基于全部卷积特征融合的显著性检测

张松龙，谢林柏*

江南大学物联网工程学院，江苏 无锡 214122

摘要 如何充分利用各级卷积特征是当前显著性检测研究的关键问题。就此提出一种基于融合全部卷积层特征的全卷积神经网络显著性检测方法。首先，将全部卷积特征映射到内部的多个尺度中，在每个尺度上联合各级卷积特征预测显著图；然后，融合各尺度的显著图，得到融合显著图；最后，通过全连接条件随机场平滑显著图和优化显著边界。实验结果表明，该方法在 ECSSD 和 SED2 数据库上具有较高的准确率、召回率和较低的平均绝对误差，可为目标识别、机器视觉等应用提供更可靠的预处理结果。

关键词 机器视觉；显著性检测；全卷积神经网络；全部卷积特征；多尺度融合

中图分类号 TP391.41 **文献标识码** A **doi**：10.3788/LOP55.101502

Salient Detection Based on All Convolutional Feature Combination

Zhang Songlong, Xie Linbo*

College of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, *China*

Abstract In the current saliency detections based on deep learning, how to make full use of the convolution features at all levels is the key issue. In order to solve this problem, we propose a saliency detection method based on full convolution neural network, which is a fusion of all convolutional features. Firstly, all the convolution features are mapped to multiple internal scales, and the saliency maps are predicted by combining the convolutional features of each level on each scale. Then the fused saliency maps are obtained by fusing the saliency maps of each scale. Finally, smooth saliency maps and optimized salient boundaries are obtained through full connected conditional random fields. Experimental results show that the proposed method has higher accuracy, recall rate and lower average absolute error in ECSSD database and SED2 database, and provides more reliable pretreatment results for target recognition, machine vision and other applications.

Key words machine vision; saliency detection; fully convolutional neural network; all convolution features; multi-scale fusion

OCIS codes 150.0155; 330.7326; 110.2970

1 引 言

近年来，显著性检测作为计算机视觉领域用来降低计算复杂度的重要预处理步骤，被广泛应用于图像缩放[1]、图像压缩[2]、目标识别[3]、图像分类[4]中，并受到越来越多的关注。虽然该领域出现了很多优秀算法，但是由于很难将所有人工特征和显著性种子有效结合，显著对象检测至今仍然是一个未解决的问题。视觉显著性算法的目的是找出图像中最明显、最能吸引注意的物体或区域。受人类视觉机制的启发，早期的显著性检测方法[5-10]利用低层次视觉特征（如颜色、纹理和对比度）和启发式先验模型，近似人类视觉显著性。然而，这些低层次特征和先验知识很难捕捉到显著对象及其周围区域的高级语义信息。

过去几年，卷积神经网络（CNN）在语义分割[11]、图像识别[12]、边缘检测[13]等计算机视觉领域取得突破性进展，推动了 CNN 应用在显著性检测领域的研究。例如：Wang 等[14]通过训练两个不同深度的 CNN 来捕捉局部信息与全局对比，整合局部估计和全局搜索来预测显著性图；Zhao 等[15]通过考虑全局和局部语义信息来预测每个超像素的显

收稿日期：2018-03-19；**修回日期**：2018-04-16；**录用日期**：2018-04-23

基金项目：国家自然科学基金(61374047,60973095)

* **E-mail**：xielb@126.com

著值,并且在深度 CNN 的多级语义信息中检测显著对象。虽然这些方法比传统方法取得了更好的效果,但它们都未能有效利用低级卷积特征,上述模型都包含全连接层,导致模型计算量大和图像空间信息丢失。Li 等[16]通过结合全卷积神经网络(FCN)和多阶段卷积特征融合实现像素级显著性预测;Wang 等[17]研究了循环全卷积神经网络,将粗糙的预测作为显著先验,逐步优化每个阶段的显著值,最终融合生成显著图。这些模型主要集中在从最后一层卷积层提取高级特征的非线性组合,没有很好地利用每一层的卷积层特征。在深度 CNN 中,高级的卷积特征具有丰富的语义信息,而低级的卷积特征包含许多有用的细节信息,如边界信息等[18]。因此,有效地融合高级和低级卷积特征,不仅能有效地预测显著区域,而且能学习到显著区域边界等细节信息。Liu[13]通过实验论证了全部卷积特征对像素级分类的高效性,通过利用各阶段的全部卷积特征做阶段性融合,在各阶段尺度上预测图像边缘,在边缘检测上取得了突出的效果。然而,前述研究只是将各阶段的卷积特征做各自的阶段性整合,未考虑各阶段卷积特征之间的互补作用。鉴于此,本文提出在全卷积神经网络模型上,通过联合各阶段的全部卷积特征,在各尺度上实现低级特征细化粗略的高级语义特征,从而得到精确的显著图。

