

基于全卷积网络的 语义显著性区域检测方法研究

郑云飞^{1,2,3}, 张雄伟¹, 曹铁勇¹, 孙 蒙¹

(1. 解放军陆军工程大学, 江苏南京 210007; 2. 解放军炮兵防空兵学院, 安徽合肥 230031;
3. 安徽省偏振成像与探测重点实验室, 安徽合肥 230031)

摘 要: 基于底层视觉特征和先验知识的显著性区域检测算法难以检测一些复杂的显著性目标, 人的视觉系统能分辨这些目标是由于其中包含丰富的语义知识. 本文构建了一个基于全卷积结构的语义显著性区域检测网络, 用数据驱动的方式构建从图像底层特征到人类语义认知的映射, 提取语义显著性区域. 针对网络提取的语义显著性区域的缺点, 本文进一步引入颜色信息、目标边界信息、空间一致性信息获得准确的超像素级前景和背景概率. 最后提出一个优化模型融合前景和背景概率信息、语义信息、空间一致性信息得到最终的显著性区域图. 在 6 个数据集上与 15 种最新算法的比较实验证明了本文算法的有效性和鲁棒性.

关键词: 语义信息; 全卷积网络; 颜色外观模型; 显著性区域检测

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2017)11-2593-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2017.11.004

The Semantic Salient Region Detection Algorithm Based on the Fully Convolutional Networks

ZHENG Yun-fei^{1,2,3}, ZHANG Xiong-wei¹, CAO Tie-yong¹, SUN Meng¹

(1. The Army Engineering University of PLA, Nanjing, Jiangsu 210007, China;

2. The Army Artillery and Defense Academy of PLA, Hefei, Anhui 230031, China;

3. The Key Laboratory of Polarization Imaging Detection Technology, Hefei, Anhui 230031, China)

Abstract: The existing salient region detection algorithms based on visual stimulus and prior knowledge are difficult to detect some complicated salient regions. The human vision system can distinguish these complicated salient regions because of the rich semantic knowledge in the human visual system. We construct a semantic salient region detection network using the fully convolutional structure. Learning the mapping from the low-level features to the human semantic cognition, our network can extract semantic salient region effectively. Aiming to the defects of the semantic salient region map, we introduce the color information, object boundary information and spatial consistency information to derive accurate superpixel-level foreground and background probability. At last, we fuse the foreground and background probability, semantic information and spatial consistency information to derive the final salient region map. The experiments comparing with the state-of-the-art 15 algorithms on 6 data sets demonstrate the effectiveness of our algorithm.

Key words: semantic information; fully convolutional network; color appearance model; salient region detection

1 引言

人的视觉系统每秒钟能处理约 $10^8 \sim 10^9$ 比特的图像数据^[1], 这种强大的数据处理能力主要得益于视觉系统中的注意机制. 视觉注意机制使视觉系统能从庞

大的图像数据中提取出最感兴趣的区域, 大大提高了数据处理的效率. 近年来, 大量研究者在研究并试图模仿这种视觉注意机制, 让计算机具备迅速找到图像中显著区域的能力, 并将其应用在多种计算机视觉任务中^[2-4].

收稿日期: 2016-08-11; 修回日期: 2016-12-08; 责任编辑: 马兰英

基金项目: 国家自然科学基金(No. 61471394); 国家青年自然科学基金(No. 61402519); 江苏省自然科学基金(No. BK2012510, No. BK20140071, No. BK20140074)

现有视觉注意模型按照用途可分为视觉关注度预测模型和显著性区域检测模型^[4]. 视觉关注度预测模型侧重于预测人眼对图像中每个像素的关注程度, 显著性区域检测模型侧重于检测出图像中的显著性目标, 本文的研究属于后者.

当前大多数的显著性区域检测模型主要使用底层视觉特征结合先验知识计算显著性, 底层视觉特征如颜色、纹理等, 先验知识如稀有性^[5-8]、紧凑性^[9]、稀疏性^[10]、边界先验^[11]、中心先验^[12]等. 程等人^[6]通过颜色对比度衡量区域的稀有性, 计算显著性区域; 鲍^[7,8]等人用剪切波变换系数的概率密度衡量稀有性并计算显著度; 文献^[9]认为显著性区域比背景区域有更紧凑的颜色分布, 并以此衡量显著性; 文献^[11]认为图像边界上的区域是背景区域, 根据与边界区域在颜色、纹理上的差异大小衡量显著性.

