结合深度学习的图像显著目标检测

赵恒*, 安维胜**

西南交通大学机械工程学院, 四川 成都 610031

摘要 基于一种改进的跨层级特征融合的循环全卷积神经网络,提出了一种结合深度学习的图像显著目标检测算法。通过改进的深度卷积网络模型对输入图像进行特征提取,利用跨层级联合框架进行特征融合,生成了高层语义特征的初步显著图;将初步显著图与图像底层特征融合进行显著性传播以获取结构信息;利用条件随机场对显著性传播结果进行优化,得到了最终显著图。利用大型数据集将所提算法与其他多种算法进行了测试对比,研究结果表明,在对复杂场景图像的显著目标检测方面,所提算法稳健性更好,显著目标检测的完整性提升,背景得到了更有效的抑制。

关键词 图像处理; 显著目标检测; 神经网络; 特征融合; 显著图

中图分类号 TP391.4 **文献标识码** A **doi**: 10.3788/LOP55.121003

Image Salient Object Detection Combined with Deep Learning

Zhao Heng*, An Weisheng**

School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, *China*

Abstract An algorithm of image salient object detection combined with deep learning is proposed based on an improved recurrent deep convolutional neural network with the cross-level feature fusion. The feature extraction of input images is performed through this improved recurrent deep convolutional neural network model. The cross-level joint framework is used for the feature fusion and thus the initial salient maps with high-level semantics features are generated. The saliency propagation is applied to the fusion of initial salient maps and low-level image features, and thus the structural information is obtained. The saliency propagation results are further optimized with the conditional random field and the final salient maps are realized. With the massive datasets, the proposed algorithm is tested and compared with other algorithms. The research results show that the proposed method is more robust than the existing algorithms in the image salient object detection of the complex scenes. Moreover, the integrity of the significant target detection is improved and the background is suppressed effectively.

Key words image processing; salient object detection; neural network; feature fusion; salient map

OCIS codes 100.3010; 150.0155; 150.1135; 110.2970; 330.7326

1 引 言

在当今网络普及、图像信息爆炸的时代,人类仅通过自身的视觉感知系统处理外部图像信息变得越发困难,利用计算机进行图像信息处理成为一种有效的方法。研究人员模拟人类能够对图像中感兴趣区域进行迅速关注的视觉感知机制,提出了图像显著目标检测方法。该方法能够提取图像场景关键信息,在有限资源下进行后续处理时可大幅度减少计算量,在图像索引[1]、图像场景理解[2]、目标感知[3]、图像视频压缩[4]等领域的应用日益广泛。随着研究的进一步深入,图像显著目标检测逐渐成为计算机视觉、神经生物学、心理学等多学科交叉的一个综合研究课题[5]。

Itti 等[6]以数学计算模型描述了显著性特征,其后研究者们提出了各种新的算法模型,根据数据处理方式,主要分为自顶向下和自底向上两类模型[7]。自底向上模型从专注于视觉注视点的检测[8]逐步发展为主要对图像目标进行检测[9-10],其基于数据驱动[11],面对复杂语义场景时缺陷较为突出。

收稿日期: 2018-05-14; **修回日期**: 2018-06-05; **录用日期**: 2018-06-08

基金项目: 四川省科技支撑计划(2016GZ0194)

* **E-mail**: zhlance@foxmail.com; ** **E-mail**: anweisheng@home.swjtu.edu.cn

底层颜色对比度等特征主要是在视觉关注的早期起作用,在大脑皮层感知信息丰富的情况下,高层语义特征逐渐起主导作用[12],因此利用图像高层语义特征并结合图像底层信息能够提升显著目标的检测能力。

自顶向下的显著目标检测算法包括特征学习生成目标模型和利用目标模型生成显著图两个计算过程[13-14]。近年来,深度卷积神经网络逐步在显著目标检测领域得到应用,Li 等[15]通过融合多个尺度的深度卷积特征获取了显著图。Wang 等[16]利用循环全卷积网络模型,在第一层将原图像和底层前景先验特征作为输入,后续卷积时将前一层的输出特征和当前层特征联合作为输入计算下一层特征;Li 等[17]提出了多任务深度卷积模型。Lee 等[18]提出了高层特征和底层特征差异图联合编码模型,基于视觉几何网络 16(VGG16)模型[19]输出了高层语义特征图,再与底层先验特征图同权值相加融合。Li 等[20]提出了多尺度全卷积网络模型,将高层语义信息与图像超像素特征融合的二值掩码图重新映射到卷积网络,并联合所有的卷积特征输出了显著特征图像,最后用全连接条件随机场(CRF)[21]进行了优化。Zhang[22]等基于卷积后的 Dropout 策略提出了一个显著特征学习模型。基于学习的方法相比于传统方法性能有进一步的突破,但也存在高层语义特征不突出、特征冗余、结构信息缺失、离散噪声等问题。针对现有显著性检测算法模型存在的不足,为提升图像显著目标的完整性,降低背景离散噪声,本文提出了一种新的结合深度学习的显著目标检测算法模型,利用改进的神经网络提取高层语义特征,并结合图像底层特征信息进行显著性优化,获取了图像结构和边缘细节信息。