2 多尺度全部卷积层特征

2.1 网络架构

本文在 VGG16[19]的基础上建立网络架构。由于 VGG16 模型简单、高效的特性,它在图像分类[19]、语义分割[11]等方面都有突出表现。VGG16 模型由 13 个卷积层和 3 个全连接层组成,它的卷积层被分为 5 个阶段,在每个阶段后连接内核大小为 2、步长为 2 的最大池化层。当向 VGG16 输入尺寸大小为[$W\times H$]的图片时,每一个阶段的输出尺度依次为输入尺度的 1/1、1/2、1/4、1/8、1/16,每一个卷积层捕获的有用信息随着尺寸增加而变得粗糙。基于全部卷积特征的新型网络如图 1 所示,其中,conv 表示卷积层,每个卷积层后设置批规范化(BN)和线性修正单元(ReLU),对神经元进行选择性激活,图中所有的池化操作均采用最大池化方式,deconv 层是反卷积层,将图像像素按照空间位置关系做选择性填充从而恢复到输入图像尺寸。和 VGG16 相比,网络做了如下修改:1) 去掉全连接层和 pool5 池化层。一方面,去掉全连接层得到全卷积神经网络;另一方面,pool5 池化层将会增加两倍的步长,不利于平衡上下文语义信息和图像细节信息。2) 在每个池化层之前引入 dropout 层[20],能够自适应地集成特定卷积层的内部特征单元,从而实现在不需要额外参数化的情况下检测不确定特征。3) 从 VGG16 模型中提取出每一阶段的全部特征图:conv1_1、conv1_2、conv2_1、conv2_2,conv3_1、conv3_2、conv3_3、conv4_1、conv4_2、conv4_3、conv5_1、conv5_2、conv5,共 13 个特征。4) 将上述全部卷积层特征通过全部卷积融合(AFC)方法映射到对应的 5 个尺度下融合,然后通过反卷积操作在该尺度下融合得到的特征图上采样,并对每个尺度上采样层连接多项损失函数层。5) 将每一个尺度下的上采样层得到的特征连接在一起,通过 $1\times1-64$ 的卷积层融合每个尺度的特征图,采用多项式损失函数计算损失值。

2.2 全部卷积特征多尺度下融合

由于上述提取到的每个阶段的卷积特征的尺度不同,因此采用 AFC 方法融合不同尺度的全部卷积特征,如图 2 所示。AFC 方法由收缩和伸展两个部分构成:收缩操作将特征缩小到目标尺度,通过卷积操作实现;伸展操作将图片扩张到目标尺度,由反卷积实现。通过收缩和伸展操作,将不同阶段的卷积特征映射到同一尺度,实现低级特征和高级特征的联合融合。假设 I 是输入的图片,$\gamma=\left[\frac{W}{2^k},\frac{H}{2^k}\right]$表示最终需要融合的目标尺度,其中 k 的取值包括 0、1、2、3、4。通过 AFC 方法融合 5 个尺度下的特征图,5 个尺度分别为输入图片的 1/16、1/8、1/4、1/2、1/1,代表不同的高级和低级特征。AFC 生成特征图的公式为

$$\begin{aligned}F_\gamma=W_\gamma * \mathrm{Cnt}\{&S_n[F_n(I);\varphi_n],\ldots,\\&S_1[F_1(I);\varphi_1],E_1[F_1(I);\psi_1],\ldots,\\&E_m[F_m(I);\psi_m]\},\end{aligned}\tag{1}$$

式中:* 代表卷积操作;S_n 表示参数 φ_n 下的收缩操作,目的是下采样高尺度的特征图;E_m 表示参数 ψ_m 下的扩展操作,目的是上采样低尺度的特征图,通过收缩和扩展操作,将全部卷积层输入的特征图融合在同一个尺度内;Cnt()表示横向通道的串联,即将收缩和扩展后统一尺度下的特征图串联起来;W_γ 表示串联的参数;F_γ 即是在 γ 尺度下联合融合得到的特征图;$F_n(I)$表示一个三维(3D)的张量,即网络中提取的 13 层卷积特征。

