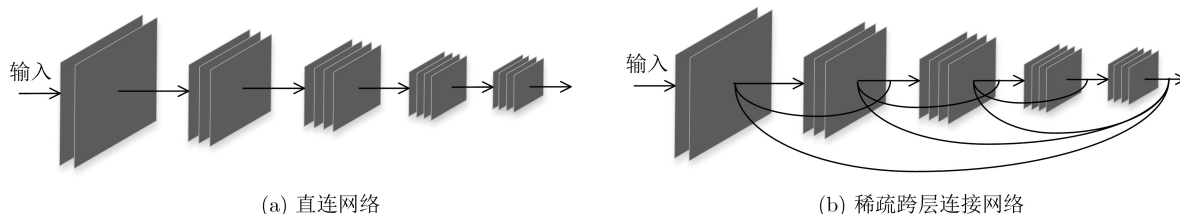


图3 直连与稀疏跨层连接网络结构对比图

为了使得每一个融合结构处可以得到不同尺度下的特征，本文在每个连接结构中加入了空洞卷积(atrous convolution)。因为后面几层可以得到前面卷积层的输入，造成融合(concat)操作后的输入通道数量较大。因此在每个卷积块的分解卷积前面都包含了一个 1×1 的卷积，进行降维操作，目的是减少输入的特征图的通道数量。通过跨层连接，充分利用上下文信息，既能降维减少计算量，又能融合各个通道的特征，进而缓解梯度消失问题。本文分别对使用跨层连接结构和无跨层连接原始结构的网络进行了训练，得到图4显著性图及表2准确率等实验数据。由图4可以看出，相比于直连结构(图3(a))，稀疏跨层连接结构可以很好地解决上下文信息不足造成的非显著目标区域误检为显著性目标的问题。同时，对比表2中准确率数据可知，本文的稀疏跨层连接网络结构，可以在保证效率的前提下提高网络的检测精度。

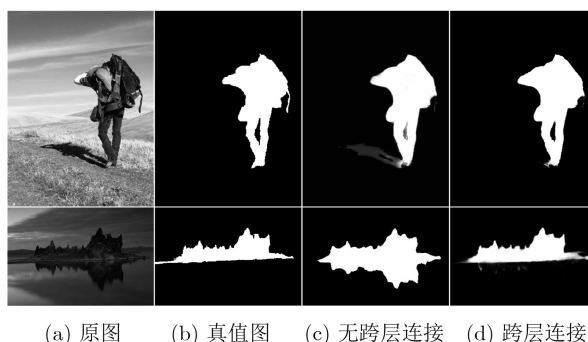


图4 不同连接结构效果对比图

结构	准确率(%)	使用时间(s)
无跨层连接	89.7	0.017
跨层连接	91.7	0.023

2.3 多尺度融合结构

在显著性检测网络中，深层次特征可以有效地

对显著性目标进行定位，而浅层次特征包含丰富的空间细节信息，将两者进行融合可以有效地提升网络对显著性目标的检测精度。基于多尺度特征融合的要求，本文采用了4次多尺度融合操作，其中前3次是两个尺度相融合，结构如图5(a)所示；最后一次融合为将所有尺度特征输出进行融合，结构如图5(b)所示。

首先，将各级特征通过不同采样率的空洞卷积进行卷积操作，增加其感受野。其次，通过金字塔池化操作，使各级特征的分辨率相同。然后，通过融合(concat)操作将各尺度特征进行融合与级联。最后，应用另一个 1×1 卷积操作，对得到的级联特征图进行降维操作，得到最终特征图。