2 算法模型

2.1 网络模型结构

所提算法模型的整体结构如图 1 所示,包括利用神经网络模型生成高层语义特征的初始显著图和结合底层特征优化两个阶段,其中 Conv 为卷积;Max pooling 为最大值池化;Ave pooing 为平均值池化;Deconv 为逆卷积;concat 为联合;ReLU 为线性修正单元层;smaps 为最终输出的显著图。图 2 所示为所提算法模型对图像进行显著性检测的处理过程示例。数据集合 $\boldsymbol{D}=\{(\boldsymbol{X}^n,G^n)\}$,其中 $\boldsymbol{X}^n$ 为训练的输入图像,G^n 为输入图像对应的真值图像,$n=1,2,\cdots,N$。基于 VGG16 模型的循环全卷积算法模型(RFCN)的特征传递过程为

$$f_s(\boldsymbol{X}_{c+1}^n)=\boldsymbol{W}*(\boldsymbol{X}_c^n\oplus\boldsymbol{X}_{c-1}^n)+\boldsymbol{b},\tag{1}$$

式中 $\boldsymbol{X}_{c-1}^n$ 和 $\boldsymbol{X}_c^n$ 分别为输入图像处于第$(c-1)$和第(c)阶段的特征;$f_s(\boldsymbol{X}_{c+1}^n)$为输出的第 $c+1$ 阶特征,即将前一层特征和当前层特征联合作为下一层卷积计算层的输入;下标 s 标记特征;$c=1$ 时,前一层特征由底层显著先验特征替代;$\boldsymbol{W}$ 为卷积核;$\boldsymbol{b}$ 为偏置;$*$ 代表卷积运算过程;$\oplus$代表跨层级特征联合时所需的一系列运算过程,包括逆卷积层、裁剪层和特征联合层等。

由(1)式可知,RFCN 若从第一层就开始循环卷积,逆卷积层、裁剪层和联合层等增多必然导致网络结构的复杂性增大,特征冗余。由深度对比学习(DCL)算法[20]可知,浅卷积层可以较多地获取图像底层特征信息,深卷积层可以较好地定位目标区域,但 DCL 算法模型在 VGG16 模型前四个阶段的池化层后都增加了额外的卷积层,并提升了算法的整体性能,说明在显著目标的检测性能提升中深卷积层特征占据着主导作用。此外,不同语义层级的深度卷积层能对浅卷积层的特征信息进行共享[17]。因此,为避免在联合不同阶段特征信息时造成网络特征冗余,并在突出高层语义特征时获取一定底层信息,在(1)式的基础上设定 $c=6$,即在原 VGG16 模型第 5 阶段的基础上进一步卷积生成第 6 阶段特征,然后将第 5 阶段特征作为前景先验与第 6 阶段特征进行跨层级的联合,计算生成第 7 阶段特征信息;采用深卷积层和浅卷积层的中间层第 4 阶段的特征信息作为底层信息的补充,让整个网络模型在突出高层语义信息的同时共享一定的底层信息。联合第 4,6,7 阶段的特征信息作为最终的特征输出。计算过程为

$$f_s(\boldsymbol{X}_6^n)=\boldsymbol{W}*\boldsymbol{X}_5^n+\boldsymbol{b},\tag{2}$$

$$f_s(\boldsymbol{X}_7^n)=\boldsymbol{W}*(\boldsymbol{X}_5^n\oplus\boldsymbol{X}_6^n)+\boldsymbol{b},\tag{3}$$

$$f_s(\boldsymbol{X}_r^n)=\boldsymbol{W}*(\boldsymbol{X}_4^n\oplus\boldsymbol{X}_6^n\oplus\boldsymbol{X}_7^n)+\boldsymbol{b},\tag{4}$$

$$f^n=\mathrm{sig}\{H_s[f_s(\boldsymbol{X}_r^n;\boldsymbol{\theta});\boldsymbol{\alpha}]\},\tag{5}$$

式中 $f_s(\boldsymbol{X}_r^n)$为输出的最终卷积特征图;r 为输出的特征图阶次;$\boldsymbol{\theta}$ 为卷积特征传递过程中涉及到的参数集合;$f_s(\boldsymbol{X}_r^n;\boldsymbol{\theta})$为神经卷积网络结构的输出特征图;$\boldsymbol{\alpha}$ 为卷积特征图生成初步显著图所设定的参数集合;$H_s(;\boldsymbol{\alpha})$为卷积特征图生成显著图的逆卷积和裁剪等运算过程;sig$(\cdot)$为 sigmod 激活函数;f^n 为神经网络模型输出的全分辨率显著图。

