

DOI: 10.11992/tis.201706043

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20171109.1255.020.html>

面向社群图像的显著区域检测方法

梁晔^{1,2}, 于剑²

(1. 北京联合大学 机器人学院, 北京 100101; 2. 北京交通大学 计算机与信息技术学院, 北京 100044)

摘 要:网络技术和社交网站的发展带来了社群图像的飞速增长。海量的社群图像成为了非常重要的图像类型。本文关注社群图像的显著区域检测问题, 提出基于深度特征的显著区域检测方法。针对社群图像带有标签的特点, 在系统框架中, 本文采取两条提取线: 基于 CNN 特征的显著性计算和基于标签的语义计算, 二者的结果进行融合。最后, 通过全连接的条件随机场模型对融合的显著图进行空间一致性优化。此外, 为了验证面向社群图像的显著区域检测方法的性能, 针对目前没有面向社群图像的带有标签信息的显著性数据集, 基于 NUS-WIDE 数据集, 本文构建了一个图像结构丰富的社群图像数据集。大量的实验证明了所提方法的有效性。

关键词:显著性; 显著区域; 社群图像; 深度学习; 标签

中图分类号: TP311 **文献标志码:** A **文章编号:** 1673-4785(2018)02-0174-08

中文引用格式: 梁晔, 于剑. 面向社群图像的显著区域检测方法 [J]. 智能系统学报, 2018, 13(2): 174-181.

英文引用格式: LIANG Ye, YU Jian. Salient region detection for social images[J]. CAAI transactions on intelligent systems, 2018, 13(2): 174-181.

Salient region detection for social images

LIANG Ye^{1,2}, YU Jian²

(1. College of Robotics, Beijing Union University, Beijing 100101, China; 2. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: The development of network technology and social website has brought about the rapid growth of social images. Massive social images have become a very important image type. This paper focuses on the detection problem of salient region for social images, a method for detecting salient region and based on depth features was proposed. By considering the feature that the social image is attached with tag, in the framework of the system, the paper used two extraction lines: the saliency computing based on CNN features and the semantic computing based on tag, the results of both parts were fused. Finally, saliency maps were optimized by a fully connected conditional random field model for the spatial consistency. In addition, for verifying the performances of the saliency region detection method orienting social image, in view of the lack of saliency dataset with tags for social images, on basis of NUS-WIDE dataset, the paper constructed a social image dataset with rich image structures. Extensive experiments demonstrated the effectiveness of the proposed method.

Keywords: saliency; salient region; social images; deep learning; tag

近年来, 随着互联网和信息行业的发展, 全球已经进入了大数据时代。同时, 数码设备的普及

和社交网站的流行导致了社交图像爆炸式地增长, 已经成为了一种重要的图像类型。Yahoo 的社群图像分享网站 Flickr、社交媒体网站 Facebook、Google 的社交视频分享网站、Youtube 都是目前最具有代表性的社交网站。以 Flickr 为例, 每分钟上传到网站上的图像超过两千张, 每天上传到网站上的图像超过两百万张。据统计, 社交

收稿日期: 2017-06-11. 网络出版日期: 2017-11-09.

基金项目: 北京市自然科学基金项目 (4182022); 北京联合大学 2017 年度人才强校百杰计划项目 (BPHR2017CZ10); “十三五”时期北京市属高校高水平教师队伍建设支持计划项目 (IDHT20170511); 国家科技支撑计划项目 (2015BAH55F03)

通信作者: 梁晔. E-mail: liangye@buu.edu.cn.

网站 Facebook 平均每天新增图像数量超过 1 亿张,其总数量已经超过 1 000 亿张; Flickr 图像总量超过了 60 亿张; 图像分享移动应用 Instagram 图像数量超过 10 亿张。然而,相对于海量的图像数据,计算资源是有限的,如何对海量的图像进行有效的管理成了亟待解决的问题。显著性检测的相关研究和应用^[1]带来了很好的解决方案。显著性检测就是让计算机模拟人类的视觉注意力机制,对处理的图像进行自动信息筛选,将有限的资源分配给重要信息,提高处理速度,提升资源利用率。显著性检测是多学科交叉的研究领域,涉及心理学、神经科学、生物视觉和计算机视觉等相关领域,有重要的理论价值和实际应用意义,是目前研究的热点。

本文关注社群图像的显著区域检测问题,目前对此领域的研究不多,主要贡献有:

1) 提出了基于深度学习的显著区域检测方法。该方法针对社群图像带有标签的特点,采取两条提取线:基于 CNN 特征的显著性计算和基于标签的语义计算。较目前流行的检测方法,本文的检测精度有一定的提升。

