

面向显著性目标检测的SSD改进模型

余春艳 徐小丹 钟诗俊*

(福州大学数学与计算机科学学院 福州 350108)

摘要: 传统显著性目标检测方法常假设只有单个显著性目标,其效果依赖显著性阈值的选取,并不符合实际应用需求。近来利用目标检测方法得到显著性目标检测框成为一种新的解决思路。SSD模型可同时精确检测多个不同尺度的目标对象,但小尺寸目标检测精度不佳。为此,该文引入去卷积模块与注意力残差模块,构建了面向多显著性目标检测的DAR-SSD模型。实验结果表明,DAR-SSD检测精度显著高于SOD模型;相比原始SSD模型,在小尺度和多显著性目标情形下性能提升明显;相比MDF和DCL等深度学习框架下的方法,也体现了复杂背景情形下的良好检测性能。

关键词: 目标检测; 显著性目标检测; 去卷积; 注意力残差

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2018)11-2554-08

DOI: 10.11999/JEIT180118

An Improved SSD Model for Saliency Object Detection

YU Chunyan XU Xiaodan ZHONG Shijun

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

Abstract: Traditional saliency object detection methods, assuming that there is only one salient object, is not conducive to practical application. Their effects are dependent on saliency threshold. Object detection model provides a kind of new solutions. SSD can accurately detect multi-objects with different scales simultaneously, except for small objects. To overcome this drawback, this paper presents a new multi-saliency objects detection model, DAR-SSD, appending a deconvolution module embedded with an attention residual module. Experiments show that DAR-SSD achieves a higher detection accuracy than SOD. Also, it improves detection performance for multi-saliency objects on small scales, compared with original SSD, and it has an advantage over complicated background, compared with MDF and DCL, which also are deep model based methods.

Key words: Object detection; Saliency object detection; Deconvolutional; Attention residual

1 引言

显著性目标检测问题不仅有助于图像缩略图生成、图像编辑和总结等视觉任务的完成,还有助于图像理解系统的实现,一直是计算机视觉领域的研究热点之一。

早期研究多为基于显著性信息的检测模型,即以显著性图为基础,寻找包含显著对象的最优检测

框。Liu等人^[1]将最优检测框定义为包含足够重要程度和包含重要对象的最小矩形框,并在显著性图上使用贪心算法求解矩形框。Valenti等人^[2]组合边缘、颜色与形状等特征获得显著性图,将获得的信息用作分割算法输出的加权,得到显著性能量图后,运用高效子窗口搜索方法找到显著性目标最优矩形窗口。Luo等人^[3]将显著性目标检测视为最大显著密度区域的定位问题,首先用显著图上的最大显著密度表示显著目标,之后通过分支定界搜索加快检测并获得最优解。此类基于显著性信息的检测模型的性能极大地依赖显著性阈值的选取,且通常假设图像中有且只有单个显著性目标,与实际应用并不吻合。然而,一旦脱离该假设,仅仅依赖性显著性图,很难直接从中分离出多个显著性目标。

有研究工作尝试在原始图像上手工提取特征,由此提取显著性目标。Feng等人^[4]提出显著性窗口的概念并据此提出基于分割的表征,利用贪心算法

收稿日期: 2018-01-26; 改回日期: 2018-07-17; 网络出版: 2018-07-27

*通信作者: 钟诗俊 n160320046@fzu.edu.cn

基金项目: 福建省产学研合作重大项目(2016H6010), 福建省自然科学基金(2015J01420), 福建省引导性基金(2016Y0060), 福建省卫生教育联合攻关计划项目(WKJ2016-2-26)

Foundation Items: The Major Project in Industry-university Cooperation of Fujian Province (2016H6010), The Natural Science Foundation of Fujian Province (2015J01420), The Guiding Found of Fujian Province (2016Y0060), The Health-Education Joint Project of Fujian Province (WKJ2016-2-26)

得到最优显著性目标检测框。Yildirim等人^[5]对颜色进行量化得到CIELAB颜色空间的直方图，通过双边滤波计算量化颜色的空间中心与方差得到显著性目标检测框。以上基于手工提取特征的工作检测效果优于基于显著性信息的检测模型，但仍存在手工提取特征不够丰富、普适性欠佳等不足。

针对手工提取特征的不足，开始出现基于深度学习的检测模型，其基本思路是将显著性目标视为目标的特例，迁移应用有监督的深度学习目标检测算法^[6-8]。Zhang等人^[9]提出的SOD模型将目标检测模型Multibox^[10]迁移后，实验结果表明，检测精度比基于手工提取特征的模型提升了16%以上。

