

基于深度残差网络与边缘监督学习的显著性检测

时斐斐**, 张松龙, 彭力*

江南大学物联网工程学院, 物联网应用技术教育部工程中心, 江苏 无锡 214122

摘要 针对复杂背景下, 图像显著区域显著值低和目标边缘表现模糊的问题, 提出了基于深度残差网络和多尺度边缘残差学习的显著性检测方法。一方面提出了边缘残差块, 使用边缘残差块在深度残差网络的基础上构建边缘监督网络, 用于显著图边缘监督学习; 另一方面, 通过构建基于背景、前景和边缘的三分类模型, 训练网络学习边缘特征, 使目标边缘更加准确, 同时输出采用空洞卷积构建多尺度空洞卷积单元, 多尺度地对全局信息进行特征整合提取。最后, 将提出的算法在数据集 SED2 和 ECSSD 上进行模型简化测试, 使用公认评价指标对所提算法和当前多种算法进行评价。实验结果表明, 该方法的准确率和召回率更高, 对显著目标保持了良好的完整性, 且在边缘轮廓区域更好地区分了显著目标与背景。

关键词 机器视觉; 显著性检测; 边缘残差块; 三分类模型; 多尺度空洞卷积

中图分类号 TP391.41

文献标识码 A

doi: 10.3788/LOP56.151502

Salient Object Detection Based on Deep Residual Networks and Edge Supervised Learning

Shi Feifei**, Zhang Songlong, Peng Li*

Engineering Research Center of Internet of Things Technology Applications of the Ministry of Education, College of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract This paper proposes a saliency detection method based on deep residual networks and multiscale edge residual learning to address the problems of low salient values and blurred edges in images having complex backgrounds. Further, an edge residual block is proposed, and an edge residual network is constructed based on the deep residual network using the edge residual block for the salient graph edge supervised learning. In addition, the edge features are learned while training the network by constructing a three-category model based on the background, foreground, and edge, which can make the target edge more accurate. The output uses atrous convolutions to construct a multiscale atrous convolution unit for integrating and extracting the multiscale features based on the global information. Finally, the proposed algorithm is tested in an ablation study using two datasets (SED2 and ECSSD) and compared with various existing algorithms based on the common evaluation indicators. The experimental results demonstrate that the proposed method exhibits high accuracy and recall rate, maintains good integrity for the significant target, and distinguishes the significant target and background from the edge contour regions.

Key words machine vision; saliency detection; edge residual block; three-classification model; multi-scale atrous convolution

OCIS codes 150.0155; 330.7326; 110.2970

收稿日期: 2018-12-29; 修回日期: 2019-02-18; 录用日期: 2019-03-07

基金项目: 国家自然科学基金(61873112)、国家重点研发项目子课题(2018YFD0400900)、教育部-中国移动科研基金项目(MCM20170204)

* E-mail: 884208590@qq.com; ** E-mail: 6171913022@stu.jiangnan.edu.cn

1 引言

近年来,得益于深度学习领域的研究与应用,显著性检测作为机器视觉领域减少后续复杂视频图像处理的重要预处理步骤,取得了重大进展。目前,研究方法模型已经从传统的生物启发^[1-3]、频域残差^[4]、相位谱^[5-6]、背景检验^[7-8]、启发式先验模型^[9]以及多尺度检验^[10-11]等开始转向以深度学习理论方法为主,从卷积神经网络(CNN)到FCN网络的应用促进了显著性检测的突破性发展,如Zhao等^[12]提出了多情景深度学习显著性检测框架,综合考虑了全局背景和局部背景对显著性的影响,但其工作只针对背景,对显著目标前景与背景的关系并未进行细致区分;Zhang等^[13]受注意力模型的启发,提出基于CNN特征提取、注意力引导分配权重来实现显著目标的检测,强调FCN网络应有选择地、循序渐进地融合不同层的内容信息,以减轻背景信息的分散性,取得了良好的效果。但在融合各层内容信息的同时,并未重视前景与背景的实质区别;Li等^[14]注重显著图的边缘模糊与基于CNN时计算和存储冗余,提出了端到端深度对比网络,产生基于多尺度特征图的像素级显著图和超像素级显著图各一张,通过条件随机场将两张不同处理的显著图进行融合,得到了很好的显著图效果。但其着重进行融合对比,从而得到最显著的特征位置,并未对边缘轮廓信息量作进一步分析。