2) 构建了面向社群图像的带有标签信息的显著性数据集,新建数据集来自于 NUS-WIDE 数据集中的多个类,包含多个标签、图像前景和背景差异性小,为面向社群图像的显著区域检测方法提供了新的实验数据。

1 相关工作

在显著性快速发展的几十年里,涌现了大量的显著区域检测方法。随着 GPU 等硬件资源的发展和大规模训练图像集的涌现,基于深度神经网络的显著区域提取近年来受到广泛的关注。深度学习可以从大数据中自动学习特征的表示,学习得到的特征能够刻画问题内部本质结构,检测结果比基于人工设计特征的提取方法有较大的提高。伴随着显著区域检测方法的研究,也涌现了多个显著性数据集。就目前的研究现状来看,面向社群图像的显著区域检测方法目前的研究不多。

1.1 基于层次和深度结构的显著区域检测

已有的研究显示了层次和深度体系结构对显著区域检测建模有重要的作用。文献[2]提出了层次结构进行显著区域检测,减少了小尺寸显著区域对检测结果的影响。文献[3]在检测显著区域之前将图像进行了多尺度分割,形成层次结构,取得了较好的检测效果。多层次显著性检测的优点是考虑了图像多尺度的特点,解决了单一分割的局限性,从一定程度上考虑了显著区域大

小不一的现象。但是这些工作仍然存在缺陷,在计算显著性时采用的特征仍是人工设计的特征,并且划分层次的数量也很难有科学的解释。

随着研究的深入,研究人员将深度体系结构应用到显著区域检测中。文献[4]通过无监督的方法学习多个中层的滤波器集合进行局部的显著性检测,并且和卷积网络得到的中层检测结果进行融合。文献[5]采用卷积神经网络得到图像的多尺度特征,包括局部区域块、它的邻域区域块和整幅图像,进行显著区域的检测。文献[6]训练了两个深度卷积网络:一个用于检测局部显著图,另一个用于检测全局显著图,然后将两种显著图进行融合。文献[7]采用全局上下文信息和局部区域信息相融合的方法实现显著区域检测。深度学习除了具有层次结构之外,还能自动学习特征,学习到的特征明显优于人工设计的特征,正因如此,基于深度学习的方法已经在显著区域检测中取得了明显的效果。当然,基于深度学习的检测方法具有深度学习固有的缺点,比如网络结构尚且无法做出合理的解释、参数众多调节费时等缺点。

1.2 基于标签语义进行显著区域提取

标签的语义虽然在图像标注领域已经得到了广泛应用,但是,标签信息通常和显著区域检测任务分开处理,在显著区域检测上的应用不多。

和本文工作最相关的是文献[8]和文献[9],均用到了标签信息。文献[8]将标签排序任务和显著区域检测任务整合在一起,迭代地进行标签排序和显著区域的检测任务;文献[9]提出 Tag-Saliency 模型,通过基于层次的过分割和自动标注技术进行标注。这两个工作的共同缺点是显著区域标注效果依赖于区域标注,采用的多示例学习方法不容易泛化;而且,二者均没有考虑标签之间的上下文关系。本文的工作和这两篇文献不同,是把标签的语义信息转化为 RCNN 特征。由于 RCNN 技术是基于卷积神经网络的,所以本文的方法更能改善检测性能。

1.3 显著性数据集

从关于显著性数据集的相关文献来看,显著性数据集主要来自于两个领域:为了显著性研究而建立的数据集^[2-3, 10-12];从图像分割领域延伸过来的显著性数据集^[13-14]。这些数据集有的以矩形框方式进行显著区域标注,更多的是以像素级进行显著区域的标注,还有的通过眼动仪进行视点的标注。然而随着社群图像的快速增加,目前针对社群图像的带有标签信息的显著性数据集研究不多,本文针对此问题构建面向社群图像的显著性数据集。

2 基于深度学习的显著区域检测

2.1 方法流程

本文提出的基于深度学习的显著区域检测方法针对社群图像带有标签的特点,系统框架中采取两条提取线:基于CNN特征的显著性计算和基于标签的语义计算,并将二者的结果进行融合,融合的显著图通过全连接的CRF模型进行一致性空间优化,获得最终显著图。主要处理流程如图1所示。

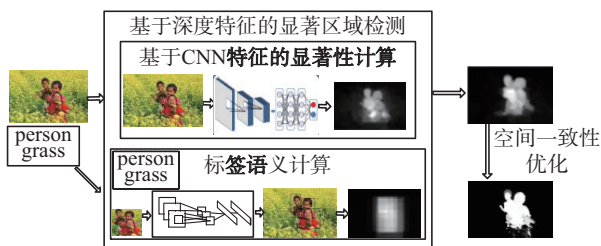


图1 系统流程图

Fig.1 System flow chart

在处理流程中,要解决的重点问题是如何提取图像的CNN特征、如何计算标签的语义特征。

2.2 基于CNN特征的显著性计算

2.2.1 网络结构

负责CNN特征提取的深度网络采用Hinton的学生Alex Krizhevsky在2012年ILSVRC竞赛中的8层卷积神经网络^[15],包括5个卷积层、2个全连接层和1个输出层。网络结构如图2所示。网络的最下面为检测的图像,图像的上面表示要提取特征的区域。

