

基于双层多尺度神经网络的显著性对象检测算法

李 鑫¹, 陈雷霆^{1,2,3}, 蔡洪斌^{1,2}, 李建平¹, 杨 帆¹

(1 电子科技大学 计算机科学与工程学院, 四川 成都 611731; 2 电子科技大学 数字媒体技术
四川省重点实验室, 四川 成都 611731; 3 电子科技大学 广东电子信息工程研究院, 广东 东莞 523000)

摘 要: 为了提高显著性对象检测的准确率, 本文提出一种基于双层多尺度神经网络的深度模型. 不同于现有的深度神经网络模型, 首先, 该模型以由精到粗的方式进行深度特征学习, 并且定位显著性对象的初始位置; 然后, 以由粗到精的方式整合多尺度上下文语义信息, 从而精确检测整个显著性对象区域, 输出相应的显著性图; 最后, 为了进一步提高检测结果的准确率, 利用全连接条件随机场对输出的显著性图进行优化, 得到最终的显著性对象检测结果. 在多个显著性对象检测公共数据集的验证结果表明, 本文算法在运行效率和准确率上均优于当前传统显著性对象检测算法以及现有的基于深度学习的显著性对象检测算法.

关键词: 显著性对象检测; 深度学习; 深度卷积网络; 条件随机场

中图分类号: TP 391. 41 **文献标识码:** A **文章编号:** 1000—7180(2018)11—0001—07

Salient Object Detection Algorithm Based on Dual-Layer Multi-Scale Neural Network

LI Xin¹, CHEN Lei-ting^{1,2,3}, CAI Hong-bing^{1,2}, LI Jian-ping¹, YANG Fan¹

(1 School of Computer Science & Engineering, University of Electronic Science and
Technology of China, Chengdu 611731, China;

2 Digital Media Technology Key Laboratory of Sichuan Province, University of Electronic Science and
Technology of China, Chengdu 611731, China;

3 Institute of Electronic & Information Engineering in Guangdong, University of Electronic Science and
Technology of China, Dongguan 523000, China)

Abstract: To further improve the accuracy of salient object detection, a novel Dual-Layer Multi-Scale Neural Network (DLMSNN) was proposed in this work. Different from existing deep model, the proposed deep model first learned deep features in a fine-to-coarse manner, and roughly located the salient object regions. Then, the model integrated multi-scale contextual information in a coarse-to-fine manner to precisely detect the entire salient object regions, and generated accurate saliency map of the input image. Finally, to further improve the performance, the dense conditional random field algorithm was used to refine the saliency map and produce the final result. The experimental results on several public benchmarks showed that the proposed algorithm outperformed traditional salient object detection methods and existing deep learning-based algorithms.

Key words: salient object detection; deep learning; deep neural network; conditional random field

1 引言

人类的视觉注意力通常被一个场景中少数的内

容吸引, 从而自动忽视大量无关紧要的信息, 使得人类大脑可以更集中、细致地处理重要区域的视觉信号. 人类的这种视觉选择机制事实上也可极大地帮

助各种复杂的计算机任务^[1],包括:人体姿态估计、语义分割、视频标注、稠密语义匹配等. 尽管如此,利用计算机模拟人类的这种视觉机制,快速准确地检测给定场景中视觉重要的物体仍然是一个极具挑战的难题. 近 20 年来,视觉显著性一直是计算机领域、神经科学领域以及心理学领域的研究热点. 在计算机和人工智能领域,视觉显著性研究主要包括两个方向:(1)显著性对象检测;(2)眼动估计.