由此可见,虽然基于深度神经网络层次的研究目前取得了巨大突破,但怎样通过深度神经网络学习足够的显著性特征仍是一项具有挑战性的任务。一般神经网络算法常采用背景和前景二分类模型,利用人工标注二值图(GT图)对图像前景和背景区域进行二值学习^[15],而卷积大多基于块集级别而非像素级别操作^[14],导致检测有多种纹理和颜色的图像时得到的目标显著值不高,产生的显著图边缘不足以细化,致使物体边缘模糊,前景与背景区分度不大。为解决这一问题,本文在深度残差网络模型的基础上,自底向上提取每阶段卷积特征,并将其输入边缘残差块(ERB)组成的边缘残差网络中进行边缘监督,利用低级细节特征细化高级的语义特征,获取更清晰的显著区域与边缘轮廓;通过修改人工标记二值图,将轮廓信息与原有的前景和背景信息结合组成三维类别,增强网络对边缘的区分与学习,同时加入空洞卷积组建多尺度空洞卷积单元(MACU)学习大范围轮廓信息,从而得到更精确的显著图。

2 基于深度残差网络与边缘残差学习的网络模型

2.1 网络架构

本文采用ResNet-101(Residual Network with 101 parameter layers)^[16]作为基础建立网络架构。ResNet-101具有良好的残差学习能力,在加深网络的同时,也解决了梯度退化的问题。考虑到计算成本,针对其残差块作了计算优化,将两个 1×1 的卷积层替换为 $1 \times 1 + 3 \times 3 + 1 \times 1$ 卷积层。在该结构中,两个 1×1 的卷积层分别用来减小计算量与还原维度,使用 3×3 卷积层可保持精度,其在目标检测^[17]、语义分割^[18]和行为识别^[19]等方面得到了广泛应用。ResNet-101模型由100个卷积层和1个全连接层组成,100个卷积层根据输出特征可分为5个阶段,每个阶段对应的总层数分别为3层、4层、23层和3层;第1阶段和第5阶段输出分别连接最大池化层与平均池化层。本文网络框架如图1所示,基于ResNet-101网络模型作了如下改进:1)去掉最后的平均池化与全连接层,从而使原CNN变成全卷积网络,同时更多地保留细节特征供后续网络学习;2)从ResNet-101提取第2到第5各阶段最终输出的特征图,并依次命名为Conv2_3、Conv3_4、Conv4_23和Conv5_3,对应图1中Res-2至Res-5的输出;3)将上述输出分别连接边缘残差块(ERB),对ERB-2与ERB-3输出特征进行相加融合,作为ERB-2p的特征输入,再将ERB-2p越阶段与ERB-4进行特征相加融合后,送入ERB-3p进行特征提取。同理连接至ERB-4p,组建边缘残差网络,如图1所示。将最后的ERB-4p输出连接SoftmaxWithLoss,对图像进行边缘监督学习;4)Res-5层输出连接MACU,该单元由3个空洞卷积层组成,使用空洞扩张度为6、12和18的三种空洞卷积核从多尺度提取显著特征,使用 1×1 卷积操作整合信息,并进行三分类输出;5)采用多项式损失函数计算损失值。

2.2 边缘残差块

为实现对前景与背景的显著特征区别以及对边缘特征的提取,设计自底向上的边缘监督网络,对原有的ResNet-101^[16]进行深度监督,以扩大边缘两侧特征的变化。由于低阶特征具有更详细的空间信息,而高阶特征具有高语义特征^[20],为了从低层次网络获取准确的边缘信息,并从高层次网络获取图像语义信息作补充,设计了与ResNet-101各阶段连