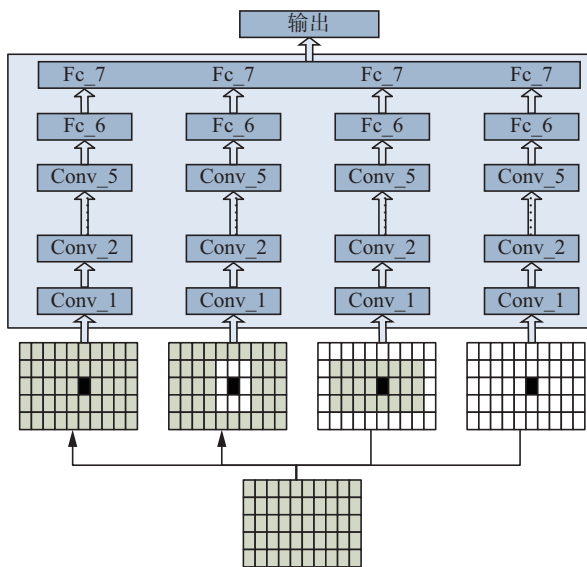


图2 网络结构

Fig.2 Network structure

5个卷积层负责多尺度特征的提取,为了实现平移不变性,卷积层后面采用最大池化操作;

特征自动学习的特征均包含4 096个元素;第1层全连接层和第2层全连接层后均通过修正线性单元ReLU(rectified linear units)进行非线性映射。修正线性单元ReLU^[15]对每个元素进行如下操作:

$$R(x^i) = \max(0, x^i) \quad (1)$$

式中 $1 \leq i \leq 4096$ 。

在输出层采用softmax回归模型得出图像块是否显著的概率。

2.2.2 网络训练与测试

采用公开的Caffe^[16]框架,利用研究工作^[15]的网络参数进行初始化,然后利用来自目标数据集的图像对参数进行微调。

在图像中,显著区域具有独特性、稀缺性,和邻域具有明显差异的特点。受文献[5]启发,为了有效地计算显著性,本文考虑了3种图像区域块间的差异:图像块和邻域的差异;图像块和图像边界的差异;图像块和整幅图像的差异。在利用卷积神经网络特征提取时,提取了4种区域的特征:1)采样的矩形区域;2)矩形区域的邻接区域;3)图像的边界区域;4)图像中去除矩形区域的剩余区域。4种区域的示例图如图3所示。

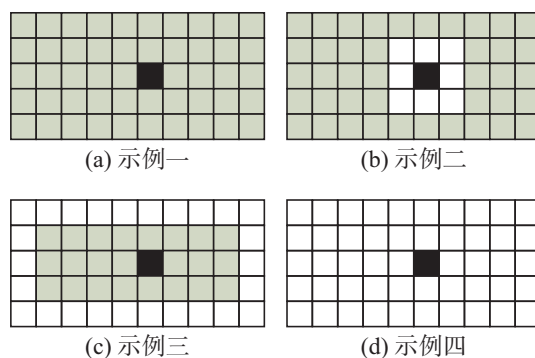


图3 4种区域示例图

Fig.3 Four regional cases

图3(a)中的黑色区域代表当前区域;图3(b)中的白色区域代表黑色区域的相邻区域;图3(c)中的白色区域代表图像的边界区域;图3(d)中的白色区域代表去掉黑色区域后的剩余区域。

对训练集中的每幅图像,采用滑动窗口方式进行采样,采样为 51×51 的区域块,采样步幅为10像素,得到用于网络训练的区域块,并采用和文献[4]相同的标注方法对采样区域块进行标注。如果图像块中至少70%的像素在基准二值标注中为显著,则这个图像块被标注为显著,否则标注为不显著。

利用这种标注策略,获得训练区域块的集合 $\{B_i\}$ 及其相应的显著标签 $\{l_i\}$ 。在微调过程中,采用权重衰减作为softmax回归模型的正则项,损失函数的定义如式(2):

$$L(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=0}^1 1\{l_i = j\} \log P(l_i = j|\theta) + \lambda \sum_{k=1}^8 \|w_k\|_F^2 \quad (2)$$

式中: θ 是可学习的参数,包括卷积神经网络各层的权重和偏置; $1\{\cdot\}$ 是符号函数; $P(l_i = j|\theta)$ 表示区域 i 预测为显著区域的概率; λ 是权重衰减参数; w_k 代表第 k 层的权重。卷积神经网络通过随机下降的方法进行训练。