SSD模型^[11]可以同时精确且快速地检测出多个不同尺度的目标对象，本文尝试将其迁移应用于显著性目标检测。针对SSD模型检测小尺寸目标精度不足的缺陷，引入去卷积模块与注意力残差模块以增加上下文信息，进而构建了DAR-SSD模型。实验结果表明将SSD模型迁移至显著性目标检测可行，且去卷积与注意力残差模块的引入可有效提高检测精度。

2 显著性目标检测与目标检测

2.1 问题描述

给定大量自然图像数据集 $S = \{X_1, X_2, \dots, X_N\}$ 及其显著性目标边界框(bounding box)真值集合 $Y = \{Y_1, Y_2, \dots, Y_N\}$ ， N 为数据集大小。若图像 X_i 中包含 l 个显著性目标， $Y_i = \{[(x_{j1}^i, y_{j1}^i), (x_{j2}^i, y_{j2}^i)] | 1\}^l$ 为图像 X_i 的显著性目标边界框真值，其中， (x_{j1}^i, y_{j1}^i) 为显著性目标 j 边界框左上角坐标， (x_{j2}^i, y_{j2}^i) 为右下角坐标。

给定卷积神经网络 D ，输入 X_i ， Y_i 成对训练，学习得到特征函数 $F_{\text{final}}(\cdot)$ ，即可对任意输入图像进行显著性值对象预测。对 S 中的图像 X_i 进行显著性目标检测时，输出候选显著性目标检测框集合 $Y_i' = \{[(x_{j1}^i, y_{j1}^i), (x_{j2}^i, y_{j2}^i), \text{conf}_j^i] | 1\}$ ，其中 (x_{j1}^i, y_{j1}^i) 为候选显著性目标 j 检测框左上角坐标， (x_{j2}^i, y_{j2}^i) 为右下角坐标， conf_j^i 是候选对象为真值显著性目标的置信度。

2.2 SOD模型

SOD模型使用CNN预测100个先验窗口的显著性目标置信度，其中先验窗口由训练集的真实边界框聚类得到；置信度越高，则该先验窗口为显著性目标检测框的概率越大；随后利用基于最大后验(Maximum A Posteriori, MAP)原理的子集优化公式产生一组紧凑的显著性目标检测结果。

该模型采用VGG网络基础架构，初始生成的100个候选检测框仅能提供显著性目标的粗略定位。受注意力机制的启发，从高到低选取前 M 个高置信度的先验窗口，根据每一个选中的先验窗口，对原图像相应位置进行2倍大小采样后再次输入到CNN中，得到 M 组先验窗口置信度，每组筛选置信度最高的前10个先验窗口，与初始生成的置信度最高的前20个先验窗口，合并为最终检测框候选集合。

SOD模型直接使用先验窗口作为候选边界框，得到的检测框通常不够精确，且其仅在单一尺度上生成初始候选检测框，当图像存在多个不同尺度显著性目标时，检测精度欠佳。

3 显著性目标检测模型DAR-SSD

Liu等人^[11]于2016年提出的SSD模型基于前馈卷积网络生成边界框集合与对应类别置信度，用非极大值抑制方法(Non-Maximum Suppression, NMS)产生最终结果。该模型使用多个尺度的特征图，且预测默认框偏移量，检测结果更加精确，而且能够同时预测不同尺度的目标，相较于SOD，更适合迁移应用于显著性目标检测问题。

3.1 基于SSD模型的显著性目标检测基本思路

图1给出了SSD模型迁移后的网络结构，基础网络结构SSD_Base共包含9个卷积模块，分别包含2, 2, 3, 3, 3, 4, 2, 2, 2个卷积，其中第4, 6, 7, 8, 9个卷积模块的最后一个卷积同时侧输出用于预测模块。预测使用一组 3×3 的卷积核，对特征图的每个位置预测了分属显著性目标和背景2个类的置信度，以及相对于默认框的偏移量。

默认框定义如式(1)：

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1} (k - 1), k \in [1, m] \quad (1)$$

其中， s_{\min} 和 s_{\max} 设为0.2和0.9，分别表示最底层和

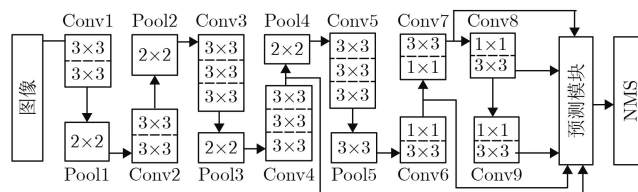


图1 SSD模型迁移网络结构图

最高层尺度, 中间层尺度均匀间隔, m 为用于预测的卷积层个数, 值为5。默认框可设定不同宽高比, 宽高比为1时增加一个尺度为 $s_k' = \sqrt{s_k s_{k+1}}$ 的默认框^[4]。每个默认框的宽和高分别为 $w_k^a = s_k \sqrt{a_k}$ 和 $h_k^a = s_k / \sqrt{a_k}$, 中心为 $\left(\frac{i+0.5}{|f_k|}, \frac{j+0.5}{|f_k|}\right)$ ($i, j \in [0, |f_k|]$), $|f_k|$ 为第 k 个特征图大小。结合所有尺度和宽高比, 默认框可产生涵盖输入各种大小和形状的预测。