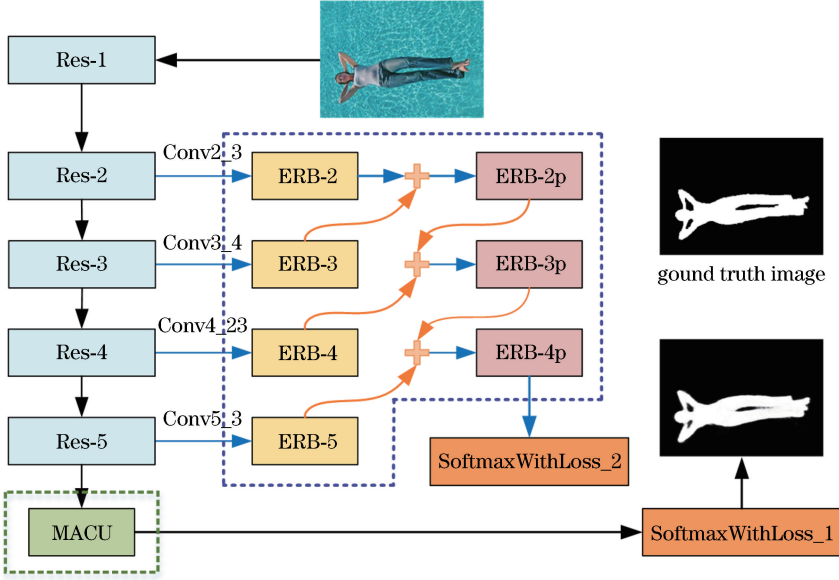


图1 网络结构图

Fig. 1 Network structure

接的 ERB, 其子模块结构如图 2 所示。在该结构中, 首先通过 1×1 的卷积操作, 对输入图像的通道数作统一处理, 分支网络使用两层 3×3 卷积提取细节信息, 与原通道特征融合, 从而达到区分细化特征的效果。在每个阶段的 ERB 输出后, 对相邻阶段特征进行自底向上的特征融合。对于数据集 $D = \{(X^n, G^n, L^n)\}$, 式中, X^n 为输入训练图像, G^n 为对应真值图, L^n 为数据标签, 编号 $n = \{1, 2, \dots, N\}$, N 为数据集训练图像总量, ResNet-101 模型传递过程可表示为

$$S(X_c^n) = F\{X_c^n, \{W_i^c\}\} + X_c^n, \quad (1)$$

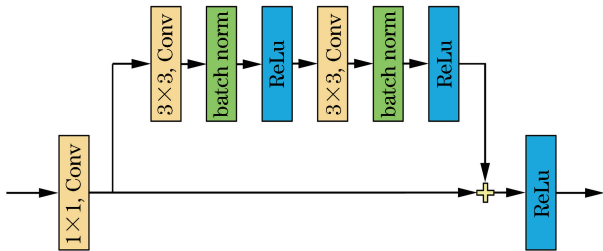


图2 边缘残差块结构图

Fig. 2 Edge residuals block structure diagram

式中: X_c^n 为当前阶段特征输入; c 为阶段数, 取值 $c \in \{1, 2, 3, 4, 5\}$; $S(X_c^n)$ 为该阶段 ResNet-101 的输出特征; W_i^c 为该阶段权重, i 为当前阶段对应的卷积层; 映射 $F = W_{i+1}\sigma(W_i X_i)$, 为需要学习的残差映射, 其中 σ 为简化忽略的线性整流函数 (ReLU) 以及偏差, X_i 为当前阶段对应的卷积层输入。对于 ERB- p 融合策略, 有

$$f^p(X_c^n) =$$

$$\begin{cases} F\{[f^{p-1}(X_{c-1}^n) + f^{p+1}(X_{c+1}^n)], \{W_i\}\} + \\ [f^{p-1}(X_{c-1}^n) + f^{p+1}(X_{c+1}^n)], \\ p = 2 \\ F\{[f^p(X_c^n) + f^{p+1}(X_{c+1}^n)], \{W_i\}\} + \\ [f^p(X_c^n) + f^{p+1}(X_{c+1}^n)], \\ p = 3, 4 \end{cases}, \quad (2)$$

式中: $f^p(X_c^n)$ 为该融合 ERB (结构中的 ERB- p) 的特征输出, p 为阶段数。逐层结合各阶段特征信息进行特征提取, 将最后的 ERB 输出连接交叉熵损失函数, 使网络收敛, 交叉熵损失函数定义为

$$H = -\left\{ \alpha_i \sum_{i=1}^{[X^n]} G_i^n \ln f_i^n \{L_i^n = 0 \mid \{X^n; W_i, \sigma\}\} + \beta_i \cdot \sum_{i=1}^{[X^n]} G_i^n \ln \{f_i^n L_i^n = 1 \mid \{X^n; W_i, \sigma\}\} + (1 - \alpha_i - \beta_i) \cdot \sum_{i=1}^{[X^n]} G_i^n \ln f_i^n \{L_i^n = 2 \mid \{X^n; W_i, \sigma\}\} \right\}, \quad (3)$$

