

全局模型和局部优化的深度网络显著性检测

刘峰¹, 沈同圣², 娄树理¹, 韩冰³

¹海军航空工程学院控制工程系, 山东 烟台 264001;

²中国国防科技信息中心, 北京 100142;

³海军 92493 部队 98 分队, 辽宁 葫芦岛 125000

摘要 设计有效的特征向量是显著性检测方法的关键, 决定了模型效果的上限, 基于深度卷积神经网络和手工提取特征相结合的思路, 提出了一种新的基于全局模型和局部搜索的显著性检测方法。在全局模型中, 通过对 VGG-16 网络设计额外的卷积层进行训练, 生成初始显著图, 达到了从图像整体角度预测每一个候选区域显著性的目的。在局部优化模型中, 设计区域对比度描述子和区域特征描述子对多级分割的超像素点进行描述, 预测每一个区域的显著性值。最后, 利用线性拟合的方法将两种模型中产生的显著图进行融合, 得到最终的显著图。对 4 个数据集进行对比测试实验, 实验结果表明, 本文方法具有最高的准确率。

关键词 机器视觉; 显著性检测; 卷积神经网络; 超像素分割; 区域对比度; 区域特性

中图分类号 TP391.4 文献标识码 A

doi: 10.3788/AOS201737.1215005

Deep Network Saliency Detection Based on Global Model and Local Optimization

Liu Feng¹, Shen Tongsheng², Lou Shuli¹, Han Bing³

¹Department of Control Engineering, Naval Aeronautical and Astronautical University,
Yantai, Shandong 264001, China;

²China Defense Science and Technology Information Center, Beijing 100142, China;

³Element 98 of Unit 92493, PLA, Huludao, Liaoning 125000, China

Abstract The design of the effective feature vectors is the key to the saliency detection algorithm, which determines the upper bound of the model effect. A new saliency detection algorithm based on global model and local search is proposed by combining the deep convolution neural networks and the hand-crafted features. In the global model, the initial saliency map is generated from designing the extra convolution layers for VGG-16 network training, and thus the saliency value of each object candidate region can be predicted from a global perspective. In local optimization model, the super-pixel region with multi-degree segmentation is described by designing the contrast descriptors and region characteristic descriptors, and the saliency score of each region is predicted. Finally, a linear fitting method is used to fuse the result generated from two models, and the final saliency map is obtained. Contrast experiments for four data sets are demonstrated and the results show that the proposed algorithm has the highest precision.

Key words machine vision; saliency detection; convolution neural network; super-pixel segmentation; regional contrast; region characteristic

OCIS codes 150.0155; 110.2970; 110.4234; 100.3008

收稿日期: 2017-07-10; 收到修改稿日期: 2017-07-26

基金项目: 国家自然科学基金(61303192)

作者简介: 刘峰(1988—), 男, 博士研究生, 主要从事计算机视觉、目标检测及识别方面的研究。

E-mail: liufeng_cv@126.com

导师简介: 沈同圣(1966—), 男, 博士, 教授, 博士生导师, 主要从事精确制导及智能化、红外图像分析等方面的研究。

E-mail: shents@126.com

1 引 言

基于视觉注意机制的显著性检测对图像分析具有非常重要的意义。注意机制是人类信息加工过程中的一项重要心理调节机制,能够对有限的信息资源进行分配,使感知具备选择能力。若能将这种机制引入图像分析领域,将计算资源优先分配给容易引起观察者注意的区域,必将极大提高现有图像处理分析方法的工作效率。因此,显著性检测已经受到国内外学者的广泛关注,然而,随着应用场景的日益复杂和准确率要求的不断提高,显著性检测算法仍在持续的发展和完善之中。

传统的显著性检测方法,大多采用特定的先验知识来近似人眼视觉注意机制并结合手工提取目标的颜色、纹理、形状等特征作为显著性判断的依据,如对比度测量^[1-2]、中心先验^[3]、背景先验^[4]等。Kim等^[1]利用目标与背景的颜色对比差异,将目标映射到高维线性空间进行显著性区域的划分。Zhu等^[3]将图像中心物体设定为目标,边界先验的判断为背景,提出边界连通性的概念进行定量分析,再对目标进行显著性分割。Yan等^[5]利用超像素对图像进行多尺度分析,通过不同尺度之间的继承性分析图像全局与局部的特性,再利用随机森林回归生成显著图。此外,还有基于微分方程求解的方法^[6]、基于频域分析的方法^[7]等。

为了更加精确地表达显著性目标的特征,研究方法趋向于对图像的多层级分解、局部与全局特征的结合,以及对高维特征向量的表述等。然而,这些基于先验知识和人工提取的特征通常基于对目标数据库的观察而设计,在未知的自然图像和日益复杂的场景应用中容易受到限制。近年来,深度学习技术在机器学习和人工智能领域产生了革命性的变化,如图像分类^[8]、目标检测^[9]、人脸识别^[10]等。此类方法可以自动地在原始图像上提取分层特征,考虑图像全局因素,并且具有较好的通用性和移植性,在显著性检测领域也逐渐得到应用。