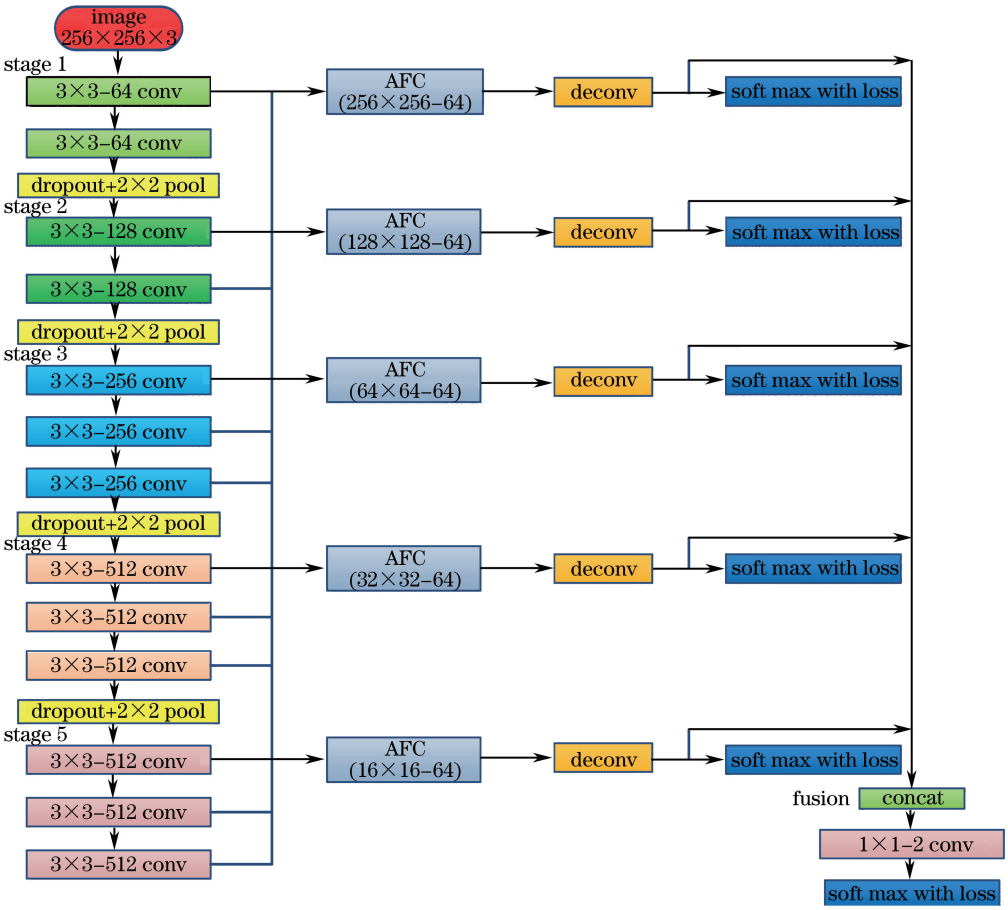

图 1 网络结构图

Fig. 1 Network structure

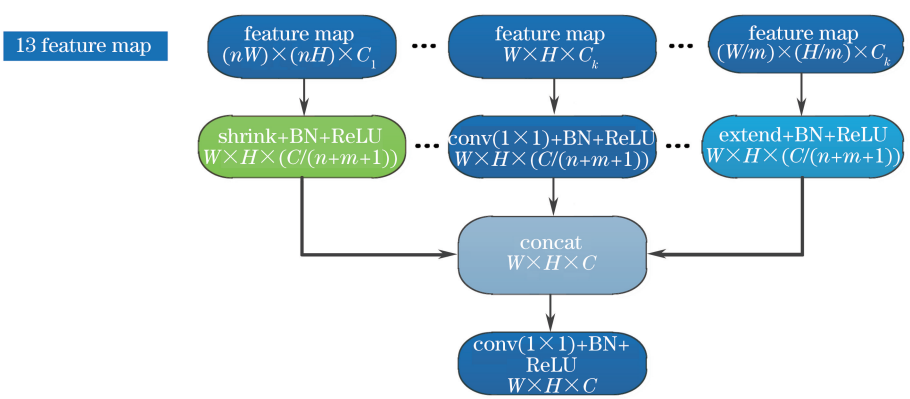

图 2 全部卷积特征融合策略

Fig. 2 All convolution feature fusion strategy

2.3 全连接条件随机场边界优化

因为本文网络是一个全卷积的网络，所以测试时能以任意大小的图像作为网络输入。输入的图像在前向传播后，网络的输出由前景激励图(P^{fe})和背景激励图(P^{be})组成，利用二者的差异，同时修剪掉错误的显著值，得到最终的显著值：

$$S=\max(P^{fe}-P^{be},0)。\tag{2}$$

这种减法策略不仅提高显著像素识别的准确率，而且捕获了上下文对比信息。

带有最大池化层的更深的模型在分类任务上非常成功，但是顶层结点的大感受野只能产生平滑响应，而不能描述显著区域的精确边界。因此，采用全连接条件随机场(CRF)[21]优化 FCN 得到的显著区域的边界。全连接条件随机场的能量函数为

$$E(x)=\sum_i\theta_i(x_i)+\sum_{i,j}\theta_{ij}(x_i,x_j),\tag{3}$$

$$\theta_i(x_i)=-\frac{\ln S_i}{\tau h(x_i)},\tag{4}$$