式中: $[X^n]$ 为输出显著图像的素个数; α_i 为真值图像中像素标签为 0 的占比; β_i 为对应标签为 1 的占比; G_i^n 为对应真值图的像素 i ; $f_i^n (L_i^n = 0 \mid \{X^n; W_i, \sigma\})$ 为原图像在超参数 W_i, σ 下输出属于标签 0 的概率; 对应 $f_i^n (L_i^n = 1 \mid \{X^n; W_i, \sigma\})$ 和 $f_i^n (L_i^n = 2 \mid \{X^n; W_i, \sigma\})$ 为原图像在超参数 W_i, σ 下输出属于标签 1 与标签 2 的概率。通过对以上计算过程不断迭代, 使损失函数值 H 收敛, 可以得到最终的神经网络参数集合 $W\sigma = \{W_i, \sigma\}$, 对应生成网络模型。

2.3 多尺度空洞卷积单元

卷积作为卷积神经网络计算的主要方式,在数字图像中通过运用不同的卷积核(滤波矩阵)可以完成平滑、模糊、去噪、锐化和边缘提取等工作。传统卷积限制了图像感受野的大小^[21],需使用池化层对 Feature Map 降维,增大感受野;而池化降维会损失图像的细节信息,造成不可逆的损失。因此,为了提取更多的显著特征,采用空洞卷积^[22]构建多尺度空洞卷积单元 MACU。输入二维图像 $I(a, b)$,使用二维卷积核 $K(m, n)$ 对像素 (a, b) 作逐像素卷积运算,对于一般的卷积操作,有

$$(I * K)(m, n) = \sum_m \sum_n I(a + m, b + n)K(m, n), \quad (4)$$

式中: $*$ 为卷积运算符。

空洞卷积在此基础上增加了扩张度参数 r ,则

$$(I * K)(m, n) =$$

$$\sum_m \sum_n I(a + m * r, b + n * r)K(m, n). \quad (5)$$

当扩张度 $r \geq 2$ 时,在不改变卷积核大小、增加额外参数的条件下,可以增加卷积操作的感受野,此时感受野的大小可以表示为

$$R_F = [(k_s + 1) \times (r - 1) + k_s]^2, \quad (6)$$

式中: k_s 为卷积核尺寸。在不损失图像输入原有细节信息的同时,通过大范围的感受野,可以提取更广泛的图像语义特征。为获取不同尺度的特征,从扩张度 r 为 6、12 和 18 的 3 个尺度下进行卷积提取图像上下文信息,其结构如图 3 所示,其中 dilation 为空洞卷积单元,通过多尺度重采样,可以有效增强特征图效果^[21],然后经 1×1 卷积进行通道聚合,配合边缘特征学习网络,从而细化特征、增强三分类尤其是边缘的区分度。

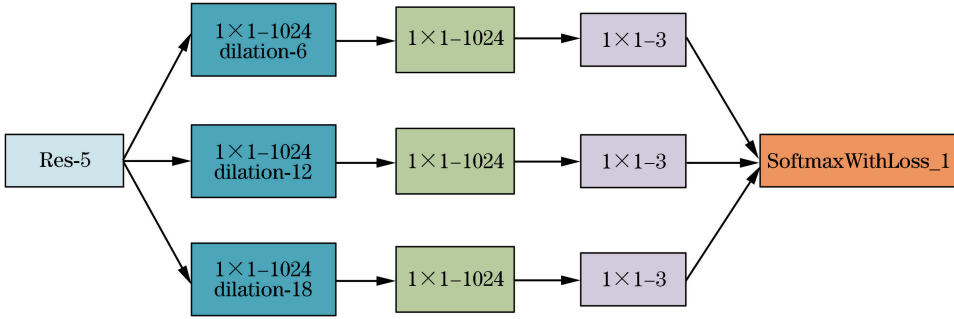


图 3 多尺度空洞卷积单元

Fig. 3 Multi-scale atrous convolution unit

2.5 基于前景、背景与轮廓的三分类学习

传统的用于显著性检测的神经网络模型,一般利用 GT 图对图像的前景与背景区域进行训练学习,得到二分类输出神经网络模型,该模型可直接检测图像的前景,从而得到显著性区域。由于背景中偶尔存在类前景像素,在一些场景中容易发生误检;另外,由于二分类训练得到的前景显著值有限,当图像中存在方向性纹理、多种颜色等细节时,显著性检测效果不佳^[15]。而在一幅图像中,轮廓作为区分目标与背景、目标与目标的重要依据,虽然占用像素少、区域小,但其中包含了丰富、重要的信息,完全可以作为区分显著与非显著区域的重要判据。基于此,提出了基于前景、背景和轮廓特征三分类学习的思路,通过修改 GT 图,将轮廓信息单独分为一类且标记为 2,与前景为 1、背景为 0 组成 $(0, 1, 2)$ 三分类学习,对网络模型进行训练。为达到这一目的,首先使用 Canny 轮廓检测算子对 GT 图进行轮廓检测,得到细小轮廓信息;为使轮廓信息更饱满,并形成完整的连通域,从

