

• 物联网 •

DOI:10.15961/j.jsuese.202000771



本刊网刊

# 多尺度特征提取和多级别特征融合的显著性目标检测方法

黎玲利<sup>1</sup>, 孟令兵<sup>1</sup>, 李金宝<sup>2\*</sup>

(1. 黑龙江大学 计算机科学技术学院, 黑龙江 哈尔滨 150080; 2. 齐鲁工业大学(山东省科学院) 山东省人工智能研究院, 山东 济南 250014)

**摘 要:**显著性目标检测已经被广泛应用到图像检索、图像分割、行人重识别等领域。目前主流的显著性目标检测方法通常采用短连接加权的方式融合多级别特征信息, 这种方式无法精准有效地控制信息流的传递。而且, 现有的检测方法通常采用单一的特征检测, 导致显著性目标区域与背景的边界不连续、易模糊。因此, 本文提出一种多尺度特征提取和多级别特征融合的显著性目标检测方法。首先, 利用不同扩张率的空洞卷积获取多尺度的上下文信息, 弥补单一特征检测带来的不足。其次, 提出一个多级别特征融合模块, 该模块有效地利用浅层特征信息、深层特征信息和全局上下文特征信息之间的分布特性进行融合, 不仅可以抑制噪声的传递, 而且可以更有效地恢复显著性目标的空间细节结构信息。同时构建一个简洁的注意力模块, 该模块有效保留特征图融合后的通道信息。本文对综合指标、平均绝对误差、结构化度量、精确率-召回率曲线和F-measure曲线进行了实验评估, 在5个公开的数据集上进行的实验结果表明: 相比于其他13种主流的检测方法, 本文方法在不同的评估指标上均有明显的提升, 在4个数据集上的综合指标和结构化度量指标均超过其他方法; 并且, 本文方法的可视化检测的显著图边缘轮廓连续性更好, 空间结构细节信息更清晰。

**关键词:**显著性检测; 多尺度特征提取; 多级别特征融合; 显著图; 深度学习

中图分类号: TP389.1

文献标志码: A

文章编号: 2096-3246(2021)01-0170-08

## Salient Object Detection Based on Multi-scale Feature Extraction and Multi-level Feature Fusion

LI Lingli<sup>1</sup>, MENG Lingbing<sup>1</sup>, LI Jinbao<sup>2\*</sup>

(1. College of Computer Sci. and Technol., Heilongjiang Univ., Harbin 150080, China;

2. Shandong Artificial Intelligence Inst., Qilu Univ. of Technol. (Shandong Academy of Sciences), Ji'nan 250014, China)

**Abstract:** Salient object detection has been widely used in image retrieval, image segmentation, pedestrian recognition and other fields. Current mainstream detection methods fuse multi-level feature information through short connection to add feature maps, which cannot accurately and effectively control the transmission of information flow. In addition, existing salient detection methods usually use single feature detection, which results in discontinuous and fuzzy boundary between the saliency object region and the background. A new salient object detection method based on multi-scale feature extraction and multi-level feature fusion was proposed in this paper. Firstly, the multi-scale context information was obtained by using the dilated convolution of different expansion rates to make up for the deficiencies caused by single feature detection. Secondly, a multi-level feature fusion module was designed, which fuses low-level feature, high-level feature and global context feature information for different distribution characteristics of them. It can not only restrain the transmission of noise, but also restore the spatial detail structure information of the saliency object effectively. At the same time, a concise attention module was constructed, which can effectively retain the channel information after feature map fusion. The F-score, mean absolute error, structural measurement, precision-recall rate curve and F-measure curve have been evaluated experimentally. Experiments on five public datasets show that compared with other thirteen mainstream detection methods, the proposed method in this paper achieves significant improvements in different evaluation indicators, of which the F-score, and structural measurement on four datasets are better than other methods. Meanwhile, the saliency map predicted by the proposed method in this paper has better con-

收稿日期: 2020-09-07

基金项目: 黑龙江省自然科学基金优秀青年项目(YQ2019F016); 黑龙江省自然科学基金项目(ZD2019F003)

作者简介: 黎玲利(1986—), 女, 副教授, 博士。研究方向: 大数据管理与分析。E-mail: lilingli@hlju.edu.cn

\* 通信作者: 李金宝, E-mail: lijnb@sds.org

网络出版时间: 2021-01-13 13:42:41

网络出版地址: <https://kns.cnki.net/kcms/detail/51.1773.TB.20210112.0925.001.html>



为解码阶段的输入。

多尺度特征提取模块如图2所示。首先,输入为经过4次残差块提取后的特征图 $M_m$ ;然后,采用不同扩张因子对其进卷积操作,其中,卷积核设定的大小为 $3 \times 3$ ,扩张因子 $r$ 设置为1、3、5、7;同时,采用自适应平均池化操作,获取到相同分辨率、相同通道数不同尺度的上下文特征信息,将各个尺度的特征信息进行级联,得到最终融合的特征图 $M_c$ 。