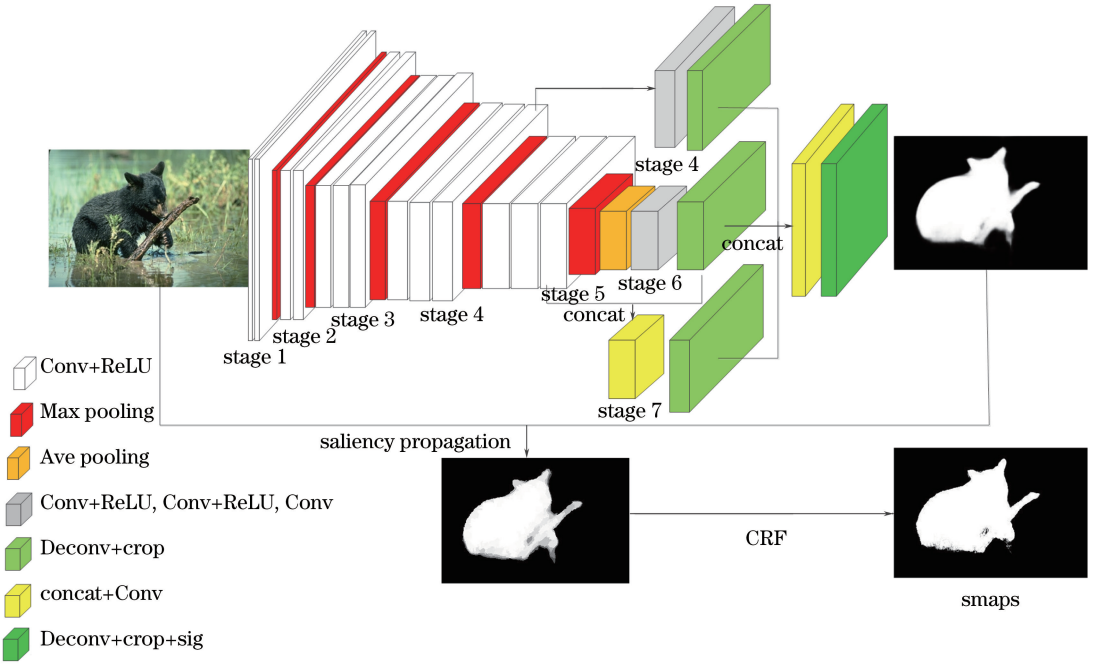

图 1 算法模型

Fig. 1 Schematic of algorithm

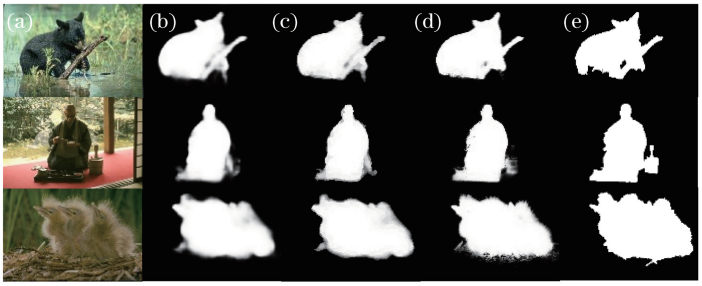

图 2 图像处理示例。(a)原图像;(b)卷积图像;(c)聚类图像;(d) CRF 优化图像;(e)真值图

Fig. 2 Examples of image processing. (a) Initial image; (b) convolved images; (c) clustered images; (d) CRF optimized images; (e) truth value chart

(2)、(3)式为三个跨阶段特征的传递过程,每个阶段都包含三个卷积层和两个 ReLU。(5)式表示对卷积特征图进行逆卷积和裁剪等运算以获取全分辨率显著图像。神经卷积网络模型训练的过程主要是提取特征并迭代求解收敛参数集合 $\boldsymbol{\theta}$,利用交叉熵损失函数评价迭代优化过程是否收敛,其定义为

$$L=-\beta_i\sum_{i=1}^{[\boldsymbol{X}^n]}G_i^n\lg f_i^n(G_i^n=1\mid\boldsymbol{X}^n;\boldsymbol{\theta})-(1-\beta_i)\cdot\sum_{i=1}^{[\boldsymbol{X}^n]}(1-G_i^n)\lg f_i^n(G_i^n=0\mid\boldsymbol{X}^n;\boldsymbol{\theta}),\tag{6}$$

式中 $[\boldsymbol{X}^n]$ 为图像像素个数;β_i 为真值图像中像素标签为 0 的数目占比;G_i^n 为某一像素 i 的真值标签;$f_i^n(G_i^n=1|\boldsymbol{X}^n;\boldsymbol{\theta})$为原图像在网络模型参数集合为 $\boldsymbol{\theta}$ 的情况下输出图像像素 i 属于标签 1 的概率;$f_i^n(G_i^n=0|\boldsymbol{X}^n;\boldsymbol{\theta})$为像素 i 属于标签 0 的概率。对计算过程进行不断迭代并求解更新 $\boldsymbol{\theta}$,当 L 值没有产生大的波动时,迭代运算便达到收敛的状态,可求得最终的参数集合 $\boldsymbol{\theta}_z$,生成目标神经网络模型。对于单幅图像输入 I,利用求解的参数 $\boldsymbol{\theta}_z$,代入(5)式便可求得初始显著图 f。

2.2 底层特征优化

图 2(b)所示为卷积神经网络输出的初始显著图,和图 2(e)所示的真值图相比,其能够对目标区域进行较好的定位,但目标边界结构边缘的细节较为模糊,需要进一步处理增加图像结构的细节信息。简单线性迭代聚类(SLIC)[23]将图像根据颜色特征分割为大小较均匀的子区域,能够去除区域不太重要的颜色特征冗余,并保留图像目标与背景的结构边界细节信息。文献[20]虽考虑了图像目标形状细