$$\theta_{ij}(x_i,x_j)=\mu(x_i,x_j)\times\left[w_1\exp\left(-\frac{\|p_i-p_j\|^2}{2\sigma_\alpha^2}-\frac{\|I_i-I_j\|^2}{2\sigma_\beta^2}\right)+w_2\exp\left(-\frac{\|p_i-p_j\|^2}{2\sigma_\gamma^2}\right)\right],\tag{5}$$

式中 x 代表每个像素预测的标签值。为了使全连接条件随机场更加适应显著性检测，本文未直接使用全卷积神经网络计算像素 i 处的显著值 S_i 作为一元项 $\theta_i(x_i)$ 的输入，$\theta_i(x_i)$ 采用(4)式得到，引入尺度参数 τ 和 $h(\cdot)$sigmoid 函数，使能量函数自适应地增加一元项的权重，从而抑制非显著点的干扰。$\theta_{ij}(x_i,x_j)$ 由(5)式可知，当 $x_i\neq x_j$ 且不为 0 时，$\mu(x_i,x_j)=1$。p_i 和 I_i 分别为像素 x_i 对应的位置和像素值，通过 w_1、w_2、σ_β、σ_α、σ_γ 控制高斯核的大小。最后，当两个类(前景和背景)的能量函数最小时，得到的像素值最稳定。如图 3 所示，经过 CRF 处理的显著图更平滑，边界更明确。

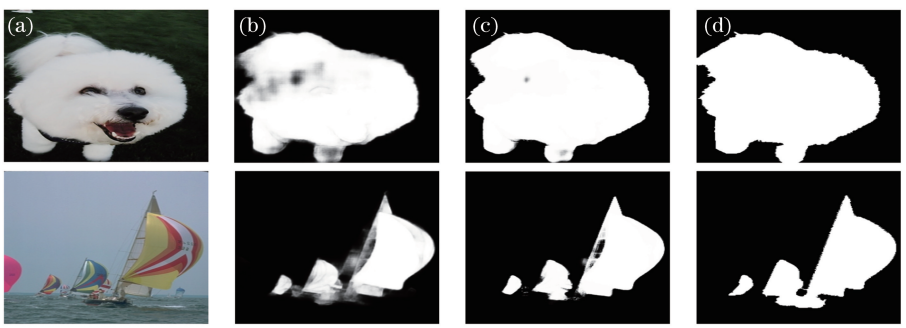

图 3 CRF 显著图效果比较。(a)原图；(b)未经 CRF 处理；(c) CRF 处理；(d)人工标注图

Fig. 3 Comparison of saliency maps results. (a) Source; (b) without CRF; (c) with CRF; (d) ground truth

