

双注意力循环卷积显著性目标检测算法

谢学立, 李传祥, 杨小冈, 席建祥*

火箭军工程大学导弹工程学院, 陕西 西安 710025

摘要 显著性目标检测是机器视觉领域的研究热点, 具有广泛的应用前景。针对现有显著性目标检测算法存在的显著区域检测不均匀、边缘表示模糊等问题, 提出一种双注意力循环卷积显著性目标检测算法。在 U-Net 全卷积骨干网络中添加像素间-通道间双注意力模块, 在跨层连接前对底层特征进行预处理, 减小噪声和杂波干扰, 提高显著区域检测性能。在骨干网络后端使用循环卷积模块, 将最后的预测图与底层卷积层特征进一步结合, 增强预测区域边缘的表示效果。在三个公开数据集上进行实验评测, 并与相关算法进行对比, 结果表明所提算法能更好地均匀突显显著区域和细化区域边缘。

关键词 机器视觉; 显著性目标检测; 全卷积神经网络; 注意力机制; 循环卷积网络

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/AOS201939.0915005

Salient Object Detection Algorithm Based on Dual-Attention Recurrent Convolution

Xie Xueli, Li Chuanxiang, Yang Xiaogang, Xi Jianxiang*

College of Missile Engineering, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China

Abstract Salient object detection has attracted considerable attention in the field of the machine vision, with a wide range of applications. This study proposes a salient object detection algorithm based on the dual-attention recurrent convolution to overcome the limitations associated with the existing algorithms, i. e. , uneven salient region detection and fuzzy edge representations. A dual-attention module consisting of pixel- and channel-wise attentions is added to a backbone U-Net fully convolutional network to preprocess the shallow convolutional features before skip-layer connection, and reduce noise and clutter interference. This improves its salient region detection performance. Then, following the backbone network, a recurrent convolutional module enhances the edge representation of the prediction region by combining the final prediction map with the shallow convolutional features. The results of experiments on three open datasets show that the proposed algorithm is better able to highlight salient regions and refine their edges than other correlation algorithms.

Key words machine vision; salient object detection; fully convolutional neural network; attention mechanism; recurrent convolutional network

OCIS codes 150.0155; 140.1135; 110.2970; 100.3008

1 引 言

人类的视觉注意力机制能帮助人眼在复杂环境中快速定位最感兴趣的目标或区域。在机器视觉领域, 常使用显著性目标检测算法模拟人类这一视觉特性: 使用显著性目标检测算法对图像像素点进行显著度计算, 突出图像前景区域像素, 抑制背景区域像素, 最终达到数据降维和减少背景干扰的目的。生成的显著图将有助于合理分配有限的计算资源,

为后续复杂的视觉任务提供先验信息。显著性目标检测被广泛应用于图像检索^[1]、图像/视频压缩^[2]、图像质量评估^[3]和虚拟视觉^[4]等领域。

传统的显著性目标检测算法主要依靠人工提取的特征来计算区域显著性。Itti 等^[5]最先通过融合图像的颜色、方向、强度等浅层特征来度量图像显著区域; 刘冬梅等^[6]利用非下采样轮廓小波变换分解的高低频分量进行“由粗到精”的显著性检测; 李德峰等^[7]利用区域协方差理论融合多尺度空间特征,

收稿日期: 2019-03-11; 修回日期: 2019-04-11; 录用日期: 2019-05-23

基金项目: 国家自然科学基金(61503012, 61703411, 61503009)

* E-mail: xijx07@mails.tsinghua.edu.cn

再结合全局核密度估计实现全局和局部特征融合估计目标显著性。上述算法在背景简单、对比度高的情况下能有效反映出目标的显著性,但由于未能利用深层语义的特征,检测结果的稳健性较差,显著区域突显不充分,算法泛化能力较弱。

近年来,卷积神经网络(CNN)被广泛应用于机器视觉领域,其具有自动提取图像特征的能力,通过堆叠卷积块,可以获得丰富的高层次特征。刘峰等^[8]利用线性拟合将卷积网络全局预测模型和区域特征描述子优化模型的结果进行融合,生成高质量显著图;张松龙等^[9]首先密集地融合尺度内卷积特征输出各尺度显著图,再融合多尺度显著图生成最终显著图;Hou等^[10]在整体嵌套边缘检测(HED)网络^[11]基础上提出了一种特征“自顶向下”传播的全卷积神经网络架构。卷积神经网络能自动提取丰富的图像特征,包括浅层的结构特征和高层的语义特征,采用监督学习的方式

对特征进行筛选、组合,增强了特征表示能力。特别是全卷积神经网络的使用,极大提高了显著性目标检测的性能。但仍存在特征利用效益不足,显著目标区域突出不均匀,背景错误检测和目标轮廓模糊等问题。

本文提出一种双注意力循环卷积网络显著性目标检测算法,选取 U-Net^[12]作为骨干网络,设计像素间-通道间双注意力模块(DAM),增强特征利用效益,减弱噪声和背景像素干扰;设计一种循环卷积模块(RCM),通过循环迭代细化显著区域边缘轮廓;使用旁路输出策略对预测结果进行多阶段约束。