Wang等^[11]首先利用卷积神经网络(CNN)对每个像素的局部区域预测显著性得分,再从全局角度对候选区域的显著性进行精细化微调。Lee等^[12]利用VGG-16网络提取图像的高级特征,再利用手工设计的特征对规则化的超像素区域提取低级特征,最后添加全连接层进行联合训练,判断区域的显著性。Liu等^[13]利用递归神经网络(RNN)将不同图像区域进行上下文联系,再设计与CNN对称的反卷积结构,端到端地生成全分辨率显著图。Li等^[14]使用多尺度的CNN特征预测每一个超像素候选区域的显著性得分。相似地,Zhao等^[15]利用两个独立的卷积网络分别从全局和局部的角度对多级上下文信息进行联合建模,生成显著图。

相比于上述提到的诸多方法,本文利用卷积神经网络分别提取图像的全局特征和局部特征。其中,全局特征利用改进的VGG-16模型从图像的整体角度对潜在的目标进行定位,局部特征利用手工设计的特征向量构造初始特征图,再构造神经网络进行训练,判断每个分割区域的显著性,结合多级分割的方法可以实现更加精细的划分,最后对多个显著图进行融合得到最终结果。

2 算法模型

2.1 全局模型

在全局模型的构建上,设计了端到端的卷积网络,通过网络训练可以将输入图像直接映射为显著图,模型结构如图1所示。首先将 $224\text{ pixel}\times 224\text{ pixel}$ 的图像输入VGG-16网络,该模型共有5组13个卷积层和相应的映射单元,5个池化层,2个全连接层和1个输出层。输入图像经过5层池化操作之后,输出 $14\times 14\times 512$ 维的特征向量,再通过2个全连接层将其转换为 1×1 的卷积核。对该模型结构进行改进,使其不同层能够输出具有相同尺寸且分辨率较高的显著图,省略最后2层的最大池化操作,并设计增加额外的卷积层和全连接层,具体操作为:针对VGG网络每个卷积层的特征向量,提取经过conv1卷积后的图像,在网络中添加 5×5 的卷积核,设定网络参数,步长为4,边缘为2,生成 $56\times 56\times 128$ 维的特征向量,再增加3136维的全连接层sigmoid激活函数,将其映射为 $56\text{ pixel}\times 56\text{ pixel}$ 的显著图。同理,在conv2、conv3和conv4的卷积层之后添加类似的卷积层,卷积核大小均为 3×3 ,步长分别为2、1和1,再利用额外添加的包含3136个神经元的全连接层,生成相同尺寸大小的显著图。在原网络中,由于省略了池化层Pooling4和Pooling5,卷积层conv5之后可生成 $56\text{ pixel}\times 56\text{ pixel}$ 的显著图。