令 N 为默认框个数, 目标损失函数为置信度损失与定位损失的加权和:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (2)$$

定位损失为预测框 l 与真实边界框 g 的平滑L1损失, 用于回归中心为 (cx, cy) , 宽为 w 高为 h 的默认框 d 的偏置, 定义如式(3):

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k f_{L1}(l_i^m - g_j^m) \quad (3)$$

其中, c 表示类别, f_{L1} 表示平滑函数, 定义为 $\|l_i^m - g_j^m\|_1$ 。 $g_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^w$, $g_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^h$, $g_j^w = \lg\left(\frac{g_j^w}{d_i^w}\right)$, $g_j^h = \lg\left(\frac{g_j^h}{d_i^h}\right)$ 。

置信度损失为多个类别的softmax损失, 定义如式(4):

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \lg(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \lg(\hat{c}_i^0) \quad (4)$$

其中, $\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$, 若默认框 i 与类别 p 的真实边界框 j 相匹配, 则 x_{ij}^p 为1, 反之为0。

3.2 去卷积模块

SSD尺度敏感且在小尺度目标检测上表现较差^[4]。由SSD_Base结构可知, 目标对象尺寸较小时, 其真实检测框大概率地对应第4个卷积层上的特征。需要指出的是, 该层特征相对底层, 包含的语义特征不足。倘若对底层特征增添高层语义特征, 将有助于提高小尺度目标的检测率。

文献[12]提出将高层与底层特征图对齐后求和, 可实现高层语义与底层外观信息的融合, 并在图像语义分割上取得了不错的效果。由此, 本文引入去卷积模块, 将高层特征图进行去卷积操作, 融合底层与高层语义特征以增强上下文信息。具体地, SSD_Base结构直接使用不同尺度下长宽逐层缩小2倍的特征层实现多尺度目标检测, 因此先使用卷积核大小为 4×4 , 步长为2的卷积对高层特征图去卷积, 使其大小扩展两倍; 进一步地, 如图2黄色部分所示, 加入crop层以适应输入任意大小, 使去卷积后的特征图与其前一层特征图大小对齐; 将两张特征图逐像素相乘, 得到去卷积模块后高层语义与底层的融合特征图。

3.3 注意力残差模块

文献[13]指出尺寸相关的目标检测器后添加面向特定任务的子网络, 可有效提升检测准确率。另一方面, 文献[14]在VGG网络上使用空间金字塔模型实现对象检测, 其第5个卷积层的可视化结果表明, 对照原始图像, 每个特征图都有其特定的最强激活区域。

考虑到显著性目标检测通常通过视觉注意机制对候选目标进行显著性分析, 本文提出在SSD_Base与预测模块间引入注意力残差模块, 强化选取注意力集中的感兴趣区域。倘若某一特征图的最强激活区域非显著性区域, 降低其显著性权重, 显然有利于显著性目标检测。具体地, 如图2绿色部分所示, 通过卷积核大小为 1×1 的卷积和sigmoid函数得到0到1之间的注意力值, 学习各卷积层特征图的注意力图, 并与原始特征图加权组合。

同时, 将注意力残差模块嵌入去卷积模块中, 即可得去卷积注意力残差模块(Deconvolution Attention Residual module, DAR), 其完整结构如图2所示。

3.4 整体结构

针对SSD模型用于多显著性目标检测面临的小尺度对象检测性能不佳和缺乏显著性信息的问题, 本文提出的DAR-SSD模型, 于SSD_Base之后, 第4, 6, 7, 8, 9卷积模块最后一个特征卷积层侧输出至一个DAR模块, 最后输出的候选显著性目标检测框和置信度通过NMS进行筛选, 如图3所示。

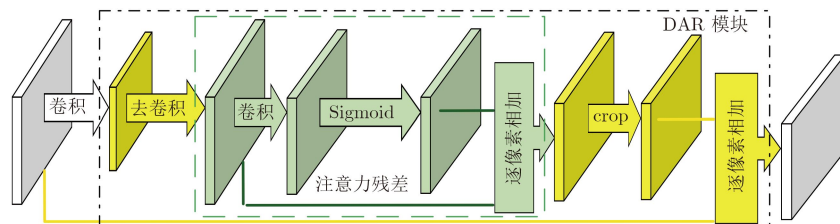


图2 DAR模块

4 实验结果与分析

4.1 数据集

本文使用Salient Object Subitizing (SOS)数据集^[5]作为训练集, 包含图像5244张, 其中无显著性目标的背景图像1293张, 单显著性目标图像2573张, 含2个和3个显著性目标的图像分别为777和601张。