本文主要研究显著性对象检测,即如何快速检测出给定场景中的视觉显著对象. 传统的显著性对象检测方法主要利用低级的图像特征,包括颜色、梯度、角度等,并结合简单的先验知识(如中心先验,边缘先验等)和观察进行建模^[2-3]. 然而,传统显著性对象检测算法提取的图像特征缺乏足够的语义信息,因而不能很好地应对复杂的场景. 近年来,基于大量的训练样本,人们利用深度神经网络来自动学习具有语义信息的深度特征,从而大大提高了显著性对象检测的准确率^[4-5]. Lee 等^[6]提出了一种深度网络模型训练图像的深度语义特征,实现显著性对象的检测. Li 等^[7]通过多任务神经网络模型提高了显著性对象检测算法的准确性. 上述算法均利用深度神经网络学习强语义特征处理结构复杂、背景杂乱的场景. 但由于对象的多样性、训练数据库的样本偏差(Biases)等,在显著性对象结构和背景复杂的情况下,现有基于深度模型的显著性对象检测算法仍然无法很好的解决二义性问题,导致生成的显著性图模糊,与用户标注存在一定差距. 如何更好地解决复杂场景中显著性对象的检测问题仍然是一个有待探索的问题.

2 双层多尺度神经网络模型

为了让深度神经网络在训练和检测的过程中感知图像的上下文语义信息,并利用这些信息更好学习深度特征,本文提出双层多尺度神经网络模型(Dual-Layer Multi-Scale Neural Network, DLM-SNN). 如图 1 所示,该模型包括三个主要部分:特征提取网络,多尺度金字塔,双层级联神经网络(浅层级联神经网络和深层级联神经网络). 特征提取网络用于提取深度特征,构建输入图像对应的特征图. 多尺度金字塔用于将特征图池化为不同尺度,构建一系列不同尺度的特征图. 双层级联神经网络则用于整合上下文语义信息:浅层级联神经网络以由精

到粗的方式整合多尺度上下文信息,并且大致定位显著性对象区域;深层级联神经网络基于浅层级联神经网络的检测结果,以由粗到精的方法进一步优化中间结果,生成准确的显著性图. 本章主要从上述三个主要部分介绍双层多尺度神经网络模型.

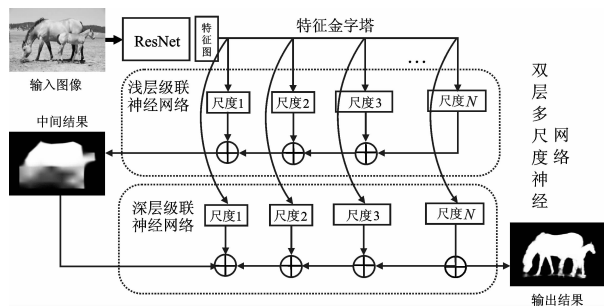
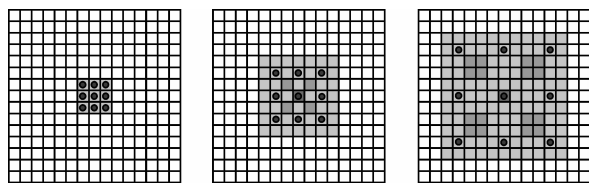


图 1 基于双层多尺度神经网络结构的深度模型

2.1 特征提取网络

对于任意输入图像 I , 特征提取网络构建具有足够精度的特征图 $F(\theta, I)$ (后文简称为 F), 其中 θ 表示特征提取网络的网络参数. 本文运用 Res-Net^[8] 网络作为模型的底层结构, 并且为了保证特征图具有足够的精度, 采用扩张卷积技术^[9]. 来取代传统 Res-Net 后三组的普通卷积操作. 如图 2 所示, 扩张卷积通过扩大感受野范围获取更多的信息, 由于与普通卷积网络卷积核大小一致, 不会增加额外的参数. 特征提取网络构建的特征图 F 具有较高精度, 从而可以构建多尺度金字塔.



(a) 一倍扩张卷积 (b) 二倍扩张卷积 (c) 四倍扩张卷积

图 2 扩张卷积示意图

2.2 多尺度金字塔

为了使得深度神经网络能够整合上下文语义信息, 本文利用深度特征图构建多尺度多尺度金字塔. 具体来说, 基于特征提取网路输出的特征图 F , 首先将其池化为 N 个不同尺度的特征图 $F_S \in \{F_{S1}, F_{S2}, \dots, F_{SN}\}$. 其次, 运用一个 1×1 的卷积层进一步降低不同尺度的特征图 F_S 的维度, 使其维度由原本的 2 048 维降低到 512 维, 从而减少后续操作的运算量. 最后, 把降维后的特征图作为后续双层级联神经网络的输入, 学习高级语义特征并完成显著性对象的提取. 降维后的多尺度金字塔可记为一个多尺

度特征图集合,记为 $f = \{f_{s_1}, f_{s_2}, \dots, f_{s_N}\}$.