节的信息缺失,但利用底层超像素信息生成掩码映射到卷积特征图时,卷积核是大小较为规则的 $k\times k$ 格窗,结构信息仍会丢失。为增加显著目标的结构信息并避免产生过分割,在利用 SLIC 时融合初始显著图信息,将图像底层特征与高层语义显著特征结合用于图像聚类分割并进行显著性传播,可以突出高层语义特征的作用,最后利用 CRF 进行像素级优化。

显著性传播的基本过程为:1) 提取(l,a,b,f_i,x,y)六维特征,其中(l,a,b)为国际照明委员会(CIE)-Lab 空间的像素颜色特征,f_i 为像素显著值,(x,y)为像素坐标;2) 设图像像素数目为 K,超像素初始聚类中心点数目为 m,则初始聚类中心点间距 $S=\sqrt{K/m}$;3) 计算聚类中心点 $2S\times 2S$ 区域内像素特征的差异,将与聚类中心特征差异较小的像素合并到聚类中心区域,更新聚类中心的六维平均特征;4) 重复上述计算过程直至收敛。

两个像素点的特征差异为

$$D_{\mathrm{t}}=d_{\mathrm{cy}}+\frac{h}{S}d_{\mathrm{k}}, \tag{7}$$

式中 D_{t} 值越小说明越相似,反之差异较大,其中下标 t 标记颜色与空间的特征差异;h 为常数,综合代表颜色显著性和空间特征在相似性度量中的权重;d_{cy}为任意两像素点 i、j 的颜色和显著性特征的欧式距离,下标 cy 标记颜色特征;d_{k} 为像素点 i、j 的空间距离;k 非变量仅标记空间特征。其中

$$d_{\mathrm{cy}}=\sqrt{(l_i-l_j)^2+(a_i-a_j)^2+(b_i-b_j)^2+\zeta^2(f_i-f_j)^2}, \tag{8}$$

$$d_{\mathrm{k}}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}, \tag{9}$$

式中(x_i,y_i)、(x_j,y_j)分别为像素点 i、j 的空间坐标;(l_i,a_i,b_i)、(l_i,a_i,b_i)分别为像素点 i、j 在 CIE-Lab 空间的颜色特征;ζ 为显著性常系数,衡量显著性特征在聚类中的作用;f_i,$f_j\in[0,1]$分别为卷积显著图像素点 i、j 的显著值。根据特征差异进行迭代聚类收敛后,图像像素显著值 f_i^{s} 的计算公式为

$$f_i^{\mathrm{s}}=\frac{1}{N^C}\sum_{j\in C}^{N^C}f_{cj},(i\in C), \tag{10}$$

式中 N^C 为 C 类像素的个数;f_{cj} 为 C 类像素的显著值,其表达式为

$$f_{cj}=\begin{cases}f_i^{\mathrm{s}}, & \text{else}\\ 0, & f_i^{\mathrm{s}}<0.3\end{cases}。 \tag{11}$$

当超像素的平均显著特征小于 0.3 时,超像素区域内大部分像素点的显著性都较低,应当属于强背景,因此区域内所有像素值置为 0。由图 2(c)可知,显著目标区域没有大的变化,但观察图像细节发现,经过显著性传播后离散噪声得到了较好的清除,并且显著目标的边界变得更加清晰明确,是目标结构更加完整的表现。CRF[21] 是一个二值标签分类框架,在图像目标结构信息较为完整的情况下能够在像素级水平增加边缘细节信息。CRF 优化过程为能量函数 $E(S')$的最优解计算过程:

$$E(S')=-\sum_i\lg P(f_i^{\mathrm{s}})+\sum_{i,j}w_{ij}(f_i^{\mathrm{s}},f_j^{\mathrm{s}}), \tag{12}$$

$$w_{ij}(f_i^{\mathrm{s}},f_j^{\mathrm{s}})=\mu(f_i^{\mathrm{s}},f_j^{\mathrm{s}})D_{\mathrm{Tc}}, \tag{13}$$

$$D_{\mathrm{Tc}}=\left[\omega_1\exp\left(-\frac{\|p_i-p_j\|^2}{2\delta_\alpha}-\frac{\|I_i-I_j\|^2}{2\delta_\beta}\right)+\omega_2\exp\left(-\frac{\|p_i-p_j\|^2}{2\delta_\gamma}\right)\right], \tag{14}$$

式中 S'代表输出图像的标签值,即图像中各个像素属于显著标签 1 或背景标签 0;$P(f_i^{\mathrm{s}})$为像素 i 属于 1 或 0 标签的概率;$w_{ij}(f_i^{\mathrm{s}},f_j^{\mathrm{s}})$为二元势函数联合代表颜色和空间特征的差异;一般 $\mu(f_i^{\mathrm{s}},f_j^{\mathrm{s}})=1$,只有当 $i=j$ 时,$\mu(f_i^{\mathrm{s}},f_j^{\mathrm{s}})=0$;$D_{\mathrm{Tc}}$为高斯核能量项;$p_i$ 和 I_i 分别为像素 i 的空间和颜色特征;ω_1、ω_2、δ_α、δ_β、δ_γ 联合控制空间和颜色特征在势函数中的作用。如图 2(d)所示,经过 CRF 优化的显著图像,整个目标区域内部均匀一致高亮,边界清晰明确,边缘细节信息丰富,与真值图的接近程度进一步提升。