最新的显著性检测算法^[6,11-17]在简单的图像上已经取得了较理想的效果, 但它们难以检测图像中的一些复杂显著性目标, 如颜色非均质的目标、低对比度目标、大尺度目标. 心理学的研究^[18]表明人的视觉系统更趋向于将目标区域作为整体关注, 人的视觉系统包含大量高层次的目标语义知识, 这是人比机器更擅长迅速找到显著目标的重要原因之一, 但如何从图像中提取语义信息并应用到显著性区域检测中是一个难题.

语义信息反映的是从图像底层特征到符合人类认知的目标的关系, 因此语义信息抽取的关键是建立从原始图像特征到符合人类认知的目标含义的映射^[19,20]. 本文受基于全卷积网络的语义分割模型^[21] (Fully Convolutional Semantic Segmentation Networks, FC-SSN) 的启发, 构建了基于全卷积结构的语义显著性区域检测网络 (Semantic Salient Region Detection Network, SSRDN), 用数据驱动的方式学习从图像底层特征到人类语义认知的映射, 提取图像中的语义显著性区域, 并对其进一步优化, 计算出最终的显著区域图. 本文的贡献主要有两点: (1) 提出了一种基于全卷积结构的语义级显著性区域提取方法. (2) 提出了一种基于语义显著性区域的显著性优化方法.

2 本文模型结构

本文模型结构如图 1 所示, 先用本文的 SSRDN 网络提取图像中的语义显著性目标区域, 根据语义显著性区域图用混合高斯模型 (Gaussian Mixture Model, GMM) 建立前景和背景的颜色外观模型, 计算像素级的前景概率和背景概率. 在对前景概率图和背景概率图优化后, 将前景概率信息、背景概率信息、语义信息、空间一致性信息用一个新的优化模型融合得到最终的显著图.

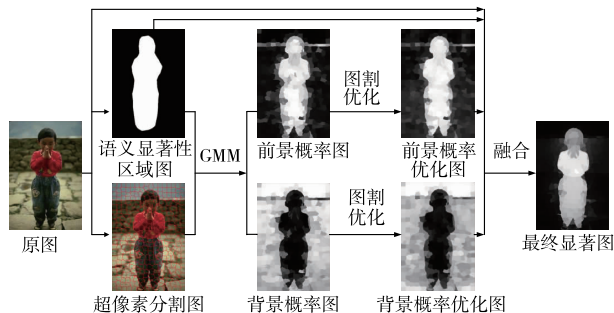


图1 语义显著性区域检测算法框架

3 语义显著性区域检测网络

3.1 网络结构

以往模型难以提取语义显著性区域的主要原因是它们在预处理时先对图像进行过分割, 导致模型无法从语义的角度分析区域的显著性. 文献^[21]提出的 FC-SSN 模型, 不需要对图像进行预分割, 用端到端的方式 (end-to-end) 将整个图像作为处理对象进行训练和预测. 本文受上述思想的启发, 构建了端到端的语义显著性区域检测网络. 网络构造的基本思想是将 VGG^[22] 识别网络改造成全卷积结构并添加反卷积层, 使网络可以直接将原始图像和对应的显著性区域标准图作为训练数据训练, 建立从图像底层特征到语义认知的映射, 提取语义显著性区域. VGG 网络是层次化的图像表示与特征提取模型, 图 2 为 VGG 网络第 10、20、30、38 层输出的部分特征图, 可以看出越浅卷积层提取的特征越接近强度、边缘等底层空间特征, 越深卷积层提取的特征越接近图像中的语义目标区域. 因此想要有效提取图像中的语义信息, 需要利用最深卷积层的输出特征图. 然而, 网络输出特征图的尺寸比原始图像小很多, 因此本文在 VGG 网络最深卷积层后面加上反卷积层, 将网络输出的小尺寸语义特征图插值到原始图像的尺寸.