测试集为MSO, ECSSD, DUT-OMRON, HKU-IS 4个数据集。MSO分离自SOS, 共计1224张图像, 包含背景图像338张。ECSSD和DUT-OMRON多为单显著性目标图像, HKU-IS多为多显著性目标图像; 我们对后三者根据显著性图进行人工标注, 人工无法判断不计入测试集; 最终3个测试集分别包含945张、4893张、3938张图像。表1对测试集依据显著性目标数目和尺度进行了分类统计。其中, 小尺度的定义与文献^[9]一致, 为真

实边界框大小小于图像1/4的显著性目标。

4.2 实验设置

本文对SOD, SSD与DAR-SSD针对不同尺度和多个显著性目标的情形进行主观对照分析, 并以PR曲线和平均准确率(Average Precision, AP)^[9]为定量指标进行客观对照分析。

与文献^[9]一致, DAR-SSD与SSD训练时, 输入图像大小为 224×224 , 初始学习率为0.001, 动量为0.9, 批次为32, 权重衰减为0.0005, 训练迭代至 3×10^4 次时终止。测试时, 3种模型各取80个候选检测框, 分别运用NMS和MAP筛选得到最终显著性目标检测框, MAP值的设定与文献^[9]一致, NMS的阈值为0.5, 主观实验中, 置信度阈值为0.5。

4.3 主观实验结果

图4~图6为3种情形下的检测结果对照。图4中SOD与DAR-SSD都准确地检测出小船, SSD未能检出。图5中3种方法无一例外地检测成功, 但SOD

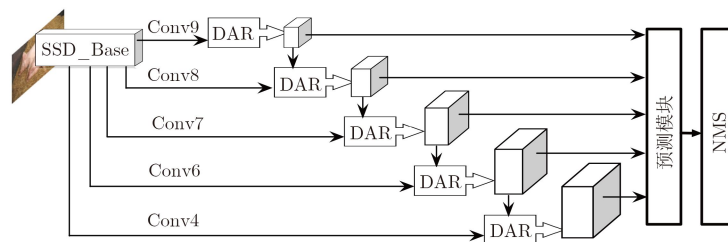


图 3 DAR-SSD网络结构图

表 1 数据集构成(张)

| 数据集 | MSO | ECSSD | DUT-OMRON | HKU-IS |
|--------------|-----|-------|-----------|--------|
| 含显著性目标图像数 | 886 | 945 | 4893 | 3938 |
| 含单显著性目标图像数 | 611 | 807 | 4121 | 1276 |
| 含多显著性目标图像数 | 275 | 138 | 772 | 2662 |
| 含小尺度显著性目标图像数 | 446 | 323 | 3067 | 3020 |
| 含大尺度显著性目标图像数 | 440 | 622 | 1826 | 918 |

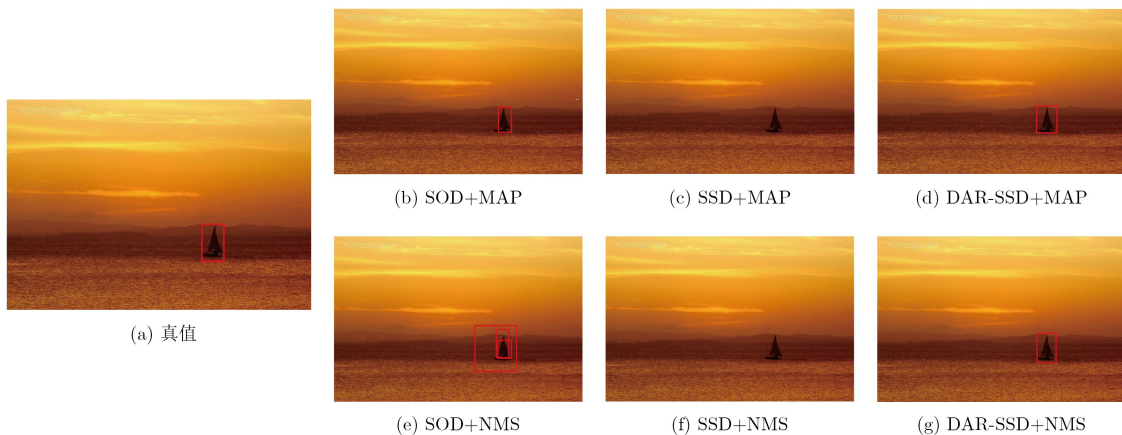


图 4 单个小尺寸显著性目标检测实例

得到的置信度最低。图6中多个显著性目标尺度不同，SSD与DAR-SSD都准确地检测出3个显著性目标，体现了多尺度检测的优势，而SOD丢失了尺度最小的显著性目标，且检测框定位偏差较大。