而将前景和背景区域完全区分,对得到的轮廓进行形态学膨胀处理,处理后的 GT 图如图 4 所示。

最后,将此方法得到的显著图与一般二分类模型得到的显著图进行对比,通过此方法得到的显著图整体效果比二分类方法得到的前景、背景更平滑,尤其边缘区域更清晰明了,如图 5 所示。

3 仿真实验分析

3.1 运行环境和数据集

1) 运行环境。本实验基于 64 位的 Ubuntu 16.04 LTS Server 操作系统和 NVIDIA GTX Genforce 1080Ti GPU 的深度学习服务器,显存为 11 GB,计算机内存 128 GB,深度学习框架为 Caffe^[23],主要使用软件为 MATLAB2014b 和 Python2.7。本文使用随机最速下降(SGD)方法训练网络,设置动量为 0.9,权重衰减为 0.0005,初始学习率为 0.001,最后分类器层为 0.01,每 2000 次迭代将学习率乘以 0.1,训练过程共花费 19 h,迭代 30000 次后达到收敛。

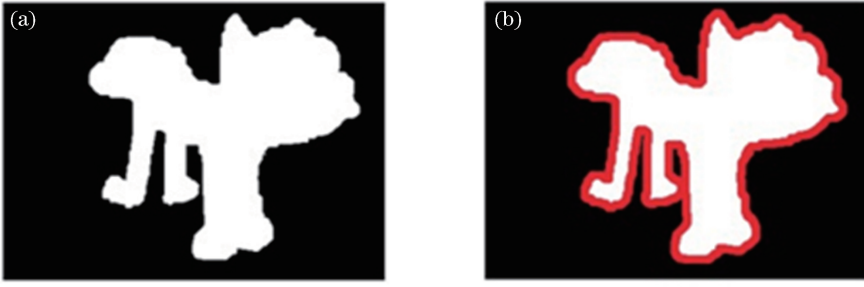


图4 真值图对比。(a)原真值图;(b)修改过的人工标记图

Fig. 4 Contrast of ground truth image. (a) Original ground truth image; (b) modified manual marking diagram

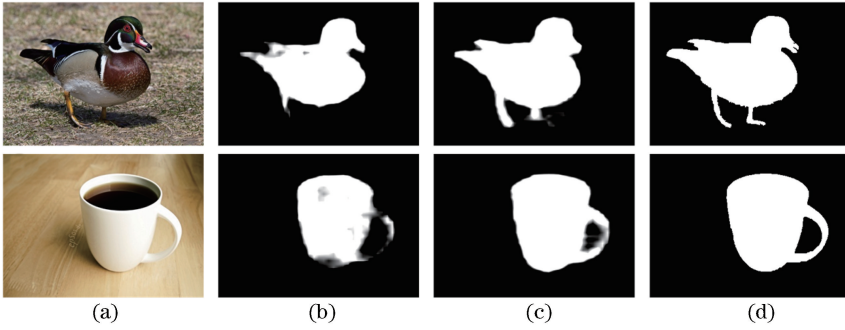


图5 三分类模型与传统二分类模型对比。(a)原图;(b)传统二分类模型结果;(c)本文三分类模型结果;(d)真值图
Fig. 5 Contrast of three-category model proposed in our paper and the traditional two-category model. (a) Original image; (b) result of traditional two-classification model; (c) results of proposed three-classification model; (d) ground truth image