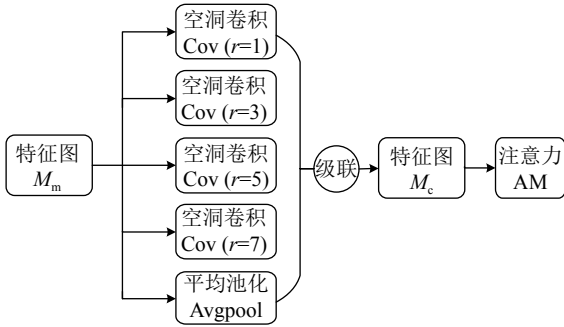


图2 多尺度特征提取模块

Fig. 2 Multi-scale feature extraction module

计算公式如(1)~(3)所示:

$$M_i = \text{Cov}(M_m, r), r \in \{1, 3, 5, 7\} \quad (1)$$

$$M_{ag} = \text{Avgpool}(M_m) \quad (2)$$

$$M_c = C(M_1, M_3, M_5, M_7, M_{ag}) \quad (3)$$

式中,  $\text{Cov}(M_m, r)$ 为特征图 $M_m$ 进行扩张率为 $r$ 的空洞卷积,  $\text{Avgpool}(\cdot)$ 为对特征图 $M_m$ 进行平均池化运算,  $C(\cdot)$ 为对所有的特征图进行级联,  $M_i$ 为不同扩张因子卷积后的特征图,  $M_{ag}$ 为平均池化后的特征图,  $M_c$ 为级联后的特征图。

级联后的特征图具有不同尺度的上下文信息,对级联后的特征图通道关系建模,衡量特征图通道的重要性进而校准通道信息,更加有利于定位显著性目标的区域。校准特征图通道模块如图3所示。首先,对输入的特征图 $M_c$ 进行2次卷积操作分别得到 $M_{c1}$ 、 $M_{c2}$ ;其次,将 $M_{c1}$ 与 $M_{c2}$ 做相乘和相加运算;最后,得到校准后的特征图 $M_{oc}$ 作为解码阶段的输入。

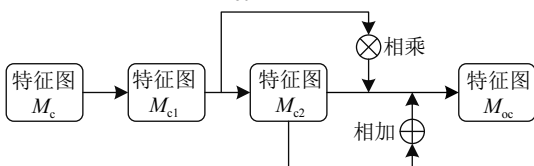


图3 注意力模块

Fig. 3 Attention module

### 1.3 多级别特征融合模块

不同层次特征包含不同信息的分布,浅层的特征包含丰富的空间结构信息,但是其分辨率较高,所以全局上下文信息较为薄弱。深层的特征包含丰富

的语义信息,可以有效地对显著性目标进行精准定位,但是,其分辨率较低,缺少显著性目标的空间细节信息。除此以外,全局上下文特征信息能够从全局的角度推断显著性目标与背景之间的关系,能够突出目标的区域,淡化背景信息的干扰,所以将这些不同级别的特征信息融合可以有效地提高显著性检测的精准度。由于深层的语义特征信息在自顶向下的传递过程中逐步被淡化,使得显著性目标在层层卷积上采样后丢失了高级别语义信息的指导,导致模型检测性能下降。因此,本文在每层卷积进行上采样时将融合的浅层特征信息和深层特征信息都加入全局上下文特征信息,可以弥补深层的语义信息的淡化,有效抑制背景信息的干扰,并在每层卷积都能实现对显著性目标的精准定位。

多级别特征融合模块如图4所示。首先,由上述多尺度模块提取的特征图经过全局平均池化(GAP)得到全局上下文特征信息;然后,校准全局上下文特征图通道信息,生成具有全局信息的掩码;浅层特征经过卷积运算与具有全局信息的掩码相乘,再经过卷积运算得到特征图的输出。全局上下文特征信息与浅层特征信息的融合,弥补了高级语义信息的淡化,同时,可以抑制浅层的背景噪声,更精准地实现显著性目标的定位。与上述融合方法类似,浅层特征信息经过卷积运算生成对应的掩码,并与深层特征信息生成的掩码相乘;同时,深层特征信息经过卷积运算生成对应的掩码,再与浅层特征信息生成的掩码相乘;由此实现浅层特征信息和深层特征信息的互补,可以有效地利用二者之间的有用信息生成精准的掩码;并通过融合操作将这些特征信息进行级联;最后,应用 $3 \times 3$ 的卷积运算得到特征图 $M_1$ 。此外,同时级联浅层特征信息、深层特征信息和全局上下文特征信息;之后,采用 $3 \times 3$ 的卷积运算,进一步校准特征图通道信息;再经过 $3 \times 3$ 的卷积运算输出特征图 $M_2$ 。将输出的2种不同级联方法的特征图 $M_1$ 、 $M_2$ 相加,再次校准融合后特征图通道信息,得到最终输出的特征图,将其作为输入进入下一阶段的解码过程。由此经过层层的上采样逐步生成高分辨率的显著图。整个过程计算公式如(4)~(9)所示:

$$M_g = \text{Cov}(M_{low}) \times \text{Mask}(M_{global}) \quad (4)$$

$$M_l = \text{Cov}(M_{high}) \times \text{Mask}(M_{low}) \quad (5)$$

$$M_h = \text{Mask}(M_{high}) \times \text{Cov}(M_{low}) \quad (6)$$

$$M_a = C(M_g, M_l, M_h) \quad (7)$$

$$M_b = C(\text{Cov}(M_{global}, M_{low}, M_{high})) \quad (8)$$

$$M_w = \text{Add}(\text{Cov}(M_a, M_b)) \quad (9)$$

式中:  $M_{low}$  为浅层特征信息,  $M_{high}$  为深层特征信息,  $M_{global}$  为全局上下文特征信息,  $M_g$  为全局上下文信息生成的掩码与浅层特征融合的结果,  $M_l$  为浅层特征生成的掩码与深层特征融合的结果,  $M_h$  为深层特征生成的掩码与浅层特征融合的结果,  $Cov(\cdot)$  为卷积运算,  $Mask(\cdot)$  为不同级别特征生成对应的掩码,  $Add(\cdot)$  为特征图对应元素相加,  $M_a$  为  $M_g$ 、 $M_l$  和  $M_h$  级联后的特征图,  $M_b$  为  $M_{global}$ 、 $M_l$  和  $M_h$  经过卷积后级联的特征图,  $M_w$  为  $M_a$  和  $M_b$  相加后的特征图。

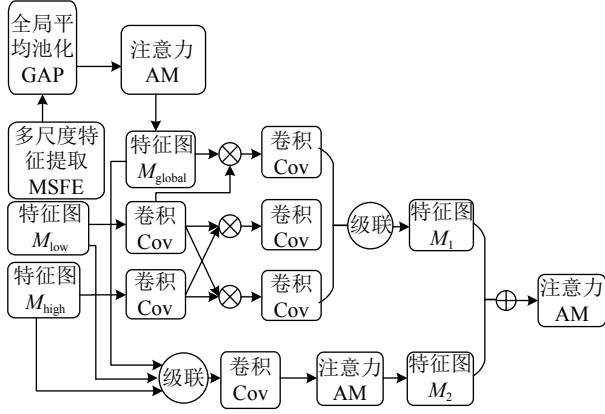


图4 多级别特征融合模块

Fig. 4 Multi-level feature fusion module

## 2 实验与结果

### 2.1 数据集

为了验证本文方法的有效性,在5个公开的数据集上进行了评估,分别是DUTS<sup>[17]</sup>、ECSSD<sup>[18]</sup>、DUT-OMRON<sup>[19]</sup>、HKU-IS<sup>[20]</sup>和PASCAL-S<sup>[21]</sup>。其中: DUTS数据集包括10 553张训练图像和5 019张测试图像,是目前为止最大的显著性目标检测数据集。HKU-IS数据集包含4 447张图像,这些图像大多数存在不连续的显著性目标,前景与背景较相似,并且整张图像的对比度较低。DUT-OMRON数据集包含5 168张自然图像,每张图像几乎都有复杂的背景信息,是最具有挑战性的数据集。ECSSD数据集包含1 000张具有复杂空间结构的图像。PASCAL-S数据集包含850张图像。每个数据集都是由原图和与之对应的真值图组成。本文采用DUTS的10 553张图片作为训练集,使用目前普遍采用的训练策略<sup>[22-25]</sup>,水平翻转训练图像用于扩充数据集,防止模型过拟合。采用Adam优化器,初始学习率设置为0.000 1,设置每2轮迭代更改学习率,其中,衰减率设置为0.9。设置数据批次大小为18,实验的显卡是RTX 2080 Ti,采用深度学习框架pytorch和python语言在Linux系统完成本文实验。

### 2.2 评估指标

本文采用F-measure、平均绝对误差和S-measure

作为评价指标。F-measure是衡量模型的整体指标,由精确率和召回率共同计算得出:

$$F_\beta = \frac{(1 + \beta^2) \times N_{\text{precision}} \times N_{\text{recall}}}{\beta^2 \times N_{\text{precision}} + N_{\text{recall}}} \quad (10)$$

式中:  $N_{\text{precision}}$  为精确率,即正确检测的显著像素的数量与检测到的显著像素总数之比,反映检测准确性;  $N_{\text{recall}}$  为召回率,即正确检测的显著像素的数量与真值图中显著像素的总数比值,反映检测全面性;通常情况下,为突出方法检测的精确率,设置  $\beta^2 = 0.3$ 。本文选取精确率-召回率(PR)曲线计算的最大值作为F值指标,记为  $F_{\text{max}}$ 。精确率和召回率计算公式为:

$$N_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (11)$$

$$N_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (12)$$

式中,  $N_{\text{TP}}$  为检测出正确的像素,  $N_{\text{FP}}$  为检测出不正确的像素,  $N_{\text{FN}}$  为未检测出正确的像素。