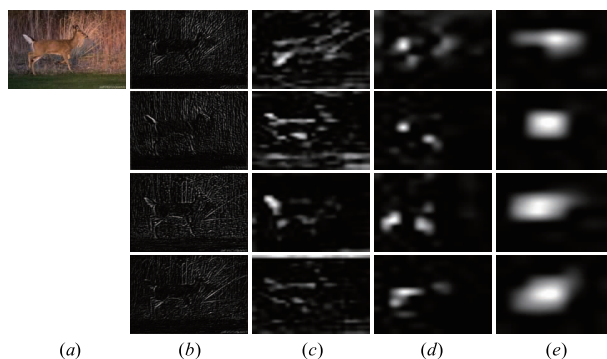


图2 (a)原始图像;(b)VGG第10层部分输出特征图;(c)VGG第20层部分输出特征图;(d)VGG第30层部分输出特征图;(e)VGG第38层部分输出特征图

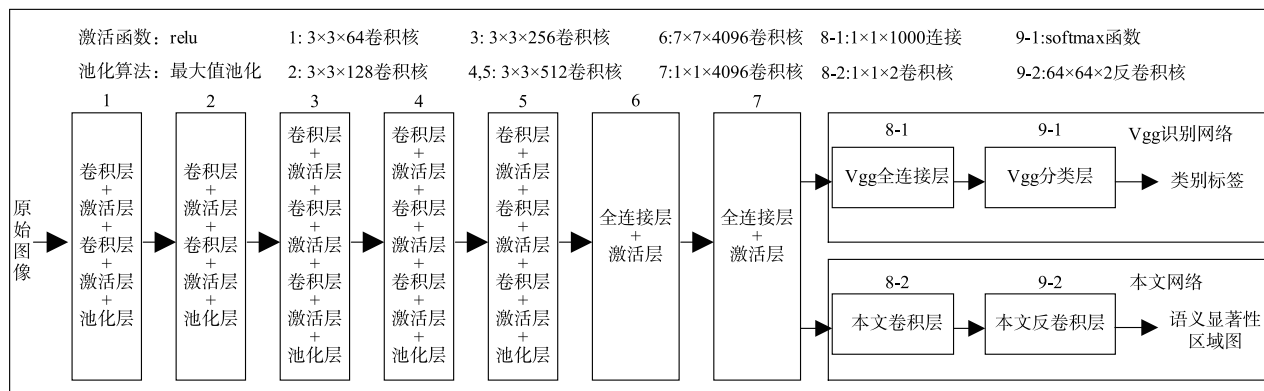


图3 VGG网络和语义显著性区域检测网络结构图

图3将本文网络结构简化表示为9个阶段,本文对VGG网络的改造在最后两层:(1)第8阶段,将VGG网络的全连接层转换为卷积层,使网络可以处理不同大小的图像.VGG网络处理的图像包含1000种目标类别,其全连接为 $1 \times 1 \times 1000$.本文模型处理的对象包括显著性区域和背景区域两种,因此卷积核设置为 $1 \times 1 \times 2$.(2)在第9阶段在网络中添加反卷积层,用反卷积核对输出语义特征图上采样,获得与原始图像相同大小的语义特征图,进而得到语义显著性区域图,其中反卷积核的大小设置为 $64 \times 64 \times 2$,上采样因子设置为32.

3.2 卷积与反卷积运算

卷积与反卷积运算是本文网络的主要算法,以下对这2种运算作简要说明.

(1)卷积运算用卷积核对原始图像或中间层特征图做空间卷积运算提取特征.当输入特征图为 $X = [x_{ij}]_{H \times W}$,卷积核为 $F = [f_{i'j'}]_{H' \times W'}$,输出特征图为 $Y = [y_{i'j'}]_{H' \times W'}$,卷积运算可表示为 $Y = X \otimes F$,输出特征图中的元素 $y_{i'j'}$ 可表示为:

$$y_{i'j'} = \sum_{s=1}^H \sum_{j=1}^{W'} f_{i'j'} x_{s(i'-1)+i'-1, s(j'-1)+j'-1} \quad (1)$$

公式(1)中 s 为卷积操作的降采样因子,即卷积操作的滑动间隔.卷积运算如图4(a)所示,图中为 4×4 的特征图与 3×3 的卷积核进行卷积运算得到 2×2 的特征图,输出特征图中的每一个数值由卷积核与输入特征图中对应位置的数值进行一次卷积运算得出,每一次卷积运算后卷积核以间隔 s 进行滑动计算下一个输出特征值.