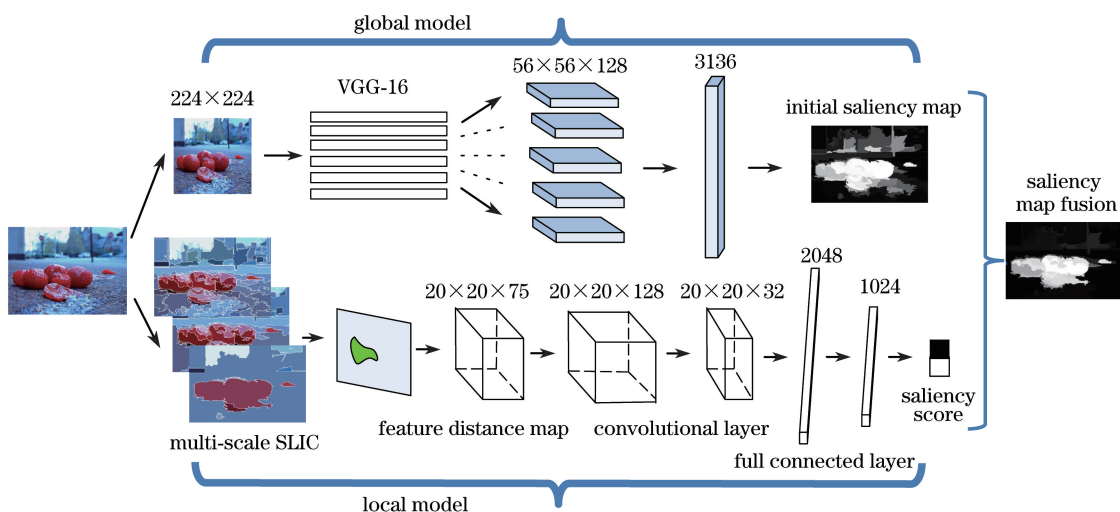


图 1 模型结构

Fig. 1 Structure of model

该网络初始化的设定如文献[8]所述,采用在大数据集 ILSVRC12 上得到的训练模型参数,在训练阶段,利用标准显著图和随机梯度下降法进行有监督的学习,再利用额外增加的卷积层和全连接层合并分辨率相同的 5 个图像 $\{S_G^1, S_G^2, S_G^3, S_G^4, S_G^5\}$, 生成初始显著图 S_G , 其分辨率为 $56 \text{ pixel} \times 56 \text{ pixel}$ 。全局模型网络结构图如图 2 所示,通过改进的 VGG 网络可以从图像的全局结构获得显著图,不同卷积层的信息提取也在不同尺度上对原始图像进行内容表述。此外,由于该网络结构缺乏对原始图像细节信息的描述,单独的全局模型无论从定量分析还是视觉观察上都无法满足需要,因此需要结合局部信息进行调整,生成更加精细的显著图。

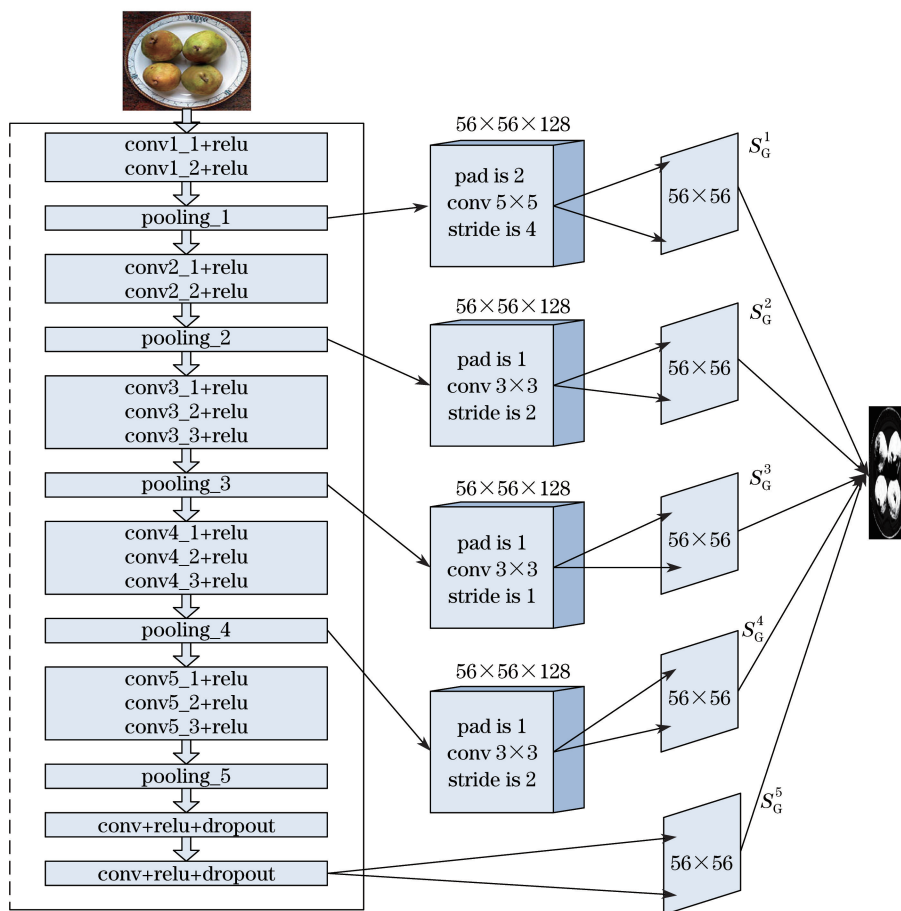


图 2 全局模型网络结构图

Fig. 2 Network structure of global model

2.2 局部优化模型

局部模型利用邻域像素之间的颜色、纹理等特征提取吸引人的点或区域,采用简单线性迭代聚类(SLIC)分割方法对图像进行多级划分。针对图像 I ,对其进行 M 级的超像素分割, $I = \{I_1, I_2, \dots, I_M\}$,每一级分割都是对图像 I 的分解,并且由 K_m 个区域组成。其中, I_1 包含数量最多的超像素点,是对原图像最精细的划分, I_M 包含最少的超像素点。 I_1 级的分割可表示为 $I_1 = \{I_1^1, I_1^2, \dots, I_1^{K_1}\}$,其他每一级分割 I_m 均是通过聚合前一级分割 I_{m-1} 获得。分割后的图像块具有相似的纹理颜色特征和空间一致性,更加利于网络的训练和分类。针对划分后的区域,使用区域对比度描述子和区域特性描述子两种形式的特征对每一个区域进行描述。