在测试的时候,通过卷积神经网络提取图像区域相关的4种特征,然后通过训练好的模型预测各个区域为显著区域的概率。

2.3 基于标签的语义特征计算

图像的标签可以分为两大类:场景标签和对象标签。对象为图像中显著区域的可能性非常大。基于此,在基于标签的语义计算中关注对象标签。

估计一个区域属于特定对象的概率从一定程度上反映出此区域为显著区域的可能性。因此,区域属于特定对象的概率可看作显著性的先验知识。

RCNN技术是一种简单、可扩展的对象检测方法,基于卷积神经网络的,在图像分类、对象检测和图像分割等领域取得了优异的性能。本文采用RCNN^[17]技术检测区域属于对象的概率。

标签语义特征计算的主要思路为利用RCNN抽取的特征计算每个像素的语义特征。

假设有 X 个对象检测子,对于第 k 个对象检测子,具体计算过程如下。

1) 选取最可能包含特定对象的 N 个矩形框;

2) 第 i 个矩形框包含特定对象的概率为 p_k^i , $1 \leq k \leq X, 1 \leq i \leq N$ 。区域内的每个像素和包含它的矩形框区域具有相同的概率值 p_k^i 。

3) 第 k 个对象检测子检测完毕后,图像中的像素包含检测子对象的可能性 $\sum_{i=1}^N p_k^i \times f_k^i$,如果像素被包含在第 i 个矩形框里,则 $f_k^i = 1$,否则 $f_k^i = 0$ 。

X 个对象检测子都检测完毕后,每个像素得到 X 维特征。 X 维特征归一化后表示为 f , $f \in R^X$ 。 f 的每一维代表像素属于每一类特定对象的概率。

计算得到的基于标签的语义特征用于辅助显著性的计算。

2.4 显著图和标签语义图的融合

假设基于CNN特征的显著图为 S_D ,基于RCNN技术检测到的标签语义为 T ,二者融合如下。

$$S = S_D \cdot \exp(T) \quad (3)$$

式中 S 表示融合后的显著图。

在融合过程中,标签语义相当于先验,对显著值起到加权的作用。

2.5 空间一致性优化

在图像分割领域,研究人员采用全连接的CRF模型对分割结果进行分割区域和边缘的平滑。借鉴文献[20]中的解决方法,本文采用全连接的CRF模型对显著图进行空间一致性优化。

能量函数定义如式(4):

$$S(L) = - \sum_i \log P(l_i) + \sum_{i,j} \theta_{ij}(l_i + l_j) \quad (4)$$

L 具有两个值,显著为1,不显著为0。 $P(l_i)$ 代表像素 x_i 具有标签 l_i 的概率,初始化时, $P(1) = S_i$, $P(0) = 1 - S_i$ 。 S_i 是像素融合后的显著值。

$\theta_{i,j}$ 定义如下:

$$\theta_{i,j} = u(l_i, l_j) \left[\omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_1^2} - \frac{\|I_i - I_j\|^2}{2\sigma_2^2}\right) + \omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_3^2}\right) \right] \quad (5)$$

当 $l_i \neq l_j$ 时, $u(l_i, l_j) = 1$, 否则为0。

$\theta_{i,j}$ 考虑了两方面的信息:位置信息和颜色信息。 $\omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_1^2} - \frac{\|I_i - I_j\|^2}{2\sigma_2^2}\right)$ 表明相近的像素且具有相似的颜色,应该具有相近的显著值。

$\omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_3^2}\right)$ 只考虑位置信息,目的在于移除小的孤立区域。

3 带有标签信息的显著性数据集构建

3.1 图像来源

本文从NUS-WIDE数据集采用如下筛选条件进行图像的筛选。

1) 显著区域和整幅图像的颜色对比度小于0.7;

2) 显著区域尺寸丰富,要求占整幅图像的比例覆盖10个等级, $[0, 0.1)$ 、 $[0.1, 0.2)$ 、 $[0.2, 0.3)$ 、 $[0.3, 0.4)$ 、 $[0.4, 0.5)$ 、 $[0.5, 0.6)$ 、 $[0.6, 0.7)$ 、 $[0.7, 0.8)$ 、 $[0.8, 0.9)$ 、 $[0.9, 1]$;

3) 至少有10%的图像中的显著区域和图像边界相连。

最终确定5429幅图像作为数据集,来自于38个文件夹,包括carvings、castle、cat、cell_phones、chairs、chrysanthemums、classroom、cliff、computers、cooling_tower、coral、cordless cougar、courthouse、cow、coyote、dance、dancing、deer、den、desert、detail、diver、dock、close-up、cloverleaf、cubs、dall、dog、dogs、fish、flag、eagle、elephant、elk、f-16、facade、fawn。