2 基本原理

所提检测算法模型主要包括全卷积骨干网络、双注意力模块和循环卷积模块。采用跨层连接策略融合深层语义特征和浅层结构特征。算法总体结构如图 1 所示。

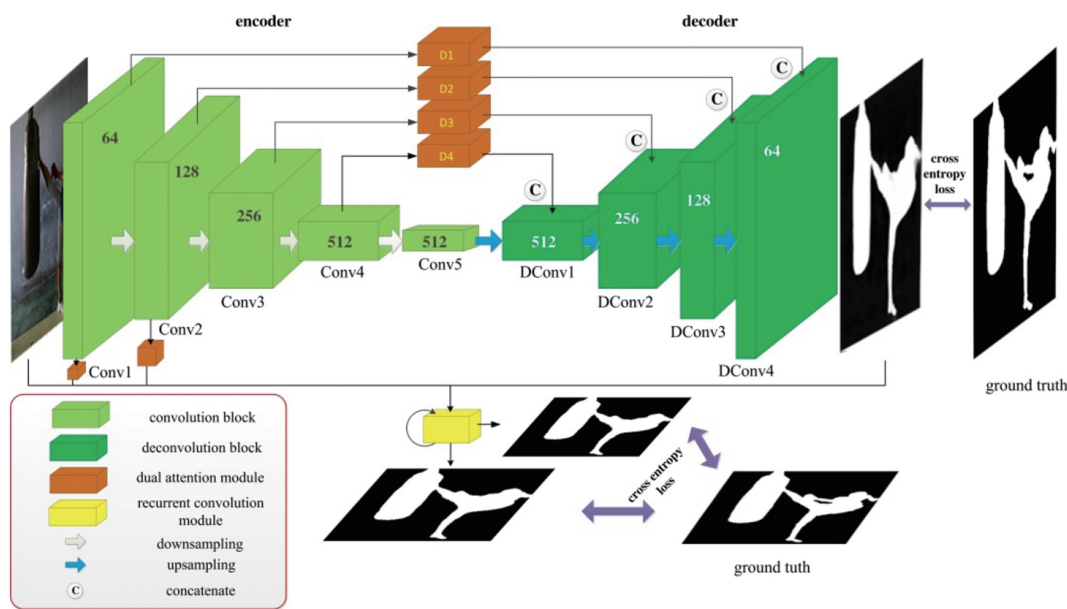


图 1 算法总体结构图

Fig. 1 Overall structure of algorithm

2.1 基于 U-Net 的全卷积骨干网络

文献^[13]的实验结果表明只利用最后一层语义特征进行解码生成的图像具有严重“马赛克”现象,这是因为编码器中最大池化层会使图像信息损失,解码器不能完全还原细节信息。卷积神经网络中深层特征学习更多的语义信息,浅层特征描述图像局部结构特征,通过跨层连接将浅层细节特征与深层语义特征进行融合,能提高编码器预测图的细节表征能力。U-Net 是图像分割领域的经典网络模型,其采用对称“金字塔形”的编码器-解码器结构进行

逐像素点预测,编码器与解码器之间采用跨层连接的方式进行多层次、多尺度特征融合,用于实现目标区域的准确分割。

本研究采用基于 U-Net 的改进全卷积神经网络作为特征提取骨干网络,用于实现显著性目标检测。编码器采用五层卷积块提取输入图像的高层语义特征,用于描述图像目标的显著性。其中卷积块由两个包含 3×3 卷积层、BN(batch normalization)层和 ReLU 激活层的子卷积模块组成。卷积块之间使用滑步为 2 的最大池化层进行下采样,增大深

层特征感受野,确保特征描述具有一定的全局性。对每次下采样后的特征通道进行翻倍,最后的下采样通道不进行加倍。解码器采用反卷积块对高层语义特征进行上采样。反卷积块具有卷积块相同对卷积子模块,只是模块之间采用反卷积层进行上采样。上采样后通道数减半,并与对应的具有相同分辨率的编码器特征进行通道堆叠(channel concatenate)。

2.2 双注意力模块

前景区域和背景区域对显著性目标检测的影响不同,各特征通道对预测结果的贡献度存在差异。U-Net 直接采用跨层连接将编码端特征与解码端特征进行 concatenate 融合,该操作会引入大量的杂波,干扰最后预测图生成。针对此问题,引入注意力

机制对编码器特征进行预加权处理,过滤有害杂波干扰。以往的注意力机制只片面考虑通道间或像素间特征的影响,未能全面地对整个特征图中各像素点的重要度进行衡量。而显著性目标检测是通过逐像素点预测来对显著性目标区域进行分割,对各像素点的质量要求较高。基于此设计中的双注意力模块,在融合编码端特征与解码端特征之前,先将编码端特征接入设计的通道间-像素间双注意力模块,标定通道间响应权重,突出前景区域像素响应强度,增强任务相关特征表征能力,弱化背景和噪声影响。为避免由于多级注意力模块串联造成的特征信息弱化与损失,双注意力模块采用并联方式融合两类注意力。双注意力模块内部结构如图 2 所示。

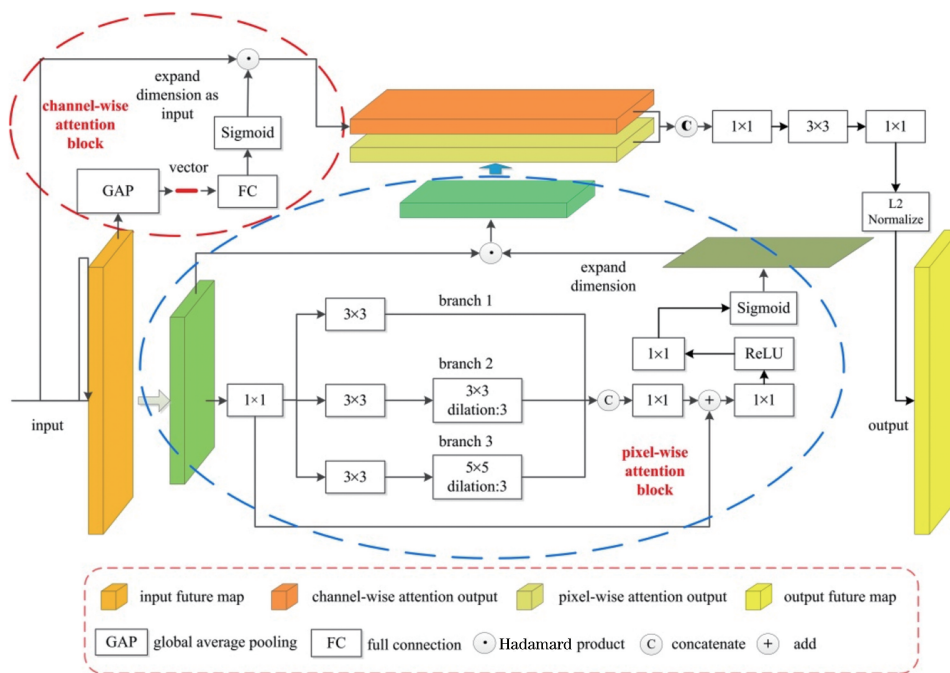


图 2 双注意力模块内部结构图

Fig. 2 Internal structure of DAM

2.2.1 像素间注意力模块

显著性目标一般存在于图像某个区域,该区域像素点对目标显著值预测的影响大于其他像素点^[14]。引入像素间注意力模块,为图像各像素点赋予权值,更多地关注前景区域像素,有助于生成有效的特征,减弱非显著区域的干扰。