对特征映射进行空洞卷积操作，针对输出特征图上的每个位置 i ，对应的输出为 y 和滤波器 w ，在输入特征映射 x 上应用空洞卷积

$$y[i] = \sum_k x[i + r \cdot k] w[k] \quad (3)$$

其中， r 为采样率，在本文中 $r = [4, 6, 8, 12, 16, 18, 24]$ 。

然后再将经过空洞卷积操作后的输出进行全局平均池化操作，最后将特征图进行级联

$$y = \text{concat}(y_1, y_2, \dots, y_n) \quad (4)$$

其中， n 为特征图数量。通过融合(concat)操作，将多尺度特征进行级联后由 1×1 卷积操作恢复其通道数并得到最终融合特征图，将其作为输入进入到下一层，最终由反卷积操作后得到显著性目标预测图。

2.4 网络结构

整体网络详细结构见表3。reconv为卷积分解结构，reconv $\times 2$ 表示2组卷积分解结构，deconv为反卷积结构，conv3代表卷积核为3的卷积， $1'$ 、 $3'$ 、 $5'$ 、 $7''$ 为经过上支路空洞卷积操作后的输出， $1''$ 、 $3''$ 、 $5''$ 为经过下支路空洞卷积操作后的输出。

3 实验结果与分析

3.1 网络函数、参数设置

Relu函数的单侧抑制特性使得神经网络中的神经元具有稀疏激活性，稀疏后的模型能够更好地挖掘相关特征，拟合训练数据。因此，本文选择Relu

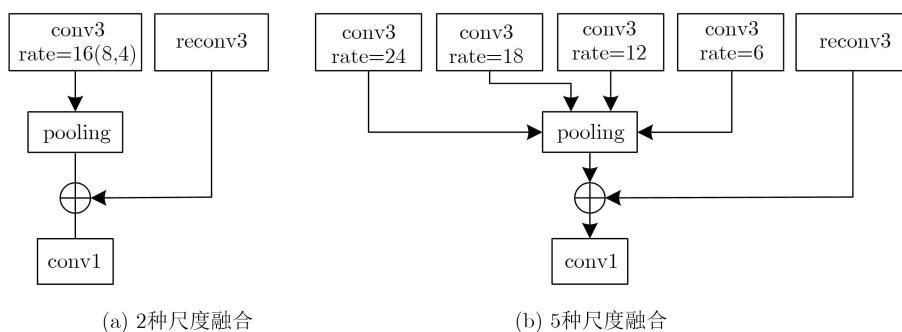


图5 多尺度融合示意图

表3 整体网络结构详表

结构	名称	类型	输出尺寸	输出编号	结构	名称	类型	输出尺寸	输出编号
convblock1	reconv $\times 2$		448 \times 448 \times 16	1	cross-layer	conv3	rate=12	224 \times 224 \times 256	5''
cross-layer	conv3	rate=16	448 \times 448 \times 32	1'	convblock4	maxpool	下采样		
cross-layer	conv3	rate=24	448 \times 448 \times 256	1''		reconv $\times 3$		56 \times 56 \times 128	6
convblock2	maxpool	下采样				concat3	融合	56 \times 56 \times 256	(5' + 6)
	reconv $\times 2$		224 \times 224 \times 32	2		conv1	降维	56 \times 56 \times 128	7
	concat1	融合	224 \times 224 \times 64	(1' + 2)		conv3	rate=6	56 \times 56 \times 256	7''
	conv1	降维	224 \times 224 \times 32	3	convblock5	maxpool	下采样		
cross-layer	conv3	rate=8	224 \times 224 \times 64	3		reconv $\times 3$		28 \times 28 \times 256	8
cross-layer	conv3	rate=18	224 \times 224 \times 256	3''		concat4	融合	28 \times 28 \times 1280	(1'' + 3'' + 5'' + 7'' + 8)
convblock3	maxpool	下采样				conv1	降维	28 \times 28 \times 256	9
	reconv $\times 3$		112 \times 112 \times 64	4	upblock1	deconv	上采样		
	concat2	融合	112 \times 112 \times 128	(3' + 4)		reconv $\times 3$		112 \times 112 \times 64	
	conv1	降维	112 \times 112 \times 64	5	upblock2	deconv	上采样	448 \times 448 \times 2	final
cross-layer	conv3	rate=4	224 \times 224 \times 128	5'					