多尺度金字塔类似于传统的特征金字塔,即构建多尺度的图像特征. 本文模型将其以网络结构的形式集成到深度神经网络中. 因此,相比传统的多尺度金字塔,多尺度金字塔的优点在于可以通过大量的样本自动学习,自动地调整融合的权重,从而更好地利用多尺度信息,排除干扰因素.

2.3 双层级联神经网络

双层级联神经网络用于整合图像的上下文语义信息. 该网络包括两个子网络:(1)浅层级联神经网络;(2)深层级联神经网络. 浅层级联神经网络用于生成显著性对象的初始位置,生成中间结果(初始显著性图). 深层级联神经网络基于中间结果进一步以由粗到精的方式整合多尺度深度特征,排除错误信息,生成最终的显著性图.

2.3.1 浅层级联神经网络

浅层级联神经网络的输入为降维后的多尺度特征图集合 $f = \{f_{s_1}, f_{s_2}, \dots, f_{s_N}\}$, 输出为初始显著性图 $S^{\text{init}}(\theta_1, f)$. 浅层级联神经网络包含 N 个不同尺度的特征图,通过由精到粗的方式学习图像的显著性特征,并定位显著性对象的大致位置,如图1所示. 该网络的每个尺度的特征学习基于相邻更精尺度学习的深度特征,

$$F'_{s_i} = \begin{cases} f_{s_i} + \partial_{s_i} \mathbb{R}_{s_i}(F_{s_{i-1}}^L), & i \leq N-1 \\ f_{s_i}, & i = N \end{cases} \quad (1)$$

式中, $\mathbb{R}_{s_i}(\cdot)$ 表示插值操作,用于保证前一尺度的学习的特征 $F_{s_{i-1}}^L$ 与该尺度(s_i)的输入深度特征 f_{s_i} 具有相同的精度; ∂_{s_i} 表示对应尺度的整合权重.

网络参数 θ_1 的学习可以视为一个回归问题,即基于该参数预测的结果需要向用户标注结果回归. 本文定义如下目标函数用于训练网络参数:

$$L_{\text{init}}(\theta_1) = \min_{\theta_1} \sum_j l(G^j, S^{\text{init}}(\theta_1, f_{s_i})) \quad (2)$$

式中, G^j 表示第 j 张训练样本对应的用户标注; $l(\cdot)$ 表示损失函数,用于计算用户标注和网络预测结果的差异,具体计算方法如下:

$$l(G, S^{\text{init}}; \theta_1) = - \sum_{p=1}^{|I|} G(p) \log \Pr(S^{\text{init}}(p) = 1 \mid I; \theta_1) + (1 - G(p)) \log \Pr(S^{\text{init}}(p) = 0 \mid I; \theta_1) \quad (3)$$

式中, p 表示显著性图的坐标位置. 损失函数用于计算每个训练图像预测结果与用户标注之间的像素级

差异. 其中, $G(p)$ 表示在坐标 p 处的用户标注,若是像素在坐标 p 处属于显著性对象区域,则 $G(p) = 1$,反之 $G(p) = 0$. 而 $S^{\text{init}}(p)$ 表示神经网络基于参数 θ_1 在坐标 p 处的预测结果.

该浅层级联神经网络通过整合多尺度深度特征来大致定位图像中显著性图的位置,并生成初始显著性图作为模型的中间结果. 这个结果虽然并不准确,但却为深层级联神经网络提供了先验知识、帮助其排除了杂质,使其能更准确的检测出完整的显著性对象区域.