区域对比度描述子对每一个超像素区域构造一个特征向量,包括颜色和纹理特征,记为 \mathbf{v}_1 ,每一维特征向量的含义如表 1 所示。对于区域 $R \in I_m$,针对其邻域 N 计算的颜色和纹理特征向量记为 \mathbf{v}_1^R ,则 R 的区域对比度描述子可通过相应特征向量的差分计算得到,记为 $\text{diff}(\mathbf{v}_1^R, \mathbf{v}_1^N)$ 。Gabor 滤波器是一个边缘提取的线性滤波器,可用于图像纹理的提取和处理,这里用其表述图像块的纹理特征,直方图的特征利用距离差异计算,其他特征通过向量间的绝对差异计算。计算后的区域对比度描述子为 34 维的特征向量。若以 d 表示向量 x_1 和 x_2 的元素数量,则可记为

$$d(\mathbf{x}_1, \mathbf{x}_2) = (|x_{11} - x_{21}|, \dots, |x_{1d} - x_{2d}|). \quad (1)$$

χ^2 定义为

$$\chi^2(\mathbf{h}_1, \mathbf{h}_2) = \sum_{i=1}^b \frac{2(h_{1i} - h_{2i})^2}{h_{1i} + h_{2i}}, \quad (2)$$

式中 b 为直方图分组数。

表 1 对比度描述子特征向量
Table 1 Feature vectors of contrast descriptor

Color texture feature			Differential feature	
Feature	Descriptor	Dimension	Definition	Dimension
Average RGB value	\mathbf{a}_1	3	$d(\mathbf{a}_1^R, \mathbf{a}_1^I)$	3
Average lab value	\mathbf{a}_2	3	$d(\mathbf{a}_2^R, \mathbf{a}_2^I)$	3
Gabor filter response	\mathbf{r}	24	$d(\mathbf{r}^R, \mathbf{r}^I)$	24
Maximum Gabor response	\mathbf{r}	1	$d(\mathbf{r}^R, \mathbf{r}^I)$	1
RGB color histogram	\mathbf{h}_1	24	$\chi^2(\mathbf{h}_1^R, \mathbf{h}_1^I)$	1
Lab color histogram	\mathbf{h}_2	24	$\chi^2(\mathbf{h}_2^R, \mathbf{h}_2^I)$	1
HSV color histogram	\mathbf{h}_3	24	$\chi^2(\mathbf{h}_3^R, \mathbf{h}_3^I)$	1

区域特性描述子为利用区域自身特性建立的描述子,记为 \mathbf{v}_2 。外观特征利用该区域的颜色和纹理分布进行描述,可以对显著性目标和背景进行区分,如背景通常具有均匀的颜色分布和相似的纹理。几何特征包括该区域的尺寸和位置,有助于描述目标和背景的空间分布,如目标通常处于图像的中心,而边缘通常为背景。根据以上描述,建立了 41 维的区域特征描述子,区域特性描述子参数如表 2 所示。

表 2 区域特性描述子参数
Table 2 Parameters of regional feature descriptor

Feature	Dimension	Feature	Dimension
Normalized x of regional center	1	Regional connectivity ^[16]	1
Normalized y of regional center	1	RGB color variance	3
Normalized are	1	Lab color variance	3
Aspect ratio of bounding box	1	HSV color variance	3
Bounding box width	1	Gabor filter response variance	24
Bounding box length	1	Normalized area of neighborhood	1

区域显著性得分是将上述两种描述子作为超像素分割区域的特征向量,则每一个区域可以得到一个75维的特征向量 v 。为了构造基于CNN的网络模型进行训练,借鉴文献[12]中的方法,首先将图像的超像素点进行规则化划分,划分的原则为超像素点落在哪个区域的面积最大则判定其属于该区域,利用构建的75维区域描述子,组成三维矩阵作为输入图像的初始特征距离图,超像素点特征向量生成过程如图3所示。图像的规则化分割取 20×20 ,共进行5级超像素分割,初始种子点数为300,其余分别为250、200、150和100。