平均绝对误差(MAE)是检测的显著图与人工标注的真值图平均绝对误差,计算公式如(13)所示:

$$N_{\text{MAE}} = \frac{1}{H \times W} \sum_{x=1}^H \sum_{y=1}^W |P(x, y) - G(x, y)| \quad (13)$$

式中,  $P$  和  $G$  为检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,  $x$  和  $y$  为像素点的横纵坐标。平均绝对误差越小,代表检测的显著图与真值图越接近,方法越好。

S-measure是目标感知和区域感知的结构相似性,计算公式如(14)所示:

$$S_m = \alpha S_o + (1 - \alpha) S_r \quad (14)$$

式中,  $\alpha$  通常设置为0.5,  $S_o$  为目标感知,  $S_r$  为区域感知。S-measure越大,表示检测的显著图与真值图在空间结构上越相似。

### 2.3 对比实验

将本文方法与13种典型的显著性检测方法进行了比较,包括ABMP<sup>[6]</sup>、AFNet<sup>[11]</sup>、BASNet<sup>[26]</sup>、CPD-R<sup>[10]</sup>、F3Net<sup>[4]</sup>、ITSD<sup>[9]</sup>、MINet<sup>[3]</sup>、MLMSNet<sup>[12]</sup>、PAGR<sup>[13]</sup>、PICA-R<sup>[8]</sup>、SMJD<sup>[14]</sup>、U2Net<sup>[5]</sup>、HRSODT<sup>[7]</sup>。各种显著性检测方法在5个测试集的结果如表1所示。由表1可知:在ECSSD数据集上,本文方法在3个评价指标上都得到最优的效果;在HKU-IS数据集上,本文方法的  $F_{\text{max}}$ 、 $S_m$  比F3Net和MINet分别提高了0.004、0.006;在PASCAL-S数据集上,本文方法的  $F_{\text{max}}$ 、 $S_m$  比F3Net分别提高了0.001、0.004。在4个数据集(除DUT-OMRON)中,本文方法的  $F_{\text{max}}$  和  $S_m$  明显高于其他对比方法,但是,在DUT-OMRON数据集上本文方法的

$F_{\max}$ 、 $S_m$  低于其他对比方法。这是由于该数据集具有极其复杂的背景, 其他方法需要设计强大的特征提取网络和多损失联合优化的策略用于检测, 虽然检测效果优于本文方法, 但计算成本较高。因此, 综合所有数据集和评价指标, 本文方法要明显优于其他

对比方法。此外, 本文还给出5个数据集的PR曲线、F-measure曲线和检测的显著图, 从客观评价指标和主观感受共同衡量不同模型检测的效果。

图5为本文方法与其他检测方法的PR曲线。其中, PR曲线包围的面积越大, 说明该方法的效果越好。

表 1 不同模型的测试结果

| 方法      | DUT-OMRON  |       |       | DUTS-TE    |       |       | HKU-IS     |       |       | ECSSD      |       |       | PASCAL-S   |       |       |
|---------|------------|-------|-------|------------|-------|-------|------------|-------|-------|------------|-------|-------|------------|-------|-------|
|         | $F_{\max}$ | MAE   | $S_m$ | $F_{\max}$ | MAE   | $S_m$ | $F_{\max}$ | MAE   | $S_m$ | $F_{\max}$ | MAE   | $S_m$ | $F_{\max}$ | MAE   | $S_m$ |
| MSML    | 0.818      | 0.054 | 0.843 | 0.891      | 0.037 | 0.893 | 0.941      | 0.029 | 0.925 | 0.953      | 0.032 | 0.933 | 0.873      | 0.063 | 0.858 |
| U2Net   | 0.823      | 0.054 | 0.847 | 0.872      | 0.045 | 0.873 | 0.935      | 0.031 | 0.916 | 0.951      | 0.033 | 0.928 | 0.859      | 0.073 | 0.838 |
| ITSD    | 0.821      | 0.061 | 0.840 | 0.883      | 0.041 | 0.885 | 0.934      | 0.031 | 0.917 | 0.947      | 0.034 | 0.925 | 0.870      | 0.065 | 0.853 |
| MINet   | 0.810      | 0.055 | 0.833 | 0.884      | 0.037 | 0.884 | 0.935      | 0.029 | 0.919 | 0.947      | 0.033 | 0.925 | 0.867      | 0.063 | 0.850 |
| F3Net   | 0.813      | 0.053 | 0.838 | 0.891      | 0.035 | 0.888 | 0.937      | 0.028 | 0.917 | 0.945      | 0.033 | 0.924 | 0.872      | 0.061 | 0.854 |
| AFNet   | 0.797      | 0.057 | 0.826 | 0.863      | 0.046 | 0.867 | 0.923      | 0.036 | 0.905 | 0.935      | 0.042 | 0.916 | 0.863      | 0.070 | 0.848 |
| BASNet  | 0.805      | 0.056 | 0.836 | 0.859      | 0.048 | 0.866 | 0.928      | 0.032 | 0.909 | 0.942      | 0.037 | 0.918 | 0.854      | 0.075 | 0.832 |
| CPD-R   | 0.797      | 0.056 | 0.825 | 0.865      | 0.043 | 0.869 | 0.925      | 0.034 | 0.905 | 0.939      | 0.037 | 0.911 | 0.859      | 0.070 | 0.842 |
| MLMSNet | 0.774      | 0.064 | 0.809 | 0.852      | 0.049 | 0.862 | 0.921      | 0.039 | 0.907 | 0.929      | 0.044 | 0.914 | 0.850      | 0.073 | 0.838 |
| SMJD    | 0.802      | 0.057 | 0.821 | 0.832      | 0.059 | 0.836 | 0.920      | 0.039 | 0.903 | 0.928      | 0.049 | 0.907 | 0.854      | 0.082 | 0.835 |
| HRSODT  | 0.743      | 0.065 | 0.762 | 0.835      | 0.050 | 0.824 | 0.910      | 0.042 | 0.877 | 0.925      | 0.052 | 0.888 | 0.846      | 0.079 | 0.810 |
| ABMP    | 0.774      | 0.064 | 0.809 | 0.851      | 0.049 | 0.862 | 0.921      | 0.039 | 0.907 | 0.928      | 0.045 | 0.911 | 0.850      | 0.073 | 0.838 |
| PICA-R  | 0.803      | 0.065 | 0.832 | 0.860      | 0.051 | 0.869 | 0.918      | 0.043 | 0.904 | 0.935      | 0.046 | 0.917 | 0.857      | 0.074 | 0.848 |
| PAGR    | 0.771      | 0.071 | 0.775 | 0.854      | 0.055 | 0.838 | 0.918      | 0.048 | 0.887 | 0.927      | 0.061 | 0.889 | 0.847      | 0.089 | 0.815 |