2.3.2 深层级联神经网络

深层级联神经网络的输入包括:(1)多尺度特征图 $f = \{f_{s_1}, f_{s_2}, \dots, f_{s_N}\}$;(2)浅层级联神经网络输出的中间结果 $S^{\text{init}}(\theta_1, f)$, 输出为模型的最终显著性图 $S^{\text{final}}(\theta_2, f, S^{\text{init}}(\theta_1, f))$. 与浅层级联神经网络类似,深层级联神经网络也包含 N 个不同尺度的特征图,但是这些特征图的整合通过由粗到精的方式进行,如图1所示,每个尺度的深度特征学习基于相邻更粗尺度学习的深度特征,通过如下方法表示:

$$\tilde{F}_{s_i} = \begin{cases} f_{s_i} + \tilde{\partial}_{s_i} \mathbb{R}(\tilde{F}_{s_{i-1}}^L), & i \geq 2 \\ f_{s_i} \oplus S^{\text{init}}, & i = 1 \end{cases} \quad (4)$$

式中, $\tilde{F}_{s_i}$ 表示尺度 s_i 对应的底层特征图; $\tilde{F}_{s_{i-1}}^L$ 表示学习的深度特征; 而 $\tilde{\partial}_{s_i}$ 表示深层级联神经网络对应的特征整合权重; $\oplus$ 表示非线性整合方法,根据如下公式计算:

$$\tilde{F}_{s_1} = \sigma\{\text{cat}(f_{s_1}, S^{\text{init}}) \otimes W + b\} \quad (5)$$

式中, $\otimes$ 表示卷积操作; W 表示卷积滤波器(convolutional filters); b 表示偏置(biases). 本文运用 Relu 作为非线性激活函数 $\sigma\{\cdot\}$. $\text{cat}(\cdot)$ 表示一个连接函数,用来连接特征图 f_{s_1} 和中间结果 S^{init} .

类似于公式(2),训练网络参数的目标函数定义如下:

$$L_{\text{final}}(\theta_2) = \min_{\theta_2} \sum_j l(G^j, S^{\text{final}}(\theta_2, f, S^{\text{init}}(\theta_1, f))) \quad (6)$$

公式(6)中的损失函数 l 与浅层级联神经网络相同.

双层级联神经网络为一个整体框架,两个子网络的深度特征学习统一在一个训练框架中. 因此,双层级联神经网络的目标函数 L_{ALL} 定义如下:

$$L_{ALL}(\theta_1, \theta_2) = L_{init}(\theta_1) + \lambda L_{final}(\theta_2) \quad (7)$$

式中, $\lambda = 1.5$ 为权重, 用于使得网络参数学习过程中倾向于最终的输出结果与用户标注更为接近。

该双层级神经网络塔模型可以有效整合图像的上下文语义信息. 从网络结构的角度看, 该网络也可以视为一个由粗到精的优化模型: 浅层级神经网络大致估计对象骨骼的初始检测结果, 而初始结果在深层级神经网络中被不断优化, 从而降低检测错误, 最终生成一个高质量的检测结果。

3 显著性图优化方法

尽管提出的深度模型已经可以获得准确的检测结果, 但随着深度模型的尺度变大, 感受野会逐渐变大, 更多的背景杂质被引入, 导致显著性对象的边缘信息模糊, 检测准确率受到影响. 针对此问题, 本文模型采用全连接条件随机场(Dense CRF), 利用颜色和亮度信息来进一步优化深度模型的检测结果 $S^{final}(\theta_2, f, S^{init}(\theta_1, f))$. 本文将显著性图的优化看作图切问题, 通过优化能量方程, 获取最优解. 定义如下能量方程:

$$E(M) = - \sum_i \log P(m_i) + \sum_{i,j} \omega_{i,j}(m_i, m_j) \quad (8)$$

上述能量方程包括两个主要部分: 数据项和平滑项. 其中 M 表示显著性值对应的标签, 且 $M \in \{0, 1\}$, 当 $M(i) = 1$ 时表示像素 i 为显著性, 反之当 $M(i) = 0$ 表示该对应像素不显著. $\log P(m_i)$ 为数据项, 其中 $P(m_i)$ 表示像素 i 为对应显著性标签的概率. 这里简单定义 $P(1) = S^{final}(i)$, 而 $P(0) = 1 - S^{final}(i)$. $\omega_{i,j}(l_i, l_j)$ 为标签平滑项, 既相邻且颜色相近的像素应该赋予相同的标签. 平滑项用于防止产生独立的小区域, 且帮助更好的保存显著性对象区域的边缘信息, 计算方法如下:

$$\omega_{i,j}(l_i, l_j) = \mu(l_i, l_j) \left\{ \gamma_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\nu^2}\right) - \frac{\|I_i - I_j\|^2}{2\sigma^2} \right\} + \gamma_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\kappa^2}\right) \quad (9)$$

式中, p_i 和 p_j 表示像素 i 和像素 j 所对应的坐标位置; I_i 与 I_j 表示像素 i 和像素 j 所对应的像素亮度(pixel intensities). 当 $l_i \neq l_j$ 时, $\mu(l_i, l_j) = 1$; 反之, $\mu(l_i, l_j) = 0$. ν 和 σ 为能量方法的控制参数, 用来分别控制位置和颜色的相似性; κ 用来防止形成独立不连续的区域.

求解公式(8)是典型的 NP-难问题. 为了获取其

最优解, 本文采用平均差近似方法(mean field approximation), 优化后的显著性图为最终结果。

4 实验与分析

4.1 实验环境

实验的硬件环境为 NVIDIA GTX Titan X GPU, 12 G 显存. 软件环境为 Ubuntu 14.04 以及 caffe 框架. 训练数据集为 MSRA-B 提供的训练数据集 $\{I_{train}, G_{train}\}$, 该数据集包含 2 500 张带有用户标注的训练图像。

与相关文献[2,3,6,7]一致, 测试集采用了三个公共数据集^[1], 包括 ECSSD, DUT-OMRON, 以及 HKU-IS. ECSSD 为目前最为广泛运用的显著性对象检测公共数据集. 该数据库包含 1 000 张图像和对象的用户标注显著性图, 这些图像具有复杂的结构和背景, 包含一个或者多个显著性对象. DUT-OMRON 是目前最大的显著性对象检测公共数据集, 包含 5 168 张非常复杂的图像以及用户标注的显著性图. HKU-IS 是最新公布的数据库, 该数据库被公认包含最少的偏差(bias), 它包含 1 447 张背景非常复杂且包含一个或者多个显著性对象的图像。

4.2 训练和测试细节

训练和测试的步骤如下:

步骤 1, 数据增强. 为了防止过度拟合问题, 训练数据集根据如下方式进行数据增强: (1) 将 $\{I_{train}, G_{train}\}$ 中所有图像每 90 度旋转生成不同方向的图像作为训练数据; (2) 再将所有图像水平方向翻转, 生成更多的训练数据. 最终生成增强后的训练数据 $\{\hat{I}_{train}, \hat{G}_{train}\}$.

步骤 2, 训练参数设置. 双层多尺度神经网络采用 5 个尺度, 分别设置为 $10 \times 10, 15 \times 15, 20 \times 20, 30 \times 30, 60 \times 60$. 网络参数训练的初始学习率设置为 10^{-7} , 并且学习率采用“poly”策略, 即学习率根据 $(1 - \frac{iter}{max_iter})^{power}$ 自动调整, power 设置为 0.9. 最大循环次数为 30 000. 所有训练图像尺度被调整为 480×480 , 以保证具有足够的精度构建多尺度金字塔网络。

步骤 3, 参数初始化. 本文所提模型的参数初始化包括两步. 第一, 由于特征提取网络使用了网络扩展技术, 所以不能直接使用 Image Net 上训练的参

数初始化. 因此,这里运用在 PASCAL-VOC 基于语义分割任务训练的网络参数初始化特征提取网络的参数 θ . 第二,双层级联神经网络的参数 $\{\theta_1, \theta_2\}$ 使用正态分布随机初始化.