文献[15]使用多尺度卷积滤波器设计上下文重加权网络,对输入特征图进行像素间加权处理,增强地建建筑区域像素响应,提高了图像地理定位精度。文献[16]利用感受野模块(RFB)模拟人类视觉感受野特性。聚合了多尺度特征信息,提升了目标检测准确率。本研究基于文献[15],参考 RFB 设计思

想,设计一种多尺度上下文重加权网络(MSCRN),作为像素间注意力模块来实现特征图像素间注意力预测,如图 2 中间虚线椭圆区域所示。

基于 Inception-ResNet^[17]结构设计一种多尺度上下文信息聚合模块,构建三路分支,分别为 3×3 卷积、 3×3 卷积加比率为 3 的 3×3 空洞卷积、 3×3 卷积加比率为 3 的 5×5 空洞卷积,实现 3×3 、 9×9 和 15×15 三个尺度特征信息的提取,并采用 1×1 卷积进行多通道特征融合。由于各级卷积特征分辨率存在差异,多尺度信息整合模块需自适应地进行调整。Conv1 和 Conv2 的特征图分辨率较大,采用三支路提取多尺度特征;由于 Conv3 特征图经过两

次下采样后分辨率较小,只采用 3×3 、 9×9 尺度特征提取分支;对于 Conv4 特征图则只取 3×3 尺度特征进行分支提取。各支路感受野如图 3 所示,圆点代表卷积核采样点,三个不同颜色的区域分别表示三路分支感受野区域。

输入特征图通过多尺度信息整合模块和激励函数,生成像素间注意力分布图。分布图分别与各通道特征图像素对应相乘,得到加权后的特征图。设输入特征图 $\mathbf{X} \in \mathbf{R}^{W \times H \times C}$,下采样特征图 $\hat{\mathbf{X}} \in \mathbf{R}^{W' \times H' \times C}$,注意力模块输出 $\tilde{\mathbf{X}} \in \mathbf{R}^{W \times H \times C}$,该模块可以表述为

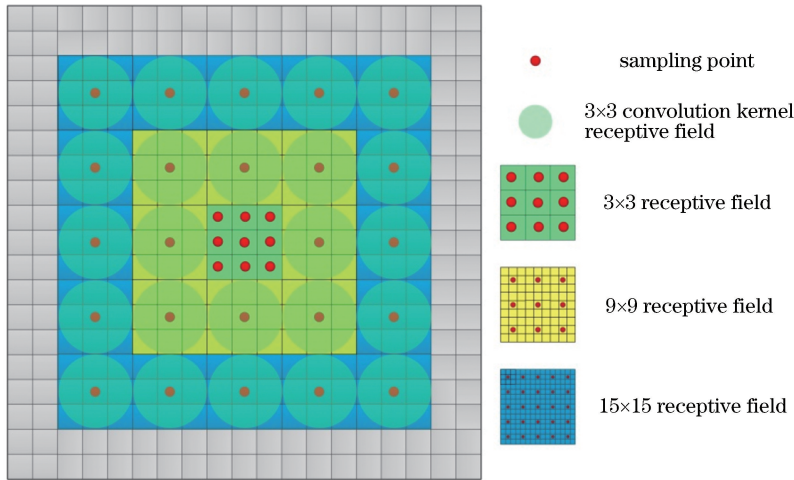


图 3 多尺度感受野示意图

Fig. 3 Schematic of multi-scale receptive field

2.2.2 通道间注意力模块

不同的特征通道往往具有不同的特征响应,不同的特征响应对于目标任务的影响存在差异。而卷积算子对输入特征图各通道的处理是等权重的,这与实际情况不相符。引入通道间注意力模块,实现通道间差异化处理,可以增大任务有利的特征通道权值,降低无用的特征通道权值,从而减弱无关信息干扰。本研究选用文献[18]提出的挤压-激励(SE)模块作为通道间注意力模块,如图 2 左上角椭圆区域所示。其采用挤压-激励操作提取通道间注意力,在监督学习下,计算预测值与真值之间的偏差,以偏差量为导向,给输入特征图各特征通道赋予权重。设输入张量为 $\mathbf{X} \in \mathbf{R}^{W \times H \times C}$ 。挤压操作是对特征图中各个通道进行全局平均池化,得到一个 $1 \times 1 \times C$ 的全局信息描述符,该描述符表示各通道特征整体响应的相对强弱,可表示为

$$\mathbf{Z} = \sum_{k=1}^C \left[\frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H \mathbf{X}(i, j) \right], \quad (5)$$

$$\hat{\mathbf{X}} = f_d(\mathbf{X}), \mathbf{X}' = f_{\text{conv}1 \times 1}(\hat{\mathbf{X}}), \quad (1)$$

$$\mathbf{X}'' = f_c[f_{\text{br}1}(\mathbf{X}'), f_{\text{br}2}(\mathbf{X}'), f_{\text{br}3}(\mathbf{X}')], \quad (2)$$

$$\mathbf{Y}' = f_s\{f''_{\text{conv}1 \times 1}[\mathbf{X}' + f'_{\text{conv}1 \times 1}(\mathbf{X}'')]\}, \quad (3)$$

$$\tilde{\mathbf{X}} = f_u[\mathbf{X}' \cdot f_c(\mathbf{Y}')], \quad (4)$$

式中: $f_{\text{conv}1 \times 1}$ 表示 1×1 卷积; f_d 表示下采样操作; f_c 表示全连接层; $f_{\text{br}1}, f_{\text{br}2}, f_{\text{br}3}$ 分别表示三个尺度特征提取分支的卷积操作; f_s 表示 Sigmoid 激活函数; f_u 表示进行上采样操作; f_c 表示复制扩展 \mathbf{Y}' 维度,使其与 \mathbf{X}' 一致; $f'_{\text{conv}1 \times 1}$ 和 $f''_{\text{conv}1 \times 1}$ 分别表示具有不同卷积核参数的 1×1 卷积。