3 仿真实验分析

3.1 数据集和运行环境

使用公开数据集 MSRA10K[8] 作为训练集。该数据集包含 10000 张高像素图片，并且大多数图片中只有一个显著目标。为了增加训练图片的多样性，通过镜像和旋转 0°、90°、180°、270°产生了 80000 张图片作为训练集。测试集采用 ECSSD[14] 和 SED2[17]。ECSSD 数据集包含了 1000 张自然图片，这些图片在真值图分割上都具有丰富的语义和复杂的结构；SED2 数据集共 100 张图片，含有像数级别的真值标注，该数据集虽然图片规模较小，但是图片背景相对复杂，而且均为多显著目标，具有极高的挑战性。

实验平台：64 位的 Ubuntu 16.04 操作系统，英伟达 GTX Genforce 1080 GPU，内存为 8 GB，采用的软件有 Matlab2014a、Python2.7，采用的深度学习框架为 Caffe[22]。使用梯度下降(SGD)方法训练网络，设置动量为 0.9、权重衰减为 0.0001、基础学习率为 10^{-8}。训练过程共花费 14 h，在迭代 220000 次后达到收敛。在实验中，τ 设置为 1.05，w_1、w_2、σ_β、σ_α、σ_γ 分别设置为 3.0、3.0、60.0、8.0、5.0。

3.2 定量评价

选择 7 种主流方法与本文算法(AFC)进行比较，包括循环全卷积神经网络(RFCN)[17]、元胞自动机(BSCA)[7]、引导学习(BL)[6]、稳健背景检测(wCO)[9]、显著滤波器(SF)[23]、测地线(GS)[24]、流行排序(MR)[10]。在实验中，采用准确率、召回率、F-measure 的数据评价方式全面评价本文方法。图 4 所示为本文算法和上述 7 种主流算法，以及人工标注(GT)图的直观比较。

为了更进一步验证本文算法的有效性，采用查准率 P、查全率 R 和 F 值三个指标来评价算法性能。通过自适应阈值 Y_u 对算法得到的显著图进行二值分割，并将得到的二值图与人工标注图 GT 对比，得到图像的查全率和查准率，进一步计算 F 值，综合评价算法性能，如图 5 所示。F 值的具体数值如表 1 所示。F 值、自适应阈值 Y_u 的计算公式为

$$F=\frac{(1+\beta^2)P\times R}{\beta^2\times P+R},\tag{6}$$

$$Y_u=\frac{2}{W\times H}\sum_{x=1}^{W}\sum_{y=1}^{H}S(x,y)。\tag{7}$$

与文献[25-26]一致，式中 β^2 取值为 0.3，W 和 H 分别为显著图的长和宽。

由图 5 可知，本文方法的 P 值、R 值、F-measure 明显高于其他方法，综合这 3 个指标分析，本文算法性能要优于对比的 7 种主流算法。

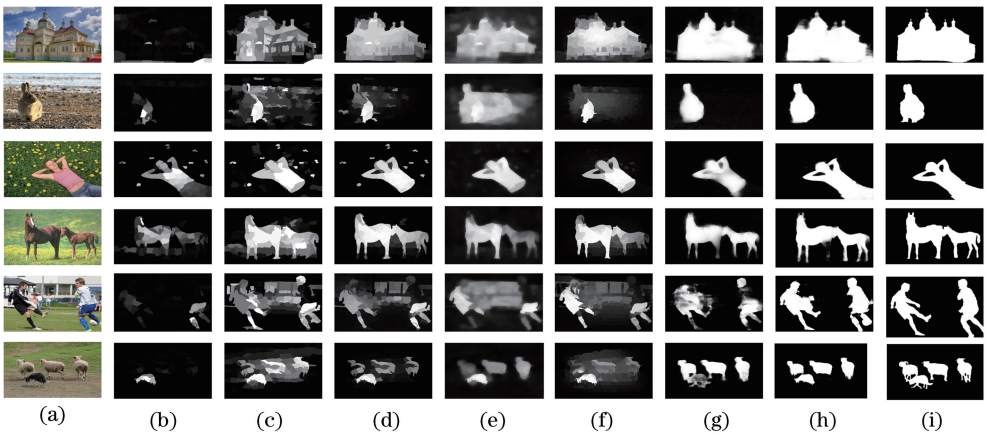

图 4 显著图比较。(a)原图;(b) SF 算法;(c) GS 算法;(d) wCO 算法;(e) BL 算法;(f) BSCA 算法;(g) RFCN 算法;(h) AFC 算法;(i)人工标注图