激活函数为本文网络的激活函数，其计算如式(5)。使用交叉熵损失函数可以加快网络训练时的收敛速度，因此，本文选用交叉熵损失作为损失函数，其计算公式如式(6)。

$$\text{Relu}(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (5)$$

$$C = -\frac{1}{n} \sum_{x=1}^n [y \ln x + (1-y) \ln(1-a)] \quad (6)$$

式(6)中， x 表示样本， n 表示样本总量， y 为真实值， a 为神经元的实际输出。在实验中，分别设置学习率为 $6e-5$ ，权重衰减为 $2e-4$ ，动量为 0.9 ，批量大小为 16 。使用随机梯度下降算法优化训练，其可以根据梯度估计矩阵动态调整学习率，提高网络计算效率。

3.2 数据集与评价指标

一个好的显著性检测模型应该在多种经典的数据集上都具有良好的表现。为此，本文利用5个最有代表性的数据集对本文的显著性目标检测模型进行了评估，分别是：MSRA^[13]，ECSSD^[14]，HKU-IS^[1]，PASCALS^[15]和SOD^[16]。所有这些数据均包含大量的图像，在显著性目标检测方面有着广泛的应用，且每个数据集都由原图和真值图组成。同时，本文选择MSRA数据集作为训练集以及验证集来训练和测试本文提出的模型。

在客观评价方面，利用F-measure和平均绝对误差(Mean Absolute Error, MAE)两种常用的度量指标，分别对本文提出的模型与其他先进的显著目标检测方法进行评价。F-measure值是一个总体性能指标，它是由精确度和召回率的加权谐波计算得到

$$F\text{-measure} = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \quad (7)$$

其中， β^2 设置为 0.3 。精度值 P 是预测显著区域中真实显著像素的比率，并将召回率 R 定义为检测到的显著像素和所有标注真实区域的百分比。通过对预测的显著图进行阈值化，并将其与相应的标注真值结果进行比较来计算精度和召回率。精度 P 以及召回率 R 的计算公式如式(8)和式(9)所示。

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

其中TP为检测出的正确的像素，FP为检测出的不正确的像素，FN为未检测出的正确的像素。平均

绝对误差(MAE)衡量预测显著性图与真值图之间的平均差值，

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)| \quad (10)$$

其中 S 和 G 是预测的显著性图和人工标注的真值图， W 和 H 为图像的宽和高， x, y 为像素点的横纵坐标。

3.3 与先进模型比较

将本文提出的显著性模型与5种当前先进的模型进行了比较，其中3种是基于深度学习的方法：DCL(Deep Contrast Learning)^[17]，ELD(Encoded Low-level Distance)^[3]，NLDF(Non-Local Deep Features)^[18]；另外2种为传统非深度学习方法：MST(Minimum Spanning Tree)^[19]，DSR(Dense and Sparse Reconstruction)^[20]。分别在处理时间、视觉效果、F-measure以及MAE等方面进行了评价比较。

3.3.1 定性评价

从数据库中随机选取了5幅图片对不同算法进行定性分析，检测结果见图6。由图6可以看出，与其它5种方法相比，直观上本文算法结果边缘更加清晰、连贯，显著性区域更加突出明显，检测结果更加接近真值图。

3.3.2 F-measure和MAE值定量评价

不同方法的F-measure和MAE对比数据如表4所示。F-measure值作为准确率与召回率的综合性评价指标，其值越大代表模型的效果越佳，MAE作为结果图与真值图之间的平均误差，其值越小代表着模型的效果越佳。由表4中的数据可以看出本文提出的算法在5个经典数据集上，不论是F-measure指标得分还是MAE指标得分都处在前三位的水平，其中在MSRA以及SOD数据集上两个指标的得分都处于领先的水平，优于其他几个算法。通过数据对比，前7种基于深度学习的显著性目标检测算法在得分上明显高于后3种传统的显著性目标检测算法。DCL，ELD算法虽然存在对多尺度特征的利用，但多尺度间相互独立，或特征融合单一，且不存在跨层结构，因此准确度要低于本文算法。NLDF算法与本文算法和性能相近，但是NLDF算法只存在一组相邻层跨层连接，而不存在非相邻层跨多层连接结构，在SOD数据集上的得分明显低于本文的算法，尤其是在MAE指标方面。