式中: \mathbf{Z} 为全局特征描述符。

激励操作是采用全连接对全局信息描述符进行仿射变换,全面捕获通道依赖性,过程可表示为

$$\mathbf{S} = f_s\{f'_c[\delta(f_c(\mathbf{Z}))]\}, \quad (6)$$

式中: δ 表示 ReLU 激活函数, f'_c 表示含有 r 个单元的全连接层, \mathbf{S} 为描述各通道权重赋值的张量, $\mathbf{S} \in \mathbf{R}^{1 \times 1 \times C}$ 。

最后,将得到的通道间权向量与各通道对应相乘,输出为

$$\tilde{\mathbf{X}} = \mathbf{X}_a \cdot f_c(\mathbf{S}_a), \quad (7)$$

式中: \mathbf{X}_a 为输入特征图的单个通道特征, \mathbf{S}_a 为对应通道特征的权重, f_c 表示复制扩展操作。

2.2.3 双注意力模块融合

由于两类注意力模块输出的权值均在 0 到 1 之间,直接采用串联方式将结果进行融合,会减小像素间灰度值差异,弱化特征响应,造成信息损失。为避免这个问题,本研究采用并联方式融合两类注意力模块。首先,让两类注意力模块分别对输入特征图

进行处理,得到两类注意力加权特征图,然后采用 concatenate 操作对结果进行通道堆叠,最后用 1×1 卷积将两类结果进行整合,并使用 L2 Normalize 对输出层进行标准化。

2.3 循环卷积模块

对高层语义特征直接进行反卷积预测得到的显著图中会存在模糊现象。采取跨层连接,引入底层特征能明显改善显著图细节表示能力,但仍存在边界模糊现象。为进一步提升预测图质量,受人类“凝视”视觉机制启发,在模型后端引入循环卷积模块,以时间迭代来增强空间分辨能力,细化目标区域边缘。

文献[19]首先提出循环卷积层(RCL),文献[20-21]将其引入显著性目标检测,用于分层细化预测图,提升了显著性目标检测分割效果。基于RCL,设计多特征循环卷积模块(MFRCM),遵循Inception思想改造卷积循环单元,聚合多尺度感受野特征;同时添加外部循环机制,结合多卷积层特征,将解码器最后一层输出与编码器底层卷积特征作为输入,输出边缘效果更好的预测图。模块内部结构如图4所示。

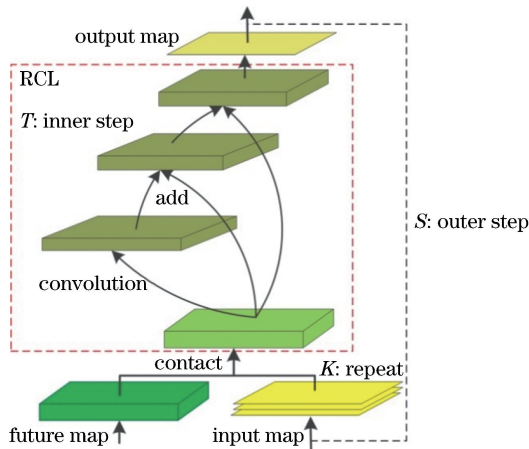


图4 多特征循环卷积模块内部结构图
Fig. 4 Internal structure of MFRCM

MFRCM将循环连接引至卷积层之间,以Inception卷积层为循环单位,逐步提取特征信息。随循环步骤 T 的增加,输出图单个像素的感受野增大,特征提取范围增大,信息描述更全面。输入图与循环单元输出之间使用捷径(shortcut)连接。设输入特征图为 $U \in \mathbf{R}^{W \times H \times C}$, $u_k^{(m,n)}$ 为第 k 通道位于 (m,n) 的像素,循环卷积过程可描述为

$$\begin{cases} z_k^{(m,n)}(t) = W_k^f \cdot u_k^{(m,n)} + b_k, t = 0 \\ z_k^{(m,n)}(t) = W_k^f \cdot u_k^{(m,n)} + W_k^r \cdot x_k^{(m,n)}(t-1) + b_k, 0 < t < T \end{cases}, \quad (8)$$

式中: W_k^f, W_k^r 分别表示前向输入与循环输入的权重; $u_k^{(m,n)}$ 表示前向输入,不随循环步长变化; $z_k^{(m,n)}$ 表示卷积输出; T 表示内循环次数; $x_k^{(m,n)}(t-1)$ 表示循环输入,即上一个循环单元输出。循环单元输出 $x_k^{(m,n)}(t)$ 由 $z_k^{(m,n)}(t)$ 变化而来,即

$$\begin{cases} f(z_k^{(m,n)}) = \text{ReLU}(z_k^{(m,n)}) \\ g(f_k^{(m,n)}) = f_{\text{LRN}}(f_k^{(m,n)}) \\ x_k^{(m,n)}(t) = g\{f[z_k^{(m,n)}(t)]\} \end{cases}, \quad (9)$$

式中 f_{LRN} 表示局部响应归一化(LRN)函数,用于模拟人脑侧抑制机制。

$$f_{\text{LRN}}(f_k^{(m,n)}) = f_k^{(m,n)} / \left[1 + \frac{\alpha}{N} \sum_{L=\max(0, k-N/2)}^{\min(C', k+N/2)} (f_L^{(m,n)})^2 \right]^\beta, \quad (10)$$

式中: N 表示参与归一化的邻近通道数; α 和 β 是用来调整标准化幅度; C' 表示总通道数, $f_k^{(m,n)}$ 表示一个计算中间值,这里表示LRN的输入特征图; L 表示表示特征图的第 L 个通道特征。