Fig. 4 Comparison of saliency maps. (a) Input images; (b) SF algorithm; (c) GS algorithm; (d) wCO algorithm; (e) BL algorithm; (f) BSCA algorithm; (g) RFCN algorithm; (h) AFC algorithm; (i) ground truth

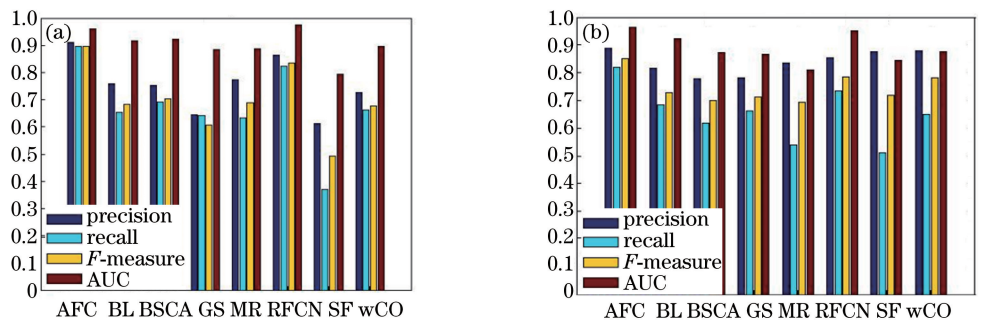

图 5 不同方法在(a) ECSSD 和(b) SED2 上的准确率、召回率、F-measure 和 AUC

Fig. 5 Precision, recall, F-measure and AUC of different methods in (a) ECSSD and (b) SED2

F 值通过设定自适应阈值而得到,为进一步验证方法有效性,通过固定阈值 $T_f=\{x \mid \forall x \in [0, 255]\}$ 将单幅显著图分割为 256 张二值图,将其与 GT 作对比得到查全率-查准率(P-R)曲线,进一步观测到在不同阈值下查准率和查全率的关系,如图 6 所示。

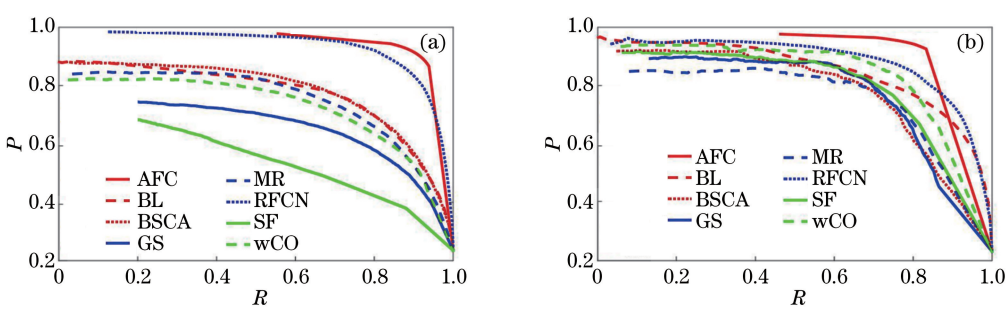

图 6 不同方法在(a) ECSSD 和(b) SED2 上的 P-R 曲线

Fig. 6 P-R curves of different methods in (a) ECSSD and (b) SED2

从图 6 可以看出,相比其他方法,本文方法的查准率和查全率与其他方法相比在整个召回率范围内都能实现更高的准确度。P-R 曲线描述的是显著图在取不同阈值下的对应值,为了对该图进行量化且更直观地评判实验效果,进一步引入 AUC 指标。AUC [8] 为 P-R 曲线下的面积,AUC 值越靠近 1,表示算法效果越精确,具体数值如表 1 所示。

为了不进行二值分割而直接分析算法所得显著图与 GT 的关系,通过平均绝对误差(MAE)值[14]来评价本文算法性能。

$$V_{\mathrm{MAE}}=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}\mid S(x,y)-G(x,y)\mid,\quad(8)$$

式中 M,N 为显著图的长和宽,$G(x,y)$表示人工标注图在(x,y)处的值,$S(x,y)$表示显著图在(x,y)处的值,MAE 值越小,说明算法得到的显著图与人工标注图的差异越小,性能越优异。在 ECSSD 数据集上,不同方法的 F 值、MAE 值和 AUC 值如表 1 所示。