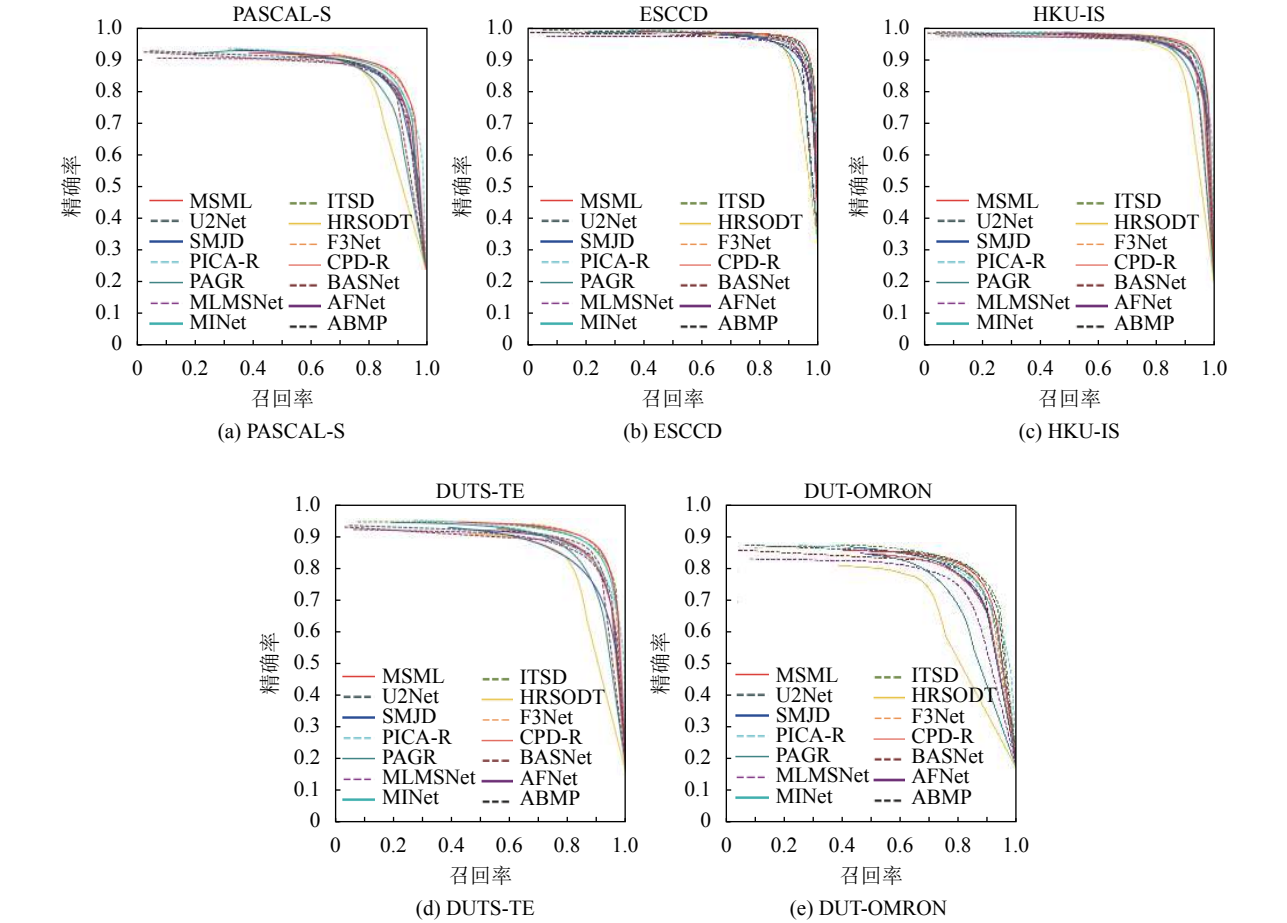


图 5 5个数据集的PR曲线

Fig. 5 Precision–Recall curves on five common saliency datasets

从图5中可以看出: 在PASCAL-S、ECSSD、HKU-IS和DUTS-TE数据集上, 本文方法的PR曲线都在其他方法的上方, 包围的面积大于其他方法, 说明本文方法的检测性能优于其他方法。在DUT-OMRON数据集上, 本文方法也优于大部分对比方法的检测效果。本文给出了F-measure曲线, 如图6所示。其中, F-

measure是精确率和召回率的调和平均数, 反映出显著性检测的综合指标。由图6可知, 在PASCAL-S、ECSSD、HKU-IS和DUTS-TE数据集中本文方法得出了更均衡的结果, 具有更好的泛化能力。图7为本文方法与其他方法的检测显著图对比, 其中, GT为真实标签。