2) 数据集。本文用于训练的数据集为 MSRA10K^[24] 公开数据集,其中含有 10000 张高像素图片,同时大多数图片只含有一个显著性目标。为了增加训练集数据量和样本的多样性,通过数据增强对原始数据集旋转 0° 、 90° 、 180° 、 270° ,并镜像得到 80000 张样本用作训练。测试集采用 SED2^[25] 和 ECSSD^[26] 两个数据集共 1100 张图片,其中 SED2 中含有 100 张测试图,虽然规模小,但含有像素级别的真值标注,且背景相对复杂,并均为多显著性目标图,极具挑战性;ECSSD 中含有 1000 张种类繁多、背景结构信息丰富的图片,具有很高的参考性和测试价值。

3.2 模型简化测试

为揭示每个方法对结果的影响,将模型分解成 3 种子结构,分别是二分类 ResNet-101、ResNet-101+边缘监督网络、三分类 ResNet-101 与本文完整模型结构进行效果对比,采用 F -measure(F 值)与平均绝对误差(MAE)值作为评价指标进行分析。 F 值基于图像的查准率和查全率得到,其计算公式为

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

式中: P 为查准率; R 为查全率;系数 β^2 设置为 0.3,

与文献[24]相同。 F 值综合考虑了 P 值和 R 值的矛盾关系,其值越大,说明分类模型越有效。MAE 值可以更好地衡量算法与 GT 图之间的关系,进而评价算法性能,其计算公式为

$$f_{\text{MAE}} = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N |S(x, y) - G_T(x, y)|, \quad (8)$$

式中: M 和 N 为显著图的长和宽; $G_T(x, y)$ 为人工标注 GT 图 (x, y) 处的值; $S(x, y)$ 为显著图 (x, y) 处的值。MAE 值越小,说明算法得到的显著图与人工标记 GT 图越接近,性能越优良^[27]。最终得到的结果如表 1 所示。由表 1 可知,与原有二分类 ResNet-101 模型相比,各模块性能均有一定优化。使用 ResNet-101 时,各指标均较理想,在上述数据集上效果良好,体现了残差网络的优越性。加入 ERB 残差监督网络后, F 值提高了 10.2%,MAE 值减小了 19.7%;使用三分类时, F 值提高了 12.5%,MAE 值减小了 10.7%;整体算法比单独使用 ResNet-101 时 F 值提高了 16.8%,MAE 值减小了 43.9%。

3.3 综合性能对比

为对比本文算法与其他显著性检测算法的实际

表1 各模块性能对比

Table 1 Performance comparison of proposed modules

Module	F-measure	MAE
ResNet-101+2 classification	0.7747	0.0856
ResNet-101+ERB	0.8537	0.0715
ResNet-101+3 classification	0.8715	0.0774
ResNet-101+ERB+3 classification	0.9048	0.0595

效果,选择7种不同的算法模型与本文算法进行比较,包括2种认可度高的传统方法——引导学习(BL)^[9]和多层元细胞机(BSCA)^[3],以及5种近年来提出的基于深度学习的显著性检测算法——多尺度深度特征(MDF)^[21]、循环全卷积神经网络

(RFCN)^[24]、深度层级显著网络(DHS)^[27]、深度对比学习(DCL)^[28]和不确定卷积特征(UCF)^[29]。采用主观评价与客观评价两种方式。图6为本文算法与其他7种算法模型的直观比较图,前4张来源于数据集 ECSSD,后2张来源于数据集 SED2。可以明显看出,本文算法在不同场景下均保持了很好的效果,比其他算法更接近真值图,处理复杂场景时,不但能够平滑均匀地显示显著目标,而且在显著目标边缘区域也表现出了良好的抑制效果,得到的显著图整体清晰,前景和背景区分度大,目标完整性好。