2.3 参数设置

基于 VGG16 模型改进的神经网络模型选用在显著目标检测中应用较为广泛的微软亚洲研究院的图像数据集(MSRA-B)[17] 作为训练集,MSRA-B 包含 2500 张自然场景图像及其对应的人工标记真值图,场景语义多样。将原图像与其对应的真值图像输入网络模型进行训练,各个初始参数设置为:基础学习率 10^{-8},权重衰减系数 0.0005,动量 0.9,批处理数量设置为 1,设置的初始最大迭代次数为 15000,采用随机梯度下降(SGD)的学习率衰减方式训练整个神经网络。通过神经网络模型设定的初始参数进行训练迭代优化求解 $\boldsymbol{\theta}$,模型迭代次数在 11000 次左右时,交叉熵损失 L 开始保持平稳;迭代次数在 12000~15000 时,L 值的迭代变化量小于

5%;损失值 L 最终保持在 15000 左右平稳波动收敛的水平,具体损失值与迭代步长的变化关系如图 3 所示,整个网络模型的迭代优化训练过程耗时 7.5 h。

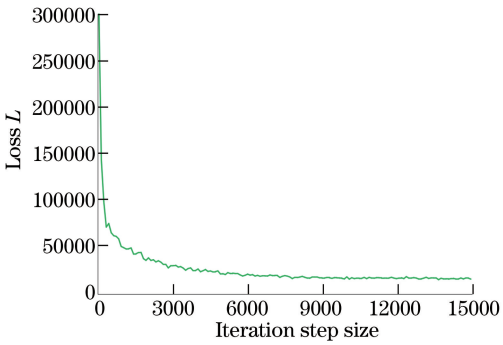

图 3 损失 L 随迭代步数的变化

Fig. 3 Loss L versus number of iterations

在实验测试中,显著性传播过程中涉及到的初始聚类数目 $m\in(800,1200)$和显著性常系数 $\zeta\in(180,250)$时,显著性检测结果没有太大的变化,最终设定 $m=900$,$\alpha=240$,其他参数根据文献[22]采取默认值。全连接 CRF 涉及到的参数 ω_1、ω_2、δ_α、δ_β、δ_γ 根据文献[20]采取默认值。

2.4 算法模型

在训练过程利用交叉熵损失函数迭代优化求解模型参数 $\boldsymbol{\theta}_z$,获得目标深度卷积神经网络模型。生成模型后,对单幅图像的处理计算流程如图 1 所示,基本步骤如下。

输入:一幅图像 I。

1) 在 VGG16 模型的基础上,根据(2)~(5)式获取全分辨率初始显著图 f。

2) 利用图像 I 底层信息和高层语义信息 f,根据(7)~(9)式获取超像素分割结果,再根据(10)、(11)式进行显著性传播获取 f^s。

3) 根据(12)~(14)式迭代优化 f^s 获取最终显著图 smaps。

输出:显著目标检测图像。

3 实验评价分析

将在显著性检测领域内应用较多的扩展复杂场景显著数据集(ECSSD)和大连理工大学的欧姆龙数据集(DUT-OMRON)两个大型数据集[17]与其他 10 种显著性检测算法进行了实验对比,通过准确率-召回率(P-R)曲线图、综合评价参数 F 值以及平均绝对误差(MAE)三种客观评价指标评测所提算法与其他算法的性能。实验基于 Intel 中央处理器(CPU)和英伟达 GTX 图形处理器(GPU),采用 Python2.7、Visual Studio 2013 软件,以 Python 和 C++ 语言进行编程处理,深度学习框架基于 Caffe[24]。

ECSSD 包含 1000 张自然场景图像及相应的人工标记真值图,DUT-OMRON 含有 5168 张自然场景图像及相应的人工标记真值图,是人工从超过 140000 张自然场景的图像中挑选出来的,每个图像中均含有多个显著目标。DUT-OMRON 图像相比于 ECSSD 图像自然场景更为复杂,显著目标检测的难度更大[17]。测试的算法包括 DCL[20],DRFI[14],DS[17],ELD[18],FT[8],MDF[15],QCUT[9],RC[10],RFCN[16](首字母排序),其中 FT 是早期经典的基于像素层面的显著性检测算法;QCUT 和 RC 是近年自底向上模型中具有代表性的算法[7];DRFI 是早期集合多特征学习的自顶向下回归森林分类算法;DCL,DS,ELD,MDF,RFCN 是基于深度学习的先进算法。所对比的显著性检测算法的图像结果由文献作者公开网页提供或网页源代码生成。