在原RCL基础上,添加外部循环机制,将上一轮MFRCM的输出值作为下一轮外部循环的输入值。连接示意图如图5所示。由于采用concatenate操作结合输入图和特征图,输入图通道占比重较小,故采用Repeat方案扩充输入图通道。

3 实验与结果分析

3.1 数据集与评价准则

3.1.1 数据集介绍

在常用公开数据集 ESSCD^[22]、HKU-IS^[23] 和 DUT-OMRON^[24] 上进行算法训练与验证。ESSCD数据集包含1000张语义结构复杂的图像。HKU-IS数据集包含4447张含有多个显著目标的图像。DUT-OMRON数据集包含对比度低、尺寸变化大的自然图像5168张,是一个十分具有挑战性的数据集。分别从ECSSD数据集、HKU-IS数据集和DUT-OMRON数据集随机抽取532张、2000张、2668张图像构成算法的训练集,共5200张。从剩余图像中抽取218张、1000张、1000张构成验证集,共2218张。其余图像用作测试集: ECSSD-250, HKU-IS-1447, DUT-OMRON-1500。

3.1.2 评价准则

选取PR(precision-recall)曲线、ROC(receiver

operating characteristic) 曲线、 F -measure 值和 MAE (mean absolute error) 值作为客观评价指标。PR 曲线是显著性目标检测常用的评价指标之一。逐渐增加阈值 $H \in [0, 255]$, 将显著图二值化与真值图进行逐像素对比, 计算出每个阈值对应下的准确率 P (precision) 与召回率 R (recall), 再以召回率

为横坐标, 准确率为纵坐标, 画出 PR 曲线。一般来说 PR 曲线越接近右上方, 算法性能越好。ROC 曲线是逐渐增加阈值下以假阳率 (FPR) 为横坐标, 真阳率 (TPR) 为纵坐标画出的曲线, 一般来说曲线越靠近左上方说明算法性能越好。

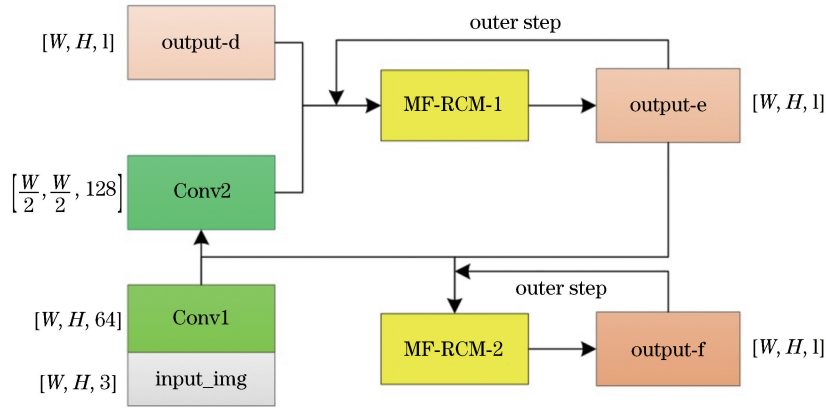


图 5 多特征循环卷积模块连接示意图
Fig. 5 Connectionschematic of MFRCM

单独使用准确率或召回率不能判断算法的优劣, F -measure 值综合考虑了准确率和召回率, 以加权平均和的形式给出一个度量值, 是一个全面的衡量指标, 该值越大算法性能越好。 F -measure 值的计算式为

$$F_{\beta} = \frac{(1 + \sigma^2) \cdot P_{\text{Precision}} \cdot R_{\text{Recall}}}{\sigma^2 \cdot P_{\text{Precision}} + R_{\text{Recall}}}, \quad (11)$$

式中 σ^2 一般取 0.3, 认为准确率比查全率更重要。在计算 F -measure 值前, 需要对预测的显著图进行二值化处理, 采用广泛使用的自适应阈值方案^[25], 以图像像素灰度平均值的 2 倍为阈值进行二值化分割, 可由 (12) 式计算, 即

$$T = \frac{2}{\hat{W} \times \hat{H}} \sum_{x=1}^{\hat{W}} \sum_{y=1}^{\hat{H}} S(x, y), \quad (12)$$

式中: \hat{W} 表示图像宽度; \hat{H} 表示图像高度; $S(x, y)$ 表示显著图的像素点灰度值。MAE 值表示二值化后显著图 S 与真值图 G 对应的像素点绝对误差, 即

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

式中: $G(x, y)$ 表示真值图像像素点的灰度值。

3.2 算法实现细节

采用深度学习框架 pytorch-0.4 进行神经网络搭建, 实验环境为 Windows 10, 使用 NVIDIA TITAN X GPU 进行神经网络训练与测试。采用二

值交叉熵 (BCE) 损失函数作为预测值与目标之间误差的度量准则。设 \mathbf{W} 表示网络参数, y_i 表示真值, \hat{y}_i 表示预测值, N' 表示图像总像素数, 则 BCE 损失函数表示为

$$L(\mathbf{W}) = \sum_{i'=1}^{N'} [y_{i'} \cdot \ln(\hat{y}_{i'}) + (1 - y_{i'}) \cdot \ln(1 - \hat{y}_{i'})]. \quad (14)$$

网络训练时使用多阶段约束方案, 约束输出值域范围, 总损失函数 L_{total} 采用各阶段损失函数加权表示, 即

$$L_{\text{total}} = \omega_d L_d + \omega_e L_e + \omega_f L_f, \quad (15)$$

式中: L_d 表示最后反卷积块的输出图与真值之间的 BCE 损失; L_e 表示循环卷积块第一次外部循环的输出图与真值之间的 BCE 损失; L_f 表示最终输出图与真值之间的 BCE 损失, $\omega_d, \omega_e, \omega_f$ 分别表示各阶段损失的权重, 训练时分别设置为 0.3, 1.0, 1.0。