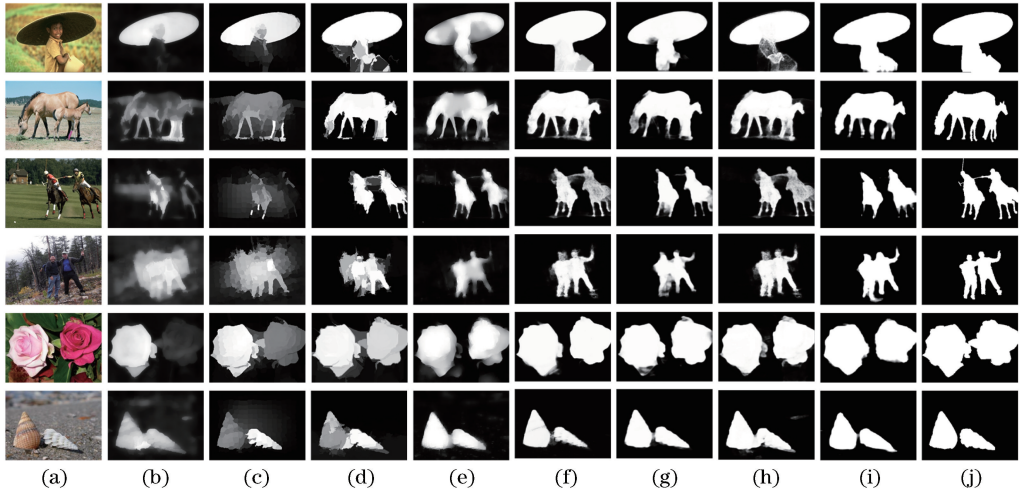
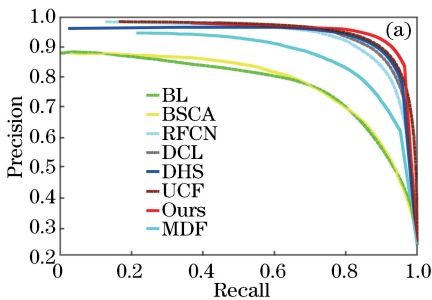


图6 显著图比较。(a)原图;(b) BL;(c) BSCA;(d) RFCN;(e) MDF;(f) DCL;(g) DHS;(h) UCF;(i)本文算法;(j)真值图

Fig. 6 Comparison of saliency maps. (a) Original image; (b) BL; (c) BSCA; (d) RFCN; (e) MDF; (f) DCL; (g) DHS; (h) UCF; (i) proposed method; (j) ground truth image

为进一步客观评价本文算法的性能,通过自适应阈值 T_u 对各算法在测试集上的显著图进行二值分割,并与人工真值图对比,根据对应的查准率 P 、查全率 R 和 F 值,自适应阈值 T_u 计算公式为

$$T_u = \frac{2}{W \times H} \sum_{x=1}^W \sum_{y=1}^H S(x, y), \quad (9)$$



式中: W 和 H 为显著图的长和宽。

固定阈值 $T_f = \{x \mid \forall x \in [0, 255]\}$,将单幅显著图分割为 256 张二值图并与 GT 图作对比,可以得到查全率-查准率(P - R)曲线,绘制上述算法得到的 P - R 曲线,得到查全率与查准率之间的关系如图 7 所示。

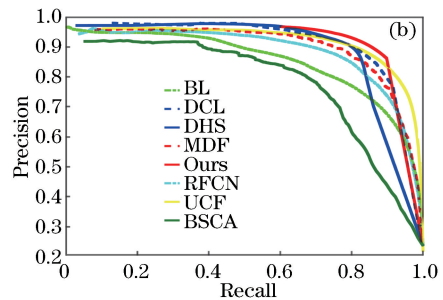


图7 本文算法及对比算法的 P - R 曲线图。(a) ECSSD 数据集结果;(b) SED2 数据集结果

Fig. 7 Precision-recall curves of the proposed algorithm and other state-of-the-art methods. (a) Results on ECSSD dataset; (b) results on SED2 dataset

从图7可以看出,相比于其他对比算法,本文方法表现更佳。为了更好地从数据层次上体现算法之间的差异,进一步采用MAE值、 F 值与曲线下方面积(AUC)^[24]指标对P-R进行分析。AUC值越接近1,说明算法的准确度越高。在ECSSD数据集上

进行测试,得到的具体数值见表2。由表2可以看出,本文方法的AUC以及 F 值明显高于其他对比算法,MAE值较其他算法更低,本文算法一定程度上优于其他算法。

表2 各算法的性能对比

Table 2 Performance comparison of each algorithm

Item	BL	BSCA	RFCN	MDF	DCL	DHS	UCF	Proposed
MAE	0.216	0.182	0.107	0.105	0.074	0.060	0.078	0.059
F -measure	0.760	0.753	0.866	0.869	0.901	0.889	0.910	0.905
AUC	0.916	0.922	0.976	0.947	0.971	0.972	0.980	0.981