综上，DAR-SSD一方面继承了原始SSD模型可以同时检测多个不同尺度对象的优点，另一方面DAR模块的引入不但改善了原始SSD模型检测小尺度对象精度欠佳的缺点，还通过注意力机制改善了定位置信度和精准性，在不同情形下均表现良好，

检测框最为精确。

4.4 客观实验结果

图7和图8展示了3种模型分别使用MAP和NMS时得到的PR(Precision-Recall)曲线，对应的AP值如表2和表3所示。3个模型的AP值趋势一致，DAR-SSD和SSD在PR和AP上性能均显著优于SOD，在NMS方法上提升了将近10%，表明SSD迁移应用后依然继承了原有目标检测的良好性能，且其能同时检测多尺度对象的特点是表现较

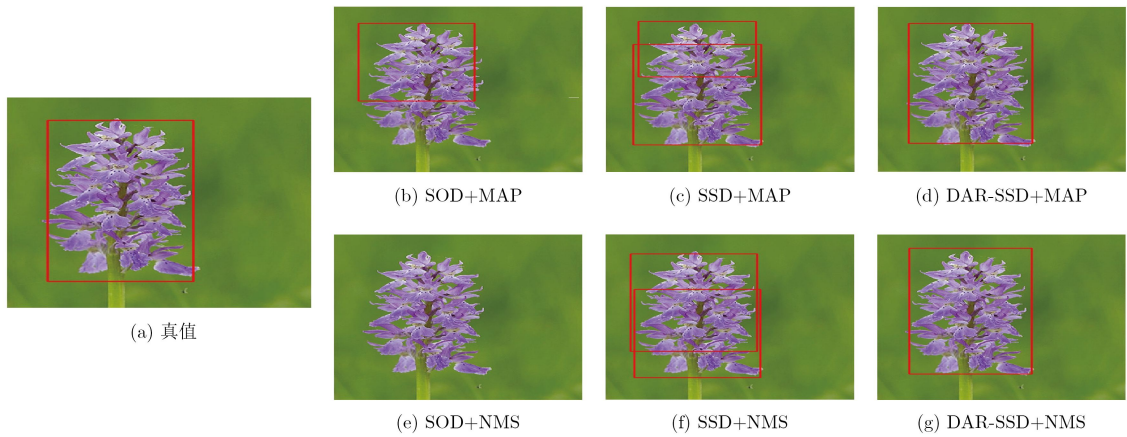


图5 单个大尺寸显著性目标检测实例

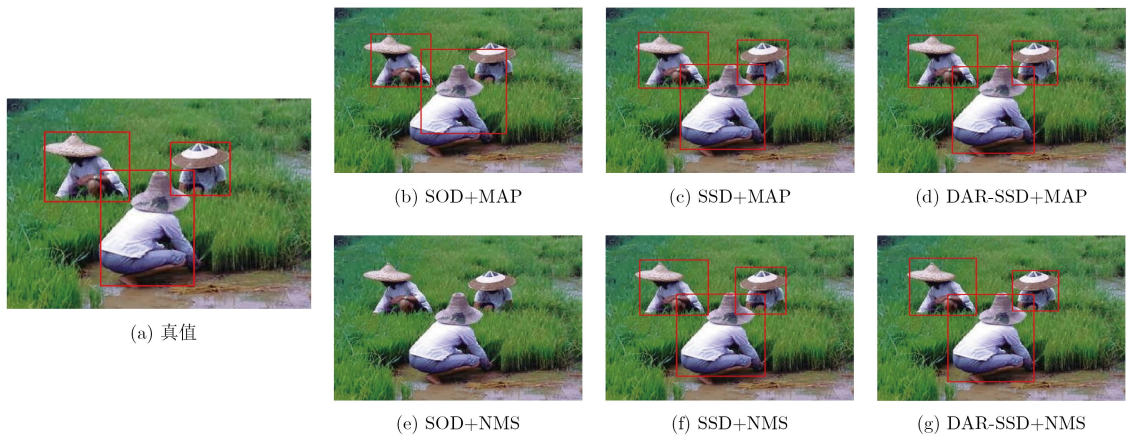


图6 多个显著性目标检测实例

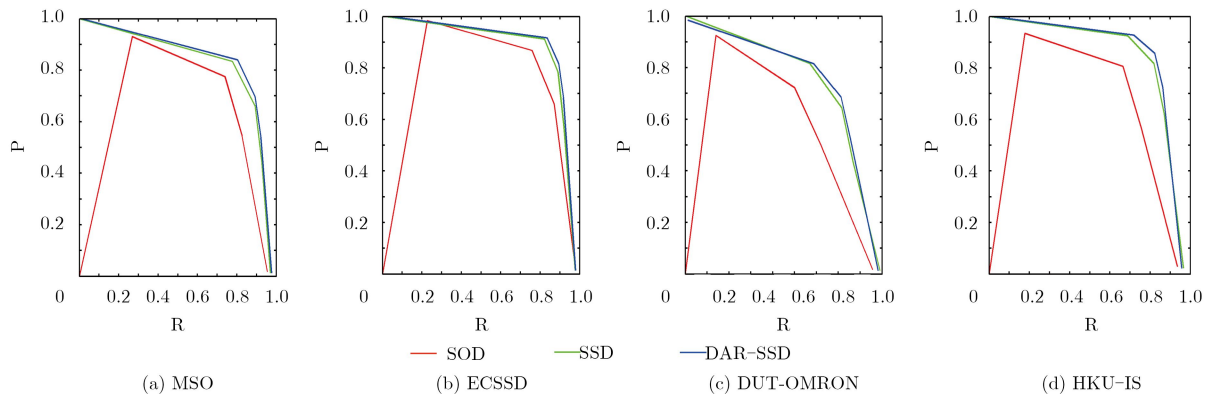


图7 MAP方法下PR曲线对照

SOD更佳的优势所在。

表2和表3中可知，相较于NMS，使用MAP时，SOD性能提升显著，SSD有一定提升，平均提升幅度1%左右，但DAR-SSD提升不明显。由于SSD检测特征图上的每个位置并输出，SOD则对置信度前5个检测框进行了重检测，得到的检测框密集程度远超SSD，所以MAP对SOD的性能提升明显强于其它两个。另外，高噪声时，MAP检测性能优于NMS，当检测框越纯净，包含噪声越少时，MAP的相对优势削弱。而DAR-SSD检测框定位最精确，SSD次之，因此，DAR-SSD下MAP的性能提升最弱，SSD其次。