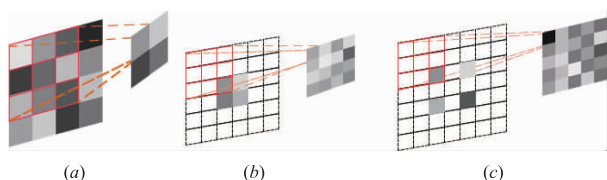


图4 卷积运算与反卷积运算示意图

(2)反卷积运算:当输入特征图为 $X = [x_{ij}]_{H \times W}$,卷积核为 $F = [f_{i'j'}]_{H' \times W'}$,输出特征图为 $Y = [y_{i'j'}]_{H' \times W'}$,输出特征图中的元素可表示为:

$$y_{i'j'} = \sum_{i=0}^{q(H',s)} \sum_{j=0}^{q(W',s)} f_{1+s \times i' + m(i'+P_s,s), 1+s \times j' + m(j'+P_s,s)} \cdot x_{1-i'+q(i'+P_s,s), 1-j'+q(j'+P_s,s)} \quad (2)$$

$$m(k, S) = (k - 1) \bmod S, \quad q(k, n) = (k - 1) / S \quad (3)$$

公式(2)中 s 为上采样因子,算子 $m(*)$ 和 $q(*)$ 如公式(3)所示.反卷积操作如图4(b)所示,输入为 2×2 矩阵,反卷积核大小为 3×3 ,输出为 4×4 矩阵.在反卷积运算中,输出特征图的分辨率大于输入特征图,因此可以看成是上采样操作.反卷积中上采样因子的作用如图4(c)所示,即输入特征图中像素间设置的间隔.

3.3 网络训练

本文的网络将原始图像和对应的显著性区域标准图作为训练数据,在给定训练集 $\chi = \{X_i\}_{i=1}^N$ 和标注 $L = \{l_i\}_{i=1}^N$ 后,训练网络的过程就是通过最小化损失函数(4)获得网络的所有参数的过程.

$$L(\chi; \theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=0}^1 \sum_{h=1}^H \sum_{w=1}^W 1\{l_i = j\} \log P_{whj}(X_i | \theta) \quad (4)$$

损失函数(4)中, W, H 表示第 i 个输入图像的宽和高, $1\{o\}$ 为指示函数, $P_{whj}\{X_i | \theta\}$ 表示输入图像 X_i 中像素 (w, h) 在参数 θ 条件下输出标注 j 的概率大小, θ 表示网络的所有参数.

网络通过随机梯度下降法训练,学习率设置为0.001,批处理数量设置为20,权重衰减系数设置为0.005,训练迭代次数设置为50.

网络的训练建立了图像特征与语义显著性目标以及背景之间的映射关系,用训练好的网络对测试图像进行前向传播运算得到的是具有 $\{0, 1\}$ 标注的语义显著性区域图.本文将像素级语义信息定义为其是否属于显著性目标区域的标注,像素 i 的语义信息值定义如下:

$$p_{ses}(i) = \begin{cases} 1, & \text{如果 } i \in \text{显著性目标区域} \\ 0, & \text{如果 } i \in \text{背景区域} \end{cases} \quad (5)$$

4 语义显著性区域优化算法

语义显著性区域在整体上标注了显著性目标区域,但存在边界不准确的缺陷,需要进一步进行优化.本文基于颜色信息建立前景和背景的外观模型计算像素级的前景和背景概率,再进一步通过超像素分割和图割算法引入目标的边界信息和空间一致性信息,获取更准确的超像素级前景概率和背景概率图.最后用一个新的优化模型融合前景和背景概率信息、语义信息、空间一致性信息获得更准确的显著图.

4.1 基于颜色外观模型的像素级前景概率与背景概率估计

语义显著性图包含了较准确的显著性区域和背景区域的颜色信息,因此本文根据语义显著性区域图利用混合高斯模型^[23]建立前景区域和背景区域的颜色外观模型.本文将语义显著性图中标注为1和0的区域分别视为前景和背景,用其对应的RGB颜色值分别训练前景和背景外观模型.参照以往高斯建模^[23]的经验,高斯分量的个数设置为8,其参数用k-means算法初始化,用EM(Expectation Maximization)算法训练确定外观模型中各个高斯分量的参数.前景和背景的外观模型训练完成后,用公式(6)计算图像中每个像素的前景概率和背景概率.

$$P_{f/b}(x) = \sum_{n=1}^8 \omega_n^{f/b} * g_{f/b}(x; \mu_n, \Sigma_n), \omega_n^{f/b} \in (0, 1) \quad (6)$$

$$\begin{aligned} & \text{且 } \sum_{n=1}^8 \omega_n^{f/b} = 1 \\ & g_{f/b}(x; \mu, \Sigma) = \\ & \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \end{aligned} \quad (7)$$

公式(6)中, $P_{f/b}(o)$ 表示像素的前景概率或背景概率, ω 表示高斯分量的权重, $g_{f/b}(x; \mu_n, \Sigma_n)$ 为前景或背景外观模型中第n个高斯分量的概率密度函数,具体如公式(7)所示,其中 μ_n 和 Σ_n 分别为前景或背景外观模型中

的均值和协方差矩阵, d 表示数据 x 的维度.