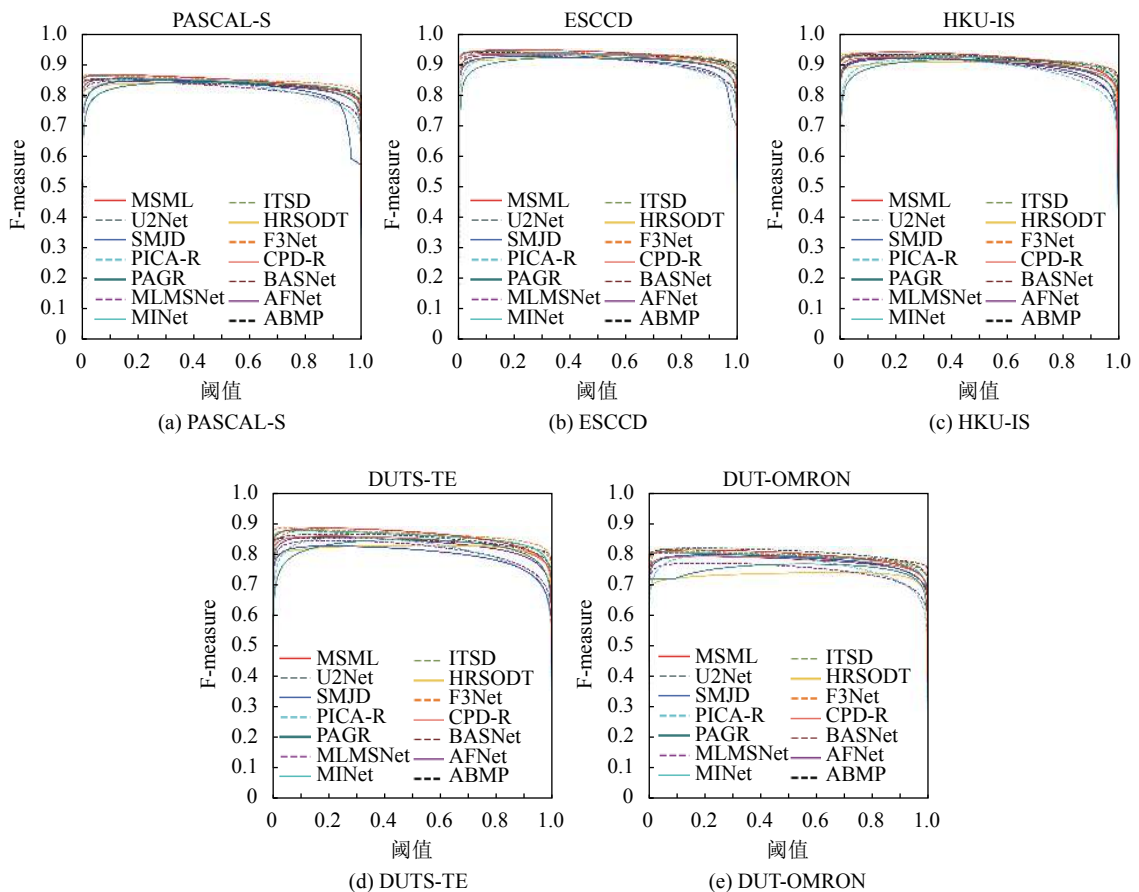


图 6 5个数据集的F-measure曲线

Fig. 6 F-measure curves on five common saliency datasets

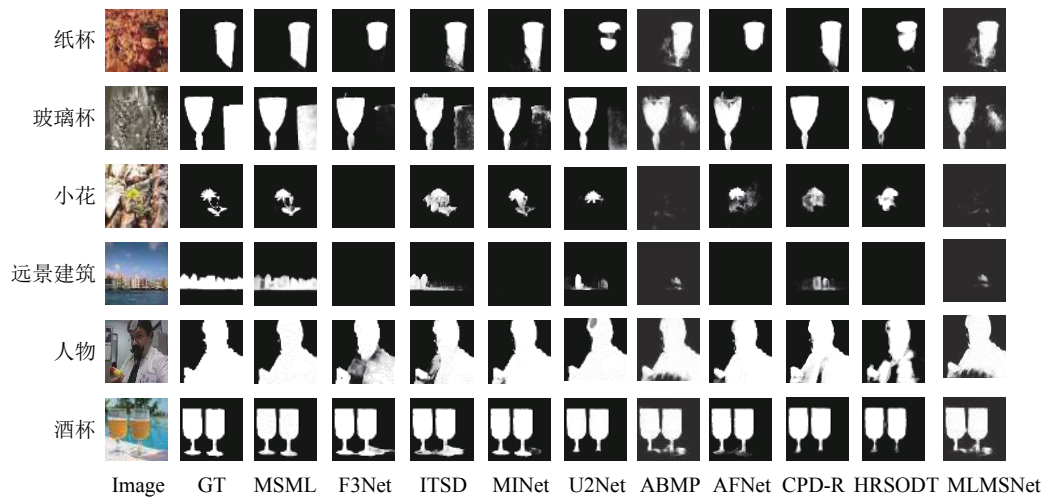


图 7 不同模型的可视化结果

Fig. 7 Comparison of visual results of different methods

由图7可知:从第5行(人物)和第6行(酒杯)结果可以看出,其他方法虽然能够较准确地检测出显著性目标区域,目标区域内部也比较均匀,但是目标区域与背景边界较为模糊,而本文方法能够准确分割出目标区域与背景边界。对于第2行(透明玻璃)、第3行(小花)、第4行(远景建筑)的图像,本文方法也能够完整准确地检测出目标区域,并且检测结果较光滑和均匀,而其他方法大多数方法得检测效果较差,甚至无法检测出显著性区域。实验结果表明,本文提出的多尺度特征提取模块有效地缓解了显著性目标与背景不连续、边界易模糊等问题,同时,本文设计的多级别特征融合模块,在抑制背景噪声的同时,能够使检测的显著性目标更完整、更均匀。综合来看,本文方法的视觉效果更好,检测结果的显著图纹理清晰、边界轮廓明显。

2.4 消融实验

为了验证模块的有效性,在PASCAL-S和ECSSD数据集上进行模块的消融实验,结果如表2所示。其中:BL 即Baseline,是原始的U-net网络架构;MSFE表示本文的多尺度特征提取模块;MLFF表示本文的多级别特征融合模块;no AM表示没有加入注意力模块。

从表2中可以看出:在PASCAL-S和ECSSD数据集上,对比于Baseline,本文的多尺度特征提取模块(MSFE)使 $F_{max}$ 分别提高了0.01、0.013,MAE分别下降了0.008、0.007;本文的多级别特征融合模块(MLFF)使 $F_{max}$ 分别提高了0.015、0.015,MAE分别下降了0.008、0.010。同时使用本文方法的多尺度特征提取模块和多级别特征融合模块在 $F_{max}$ 、MAE指标上达到了最优的效果。因此,本文提出的多尺度特征提取模块和多级别特征融合模块可以有效地提高检测的性能。

表 2 在2个数据集上的消融实验结果

Tab. 2 Ablation experiments results on two datasets