为了进一步凸显DAR-SSD中DAR模块的改进作用，本文分别就单个、多个、小尺度和大尺度显著性目标的不同情形，进行了AP对照分析，如表4~

表7所示。

单显著性目标情形下，DAR-SSD在MSO和ECSSD上性能与SSD接近，在HKU-IS和DUT-OMRON上优于SSD，NMS下提升1.5%以上。事实上，前两个数据集图像简单，后两个数据集含有大量复杂背景的图片，此时，DAR-SSD通过注意力残差模块引入注意力机制，能够学得更多感兴趣区域相关特征，有助于在复杂背景下更好地检测显著性目标。

多显著性目标情形下，DAR-SSD在4个数据集上均有1.5%的提升；小尺度显著性目标情形下，DAR-SSD性能提升更为明显，在4个数据集上至少都有2.4%的提升。实验结果表明，去卷积模块将高层语义与底层特征融合，有助于提高小尺度对象的检测精度。大尺度显著性目标情形下，DAR-SSD

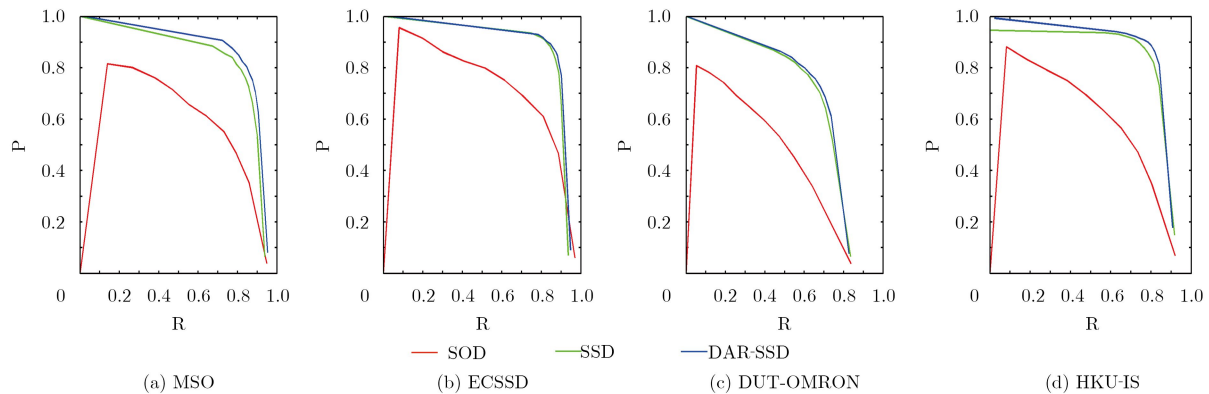


图 8 NMS方法下PR曲线对照

表 2 MAP方法下AP对照

| AP-MAP | MSO | ECSSD | DUT-OMRON | HKU-IS | 平均 |
|---------|--------|--------|-----------|--------|--------|
| SOD | 0.7338 | 0.8152 | 0.5476 | 0.6938 | 0.6976 |
| SSD | 0.8229 | 0.8645 | 0.6698 | 0.8164 | 0.7934 |
| DAR-SSD | 0.8361 | 0.8766 | 0.6774 | 0.8317 | 0.8054 |