3.2 图像标注

矩形框级别的标注不能准确地定位对象边缘,标注结果不精确。本文采用像素级别的二值

标注。

选取了 5 个观察者进行标注。不同用户标注结果通常存在不一致的现象。为了减少标注的不一致性, 计算每个像素标注的一致性分值。

$$s_x = \frac{\sum_{p=1}^N a_x^{(p)}}{N} \quad (6)$$

式中: $a_x^{(p)}$ 表示第 p 个观察者对像素 x 的标注, 如果像素 x 标注为显著, 则 $a_x^{(p)} = 1$, 否则 $a_x^{(p)} = 0$; N 为观察者的个数; s_x 为显著阈值。根据文献 [13], 如果一个像素有 50% 的观测者都标注为显著, 则认为显著。

最后, 两个观察者使用 Adobe Photoshop 手动从图像中分割出显著区域。

3.3 数据集的标签信息

NUS-WIDE 数据集提供了 81 个基准标签集。新构建的数据集标签来自于 81 个基准标签集, 共 78 个标签。每幅图像包含 1~9 个标签。

4 实验

4.1 实验设置

以本文构建的数据集为实验对象, 选择 20 个对象标签, 包括 bear、birds、boats、buildings、cars、cat、computer、coral、cow、dog、elk、fish、flowers、fox、horses、person、plane、tiger、train、zebra; 选取和对象标签相对应的 20 个对象检测子进行 RCNN 特征提取, 选取 2 000 个包含对象概率最大的矩形框。

采用 Caffe 框架 [16] 进行深度卷积神经网络的训练和测试。通过随机下降方法进行训练, 每次迭代参与的样本数量 (batch) 为 256; 冲量值 (momentum) 为 0.9; 正则化项的权重为 0.000 5; 学习率初始值为 0.01, 当损失稳定的时候学习率以 0.1 的速度下降; 对每层的输出采用比率为 0.5 的 dropout 操作来防止过拟合; 训练迭代次数为 80 次。

实验对比了 27 种流行的显著区域检测方法, 包括 HS^[2]、DRFI^[3]、SMDF^[5]、LEGS^[6]、MCDL^[7]、FT^[14]、BL^[18]、RFCN^[19]、CB^[21]、SEG^[22]、RC^[23]、SVO^[24]、LRR^[25]、SF^[26]、GS^[27]、CA^[28]、SS^[29]、TD^[30]、MR^[31]、PCA^[32]、HM^[33]、GC^[34]、MC^[35]、DSR^[36]、SBF^[37]、BD^[38] 和 SMD^[37]。这些检测方法涵盖范围特别广泛。

本文的检测方法简称为 DBS。

在量化的性能评价中, 采用当前流行的性能评价指标: 1) 查准率和查全率曲线 (PR 曲线); 2) F-measure 值; 3) 受试者工作特征曲线 (ROC Curve); 4) AUC 值 (ROC 曲线下面的面积); 5) 平均绝对误差 (MAE)。

4.2 与流行方法的比较

与 27 种流行方法比较, 实验结果如表 1 和图 4、5 所示。

表 1 本文方法与 27 种流行方法的比较

Table 1 Compare with 27 popular methods

方法	F_measure	AUC	MAE
CB ^[21]	0.547 2	0.797 1	0.266 2
SEG ^[22]	0.491 7	0.758 8	0.359 2
SVO ^[24]	0.349 8	0.836 1	0.409 0
SF ^[26]	0.365 9	0.754 1	0.207 7
CA ^[28]	0.516 1	0.828 7	0.277 8
TD ^[30]	0.543 2	0.808 1	0.233 3
SS ^[29]	0.251 6	0.671 4	0.249 9
HS ^[2]	0.557 6	0.788 3	0.274 7
DRFI ^[3]	0.589 7	0.862 3	0.206 3
HM ^[33]	0.489 2	0.794 5	0.226 3
BD ^[38]	0.544 3	0.818 5	0.195 5
BL ^[18]	0.582 3	0.856 2	0.266 0
MR ^[31]	0.508 4	0.775 3	0.229 0
PCA ^[32]	0.539 2	0.843 9	0.277 8
FT ^[14]	0.355 9	0.612 6	0.280 8
RC ^[23]	0.530 7	0.810 5	0.312 8
LRR ^[25]	0.512 4	0.795 6	0.306 7
GS ^[27]	0.516 4	0.813 6	0.205 6
SMD ^[37]	0.603 3	0.843 7	0.197 6
GC ^[34]	0.506 3	0.751 1	0.259 6
DSR ^[36]	0.503 5	0.813 9	0.210 5
MC ^[35]	0.574 0	0.842 7	0.228 7
SBF ^[37]	0.493 0	0.848 0	0.232 5
LEGS ^[6]	0.612 4	0.819 3	0.184 4
MCDL ^[7]	0.655 9	0.881 3	0.145 7
RFCN ^[19]	0.676 8	0.880 3	0.147 6
SMDF ^[5]	0.657 4	0.848 3	0.155 6
DBS	0.662 1	0.891 7	0.150 5