步骤 4,深度神经网络训练. 深度神经网络参数训练的目标函数为公式(7),损失函数计算根据公式(3). 深度神经网络的训练利用随机梯度下降(SGD)方法,训练数据集为数据增强后的训练数据集 $\{\hat{I}_{\text{train}}, \hat{G}_{\text{train}}\}$.

步骤 5,测试和后处理. 对于任意用户输入图像 I ,先利用训练好的深度神经网络模型,直接计算出相应的显著性图 $S^{\text{final}}(I; \theta_2, f, S^{\text{init}}(\theta_1, f))$. 然后,显著性图利用公式(8)优化,输出最终检测结果.

4.3 评价标准

与相关文献[2,3,6,7]一致,评价标准采用:精确率-召回曲线(PR-curves),F-measure 以及绝对方差错误(MAE).

根据检测的显著性图,基于不同的阈值,将显著性图转化为一组二值图 B ,并且比较它们和标准标注 G 的差异,获得给定显著性图的精确率 $P = \frac{|B \cap G|}{|B|}$,以及召回率 $R = \frac{|B \cap G|}{|G|}$. F-measure 指标反映了方法的准确性,基于不同阈值计算精确率-召回曲线,F-measure(F_β) 的计算如下:

$$F_\beta = (1 + \beta^2) \frac{P \times R}{\beta^2 P + R}$$

(10)

根据相关文献[2,7]的建议 β^2 设置为 0.3. 绝对方差错误是计算每个像素和用户标注 (Ground Truth, GT) 的平均误差,计算如下:

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)|$$

(11)

式中, W 和 H 分别代表图像的长度和宽度.

4.4 对比分析

为了验证本文算法的有效性,与五个显著性对象检测算法进行对比,其中包括两个最具代表性的传统显著性对象检测算法:基于鲁棒背景检测显著性优化(Robust Background Detection, RBD)^[2]和显著性区域融合算法(Discriminative Regional Feature Integration, DRFI)^[3]. 三个近期发表的基于深度学习的显著性对象检测方法:基于距离图深度显著性检测算法(Encoded Low-level Distance, ELD)^[6],基于深度网络的显著性对象检测算法

(Deep Saliency, DS)^[7],基于深度对比度的显著性检测算法(Deep Contrast Learning, DCL)^[10].

4.4.1 定量分析

根据不同验证方法,定量比较本文所提方法和其他相关方法. 如表 1 所示,根据 F-measure 的检测结果,可以得出如下结论:(1)基于深度学习的显著性对象检测算法比传统显著性对象检测算法获得更高的 F-measure;(2)本文算法的 F-measure 在三个公共数据集上都高于现有的基于深度学习的显著性对象检测算法;(3)运用显著性图优化方法可进一步提高了本文算法的 F-measure 值. 主要原因是本文方法能更好地整合图像的上下文语义信息,完整地检测整个对象区域. 另外,显著性图优化方法能够保存完整的边缘,使得检测结果的准确率进一步提高,从而使得结果更加准确.

表 1 不同方法在三个公共数据集的比较结果.

方法	ECSSD		DUT-OMRON		HKU-IS	
	F_β	MAE	F_β	MAE	F_β	MAE
RBD	0.716	0.171	0.630	0.144	0.657	0.120
DRFI	0.786	0.164	0.665	0.155	0.684	0.226
DS	0.882	0.122	0.716	0.120	0.866	0.079
ELD	0.869	0.098	0.720	0.091	0.777	0.121
DCL	0.887	0.072	0.718	0.094	0.879	0.059
Ours (w/o CRF)	0.913	0.051	0.763	0.067	0.899	0.040
Ours	0.922	0.049	0.776	0.061	0.907	0.038

注: F_β 值越高代表准确度越高;MAE 值越低代表错误越少.