4 结 论

提出了基于深度残差网络与边缘监督学习的显著性检测方法。针对目前显著性检测算法边缘区域不清晰的现象,通过构建边缘残差块组成边缘监督网络,加入多尺度空洞卷积单元细化边缘特征,同时构建三分类训练模型,以图像边缘为界对前景与背景进行差异化区分,更好地区分图像显著区域和背景信息,从而获得精确的边界和更为完整的显著目标。通过在两个公认数据集SED2和ECSSD上的测试以及与其他7种优秀显著性检测方法的对比可知,本文算法在准确率、召回率、 F 值和MAE指标上优于当前流行算法。后续工作当致力于简化训练网络、提高训练效率以及提高算法在复杂场景下的检测精度。

参 考 文 献

- [1] Koch C, Poggio T. Predicting the visual world: silence is golden[J]. *Nature Neuroscience*, 1999, 2(1): 9-10.
- [2] 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, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 13824451.
- [3] Qin Y, Lu H C, Xu Y Q, *et al.* Saliency detection via cellular automata[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 110-119.
- [4] Hou X D, Zhang L Q. Saliency detection: a spectral residual approach [C] // 2007 IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN, USA. New York: IEEE, 2007: 9738156.

- [5] Guo C L, Ma Q, Zhang L M. Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform [C] // 2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE, 2008: 10140020.
- [6] 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.
- [7] Liu T, Sun J, Zheng N N, *et al.* Learning to detect a salient object [C] // 2007 IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN, USA. New York: IEEE, 2007: 9737916.
- [8] Liu T, Zheng N N, Ding W, *et al.* Video attention: learning to detect a salient object sequence[C]//2008 19th International Conference on Pattern Recognition, December 8-11, 2008, Tampa, FL, USA. New York: IEEE, 2008: 10457751.
- [9] Tong N, Lu H C, Ruan X, *et al.* Salient object detection via bootstrap learning [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 1884-1892.
- [10] Xu L, Cui G M, Zheng C P, *et al.* Fusion method of visible and infrared images based on multi-scale decomposition and saliency region extraction [J]. *Laser & Optoelectronics Progress*, 2017, 54(11): 111003.
许磊, 崔光芒, 郑晨浦, 等. 基于多尺度分解和显著性区域提取的可见光红外图像融合方法[J]. *激光与光电子学进展*, 2017, 54(11): 111003.
- [11] Zhu W J, Liang S, Wei Y C, *et al.* Saliency optimization from robust background detection[C]//2014 IEEE Conference on Computer Vision and

- Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 2814-2821.
- [12] Zhao R, Ouyang W L, Li H S, *et al.* Saliency detection by multi-context deep learning [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 1265-1274.
- [13] 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, USA. New York: IEEE, 2018: 714-722.
- [14] Li G B, Yu Y Z. Deep contrast learning 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: 478-487.
- [15] Bi L H, Liu Y C. Plant leaf image recognition based on improved neural network algorithm[J]. *Laser & Optoelectronics Progress*, 2017, 54(12): 121102.
毕立恒, 刘云潺. 基于改进神经网络算法的植物叶片图像识别研究[J]. *激光与光电子学进展*, 2017, 54(12): 121102.
- [16] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [17] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks[C] // Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. USA: Curran Associates Inc., 2016: 379-387.
- [18] Lin G S, Milan A, Shen C H, *et al.* RefineNet: multi-path refinement networks for high-resolution semantic segmentation [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5168-5177.
- [19] Feichtenhofer C, Pinz A, Wildes R P. Spatiotemporal multiplier networks for video action recognition[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 7445-7454.
- [20] Mahendran A, Vedaldi A. Understanding deep image representations by inverting them [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 5188-5196.
- [21] 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.
- [22] Chen L C, Papandreou G, Kokkinos I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [23] Jia Y Q, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding [C] // Proceedings of the 22nd ACM international conference on Multimedia, November 3-7, 2014, Orlando, Florida, USA. New York: ACM, 2014: 675-678.
- [24] 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.
- [25] Wang L Z, Wang L J, Lu H C, *et al.* Saliency detection with recurrent fully convolutional networks [M] // Leibe B, Matas J, Sebe N, *et al.* Computer Vision - ECCV 2016. Cham: Springer, 2016: 825-841.
- [26] Wang L J, Lu H C, 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), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3183-3192.
- [27] Zhao H, An W S. Image salient object detection combined with deep learning [J]. *Laser & Optoelectronics Progress*, 2018, 55(12): 121003.
赵恒, 安维胜. 结合深度学习的图像显著目标检测[J]. *激光与光电子学进展*, 2018, 55(12): 121003.
- [28] 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.
- [29] Li G B, Yu Y Z. Deep contrast learning 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: 478-487.