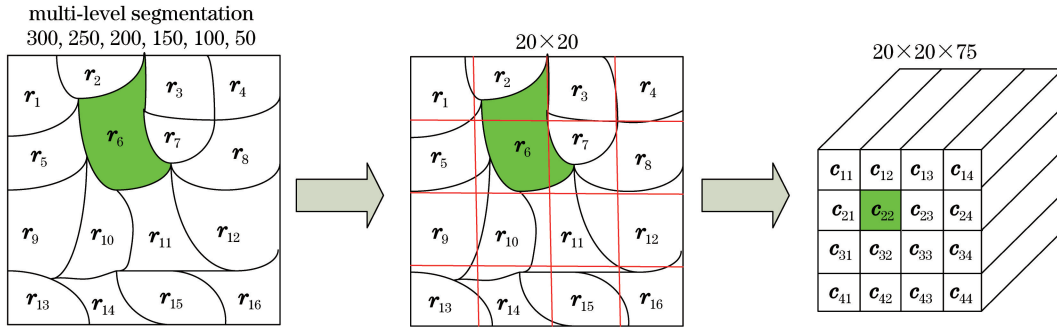


图3 超像素点特征向量生成过程

Fig. 3 Generation of feature vectors of super-pixel

每个超像素点构造成 $20 \times 20 \times 75$ 维的初始特征距离图,作为网络的输入,按照图1所示的网络,采用2个卷积层和2个全连接层进行训练。卷积层利用 1×1 的卷积核和线性修正单元(ReLU)激活函数,分别生成 $20 \times 20 \times 128$ 维和 $20 \times 20 \times 32$ 维的特征向量。 1×1 卷积核的优点在于可以在保持特征图尺寸的情况下进行卷积核通道数的降维和升维,还可以大幅增加网络的非线性特性。2个全连接层的尺寸分别为2048和1024个神经元,起到分类器和回归的作用,将分布式特征表示映射到样本的标记空间。最后利用Softmax层对进行分类,判断每一个区域的显著性,生成与原始尺寸相同的显著图。局部模型中,对图像进行了5级划分,因此,共生成5张全分辨率的显著图。训练数据的构建利用标准显著图对这些图像块进行标注,生成训练网络的正样本和负样本。正样本为当图像块 K 与 G 重叠部分满足 $|K \cap G| \geq 0.8 \times \min(|K|, |G|)$ 的样本;其他像素块均为负样本。

2.3 显著图融合

通过全局和局部模型共得到 M 个显著图 $\{S_1, S_2, \dots, S_M\}$,目标是对这些显著图进行有效的融合,生成最终的显著图 S 。各步骤显著图结果如图4所示。不同尺度下,局部模型得到的均是全尺寸显著图,在融合之前,需要在全局模型得到的显著图进行上采样,保持与原图尺寸相同。采用基于线性组合的方式对 M 个显著图进行融合,可表示为

$$S = \sum_{m=1}^M \omega_m S_m, \quad (3)$$

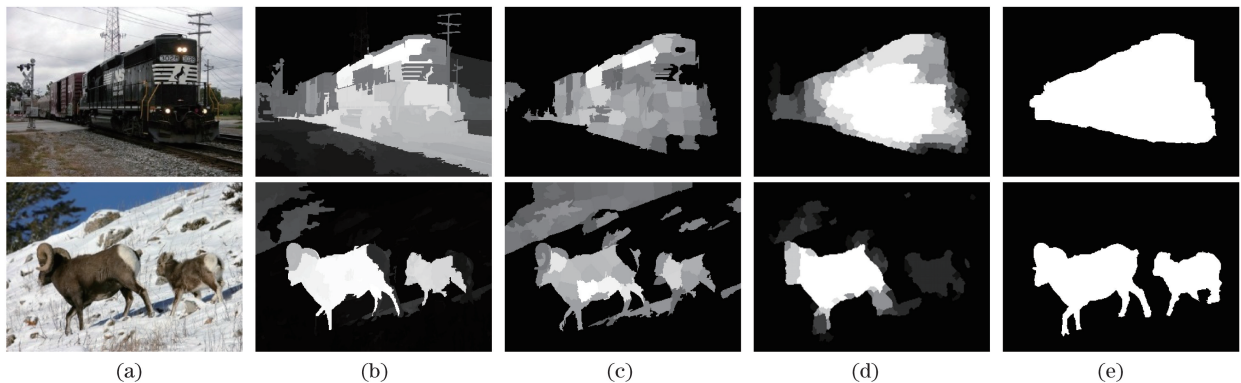


图4 各步骤显著图结果。(a)原图;(b)全局模型;(c)局部优化图;(d)最终显著图;(e)标准显著图

Fig. 4 Saliency maps of different steps. (a) Original images; (b) global models; (c) images of local optimization; (d) final saliency maps; (e) ground truth images

式中权重系数 ω 可通过在验证集上进行最小二乘估计得到。通过在所有训练图像上最小化代价函数可得

$$\{\omega_m\}_{m=1}^M = \arg \min_{\omega_1, \omega_2, \dots, \omega_M} \left(\left\| \mathbf{S} - \sum_{m=1}^M \omega_m \mathbf{S}_m \right\|_F^2 \right). \quad (4)$$

此外,还可以采用条件随机场(CRF)的方法对多级显著图进行融合^[14,17],然而,针对本文的情况,简单的线性组合即可得到较为满意的结果,生成与复杂算法近似的最终显著图。