此外,由 MAE 评估标准可知(见表 1),本文算法在三个公共数据集上都获得了更低的 MAE 值. 说明本文方法检测的显著性值更接近用户标注. 通过 MAE 的实验结果,可以得出如下结论:(1)本文方法具有更小的绝对方差错误;(2)显著性图优化算法明显提高了检测结果的准确率. 对比现有基于深度学习方法,本文算法在三个公共数据集(ECSSD, DUT-OMRON 以及 HKU-IS)上绝对方差错误分别降低了 31.9%,33.0%以及 35.6%. 本文算法可以生成更接近于用户标注的结果,主要原因在于本文方法整合了多尺度语义信息,从而获得更加完整的检测出显著性对象区域. 除此之外,利用显著性图优化算法可以帮助进一步排除杂质,降低错误.

如图 3 所示,在精确率-召回曲线(PR-curves)上,本文算法在三个公共数据集上也获得更优结果. 实验结果证明在深度网络设计中考虑图像多尺度语

义信息能够提高模型检测的准确率。

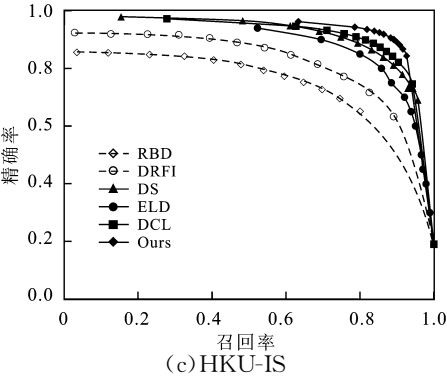
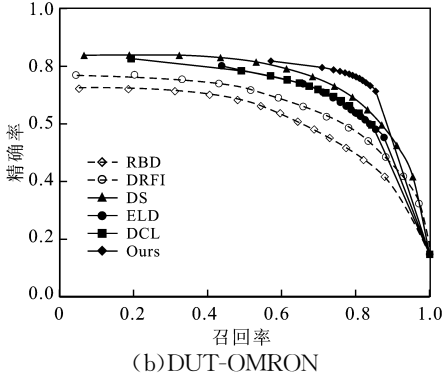
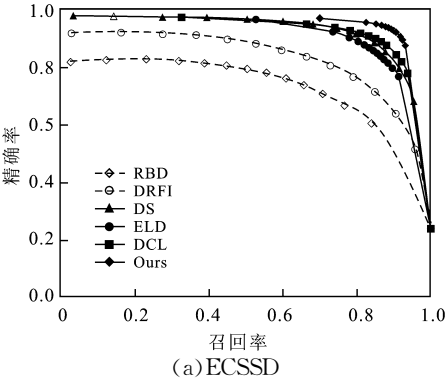


图3 不同方法在三个公共数据集的PR曲线比较结果

4.4.2 定性分析

如图4所示,本文算法在检测复杂场景时,能够更完整地检测显著性对象,并且最终生成的结果与用户标注更加接近.这些测试图像具有复杂的背景,其包含的显著性对象也具有各异的结构.对比现有基于深度学习的方法,由于本文算法引入了图像的上下文语义信息,从而可以克服二义性问题,准确的检测出完整的显著性对象区域.除此之外,运用显著性图优化方法也帮助进一步排除了错误,更好地保存了对象的边缘信息.

在复杂场景中检测显著性对象需要更强的语义信息,由于传统的显著性对象检测算法不能获得图

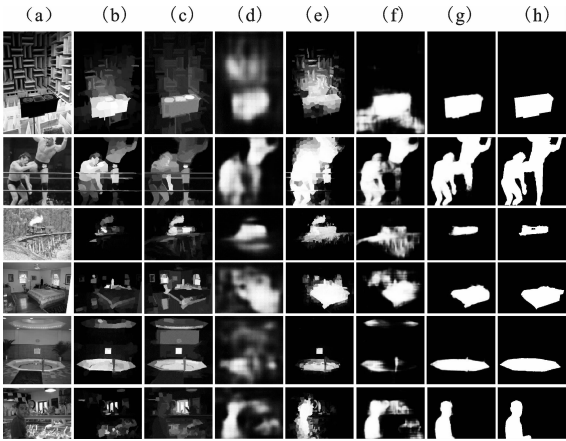


图4 不同方法定性比较

从左至右:(a)输入图像;(b)RBD;(c)DRFI;(d)DS;
(e)ELD;(f) DCL;(g) 本文所提;(h)用户标注

像的强语义特征,因而缺乏准确性.而基于深度学习的显著性对象检测算法通过深度模型语义级的特征,从而检测更加集中在有意义的对象区域.本文算法由于在模型训练过程中整合了图像的上下文语义信息,因而更能准确地发现复杂场景中的显著性对象,获取更好的检测结果.