所提算法使用 U-Net 语义分割模型卷积层权重作为预训练参数进行训练, 使用随机梯度下降优化器进行优化, 其中初始学习率设置为 0.001, 动量设置为 0.9, 权重衰减设置为 0.0001, 学习率衰减采用基于轮次下降的策略, 每 200 个轮次降为原来的 0.1。循环卷积模块中 LRN 的超参数设置 $\alpha = 0.001, \beta = 0.75, N = C/8$, 内循环次数设置为 $T = 3$, 外循环次数设置为 $S = 2$ 。在训练阶段, 保持长宽比例将图像放缩至短边为 256 pixel, 并随机裁剪使图

像输入尺寸固定为 256 pixel×256 pixel。批量训练样本数量(batch size)设为 4,训练轮次设为 1000,取最佳指标对应轮次的模型参数为最终参数。测试帧率约为 9 frame/s。

3.3 实验结果比较与分析

将所提出的算法与 6 种流行的基于卷积神经网络的显著性目标检测算法进行对比实验:SRM^[26], AMU^[27],UCF^[28],KSR^[29],MDF^[23],DSS^[10]。对比算法的评测数据,均来自算法作者公布的显著结果图和相关的模型参数。

3.3.1 主观视觉对比

为比较所提出的算法与对比算法在不同情况下的目标检测效果,选择 6 组算法效果图作为主观视觉对比(见图 6)。其中,图 6(a)和(b)反映了 7 种算法在低对比度环境下对目标区域的检测与分割,可

以看出所提算法无论是对显著性目标位置的检测,还是对目标边缘分割,都优于其他 6 种算法;图 6(c)和(d)反映对多显著目标的检测效果,虽然所提算法在检测左边啤酒杯时出现把手漏检情况,但是从总体的分割细节来看,优于其他算法。尤其在图 6(d)上所提算法效果明显优于其他算法。图 6(e)和(f)反映算法对细节检测的能力,可以看出所提算法的优势。

3.3.2 客观量化对比

根据选择的客观评测准则对 7 种算法进行量化对比。表 1 为各算法在三个测试集下 F -measure 值和 MAE 值,其中,算法性能同 F -measure 值呈正相关,同 MAE 值呈负相关。根据表 1 数据可以看出在三个测试集上所提算法性能均优于其他算法。尤其是在 DUT-OMRON-1500 数据集上,所提算法优势明显。

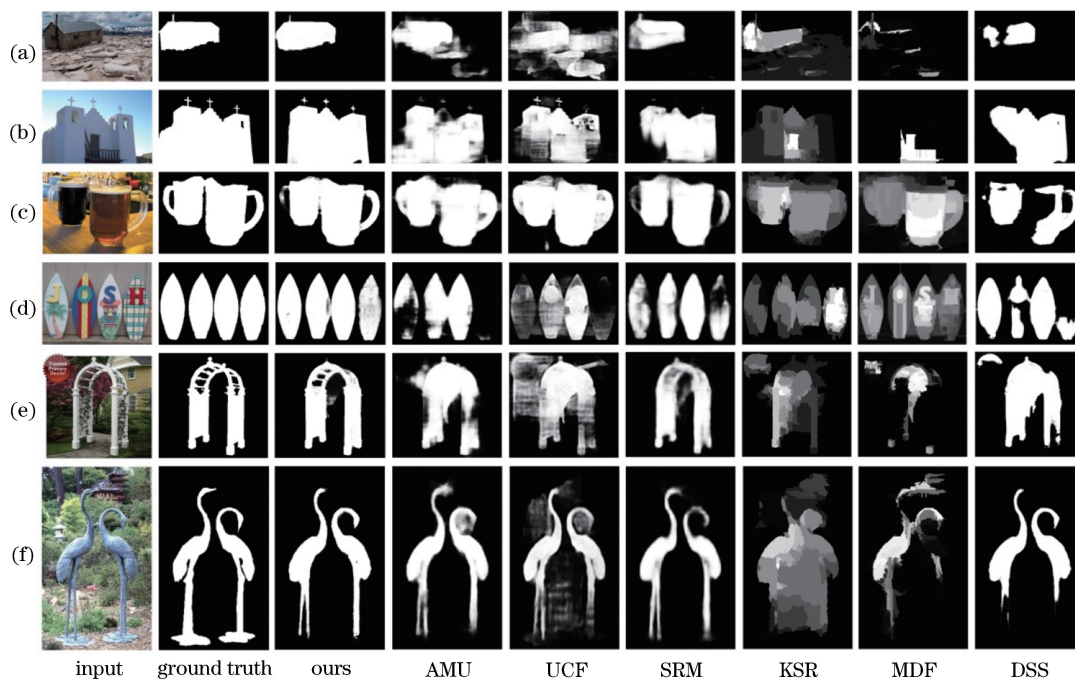


图 6 主观视觉效果对比图

Fig. 6 Comparison of subjective visual effects

表 1 客观量化指标对比

Table 1 Comparison of objective quantitative indicators

Dataset	Evaluating indicator	Value						
		Ours	AMU	UCF	SRM	KSR	MDF	DSS
ECSSD-250	F -measure	0.928	0.918	0.912	0.916	0.827	0.834	0.917
	MAE	0.043	0.054	0.071	0.054	0.117	0.104	0.056
HKU-IS-1447	F -measure	0.923	0.907	0.893	0.912	0.784	0.850	0.910
	MAE	0.038	0.049	0.058	0.042	0.127	0.118	0.040
DUT-OMRON-1500	F -measure	0.782	0.748	0.737	0.763	0.634	0.647	0.758
	MAE	0.056	0.071	0.131	0.068	0.131	0.087	0.074

图 7、8、9 中图(a)为 7 种算法在三个测试集上的 PR 曲线,可以明显看出所提算法曲线右上部更接近(1,1)点,说明通过调整阈值所提算法能取得高准确率的同时保持高召回率,对显著目标的检测效果也最好。图(b)为 7 种算法在三个测试集上的 ROC 曲线,在 ECSSD- 250 和 HKU-IS-1447

测试集上所提算法具有明显优势,在 DUT-OMRON-1500 测试集上 ROC 曲线曲率出现突变,这是因为曲线是使用点对直接画出,当算法结果具有较低假阳率和较高的真阳率时,点对会集中出现在低假阳率区域,因此可以说所提算法优于其他算法。