如图5所示,像素级的前景概率图和背景概率图更完整、准确地表示了像素属于显著性区域和背景区域的概率大小,但当背景中有一些区域的颜色与前景区域的颜色相似时,尤其当显著性区域与背景区域对比度非常低时,仍然会将许多背景区域错判为前景区域.针对这一问题,本文进一步引入超像素分割和图割算法,获取更准确的超像素级前景和背景概率.

4.2 超像素级的前景概率与背景概率

超像素分割将图像中具有相似颜色、纹理等特征的相邻像素聚合成图像块,即超像素.超像素分割较好的保留了图像中的目标边界,能有效的提高后续处理的效率和准确率.本文使用线性迭代聚类算法^[24]进行超像素分割.根据文献[11]的经验,设置初始超像素数目为250.超像素级的前景、背景概率计算如公式(8):

$$P_{f/b}(sp_i) = \frac{1}{N} \sum_{n=1}^N P_{f/b}(x_n) \quad (8)$$

其中 N 为超像素包含像素的个数, $P_{f/b}(sp_i)$ 表示超像素 sp_i 的前景概率或背景概率.

4.3 颜色外观模型优化

如图5(e)、(h)所示,超像素级的前景概率图和背景概率图在一定程度上抑制了背景区域的显著度,但对于低对比度的图像,抑制的效果仍然不明显.本文进一步用图割算法^[25]引入空间一致性先验进行优化,空间一致性指示相邻像素或超像素的显著性的平滑程度和它们特征的相似性正相关.以前景概率值的优化过程为例,根据前景、背景概率和图像的空间一致性构造能量函数,将图像分割问题转换为一个能量函数最小化问题.能量函数如公式(9)所示,其中 G 表示由图像中所有像素构成的无向图模型, $L = \{l_i\}$ 表示能量模型为所有像素分配的标注集合,其中 $l_i \in \{0, 1\}$ 表示像素 p_i 获得的标注值. $s_i(l_i)$ 表示像素 p_i 属于前景或背景的概率,具体如公式(10)所示,其中 $P_f(p_i)$ 表示像素 p_i 的前景概率. $W_{ij}(p_i, p_j)$ 为相邻像素的相似性,定义如公式(11)所示,其中 $d_{lab} \{ * \}$ 表示两个像素在 Lab 颜色空间的距离, σ_c 根据文献[23]设为10. λ 为权重参数,根据