3 实验仿真及对比分析

3.1 测试数据集及训练参数

为了对本文方法进行验证,在通用的 4 个数据集中进行测试,4 个数据集分别为 SOD、ECSSD、PASCAL-S 和 MSRA。SOD 数据集来源于伯克利大学的目标分割数据库,共包含 300 幅图像,每幅图像包含多个不同尺寸和位置的目标。ECSSD 数据集共 1000 张来源于网络的图像,其更贴近真实场景下的图像,因此具有复杂的场景。PASCAL-S 是在 PASCAL VOC 2012 数据集中选出的部分图像制作的显著性检测数据集,共 850 幅,图像包含多个目标和复杂背景,是目前最难以检测的数据集之一。MSRA 数据集由微软亚洲研究院发布,共 5000 幅图像,该数据集包含大量不同类别的目标,但多数图像仅包含一个显著性目标。以上 4 个数据集均包含手工标注和标准显著图供算法验证。在实验中,选用 MSRA 中的 3000 幅图像,以及 CSSD 和 PASCAL-S 中各 300 幅图像作为训练集,其余图像作为测试集。

为了对本文方法的性能进行评价,使用准确率-召回率(PR)曲线和 F-measure 平均测量误差(MAE)对算法进行定量分析,其中,PR 曲线通过不同的阈值对显著图进行划分,并将二值图像与标准显著图进行对比,F-measure 曲线定义为

$$F_\beta = \frac{(1 + \beta^2) \cdot h_P \times h_R}{\beta^2 \cdot h_P + h_R}, \quad (5)$$

式中 $\beta^2 = 0.3$, h_P 为准确率, h_R 为召回率。F-measure 曲线测量了显著图与标准图之间像素的平均误差。

本文方法的实现利用深度学习框架 Caffe^[18] 进行训练和测试,在全局模型中,首先将输入图像映射为 $224 \text{ pixel} \times 224 \text{ pixel}$,新添加的网络层初始学习率设为 0.01,原 VGG-16 网络层为 0.001,冲量参数为 0.9,权重延迟为 0.0005。

3.2 性能对比

将本文方法与目前较为流行的 HDCT^[1]、GC^[2]、wCtr^[3]、DRFI^[4]、MR^[5]、主成分分析(PCA)^[19]、视觉显著性(GBVS)^[20]、LEGS^[11] 和 SF^[15] 等方法进行对比,所用的实验结果和显著图均为原文中公开的数据。图 5 所示为 4 个数据集下不同方法的 PR 曲线,从图中可以看出,在 4 个数据集的测试中,本文方法在绝大多数情况下都能取得优于其他方法的准确率和召回率。其中,在 MSRA 图像集中各方法的曲线得分最高,在较为复杂的图像集 SOD 和 CSSD 中,各方法的性能略差,这也表明复杂场景下的目标更加难以检测,需要设计更有普适性的方法模型和多种特征协同检测。本文方法在 PASCAL-S 数据集中取得了较高的检测得分,主要原因是卷积神经网络的初始化参数,在训练的过程中与该数据集中的图像存在重复,在检测时具有一定的优势。

表 3 所示为不同方法下 F-measure 得分对比,加粗标出的数据为不同方法的最优得分。从表中可以看出,基于深度学习的显著性检测方法普遍优于传统的检测方法。本文方法在 SOD、PASCAL-S 和 CSSD 数据集中,可以得到较好的表现;在 MSRA 数据集中,与 LEGS 和 DRFI 方法得分相比互有高低。

图 6 为不同方法的显著图对比,列举出的图像包含低对比度、复杂背景、多目标和目标靠近边界等情况。从图中可以看出,本文方法产生的显著图细节更加突出,在多种复杂的场景中对目标的划分更完整。如前两排图像中,目标与背景之间的对比度较低,wCtr 方法与 PCA 方法受到较严重的干扰,无法对目标进行完整检测。中间两排图像为多目标场景,其中,LEGS 方法和 wCtr 方法发生了漏检的情况,GBVS 方法的检测结果对背景和目标的区分度较差。最后两排为复杂背景情况,目标融于背景之中,HDCT 方法和 DRFI 方法生