4.4.3 效率分析

运行效率测试结果如表2所示.传统方法的测试平台为 i7 2.50 GHz CPU 以及 8 GB RAM;而基于深度学习的显著性对象检测方法的测试平台为 NVIDIA GTX Titan X GPU 以及 12 G 显存.实验结果表明,传统方法的运行效率低于基于深度学习的显著性对象检测算法.本文的深度模型运行效率为 8 fps.但是由于本文算法的结果需要额外的后处理,因此处理一张输入图像的整体运行时间为0.51 s,依然优于其余对比方法.

表2 运行效率比较

方法	时间	方法	时间	方法	时间
RBD	0.52	DS	0.73	DCL	1.17
DRFI	6.34	ELD	0.59	Ours	0.51

5 结束语

本文提出一种基于深度学习的显著性对象检测算法.不同于现有基于深度学习的显著性对象检测算法,本文算法在模型设计中考虑了图像上下文语义信息,克服了检测复杂显著性对象时的二义性问题,能够生成准确、清楚的显著性图.并利用全连接条件随机场优化模型输出的显著性图,生成更接近

于用户标注的结果. 实验结果表明, 比较传统显著性对象检测算法和现有基于深度学习的检测算法, 本文算法在定量和定性指标上都有改善, 证实了在显著性对象检测中考虑图像的上下文语义信息的必要性. 如何将双层神经网络用于计算机视觉的相关任务, 包括语义分割、深度图估计等, 是未来工作值得关注的方向.

参考文献:

- [1] Borji A, Cheng M M, Jiang H, et al. Salient object detection: a benchmark[J]. IEEE Transactions on Image Processing, 2015, 24(12): 5706-5722.
- [2] Zhu W, Liang S, Wei Y, et al. Saliency optimization from robust background detection[C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014. Piscataway, NJ: IEEE, 2014:2814-2821.
- [3] Jiang H, Wang J, Yuan Z, 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 2013. Piscataway, NJ: IEEE, 2013:2083-2090.
- [4] 徐新, 穆楠, 张晓龙. 一种鲁棒的夜间图像显著性对象检测模型[J]. 软件学报, 2018, 29(9): 2616-2631
- [5] 周洋, 何永健, 刘晓琪, 等. 结合立体视觉舒适度的立体图像显著性检测[J]. 软件学报, 2017, 28(s2): 1-10.
- [6] Lee G, Tai Y W, Kim J. Deep saliency with encoded low level distance map and high level features[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern, CVPR 2016. Las Vegas, Nevada, USA, 2016 :660-668.
- [7] Li X, Zhao L, Wei L, 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.
- [8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C] // The IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016. Las Vegas, Nevada, USA, NJ: IEEE, 2016.
- [9] Chen LC, 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.
- [10] Li G, Yu Y. Deep Contrast learning for salient object detection[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016. Las Vegas, Nevada, USA, Piscataway, NJ: IEEE. 2016.

作者简介:

李鑫 男, (1986-), 博士研究生. 研究方向为计算机视觉、深度学习、显著性对象检测.

E-mail: Lxinli_uestc@hotmail.com.

陈雷霆 男, (1966-), 博士, 教授, 博士生导师. 研究方向为多媒体应用、人工智能、虚拟现实.

蔡洪斌 男, (1966-), 博士, 教授, 博士生导师. 研究方向为三维图形学、虚拟现实、计算机视觉、机器学习.

李建平 男, (1964-), 博士, 教授, 博士生导师. 研究方向为小波信号处理、模式识别、图像处理.

杨帆 男, (1987-), 博士研究生. 研究方向为深度学习、多媒体、视觉显著性、医学图像分析.