表 1 不同方法的 3 种性能指标对比

Table 1 Comparison of three indicators of different methods

Algorithm	MAE	F-measure	AUC
AFC	0.059	0.879	0.981
RFCN	0.107	0.866	0.976
BSCA	0.182	0.76	0.922
BL	0.216	0.76	0.916
wCO	0.171	0.727	0.886
MR	0.189	0.773	0.889
GS	0.206	0.645	0.884
SF	0.219	0.612	0.793

由表 1 可以看出,本文算法的 MAE 值小于其他算法,而 F 值与 AUC 值均大于其他算法,表明本文算法在显著目标检测方面优于对比算法。

4 结 论

提出一种融合全部卷积层特征的全卷积神经网络显著性检测方法。一方面,在每个尺度上联合各级的全部卷积特征,使高低级特征在各尺度上互补,不仅能有效预测显著区域,而且细化了显著区域的细节信息;另一方面,通过全连接条件随机场的后处理,获得精确边界和平滑的显著图。实验结果表明,该方法较现有算法在准确率、召回率、F 测度以及平均绝对误差测度等方面性能都得到明显提升。

参 考 文 献

[1] Fang Y M, Chen Z Z, Lin W S, *et al*. Saliency detection in the compressed domain for adaptive image retargeting[J]. IEEE Transactions on Image Processing, 2012, 21(9): 3888-3901.

[2] Gao R, Tu Q, Xu J, *et al*. Visual saliency detection based on mutual information in compressed domain [C]. Visual Communications and Image Processing (VCIP), 2015: 1-4.

[3] Ren Z X, Gao S H, Chia L T, *et al*. Region-based saliency detection and its application in object recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(5): 769-779.

[4] Sharma G, Jurie F, Schmid C. Discriminative spatial saliency for image classification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3506-3513.

[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] Tong N, Lu H, Ruan X, *et al*. Salient object detection via bootstrap learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1884-1892.

[7] Qin Y, Lu H, Xu Y, *et al*. Saliency detection via cellular automata [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 110-119.

[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] Zhu W, Liang S, Wei Y, *et al*. Saliency optimization from robust background detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2814-2821.

[10] Yang C, Zhang L, Lu H, *et al*. Saliency detection via graph-based manifold ranking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3166-3173.

[11] Dai J, He K, Li Y, *et al*. Instance-sensitive fully convolutional networks[C]// European Conference on Computer Vision, 2016: 534-549.

[12] 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.

[13] Liu Y, Cheng M M, Hu X, *et al*. Richer convolutional features for edge detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5872-5881.

[14] Wang L, Lu H, Ruan X, *et al*. Deep networks for saliency detection via local estimation and global search[C]// Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition, 2015: 3183-3192.

[15] Zhao R, Ouyang W, Li H, *et al*. Saliency detection by multi-context deep learning[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1265-1274.

[16] Li G, Yu Y. Visual saliency based on multiscale deep features[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5455-5463.

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

[18] Mahendran A, Vedaldi A. Understanding deep image representations by inverting them[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5188-5196.

[19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

[20] Zhang P, Wang D, Lu H, *et al*. Learning uncertain convolutional features for accurate saliency detection [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 212-221.

[21] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C] // Advances in Neural Information Processing Systems, 2011: 109-117.

[22] Jia Y, Shelhamer E, Donahue J, *et al*. Caffe: convolutional architecture for fast feature embedding [C] // Proceedings of the 22nd ACM International Conference on Multimedia, 2014: 675-678.

[23] Perazzi F, Krähenbühl P, Pritch Y, *et al*. Saliency filters: contrast based filtering for salient region detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 733-740.

[24] Wei Y, Wen F, Zhu W, *et al*. Geodesic saliency using background priors[C]. European Conference on Computer Vision, 2012: 29-42.

[25] Fang Z M, Cui R Y, Jin J X, *et al*. Static saliency region detection in traffic scenes [J]. Laser & Optoelectronics Progress, 2017, 54(5): 051501.
方志明, 崔荣一, 金璟璇. 交通场景静态显著性区域检测[J]. 激光与光电子学进展, 2017, 54(5): 051501.

[26] Hou Q, Cheng M M, Hu X, *et al*. Deeply supervised salient object detection with short connections[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5300-5309.