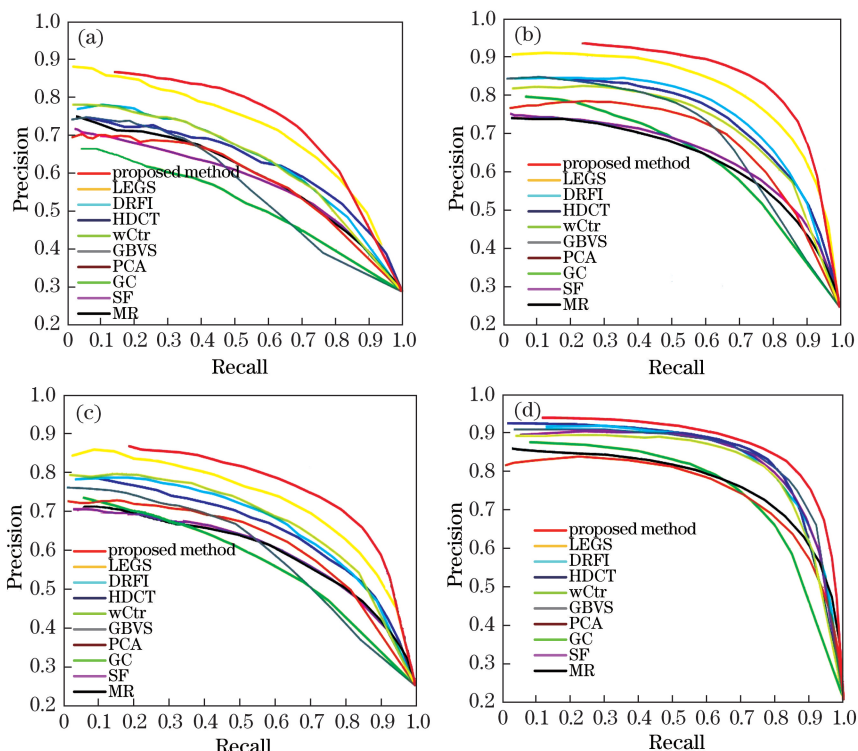


图 5 4 个数据集下不同方法的 PR 曲线。(a) SOD; (b) PASCAL-S; (c) CSSD; (d) MSRA

Fig. 5 PR curves of four data sets with different methods. (a) SOD; (b) PASCAL-S; (c) CSSD; (d) MSRA

表 3 不同方法下的 F-measure 得分对比

Table 3 Comparison of F-measure scores with different methods

%

Dataset	Curve	Proposed method	LEGS method	DRFI method	HDCT method	wCtr method	GBVS method	PCA method	GC method	SF method	MR method
SOD	F-measure	73.1	67.4	70.2	65.4	63.7	61.3	54.9	50.6	55.3	54.2
	MAE	20.4	21.2	24.1	26.6	24.5	26.9	25.3	28.8	26.7	27.4
CSSD	F-measure	84.6	83.1	78.8	70.5	66.8	65.3	57.5	55.7	54.5	67.5
	MAE	12.8	11.9	17.9	19.9	18.4	22.7	25.2	23.4	20.1	19.0
PASCAL	F-measure	75.3	74.9	69.9	60.4	61.1	69.3	53.1	61.6	57.4	58.3
	MAE	14.7	15.5	20.3	22.9	20.1	17.8	23.9	25.5	21.4	21.2
MSRA	F-measure	91.2	90.5	91.9	80.5	78.3	65.9	70.1	68.2	62.5	78.3
	MAE	10.4	8.9	14.3	11.9	16.6	13.7	18.9	14.7	16.2	13.0

成的显著图不清晰,无法定位目标中心的准确位置。本文方法在这些复杂情况下均获得了较为理想的显著性提取。分析原因,本文模型的成功之处在于包含了两种不同的注意机制,全局模型的训练关注图像的整体,使其可以检测到图像中的多个目标区域,不受孤立噪声的影响;而局部模型起到优化的作用,因此检测到的结果更加准确精细。

4 结 论

提出了一种基于深度神经网络训练的显著性检测方法,该方法通过全局模型和局部模型两种网络结构构造多级显著图,再利用线性组合的方法进行融合,生成最终的精细显著图。其中,全局模型可以有效地利用全局显著性特征预测每一个目标区域的显著性值,局部模型可以从划分的多级超像素点中学习到目标的局部对比度、纹理和形状信息。此外,局部模型的初始特征是利用手工设计的方法进行提取,再结合神经网络