| 方法                  | PASCAL-S  |       | ECSSD     |       |
|---------------------|-----------|-------|-----------|-------|
|                     | $F_{max}$ | MAE   | $F_{max}$ | MAE   |
| BL                  | 0.853     | 0.073 | 0.935     | 0.042 |
| BL+MSFE             | 0.863     | 0.065 | 0.948     | 0.035 |
| BL+MLFF             | 0.868     | 0.065 | 0.950     | 0.032 |
| BL+MSFE+MLFF(no AM) | 0.870     | 0.063 | 0.951     | 0.032 |
| BL+MSFE+MLFF        | 0.873     | 0.063 | 0.953     | 0.032 |

3 结 论

本文提出了一种多尺度特征提取和多级别特征融合的显著性检测方法。采用不同扩张率的空洞卷积获取丰富的上下文信息,缓解显著性目标区域与

背景不连续、边界易模糊等问题。本文提出的多级别特征融合模块,不仅可以抑制噪声的传递,而且可以有效地恢复显著性目标空间结构的细节信息。从定性评价和定量评价的角度评估,本文提出的方法都有显著的提升。在下一步工作中,将采用特征提取能力更强的主干网络提高算法的精度,继续优化网络减少模型的参数量。

参考文献:

[1] Zhao Rui,Ouyang Wanli,Li Hongsheng,et al.Saliency detection by multi-context deep learning[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition.Boston:IEEE,2015:1265–1274..

[2] Lee Gayoung,Tai Yuwing,Kim Junmo.Deep saliency with encoded low level distance map and high level features[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition.Las Vegas:IEEE,2016:660–668.