3.3.3 P-R曲线定量评价

在图7中，显示了本文算法与其他算法的P-R曲线对比情况。P-R曲线是由召回率(Recall)为横

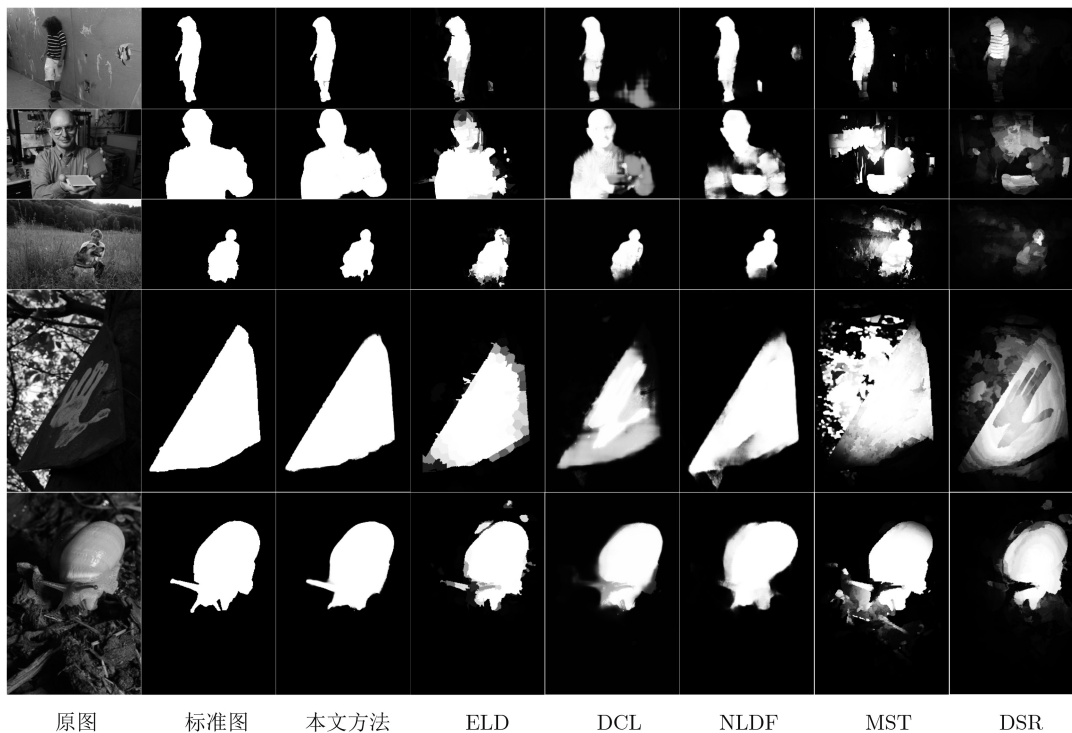


图6 不同模型视觉对比图

表4 F-measure(F-m)和MAE得分表

算法	MSRA		ECSSD		PASCAL-S		SOD		HKU-IS	
	F-m	MAE	F-m	MAE	F-m	MAE	F-m	MAE	F-m	MAE
本文方法	0.914	0.045	0.893	0.060	0.814	0.113	0.832	0.119	0.893	0.036
DCL	0.905	0.052	0.890	0.088	0.805	0.125	0.820	0.139	0.885	0.072
ELD	0.904	0.062	0.867	0.080	0.771	0.121	0.760	0.154	0.839	0.074
NLDF	0.911	0.048	0.905	0.063	0.831	0.099	0.810	0.143	0.902	0.048
MST	0.839	0.128	0.653	0.171	0.584	0.236	-	-	-	-
DSR	0.812	0.119	0.737	0.173	0.646	0.204	0.655	0.234	0.735	0.140