表 1 中, F_measure、AUC 和 MAE 排在前 3 位的为 4 种目前流行的深度学习方法 SMDF^[5]、LEGS^[6]、MCDL^[7]、RFCN^[19] 和本文的 DBS 方法。在某种程度上可以说深度学习的检测方法超过了非深度学习的检测方法, 精度上有所提高。其中, DBS 方法的 AUC 值是最高的, DBS 方法的 F-measure 值是最高的, DBS 的 MAE 值是最低的, 所以 DBS 方法的性能最佳。

PR 曲线图和 ROC 曲线图如图 4 和 5 所示。DBS 的 PR 曲线和 ROC 曲线均高于其他所有方法。

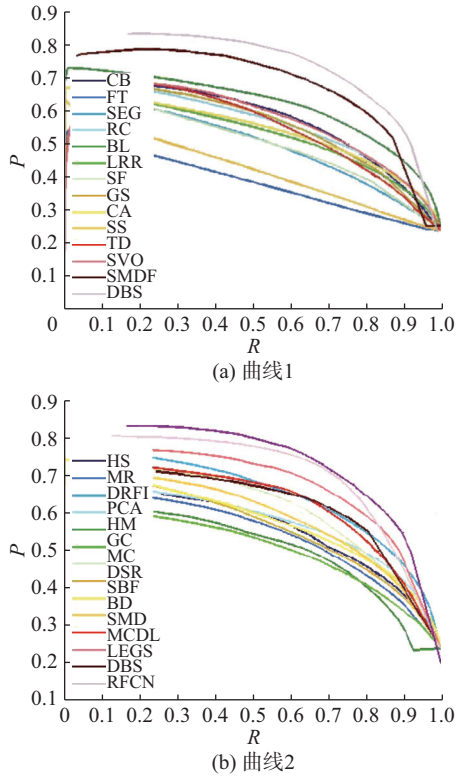


图4 PR 曲线图

Fig. 4 PR curves

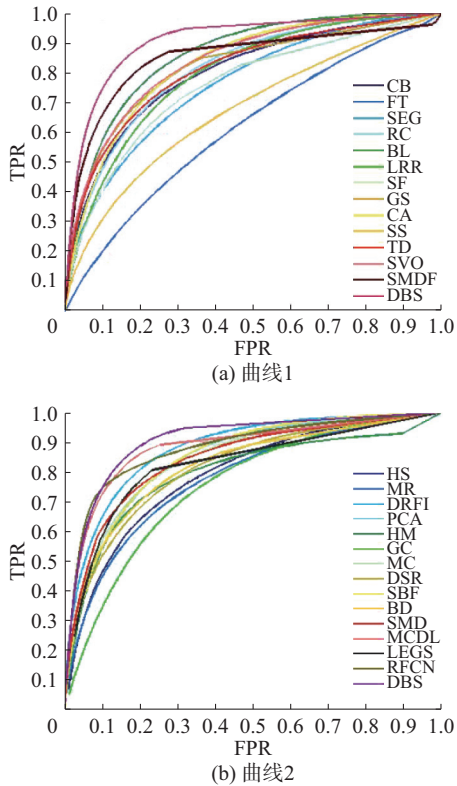


图5 ROC 曲线图

Fig. 5 ROC curves

4.3 视觉效果比较

选择典型图像和 27 种流行方法进行视觉效果对比,如图 6 所示。

27 种流行方法的检测结果存在如下问题: 1) 存不完整的现象,如 SMDF^[5]、LRR^[25]、GS^[27]; 2) 存在包含非显著区域的现象,如 LEGS^[6]、RFCN^[19]、SS^[29]、TD^[30]; 3) 存在边界模糊不清的现象,如 SEG^[22]、SVO^[24]、SS^[29]; 4) 存在只高亮地显示边缘的现象,如 CA^[27]、PCA^[31]。此外,流行的深度学习检测方法的检测性能也不尽相同,原因在于输入到网络图像块的上下文信息不同,学习到的特征不一样,导致对比度的计算结果也不同。