表 3 NMS方法下AP对照

| AP-NMS | MSO | ECSSD | DUT-OMRON | HKU-IS | 平均 |
|---------|--------|--------|-----------|--------|--------|
| SOD | 0.6104 | 0.7157 | 0.4409 | 0.5822 | 0.5873 |
| SSD | 0.8120 | 0.8619 | 0.6585 | 0.7974 | 0.7824 |
| DAR-SSD | 0.8387 | 0.8665 | 0.6737 | 0.8256 | 0.8011 |

表 4 单显著性目标情形下AP对照

| AP-one | MSO | ECSSD | DUT-OMRON | HKU-IS |
|-------------|--------|--------|-----------|--------|
| SSD+NMS | 0.8823 | 0.8841 | 0.6908 | 0.8147 |
| DAR-SSD+NMS | 0.8844 | 0.8803 | 0.7052 | 0.8323 |
| SSD+MAP | 0.8881 | 0.8870 | 0.7076 | 0.8483 |
| DAR-SSD+MAP | 0.8877 | 0.8921 | 0.7127 | 0.8571 |

表5 多显著性目标情形下AP对照

| AP-multi | MSO | ECSSD | DUT-OMRON | HKU-IS |
|-------------|--------|--------|-----------|--------|
| SSD+NMS | 0.8207 | 0.7684 | 0.5920 | 0.8122 |
| DAR-SSD+NMS | 0.8583 | 0.8055 | 0.6071 | 0.8305 |
| SSD+MAP | 0.8453 | 0.7882 | 0.6073 | 0.8198 |
| DAR-SSD+MAP | 0.8616 | 0.8149 | 0.6137 | 0.8331 |

表6 小尺度显著性目标情形下AP对照

| AP-small | MSO | ECSSD | DUT-OMRON | HKU-IS |
|-------------|--------|--------|-----------|--------|
| SSD+NMS | 0.7984 | 0.7719 | 0.5734 | 0.7886 |
| DAR-SSD+NMS | 0.8288 | 0.7951 | 0.5980 | 0.8142 |
| SSD+MAP | 0.8044 | 0.7877 | 0.5786 | 0.8056 |
| DAR-SSD+MAP | 0.8310 | 0.8067 | 0.5948 | 0.8128 |

表7 大尺度显著性目标情形下AP对照

| AP-large | MSO | ECSSD | DUT-OMRON | HKU-IS |
|-------------|--------|--------|-----------|--------|
| SSD+NMS | 0.9016 | 0.8936 | 0.8190 | 0.8680 |
| DAR-SSD+NMS | 0.9017 | 0.8888 | 0.8253 | 0.8723 |
| SSD+MAP | 0.9019 | 0.8978 | 0.8330 | 0.8807 |
| DAR-SSD+MAP | 0.9019 | 0.8997 | 0.8306 | 0.8860 |

表8 多种显著性目标检测方法AP对照

| 数据集 | RC | GLGOV | MDF | DCL | DAR-SSD+NMS | DAR-SSD+MAP |
|-----------|-------|-------|-------|-------|-------------|-------------|
| ECSSD | 0.733 | 0.773 | 0.829 | 0.897 | 0.8665 | 0.8766 |
| DUT-OMRON | 0.503 | 0.539 | 0.649 | 0.675 | 0.6737 | 0.6774 |

与SSD间性能无差别的实验结果表明去卷积模块具有小尺度对象敏感性,且不会降低大尺度对象检测性能。

本文还和RC^[16], GLGOV^[17], MDF^[18]和DCL^[19]在ECSSD和DUT-OMRON数据集上进行了AP对照,其中MDF和DCL均属于基于深度学习的方法。如表8所示,与RC和GLGOV传统方法相比,基于深度学习的方法整体性能更好;3种基于深度学习的方法中,DAR-SSD明显优于MDF,在图像简单的ECSSD上略低于DCL,在图像较为复杂的DUT-OMRON上与DCL性能一致,体现了DAR模块中的注意力机制对复杂背景较敏感。

5 结束语

本文以SSD模型为基础,提出了面向多尺度显著性目标检测的DAR-SSD模型,引入去卷积模块融合底层和高层语义特征,改善小尺度对象检测性能;引入注意力残差模块学得更多感兴趣区域相关特征,增加上下文信息,改善定位信度和精准性。主客观对照实验结果表明,DAR-SSD总体检