轴, 准确率(Precision)为纵轴绘制的曲线。 $P-R$ 曲线的评价准则为: 若一个算法的 $P-R$ 曲线完全包住了另一个算法的 $P-R$ 曲线, 则认为前者的性能优于后者。如果两个算法的 $P-R$ 曲线有交叉, 则可通过“平衡点”(Break-Event Point, BEP)来评价算法的优劣性, BEP是“查准率=查全率”的数值, 点越高则模型越好。由图7曲线可以看出, 在MSRA, ECSSD以及SOD数据集上, 本文算法 $P-R$ 曲线在其他算法曲线的上方, 说明本文算法优于其他算法; 在HKU-IS与PSACAL数据集上本文算法也取得了优于大多数对比方法的性能。

3.3.4 检测效率定量评价

现将各种算法在处理时间上进行对比, 整个实验在Intel core i7以及单NVIDIA GTX1080显卡环境下进行, 对比结果见表5。公平起见, 对于可以获取到源码的模型, 在同一环境下进行了实验, 未

获取到源码的模型, 选取作者文章中给出的处理效率最大值作为对比数据, 在未给出处理效率的且未获取到源码的模型没有参与比较。所列运行环境中, Titan显卡性能要强于本文所用的GTX1080显卡。所列方法中, ELD, DCL均使用原图与块像素分割信息或图像内反射信心来训练两个或多个主网络分支进行显著性目标检测, 多网络分支结构相比于本文的单分支主网络结构更复杂, 因此实时性弱于本文的方法。NLDF方法虽然网络结构相对简单, 但其并未采用分解卷积结构, 实时性要弱于本文方法。本文方法在检测效率方面优势明显。