3.1 显著性检测客观评价

将显著图按阈值从 0 到 255 进行固定阈值分割后,与真值图比较计算准确率 P 和召回率 R 值,并在直角坐标系进行绘制得到 P-R 曲线图。P、R 的计算公式分别为

$$P=\frac{\sum_{(x,y)}G_z(x,y)S_b(x,y)}{\sum_{(x,y)}S_b(x,y)},\tag{15}$$

$$R=\frac{\sum_{(x,y)}G_z(x,y)S_b(x,y)}{\sum_{(x,y)}G_z(x,y)},\tag{16}$$

式中 $G_z(x,y)$为真值图的灰度值;$S_b(x,y)$为显著图阈值分割后的灰度值,下标 b 标记二值分割。图 4 所示为两个数据集多种算法的 P-R 曲线图对比。可以看出,P-R 曲线在较高准确率区间持续范围越广、越靠近坐标系右上角区域,算法性能越优越。在两个数据集上,基于学习的算法比基于底层特征的算法在整体性能上更加优越。在 ECSSD 数据集上,所提算法和 DCL、RFCN 算法都较为突出,其 P-R 曲线的表现没有太明显的差异,但观察坐标细节可以看出,准确率在 90%以上时所提算法优于其他对比算法;DUT-OMRON 数据集自然场景的复杂性高于 ECSSD 的,所有算法的测试结果比在 ECSSD 数据集上的结果都更远离坐标系右上角,仍是 DCL、RFCN 和所提算法较为优越,并且所提算

法的 P-R 曲线相比于其他算法有更加突出的表现，最接近坐标右上角区域，准确率高于 70% 的高水平持续区间范围最广，明显优于 DCL、RFCN 算法。综合来看，在 ECSSD 数据集上所提算法的 P-R 曲线表现略优于其他算法的，而在难度更大的 DUT-OMRON 数据集上，所提算法的优越性较为明显，说明面对场景更为复杂的图像，所提算法的稳健性好。

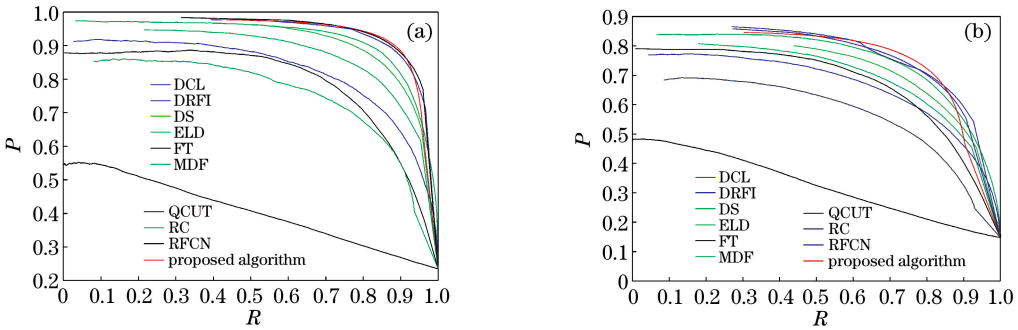

图 4 两个数据集多种算法的 P-R 曲线图对比。(a) ECSSD；(b) DUT-ORMON

Fig. 4 Comparison among P-R curves by several algorithms. (a) ECSSD; (b) DUT-ORMON

取显著图像素灰度平均值的 2 倍进行自适应阈值分割，阈值计算公式为

$$Y_z=\frac{2}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}S(x,y), \tag{17}$$

式中 $S(x,y)$ 为整个算法模型输出的显著图灰度值；M 和 N 代表输入图像的大小。

F 值的计算公式为

$$F=\frac{(1+\psi^2)PR}{\psi^2P+R}, \tag{18}$$

式中 ψ 为衡量准确率与召回率的权值。通常情况下需突出算法检测的准确率，设 $\psi^2=0.3$。F 值越大，算法性能越好。两个数据集上 F 值的对比见表 1。可以看出，在两个数据集上，所提算法的 F 值高于其他所有对比算法的结果，说明所提算法的准确率有明显的提升，进一步提高了图像显著目标检测的完整性，算法性能优于其他算法的性能。P-R 曲线图、F 值反映显著目标是否较为完整均匀一致，对于显著图也需要考虑图像离散背景的影响。MAE 反映显著图与真值图在整体上包括背景区域的相似程度，计算公式为

$$f_{\mathrm{MAE}}=\frac{1}{M\times V}\sum_{x=1}^{M}\sum_{y=1}^{V}\left|S(x,y)-G_z(x,y)\right|, \tag{19}$$

式中 M 和 V 代表输入图像的大小。MAE 值越小，显著图目标区域与真值图的相似程度越高，背景区域噪声越少，算法的整体性能也更优越。不同算法的 MAE 值见表 1。可以看出，在两个数据集上，所提算法的 MAE 值优于其他所有对比算法的，与真值最为相似，说明相对于其他算法，所提算法的目标区域均匀一致突显程度高，背景区域噪声也得到了很好的抑制。

表 1 F 值及 MAE 的对比

Table 1 Comparison of F and MAE

Algorithm	ECSSD		DUT-OMRON	
	F	MAE	F	MAE
Proposed	0.877	0.046	0.689	0.048
DCL	0.863	0.047	0.681	0.053
DRFI	0.732	0.080	0.545	0.083
DS	0.824	0.075	0.598	0.062
ELD	0.814	0.050	0.609	0.059
FT	0.379	0.121	0.311	0.106
MDF	0.795	0.079	0.603	0.064
QCUT	0.743	0.121	0.495	0.072
RC	0.680	0.098	0.678	0.110
RFCN	0.873	0.047	0.681	0.057