[3] Pang Youwei,Zhao Xiaopi,Zhang Lihe,et al.Multi-scale interactive network for salient object detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Seattle:IEEE,2020:9140–9149.

[4] Wei Jun,Wang Shuhui,Huang Qingming.F<sup>3</sup>Net:Fusion, feedback and focus for salient object detection[C]//Proceedings of the AAAI Conference on Artificial Interlligence.New York:AAAI,2020:12321–12328.

[5] Qin Xuebin,Zhang Zichen,Huang Chenyang,et al.U<sup>2</sup>-Net: Going deeper with nested U-structure for salient object detection[J].Pattern Recognition,2020,106:1207404.

[6] Liu Nian,Han Junwei,Yang Minghsuan.PiCANet:Learning pixel-wise contextual attention for saliency detection[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Salt Lake City:IEEE,2018: 3089–3098.

[7] Zeng Yi,Zhang Pingping,Lin Zhe,et al.Towards high-resolution salient object detection[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul:IEEE,2019:7233–7242.

[8] Zhang Lu,Dai Ju,Lu Huchuan,et al.A bi-directional message passing model for salient object detection[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Salt Lake City:IEEE,2018:1741–1750.

[9] Feng Mengyang,Lu Huchuan,Ding Errui.Attentive feedback network for boundary-aware salient object detection[C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Long Beach:IEEE,2019:1623–1632.

[10] Wu Zhe,Su Li,Huang Qingming.Cascaded partial decoder for fast and accurate salient object detection[C]//Proceed-

- ings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [Long Beach:IEEE](#),2019:3907–3916.
- [11] Zhou Huajun,Xie Xiaohua,Lai Jianhuang,et al.Interactive two-stream decoder for accurate and fast saliency detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [Seattle:IEEE](#),2020:9138–9147.
- [12] Wu Runmin,Feng Mengyang,Guan Wenlong,et al.A mutual learning method for salient object detection with intertwined multi-supervision[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [Long Beach:IEEE](#),2019:8150–8159.
- [13] Zhang Xiaoning,Wang Tiantian,Qi Jinqing,et al.Progressive attention guided recurrent network for salient object detection[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [Salt Lake City:IEEE](#),2018:714–722.
- [14] Xu Yingyue,Xu Dan,Hong Xiaopeng,et al.Structured modeling of joint deep feature and prediction refinement for salient object detection[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. [Seoul:IEEE](#), 2019:3789–3798.
- [15] Li Xuelong,Song Dawei,Dong Yongsheng.Hierarchical feature fusion network for salient object detection[J].[IEEE Transactions on Image Processing](#),2020,29:9165–9175.
- [16] Chen Shuhan,Tan Xiuli,Wan Ben,et al.Reverse attention-based residual network for salient object detection[J].[IEEE Transactions on Image Processing](#),2020,29:3763–3776.
- [17] Wang Lijun,Lu Huchuan,Wang Yifan,et al.Learning to detect salient objects with image-level supervision[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. [Honolulu:IEEE](#),2017:3796–3805.
- [18] 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:IEEE](#),2013:3166–3173.
- [19] Li Guanbin,Yu Yizhou.Visual saliency based on multiscale deep features[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. [Boston:IEEE](#), 2015:5455–5463.
- [20] Yan Qiong,Xu Li,Shi Jianping,et al.Hierarchical saliency detection[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. [Portland:IEEE](#), 2013:1155–1162.
- [21] Li Yin,Hou Xiaodi,Koch C,et al.The secrets of salient object segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. [Columbus:IEEE](#),2014:280–287.
- [22] Liu Jiangjiang,Hou Qibin,Cheng Mingming.Dynamic feature integration for simultaneous detection of salient object,edge,and skeleton[J].[IEEE Transactions on Image Processing](#),2020,29:8652–8667.
- [23] Mohammadi S,Noori M,Bahri A,et al.CAGNet:Content-aware guidance for salient object detection[J].[Pattern Recognition](#),2020,103:107303.
- [24] Li Haofeng,Li Guanbin,Yu Yizhou.ROSA:Robust salient object detection against adversarial attacks[J].[IEEE Transactions on Cybernetics](#),2019,50(11):4835–4857.
- [25] Zeng Yu,Feng Mengyang,Lu Huchuan,et al.An unsupervised game-theoretic approach to saliency detection[J].[IEEE Transactions on Image Processing](#),2018,27(9):4545–4554.
- [26] Qin Xuebin,Zhang Zichen,Huang Chenyang,et al.BASNet: Boundary-aware salient object detection[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [Long Beach:IEEE](#),2019:7479–7489.

(编辑 赵 婧)

引用格式: Li Lingli,Meng Lingbing,Li Jinbao.Salient object detection based on multi-scale feature extraction and multi-level feature fusion[J].Advanced Engineering Sciences,2021,53(1):170–177.[黎玲利,孟令兵,李金宝.多尺度特征提取和多级别特征融合的显著性目标检测方法[J].工程科学与技术,2021,53(1):170–177.]