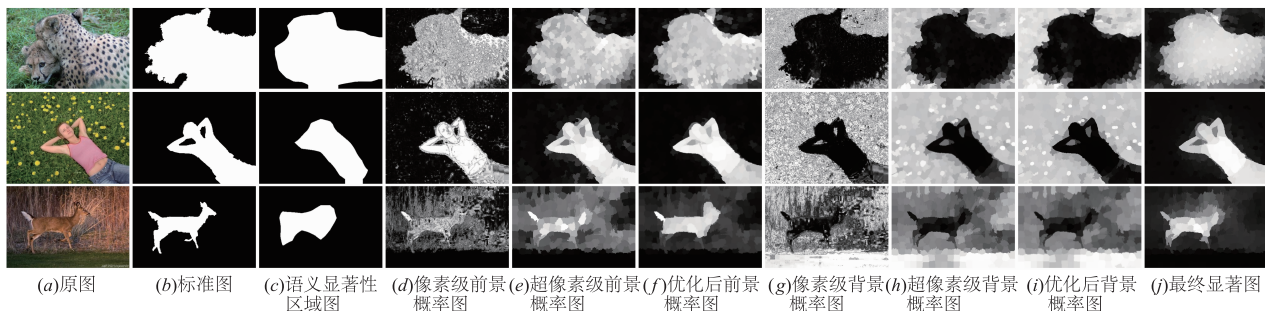


图5 语义显著性区域优化过程

文献[16]的经验设为0.3. 能量函数最小化过程鼓励有较大前景概率的像素获得标注1, 较小前景概率的像素获得标注0, 鼓励颜色相似的相邻像素获得相同的标注. 最后用最大流算法^[26]获得能量函数的解, 得到每个像素的标注值 $label(p_i)$. 优化后的超像素级前景概率的计算如公式(12)所示, 其中 $P_f(p_i)$ 分别为像素 p_i 初始的前景概率值. $label(p_i)$ 为像素 p_i 的由图割算法获得的标注值. $N\{\ast\}$ 为归一化算子, $P_{of}(p_i)$ 为优化后的前景概率值. 超像素级背景概率值的优化过程与超像素级前景概率值的优化过程类似, 不同之处在于能量函数中相应的 $s_i(l_i)$ 项用背景概率值和背景概率相反值表示, 最终获得优化后的背景概率值表示为 $P_{ob}(p_i)$.

$$E(L) = \sum_{i \in G} s_i(l_i) + \lambda \sum_{(i,j) \in G} W_{ij}(l_i, l_j) \quad (9)$$

$$s_i(l_i) = \begin{cases} P_f(p_i), & \text{若 } l_i = 0 \\ 1 - P_f(p_i), & \text{若 } l_i = 1 \end{cases} \quad (10)$$

$$W_{ij}(l_i, l_j) = \begin{cases} \exp\left(-\frac{d_{lab}(p_i, p_j)}{2\sigma_c^2}\right), & \text{如果 } l_i \neq l_j; \\ 0, & \text{如果 } l_i = l_j \end{cases} \quad (11)$$

$$P_{of}(p_i) = N\{P_f(p_i) + label(p_i)\} \quad (12)$$

如图5所示, 优化后的前景概率图更好的抑制了背景区域, 突出了前景区域且前景区域更加平滑. 优化后的背景概率图更好的突出了背景区域.

4.4 显著性信息融合模型

本文的语义信息、前景概率和背景概率信息从不同角度描述了图像的显著性. 语义显著性区域图虽然包含的显著性目标区域不完整、边界不准确, 但包含错检的显著性区域较少. 前景概率图和背景概率图有更完整的显著性区域, 但是其包含错检的显著性区域较多. 为获得更准确的显著性区域图, 本文提出一个新的无约束优化模型, 有效融合之前获得的语义显著性信息、前景和背景概率信息. 目标函数如公式(13)所示, 其中 N 表示图像中包含的超像素个数, s^* 为优化模型为每个超像素分配的显著性值.

$$s^* = \operatorname{argmin}_s \frac{1}{2} \left(\underbrace{\sum_{i=1}^N \omega_i^{bg} s_i^2}_{\text{背景信息}} + \underbrace{\sum_{i=1}^N \omega_i^{fg} (s_i - 1)^2}_{\text{前景信息}} + \underbrace{\sum_{i=1}^N \omega_{ij} (s_i - s_j)^2}_{\text{空间一致信息}} + \underbrace{\sum_{i=1}^N T_i (s_i - l_i)^2}_{\text{语义信息}} \right) \quad (13)$$

$$\omega_{ij} = \exp\left(-\frac{d_{lab}(p_i, p_j)}{2\sigma_c^2}\right) \quad (14)$$

$$l_i = \begin{cases} 1, & ses_i \geq \sigma_f \\ 0, & ses_i \leq \sigma_b \end{cases}, ses_i = \frac{1}{M} \sum_{m=1}^M pses_m \quad (15)$$

$$T_i = \begin{cases} 1, & ses_i \geq \sigma_f \text{ 或 } ses_i \leq \sigma_b \\ 0, & \text{其它} \end{cases} \quad (16)$$

目标函数包括四个约束项: 背景约束项促使有较大背景概率的超像素获得较小的显著性值, ω_i^{bg} 为超像素 i 的背景概率值. 前景约束项促使有较大前景概率的超像素获得较大的显著性值, ω_i^{fg} 为超像素 i 的前景概率值. 空间一致约束项促使颜色相似的相邻超像素获得相近的显著度值, i, j 为在空间上相邻的两个超像素, ω_{ij} 表示超像素 i, j 的颜色相似性, 具体如公式(14)所示. 语义信息约束项根据语义显著性值选取一些超像素作为前景和背景超像素, 促使这些超像素获得的显著性值与其语义显著性标注尽量相同. 公式中 l_i 为超像素 i 的语义显著性标注值. l_i 的计算方法如公式(15)所示, 其中 ses_i 表示超像素 i 的语义显著性值. σ_f 和 σ_b 分别表示选择前景和背景超像素的阈值. T_i 表示选择超像素的指示值, 其计算方法如公式(16)所示.