3.2 视觉效果对比

显著性检测算法来源于人类的视觉注意机制，显著性检测结果的好坏也应当从视觉观感上进行评判。图 5 所示为部分图像不同显著性检测算法处理生成的结果示例，其中包含原图像及真值图像。可以看出，针对各种不同复杂自然场景图像，各个显著性检测算法的结果在不同程度上都与真值图存在差异，但所提算法的结果和真值图最为接近。经典算法 FT 模型的检测结果较差，其在像素层面关注显著性，凸显像素点离散，无法形成具体的目标；QCUT、RC 算法的结果相对较好，但其基于区域对比的底层特征进行显著性检测，图像目标相对于背景特征比较杂乱，并且在不具有明显差异时，显著目标不够平滑完整，离散背景的凸显程度高；早期的学习型 DRFI 算法，虽然目标定位较为准确，但是出现了大面积的背景噪声，显著目标的完整性缺失严重；

所提算法及 DCL、DS、ELD、MDF、RFCN 算法在整体上都能够提取目标区域，但从细节上看，所提算法的检测结果的目标完整性最高，目标区域凸显高亮的程度均匀一致，边缘的细节信息丰富，而且背景噪声最少。从视觉效果看，所提算法对复杂背景图像进行处理时，不仅能够均匀一致地高亮显著目标，而且对背景区域进行了很好的抑制，图像目标的完整性提高。

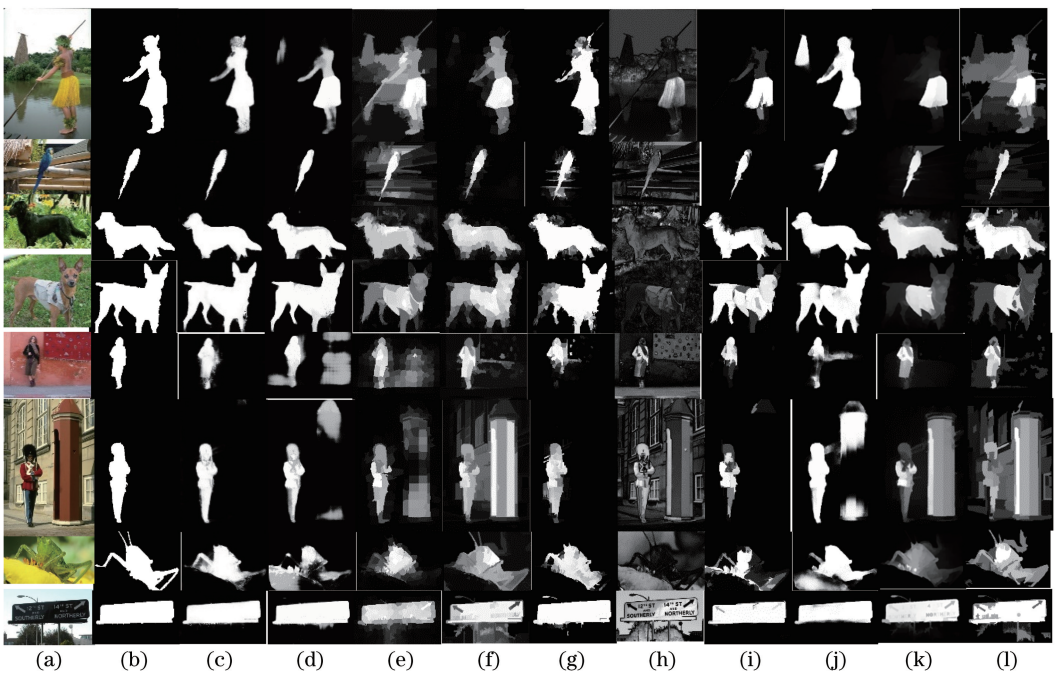

图 5 各个算法的显著性检测结果对比。(a)原图像；(b) GT；(c)所提算法；(d) DCL；(e) DRFI；(f) DS；(g) ELD；(h) FT；(i) MDF；(j) QCUT；(k) RC；(l) RFCN

Fig. 5 Comparison among image salient detection results by several algorithms. (a) Initial images; (b) GT; (c) proposed algorithm; (d) DCL; (e) DRFI; (f) DS; (g) ELD; (h) FT; (i) MDF; (j) QCUT; (k) RC; (l) RFCN

3.3 效率对比

基于深度学习的算法性能效果比其他算法的好，因此只将性能较好的算法的运行时间进行对比，处理分辨率约为 400 pixel×400 pixel 的单张图像的平均耗时见表 2。

表 2 耗时对比

Table 2 Comparison of time consuming

Algorithm	Proposed	DCL	DS	ELD	MDF	RFCN
Time /s	0.3	1.5	1.1	0.5	8	4.6

所提算法是一种端到端的显著性检测算法模型，只需输入原始图像便可生成相应的显著图。RFCN 算法存在底层先验特征、网络结构、底层空间特征置信度、颜色特征置信度、数学形态学处理等运算过程，整个算法模型的计算复杂度较高。所提算法的神经网络模型降低了冗余，采用的后处理方法基于成熟高效的以显著性引导的 SLIC 和 CRF 传播方法，算法效率提升明显，完成一张图像处理耗时只需 0.3 s。所提算法基于 RFCN 算法模型进行了改进，算法性能提升，时间效率也较高。所提算法比其他基于深度学习的算法的时间效率都要高，满足显著性检测的实时性要求，在高效率下算法性能表现也比较好。