4 结束语

本文提出了一种高效的显著性目标检测模型。相对于其他显著性目标检测模型, 本文主要是针对检测效率和精度两方面共同提升的问题进行改进。

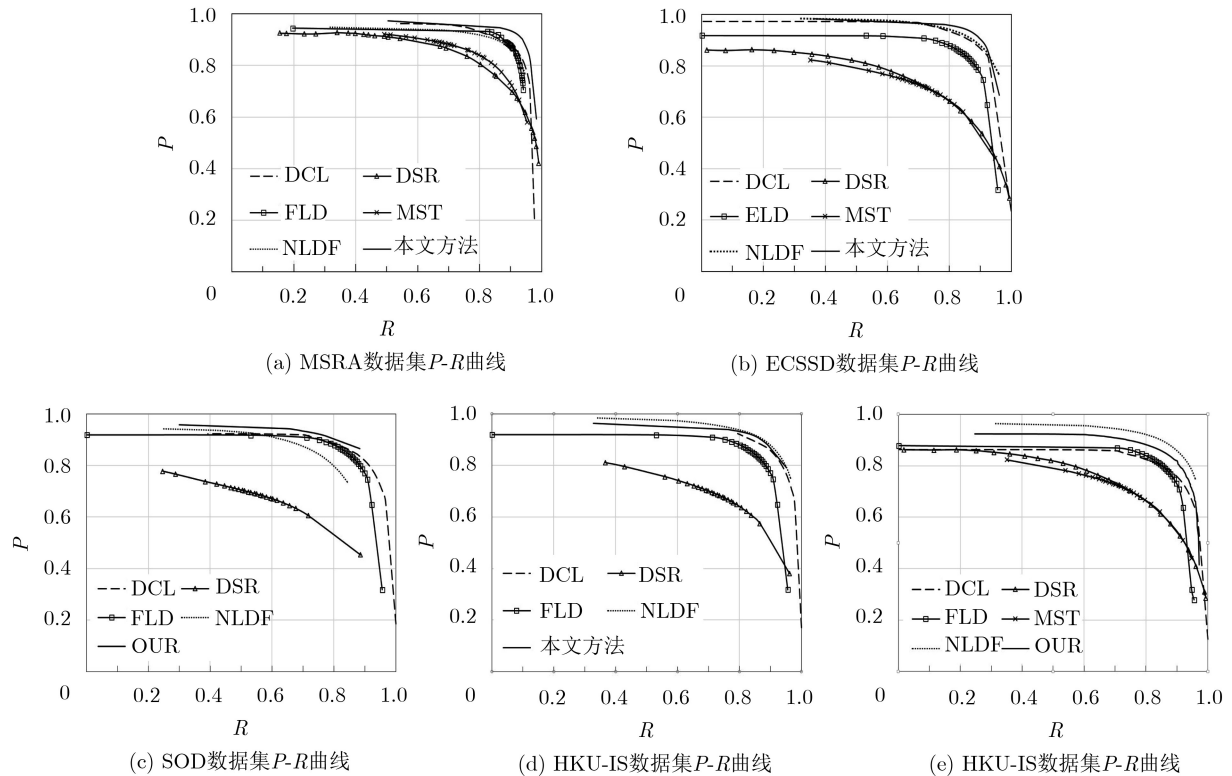


图 7 5种数据集上不同算法P-R曲线图

表 5 不同算法处理时间对比(s)

模型	本文方法	DCL	ELD	NLDF	MST	DSR
时间	0.023	1.200	0.300	0.080	0.025	13.580
环境	GTX1080	GTX1080	GTX1080	Titan X	i7 CPU	i7 CPU
尺寸	448×448	300×400	400×300	300×400	300×400	400×300

本文的模型在保证精度的同时大幅提升了模型的检测效率，节约了计算成本，提高了显著性目标检测作为预处理阶段的作用。本文首先利用1维卷积分解结构代替2维卷积结构，有效降低了模型参数量，从而使算法的检测速度得到明显提升。然后采用跨尺度连接结构，充分利用不同尺度特征信息，避免非显著性小目标对检测结果的干扰，最终通过多尺度融合结构得到高精度的显著性目标检测结果。在5个数据集上，利用不同的评价指标对本文提出的算法进行评估，结果表明，本文方法在检测精度，以及检测速度方面都有着明显的优势或处于前列。本文方法还有很大的提升空间，后续将进一步增强网络的鲁棒性，使得网络可以在当前经典数据集上都达到领先水平。

参考文献

[1] WANG Lijun, LU Huchuan, RUAN Xiang, *et al.* Deep networks for saliency detection via local estimation and global search[C]. 2015 IEEE Conference on Computer

Vision and Pattern Recognition, Boston, USA, 2015: 3183–3192. doi: [10.1109/CVPR.2015.7298938](https://doi.org/10.1109/CVPR.2015.7298938).
 [2] LI Guanbin and YU Yizhou. Visual saliency based on multiscale deep features[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 5455–5463. doi: [10.1109/CVPR.2015.7299184](https://doi.org/10.1109/CVPR.2015.7299184).
 [3] LEE G, TAI Y W, and KIM J. Deep saliency with encoded low level distance map and high level features[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 660–668. doi: [10.1109/CVPR.2016.78](https://doi.org/10.1109/CVPR.2016.78).
 [4] LIU Nian and HAN Junwei. DHSNet: Deep hierarchical saliency network for salient object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 678–686. doi: [10.1109/CVPR.2016.80](https://doi.org/10.1109/CVPR.2016.80).
 [5] WANG Linzhao, WANG Lijun, LU Huchuan, *et al.* Saliency detection with recurrent fully convolutional networks[C]. The 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 825–841. doi: [10.1007/978-3-319-46493-0_50](https://doi.org/10.1007/978-3-319-46493-0_50).
 [6] ZHANG Xinsheng, GAO Teng, and GAO Dongdong. A new deep spatial transformer convolutional neural network for image saliency detection[J]. *Design Automation for Embedded Systems*, 2018, 22(3): 243–256. doi: [10.1007/s10617-018-9209-0](https://doi.org/10.1007/s10617-018-9209-0).
 [7] ZHANG Jing, ZHANG Tong, DAI Yuchao, *et al.* Deep