本文方法综合考虑了 CNN 深度特征和标签语义特征,保证了本文方法得到的显著区域相对完整、均匀高亮。

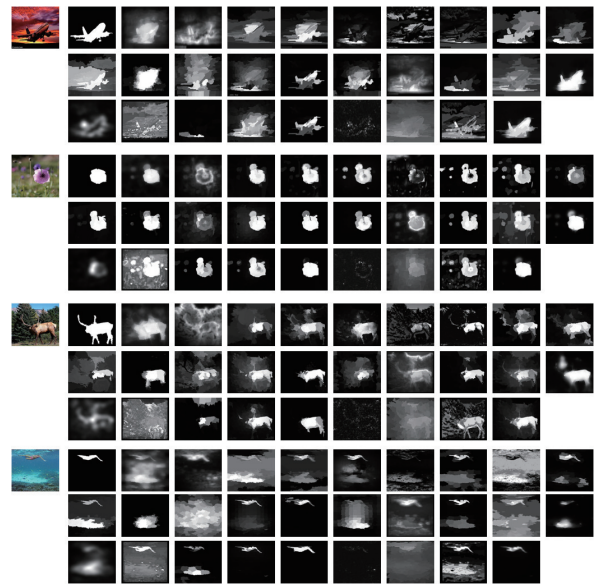


图6 DBS 方法与 27 种流行方法的视觉效果比较图

Fig. 6 Visual comparison between DBS with 27 popular methods

图 6 中图像出现的顺序为: 原始图像、标准二值标注、BL^[18]、CA^[28]、CB^[21]、DRFI^[3]、DSR^[36]、FT^[14]、GC^[34]、GS^[27]、HM^[33]、HS^[2]、LEGS^[6]、LRR^[25]、MC^[35]、MCDL^[7]、MR^[31]、PCA^[32]、BD^[38]、RC^[23]、RFCN^[19]、SBF^[27]、SEG^[22]、SF^[26]、SMD^[37]、SMDF^[5]、SS^[29]、SVO^[24]、TD^[30]、DBS。

5 结束语

本文提出了基于深度学习的显著区域检测方法,该方法面向带有标签的社群图像,将标签信息纳入到显著区域的提取方法中。基于深度学习的特征包括 CNN 特征和标签特征。此外,随着社群图像快速发展,文本构建了面向社群图像的带有标签信息的显著性数据集,为面向社群图像的显著区域检测研究提供了新的实验数据。大量的实验证明了所提方法的有效性。

参考文献:

- [1] XIAO Chuanmin, SHI Zelin, XIA Renbo, et al. Edge-detection algorithm based on visual saliency[J]. *Information and control*, 2014, 43(1): 9–13.
- [2] YAN Qiong, XU Li, SHI Jianping, et al. Hierarchical saliency detection[C]//*Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Portland, OR, USA, 2013: 1155–1162.
- [3] JIANG Huaizu, WANG Jingdong, YUAN Zejian, et al. Salient object detection: a discriminative regional feature integration approach[C]//*Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Portland, OR, USA, 2013: 2083–2090.
- [4] LIN Yuetan, KONG Shu, WANG Donghui, et al. Saliency detection within a deep convolutional architecture[C]//*Workshops at the 28th AAAI Conference on Artificial Intelligence*. Québec City, Canada, 2014.
- [5] LI Guanbin, YU Yizhou. Visual saliency based on multiscale deep features[C]//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, 2015: 5455–5463.
- [6] WANG Lijun, LU Huchuan, RUAN Xiang, et al. Deep networks for saliency detection via local estimation and global search[C]//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, 2015: 3183–3192.
- [7] ZHAO Rui, OUYANG Wanli, LI Hongsheng, et al. Saliency detection by multi-context deep learning[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA, 2015: 1265–1274.
- [8] WANG Wen, LANG Congyan, FENG Songhe. Contextualizing tag ranking and saliency detection for social images[M]//LI Shipeng, EL SADDIK A, WANG MENG, et al. *Advances in Multimedia Modeling*. Berlin Heidelberg, Germany: Springer, 2013: 428–435.
- [9] ZHU Guokang, WANG Qi, YUAN Yuan. Tag-saliency: combining bottom-up and top-down information for saliency detection[J]. *Computer vision and image understanding*, 2014, 118: 40–49.
- [10] LIU Tie, SUN Jian, ZHENG Nanning, et al. Learning to detect a salient object[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Minneapolis, Minnesota, USA, 2007: 1–8.
- [11] LI Jian, LEVINE M D, AN Xiangjing, et al. Visual saliency based on scale-space analysis in the frequency domain[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(4): 996–1010.
- [12] MARTIN D, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]//*Proceedings of the 8th IEEE International Conference on Computer Vision*. Vancouver, BC, USA, 2001, 2: 416–423.
- [13] BATRA D, KOWDLE A, PARIKH D, et al. iCoseg: Interactive co-segmentation with intelligent scribble guidance[C]//*Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. San Francisco, CA, USA, 2010: 3169–3176.
- [14] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Miami, FL, USA, 2009: 1597–1604.
- [15] JIA Yangqing, DEVELOPER L, SHELHAMER E. Caffe [N/OL]. [2016-11-12]. <http://caffe.berkeleyvision.org/>, 2013.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada, USA, 2012: 1097–1105.
- [17] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Columbus, OH, USA, 2014: 580–587.
- [18] TONG Na, LU Huchuan, RUAN Xiang, et al. Salient object detection via bootstrap learning[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, 2015: 1884–1892.
- [19] WANG Linzhao, WANG Lijun, LU Huchuan, et al. Saliency detection with recurrent fully convolutional networks [C]//*Proceedings of the 14th European Conference Computer Vision—ECCV 2016*. Cham, Germany, 2016: 151–157.
- [20] KRÄHENBÜHL P, KOLTUN V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C]//*Proceedings of the 24th International Conference on Neural Information Processing Systems*. Granada, Spain, 2012: 109–117.
- [21] JIANG Huaizu, WANG Jingdong, YUAN Zejian, et al. Automatic salient object segmentation based on context and shape prior[C]//*Proceedings of the British Machine Vision Conference (BMVC)*. Dundee, Scotland, 2011: 1–12.
- [22] RAHTU E, KANNALA J, SALO M, et al. Segmenting salient objects from images and videos[C]//*Proceedings of the 11th European Conference on Computer Vision (EC-*