4 结 论

提出了结合深度学习的图像显著目标检测算法。针对现有结合卷积神经网络的算法模型存在的不足，对深卷积层特征进行了跨层级联合，突出了高层语义特征在显著性检测中占据的主导性作用。在进行 CRF 优化前，针对结构信息的不足，将卷积网络输出的显著图与底层特征融合进行图像区域聚类显著性传播，图像目标区域和结构细节信息得到了较好的整合，边缘信息得到了有效的保留。在两个大型数据集上，与其他 10 种显著性算法进行了实验对比，三个客观评价指标、视觉效果图以及时间运行效率的对比结果证明了在复杂自然场景图像显著目标检测中，所提算法是有效的，能够在高效率下更加精准地分离前景与背景，使显著目标的完整性得到提升，背景噪声少，算法稳健性好，具有一定的使用价值。

参 考 文 献

[1] Zheng L, Wang S J, Liu Z Q, *et al*. Fast image retrieval: Query pruning and early termination[J]. IEEE Transactions on Multimedia, 2015, 17(5): 648-659.

[2] Zhu J Y, Wu J J, Wei Y C. Unsupervised object class discovery via saliency-guided multiple class learning[J]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3218-3225.

[3] 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.

[4] Hadizadeh H, Bajic I V. Saliency-aware video compression [J]. IEEE Transactions on Image Processing, 2014, 23 (1): 19-33.

[5] Chen Z H, Wang H Z, Zhang L M, *et al*. Visual saliency detection based on homology similarity and an experimental evaluation[J]. Journal of Visual Communication and Image Representation, 2016, 40: 251-264.

[6] 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.

[7] Borji A, Cheng M M, Jiang H Z, *et al*. Salient object detection: A benchmark [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5706-5722.

[8] Achanta R, Hemami S, Estrada F, *et al*. Frequency-tuned salient region detection[J]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1597-1604.

[9] Aytekin C, Kiranyaz S, Gabbouj M. Automatic object segmentation by quantum cuts[J]. Proceeding of IEEE International Conference on Pattern Recognition, 2014: 112-117.

[10] Cheng M M, Mitra N J, Huang X L, *et al*. Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569-582.

[11] Mou L, Zhang X W, Zhang Z, *et al*. Saliency detection optimization method in natural scene[J]. Laser & Optoelectronics Progress, 2016, 53(12): 121501.
牟丽, 张学武, 张卓, 等. 自然场景下的显著性检测优化方法[J]. 激光与光电子学进展, 2016, 53(12): 121501.

[12] Liu N, Han J W, Zhang D W, *et al*. Predicting eye fixations using convolutional neural networks[J]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 362-370.

[13] Lin C, He B W, Dong S S. An indoor object fast detection method based on visual attention mechanism of fusion depth information in RGB image[J]. Chinese Journal of Lasers, 2014, 41(11): 1108005.
林昌, 何炳蔚, 董升升. 融合深度信息的室内 RGB 图像视觉显著物体快速检测方法[J]. 中国激光, 2014, 41(11): 1108005.

[14] Wang J D, Jiang H Z, Yuan Z J, *et al*. Salient object detection: A discriminative regional feature integration approach[J]. International Journal of Computer Vision, 2017, 123(2): 251-268.

[15] Li G B, Yu Y Z. Visual saliency based on multiscale deep features[J]. Proceeding of IEEE Computer Vision and Pattern Recognition, 2015: 5455-5463.

[16] Wang L Z, Wang L J, Lu H C, *et al*. Saliency detection with recurrent fully convolutional networks [C]. European Conference on Computer Vision, 2016: 825-841.

[17] Li X, Zhao L M, Wei L N, *et al*. Deep saliency: Multi-task deep neural network model for salient object detection[J]. IEEE Transactions on Image Processing, 2016, 25(8): 3919-3930.

[18] Lee G, Tai Y W, Kim J. Deep saliency with encoded low level distance map and high level features[J]. Proceeding of IEEE Computer Vision and Pattern Recognition, 2016: 660-668.

[19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2018-03-02]. https://arxiv.org/abs/1409.1556.

[20] Li G B, Yu Y Z. Deep contrast learning for salient object detection[J]. Proceeding of IEEE Computer Vision and Pattern Recognition, 2016: 478-487.

[21] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[EB/OL]. (2012-10-20)[2018-03-02]. https://arxiv.org/abs/1210.5644.

[22] Zhang S L, Xie L B. Salient object 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.

[23] Achanta R, Shaji A, Smith K, *et al*. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.

[24] Jia Y Q, Shelhamer, Donahue J, *et al*. Caffe: Convolutional architecture for fast feature embedding[EB/OL]. (2014-06-20)[2018-03-02]. https://arxiv.org/abs/1408.5093.