- unsupervised saliency detection: A multiple noisy labeling perspective[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 9029–9038. doi: [10.1109/CVPR.2018.00941](https://doi.org/10.1109/CVPR.2018.00941).
- [8] CAO Feilong, LIU Yuehua, and WANG Dianhui. Efficient saliency detection using convolutional neural networks with feature selection[J]. *Information Sciences*, 2018, 456: 34–49. doi: [10.1016/j.ins.2018.05.006](https://doi.org/10.1016/j.ins.2018.05.006).
- [9] ZHU Dandan, DAI Lei, LUO Ye, *et al.* Multi-scale adversarial feature learning for saliency detection[J]. *Symmetry*, 2018, 10(10): 457–471. doi: [10.3390/sym10100457](https://doi.org/10.3390/sym10100457).
- [10] ZENG Yu, ZHUGE Yunzhi, LU Huchuan, *et al.* Multi-source weak supervision for saliency detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 6067–6076.
- [11] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. 2014, arXiv: 1409.1556.
- [12] ALVAREZ J and PETERSSON L. DecomposeMe: Simplifying convNets for end-to-end learning[J]. 2016, arXiv: 1606.05426v1.
- [13] LIU Tie, YUAN Zejian, SUN Jian, *et al.* Learning to detect a salient object[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(2): 353–367. doi: [10.1109/TPAMI.2010.70](https://doi.org/10.1109/TPAMI.2010.70).
- [14] YAN Qiong, XU Li, SHI Jianping, *et al.* Hierarchical saliency detection[C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 1155–1162. doi: [10.1109/CVPR.2013.153](https://doi.org/10.1109/CVPR.2013.153).
- [15] LI Yin, HOU Xiaodi, KOCH C, *et al.* The secrets of salient object segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 280–287. doi: [10.1109/CVPR.2014.43](https://doi.org/10.1109/CVPR.2014.43).
- [16] MOVAHEDI V and ELDER J H. Design and perceptual validation of performance measures for salient object segmentation[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 49–56. doi: [10.1109/CVPRW.2010.5543739](https://doi.org/10.1109/CVPRW.2010.5543739).
- [17] LI Guanbin and YU Yizhou. Deep contrast learning for salient object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 478–487. doi: [10.1109/CVPR.2016.58](https://doi.org/10.1109/CVPR.2016.58).
- [18] LUO Zhiming, MISHRA A, ACHKAR A, *et al.* Non-local deep features for salient object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6593–6601. doi: [10.1109/CVPR.2017.698](https://doi.org/10.1109/CVPR.2017.698).
- [19] TU W C, HE Shengfeng, YANG Qingxiang, *et al.* Real-time salient object detection with a minimum spanning tree[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2334–2342. doi: [10.1109/CVPR.2016.256](https://doi.org/10.1109/CVPR.2016.256).
- [20] LI Xiaohui, LU Huchuan, ZHANG Lihe, *et al.* Saliency detection via dense and sparse reconstruction[C]. 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 2013: 2976–2983. doi: [10.1109/ICCV.2013.370](https://doi.org/10.1109/ICCV.2013.370).
- 张文明: 男, 1979年生, 副教授, 研究方向为工业过程控制、机器视觉。
- 姚振飞: 男, 1992年生, 硕士生, 研究方向为机器视觉与图像处理。
- 高雅昆: 男, 1988年生, 博士生, 研究方向为机器视觉与图像处理。
- 李海滨: 男, 1978年生, 教授, 研究方向为工业过程控制、机器视觉、人工智能。