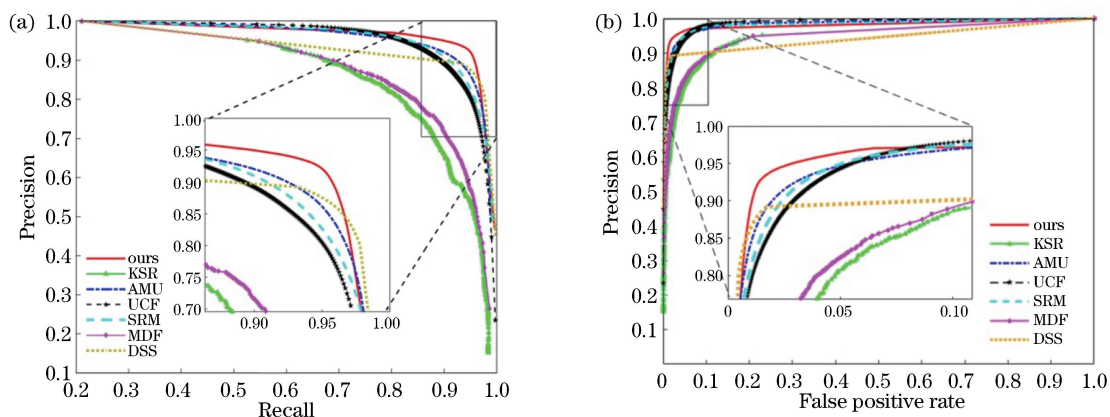


图 7 不同算法在 ECSSD-250 测试集下的特征曲线。(a) PR 曲线;(b) ROC 曲线

Fig. 7 Characteristic curves of seven algorithms on ECSSD-250 testset. (a) PR curves; (b) ROC curves

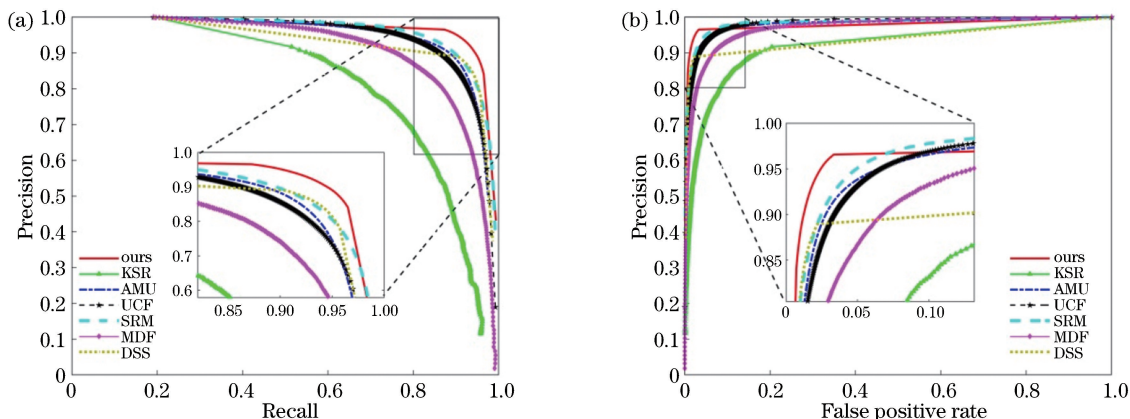


图 8 不同算法在 HKU-IS-1447 测试集下的特征曲线。(a) PR 曲线;(b) ROC 曲线

Fig. 8 Characteristic curves of seven algorithms on HKU-IS-1447 testset. (a) PR curves; (b) ROC curves

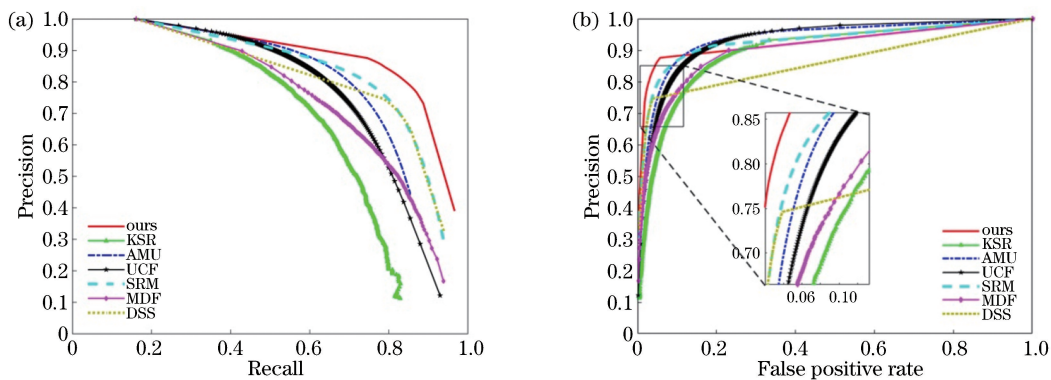


图 9 7 种算法在 DUT-OMRON-1500 测试集下的特征曲线。(a) PR 曲线;(b) ROC 曲线

Fig. 9 Characteristic curves of seven algorithms on DUT-OMRON-1500 testset. (a) PR curves; (b) ROC curves

3.3.3 算法运行时间及参数量对比

表 2 对比了 7 种算法的运行时间。实验平台配置如下: CPU: Xeon E5-1650 v4, 3.6 GHz; GPU: NVIDIA TITAN-X, 12 GB; 运行内存为 32 GB; 固态硬盘容量为 512 GB。所提算法处理速度约为 9 frame/s, 达到次实时要求。

表 3 对 AMU、UCF、SRM、DSS 和所提算法模

表 2 各算法的运行时间

Table 2 Running time of different algorithms

Algorithm	AMU	UCF	KSR	SRM	MDF	DSS	Ours
Time /s	0.08	0.08	47.20	0.10	7.64	0.48	0.11