测性能均较原始SSD, SOD, MDF等均有不同程度地提升,达到DCL等同类基于深度学习方法的性能,在小尺度和复杂背景等情形下具有一定优势。

参考文献

- [1] LIU Feng and GLEICHER M. Automatic image retargeting with fisheye-view warping[C]. Annual ACM Symposium on User Interface Software and Technology, Seattle, USA, 2005: 153–162. doi: [10.1145/1095034.1095061](https://doi.org/10.1145/1095034.1095061).
- [2] VALENTI R, SEBE N, and GEVERS T. Image saliency by isocentric curvedness and color[C]. IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009: 2185–2192. doi: [10.1109/ICCV.2009.5459240](https://doi.org/10.1109/ICCV.2009.5459240).
- [3] LUO Ye, YUAN Junsong, XUE Ping, *et al.* Saliency density maximization for efficient visual objects discovery[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(12): 1822–1834. doi: [10.1109/TCSVT.2011.2147230](https://doi.org/10.1109/TCSVT.2011.2147230).
- [4] FENG Jie, WEI Yichen, TAO Litian, *et al.* Salient object detection by composition[C]. IEEE International Conference on IEEE Computer Vision. Barcelona, Spain, 2011:

- 1028–1035. doi: [10.1109/ICCV.2011.6126348](https://doi.org/10.1109/ICCV.2011.6126348).
- [5] YILDIRIM G and SUSSTRU S. FASA: Fast, accurate, and size-aware salient object detection[C]. Asian Conference on Computer Vision, Singapore, 2014: 514–528. doi: [10.1007/978-3-319-16811-1_34](https://doi.org/10.1007/978-3-319-16811-1_34).
- [6] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al.* Object detection networks on convolutional feature maps[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(7): 1476–1481. doi: [10.1109/TPAMI.2016.2601099](https://doi.org/10.1109/TPAMI.2016.2601099).
- [7] REDMON J, DIVVAL S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779–788. doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [8] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al.* Faster r-cnn: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [9] ZHANG Jianming, SCLAROFF S, LIN Zhe, *et al.* Unconstrained salient object detection via proposal subset optimization[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 5733–5742. doi: [10.1109/CVPR.2016.618](https://doi.org/10.1109/CVPR.2016.618).
- [10] ERHAN D, SZEGEDY C, TOSHEV A, *et al.* Scalable object detection using deep neural networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 2155–2162. doi: [10.1109/CVPR.2014.276](https://doi.org/10.1109/CVPR.2014.276).
- [11] LIU Wei, ANGULEVO D, ERHAN D, *et al.* SSD: Single shot multibox detector[C]. European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 21–37. doi: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [12] LONG J, SHELHAMER E, DARRELL T, *et al.* Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683).
- [13] CAI Zhaowei, FAN Quanfu, FEIRS R S, *et al.* A unified multi-scale deep convolutional neural network for fast object detection[C]. European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 354–370. doi: [10.1007/978-3-319-46493-0_22](https://doi.org/10.1007/978-3-319-46493-0_22).
- [14] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824).
- [15] ZHANG Jianming, MA Shugao, SAMEKI M, *et al.* Salient object subitizing[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 4045–4054. doi: [10.1109/CVPR.2015.7299031](https://doi.org/10.1109/CVPR.2015.7299031).
- [16] CHENG Mingming, MITRA N J, HUANG Xiaolei, *et al.* Global contrast based salient region detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 569–582. doi: [10.1109/TPAMI.2014.2345401](https://doi.org/10.1109/TPAMI.2014.2345401).
- [17] YAN Yijun, REN Jinchang, SUN Genyun, *et al.* Unsupervised image saliency detection with gestalt-laws guided optimization and visual attention based refinement[J]. *Pattern Recognition*, 2018, 7, 9(7): 65–78. doi: [10.1016/j.patcog.2018.02.004](https://doi.org/10.1016/j.patcog.2018.02.004).
- [18] LI Guanbin and YU Yizhou. Visual saliency detection based on multiscale deep CNN features[J]. *IEEE Transaction on Image Processing*, 2016, 25(11): 5012–5024. doi: [10.1109/TIP.2016.2602079](https://doi.org/10.1109/TIP.2016.2602079).
- [19] LI Guanbin and YU Yizhou. Deep contrast learning for salient object detection[C]. 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).

余春艳：女，1976年生，副教授，主要研究方向为智能信息处理、虚拟环境与仿真技术、智能算法等。

徐小丹：女，1992年生，硕士生，研究方向为图像处理。

钟诗俊：男，1994年生，硕士生，研究方向为图像处理。