- CV). Crete, Greece, 2010: 366–379.
- [23] CHENG Mingming, ZHANG Guoxin, MITRA N J, et al. Global contrast based salient region detection[C]//Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA, 2011: 409–416.
- [24] CHANG K Y, LIU T L, CHEN H T, et al. Fusing generic objectness and visual saliency for salient object detection[C]//Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain, 2011: 914–921.
- [25] SHEN Xiaohui, WU Ying. A unified approach to salient object detection via low rank matrix Recovery[C]//Proceedings of the 2012 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island, USA, 2011: 853–860.
- [26] PERAZZI F, KRÄHENBUHL P, PRITCH Y, et al. Saliency filters: contrast based filtering for salient region detection[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island, USA, 2012: 733–740.
- [27] WEI Yichen, WEN Fang, ZHU Wangjiang, et al. Geodesic saliency using background priors[C]//Proceedings of the 12th European Conference on Computer Vision (ECCV). Florence, Italy, 2012: 29–42.
- [28] GOFERMAN S, ZELNIK-MANOR L, TAL A. Context-aware saliency detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA, 2010: 2376–2383.
- [29] HOU Xiaodi, HAREL J, KOCH C. Image signature: highlighting sparse salient regions[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34(1): 194–201.
- [30] SCHARFENBERGER C, WONG A, FERGANI K, et al. Statistical textural distinctiveness for salient region detection in natural images[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, Oregon, USA, 2013: 979–986.
- [31] YANG Chuan, ZHANG Lihe, LU Huchuan, et al. Saliency detection via graph-based manifold ranking[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA, 2013: 3166–3173.
- [32] MARGOLIN R, TAL A, ZELNIK-MANOR L. What makes a patch distinct?[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA, 2013: 1139–1146.
- [33] LI Xi, LI Yao, SHEN Chunhua, et al. Contextual hypergraph modeling for salient object detection[C]//Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, NSW, Australia, 2013: 3328–3335.
- [34] CHENG Mingming, WARRELL J, LIN Wenyan, et al. Efficient salient region detection with soft image abstraction[C]//Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia, 2013: 1529–1536.
- [35] JIANG Bowen, ZHANG Lihe, LU Huchuan, et al. Saliency detection via absorbing Markov chain[C]//Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, NSW, Australia, 2013: 1665–1672.
- [36] LI Xiaohui, LU Huchuan, ZHANG Lihe, et al. Saliency detection via dense and sparse reconstruction[C]//Proceedings of the 2013 IEEE International Conference on Computer Vision. Washington, DC, USA, 2013: 2976–2983.
- [37] PENG H, LI B, LING H, et al. Salient object detection via structured matrix decomposition[J]. IEEE transactions on pattern analysis & machine intelligence, 2017, 39(4): 818–832.
- [38] ZHU Wangjiang, LIANG Shuang, WEI Yichen, et al. Saliency optimization from robust background detection[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA, 2014: 2814–2821.

作者简介:



梁晔, 1978 年生, 女, 讲师, 主要研究方向为图像处理和机器学习, 发表中文核心和三大检索论文 10 余篇。



于剑, 男, 1969 年生, 教授, 博士生导师, 博士, 主要研究方向为 2005 年分别获得第八届北京青年优秀科技论文奖一等奖、第七届詹天佑铁道科技奖北京交通大学专项基金奖, 2006 年获得霍英东青年教师基金。发表学术论文 30 余篇。