表 3 算法参数量统计

Table 3 Parameters statistics of different algorithms

Algorithm	U-Net	Ours	AMU	UCF	SRM	DSS
Number of parameters	13395329	19062256	17740720	30406400	45484608	19130619

4 结 论

本研究提出一种双注意力循环卷积显著性目标检测算法。算法在 U-Net 骨干网络中引入像素间-通道间双注意力模块, 对跨层连接的特征进行重加权处理, 提高特征利用的有效性。其中使用空洞卷积设计像素间注意力模块, 采用三路特征提取分支模拟感受野机制, 有效整合多尺度上下文信息, 实现像素间加权。采用多特征循环卷积层将 U-Net 解码端输出与编码端卷积特征进行结合处理, 增强输出显著图边缘表示效果。实验表明, 所提算法在多显著性目标检测和细节检测能力上表现良好, 在三个测试集上的客观量化指标性能优于其他 6 类对比算法。但由于骨干网络的最大池化层会造成信息量丢失, 影响预测图重构质量, 下一步工作是改进编码器网络, 提取适合显著性检测的特征, 同时优化模型结构与参数, 进一步改善运行速度。

参 考 文 献

- [1] Gao Y, Wang M, Zha Z J, *et al.* Visual-textual joint relevance learning for tag-based social image search [J]. IEEE Transactions on Image Processing, 2013, 22(1): 363-376.
- [2] Guo C L, Zhang L M. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression [J]. IEEE Transactions on Image Processing, 2010, 19(1): 185-198.
- [3] Guo C L, Ma Q, Zhang L M. Spatio-temporal

型参数量进行统计, 并给出 U-Net 骨干网络模型参数量作为对照。由于 MDF 和 KSR 与其他算法性能相差较大, 参数量对比无实际意义, 故未进行统计。从表 3 可以看出所提算法参数量相对较小, 一定程度说明在中小规模数据集上不易出现过拟合, 更容易训练。

saliency detection using phase spectrum of quaternion Fourier transform [C] // 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2008, Anchorage, AK, USA. New York: IEEE, 2008: 4587715.

- [4] Courty N, Marchand E. Visual perception based on salient features [C] // Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), October 27-31, 2003, Las Vegas, Nevada, USA. New York: IEEE, 2003: 1024-1029.
- [5] 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.
- [6] Liu D M, Chang F L. Coarse-to-fine saliency detection based on non-sampled contourlet transform enhancement [J]. Acta Optica Sinica, 2019, 39(1): 0115003.
刘冬梅, 常发亮. 基于非下采样轮廓小波变换增强的从粗到精的显著性检测 [J]. 光学学报, 2019, 39(1): 0115003.
- [7] Li D F, Liu S T. Image target saliency detection method based on feature fusion [J]. Semiconductor Optoelectronics, 2018, 39(6): 898-902, 908.
李德峰, 刘松涛. 基于特征融合的图像目标显著性检测方法 [J]. 半导体光电, 2018, 39(6): 898-902, 908.
- [8] Liu F, Shen T S, Lou S L, *et al.* Deep network saliency detection based on global model and local optimization [J]. Acta Optica Sinica, 2017, 37(12): 1215005.
刘峰, 沈同圣, 娄树理, 等. 全局模型和局部优化的

- 深度网络显著性检测[J]. 光学学报, 2017, 37(12): 1215005.
- [9] Zhang S L, Xie L B. Salient detection based on all convolutional feature combination [J]. *Laser & Optoelectronics Progress*, 2018, 55(10): 101502.
张松龙, 谢林柏. 基于全部卷积特征融合的显著性检测[J]. *激光与光电子学进展*, 2018, 55(10): 101502.
- [10] Hou Q B, Cheng M M, Hu X W, *et al.* Deeply supervised salient object detection with short connections[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 5300-5309.
- [11] Xie S N, Tu Z W. Holistically-nested edge detection [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1395-1403.
- [12] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W, *et al.* Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [13] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.
- [14] Zhang X N, Wang T T, Qi J Q, *et al.* Progressive attention guided recurrent network for salient object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT. New York: IEEE, 2018: 714-722.
- [15] Kim H J, Dunn E, Frahm J M. Learned contextual feature reweighting for image geo-localization [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 3251-3260.
- [16] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, *et al.* Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 404-419.
- [17] Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, inception-resnet and the impact of residual connections on learning [C] // 31st AAAI Conference on Artificial Intelligence, February 4-10, 2017, San Francisco. California: AAAI Press, 2017: 4278-4284.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT. New York: IEEE, 2018: 7132-7141.
- [19] Liang M, Hu X L. Recurrent convolutional neural network for object recognition[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3367-3375.
- [20] Liu N, Han J W. DHSNet: deep hierarchical saliency network for salient object detection[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 678-686.
- [21] Tang Y B, Wu X Q, Bu W. Deeply-supervised recurrent convolutional neural network for saliency detection[C]//The 24th ACM International Conference on Multimedia, October 15-19, 2016, Amsterdam, Netherlands. USA: ACM, 2016: 397-401.
- [22] Yan Q, Xu L, Shi J P, *et al.* Hierarchical saliency detection[C] // 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 1155-1162.
- [23] Li G B, Yu Y Z. Visual saliency based on multiscale deep features[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 5455-5463.
- [24] Yang C, Zhang L H, Lu H C, *et al.* Saliency detection via graph-based manifold ranking[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3166-3173.
- [25] Achanta R, Hemami S, Estrada F, *et al.* Frequency-tuned salient region detection[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 1597-1604.
- [26] Wang T T, Borji A, Zhang L H, *et al.* A stagewise refinement model for detecting salient objects in images[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE, 2017: 4019-4028.
- [27] Zhang P P, Wang D, Lu H C, *et al.* Amulet: aggregating multi-level convolutional features for salient object detection[C] // 2017 IEEE International

- Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE, 2017: 202-211.
- [28] Zhang P P, Wang D, Lu H C, *et al.* Learning uncertain convolutional features for accurate saliency detection[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE, 2017: 212-221.
- [29] Wang T T, Zhang L H, Lu H C, *et al.* Kernelized subspace ranking for saliency detection [M] // Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9912: 450-466.