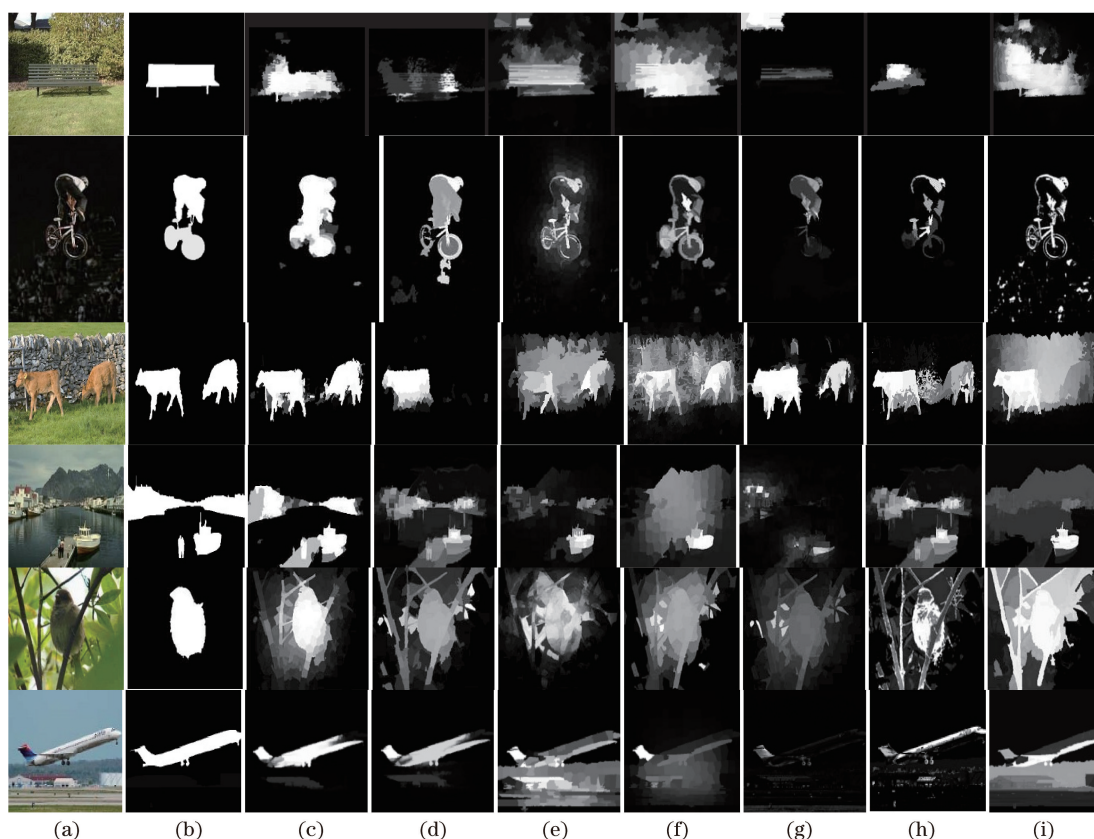


图 6 本文方法与其他方法显著图对比。(a)原图;(b)标准显著图;(c)本文方法;(d) LEGS;
(e) DRFI;(f) HDCT;(g) wCtr;(h) PCA;(i) GBVS

Fig. 6 Visual comparisons of our results and others. (a) Original images; (b) ground truth images;
(c) proposed method; (d) LEGS; (e) DRFI; (f) HDCT; (g) wCtr; (h) PCA; (i) GBVS

络进行训练,两种方法的结合有助于提取更有针对性的特征并提高网络模型的训练速度。测试数据集的实验表明,本文方法可以得到更加准确的显著性检测结果。

参 考 文 献

- [1] Kim J, Han D, Tai Y W, *et al.* Salient region detection via high-dimensional color transform and local spatial support[J]. IEEE Transactions on Image Processing, 2016, 25(1): 9-23.
- [2] Cheng M M, Mitra N J, Huang X, *et al.* Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569-582.
- [3] Zhu W J, Liang S, Wei Y C, *et al.* Saliency optimization from robust background detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2814-2821.
- [4] Jiang H Z, Wang J D, Yuan Z J, *et al.* Salient object detection: a discriminative regional feature integration approach[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2083-2090.
- [5] Yan Q, Xu L, Shi J P, *et al.* Hierarchical saliency detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 1155-1162.
- [6] Liu R S, Cao J J, Lin Z S, *et al.* Adaptive partial differential equation learning for visual saliency detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3866-3873.
- [7] Jiang P, Vasconcelos N, Peng J L. Generic promotion of diffusion-based salient object detection[C]. IEEE International Conference on Computer Vision, 2015: 217-225.
- [8] Liu Dawei, Han Ling, Han Xiaoyong. High spatial resolution remote sensing imaging classification based on deep learning[J]. Acta Optica Sinica, 2016, 36(4): 0428001.
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.
- [9] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic

- segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [10] Taigman Y, Yang M, Ranzato M A, *et al.* Deepface: closing the gap to human-level performance in face verification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1701-1708.
- [11] Wang L J, Lu H C, Ruan X, *et al.* Deep networks for saliency detection via local estimation and global search[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3183-3192.
- [12] Lee G, Tai Y W, Kim J. Deep saliency with encoded low level distance map and high level features[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 660-668.
- [13] Liu N, Han J W. DHSNet: deep hierarchical saliency network for salient object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 678-686.
- [14] Li G B, Yu Y Z. Visual saliency based on multiscale deep features[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5455-5463.
- [15] Zhao R, Ouyang W L, Li H S, *et al.* Saliency detection by multi-context deep learning[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1265-1274.
- [16] Liu Feng, Shen Tongsheng, Han Yanli, *et al.* Saliency detection via background aware and color contrast[J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(10): 1705-1712.
刘峰, 沈同圣, 韩艳丽, 等. 融合背景感知和颜色对比的显著性检测方法[J]. 计算机辅助设计与图形学学报, 2016, 28(10): 1705-1712.
- [17] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C]. Advances in Neural Information Processing Systems, 2011: 109-117.
- [18] Jia Y Q, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding[C]. 22nd ACM International Conference on Multimedia, ACM, 2014: 675-678.
- [19] Margolin R, Tal A, Zelnik-Manor L. What makes a patch distinct?[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 1139-1146.
- [20] Li Y, Hou X D, Koch C, *et al.* The secrets of salient object segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 280-287.