本文的优化模型可用公式(17)求解:

$$s^* = (D - W + B + F + T)^{-1} (Fv + Tl) \quad (17)$$

其中 $W = [\omega_{ij}]_{N \times N}$ 为颜色相似度矩阵. $D = \operatorname{diag}\{d_{11}, d_{22}, \dots, d_{NN}\}_{N \times N}$ 是一个对角矩阵, N 表示超像素的个数, d_{ii} 表示颜色相似度矩阵 W 第 i 行的累加和. $B = \operatorname{diag}\{b_{11}, b_{22}, \dots, b_{NN}\}_{N \times N}$ 中 b_{ii} 表示超像素 i 的背景概率值. $F = \operatorname{diag}\{f_{11}, f_{22}, \dots, f_{NN}\}_{N \times N}$ 中 f_{ii} 表示超像素 i 的前景概率值. $T = \operatorname{diag}\{T_{11}, T_{22}, \dots, T_{NN}\}_{N \times N}$ 中 T_{ii} 表示超像素 i 的选择标注值, 其定义如公式(16)所示. $v = [1, 1, \dots, 1]_N$ 为长度为 N 的向量. $l = [l_1, l_2, \dots, l_N]_N$ 表示长度为 N 的向量, l_i 表示超像素 i 的语义显著性标注, 定义如公式(15)所示.

5 实验与分析

为分析和评价本文算法的性能, 本文将算法在6个常用的显著性数据库上与15种最新的显著性区域检测算法进行定性和定量比较.

5.1 评价数据集

本文使用 MSRA-B^[27]、ICoSeg^[28]、HKU-IS^[29]、SOD^[30]、ECCSD^[31]、ROC^[32] 6个显著性检测数据库评价模型, 六个数据库分别包含5000、643、4447、300、999、235幅图片和对应的人工标注图. 其中MSRA-B数据集相对较简单, 其他数据集都包含了检测难度较大的图片. 实验时分别从HKU-IS、MSRA-B、ECCSD数据集中随机抽取2500、2500、400幅图像, 将其作为语义显著性区域检测网络的训练数据, 分别从HKU-IS、MSRA-B、ECCSD数据集中随机抽取500、500、100幅图像, 将其作为语义显著性区域检测网络的标定数据. 6个数据集中剩余的图像作为测试数据.

5.2 评价指标

本文采用显著性检测中常用的F值^[33] (F-measure)、准确率-召回率曲线^[33] (Precision-Recall, P-R)、平

均绝对误差值^[33]值 (Mean Absolute Error, MAE) 三种指标评价模型的性能, 三种指标具体的定义可参见文献 [33].

5.3 本文算法分析

本部分主要分析算法中两个参数 σ_f 和 σ_b 对性能的影响, 检验算法是否对超参数敏感, 并为选择合适的参数提供依据.

表 1 参数 σ_f 和 σ_b 对算法 F 指标和 MAE 值的影响

	$\sigma_f=0.9,$ $\sigma_b=0.1$	$\sigma_f=0.8,$ $\sigma_b=0.2$	$\sigma_f=0.7,$ $\sigma_b=0.3$	$\sigma_f=0.6,$ $\sigma_b=0.4$	$\sigma_f=0.5,$ $\sigma_b=0.5$
F 值	0.8532	0.8536	0.8522	0.8519	0.8504
MAE	0.1198	0.1196	0.1199	0.1203	0.1208

实验中设置了五组参数在 5.1 所述测试集上进行定量评价. F 值和 MAE 的测试结果如表 1 所示, P-R 曲线的测试结果如图 6 所示. 如表 1 和图 6 所示, 模型在不同的阈值参数 σ_f 和 σ_b 下都有较好的性能, 且差别不大, 这说明模型对超参数不敏感, 具有良好的稳定性. 根据表 1 和图 6, $\sigma_f=0.8$ 且 $\sigma_b=0.2$ 时, 模型有最好的效果, 在后续的实验中, 设置参数 $\sigma_f=0.8$ 和 $\sigma_b=0.2$.

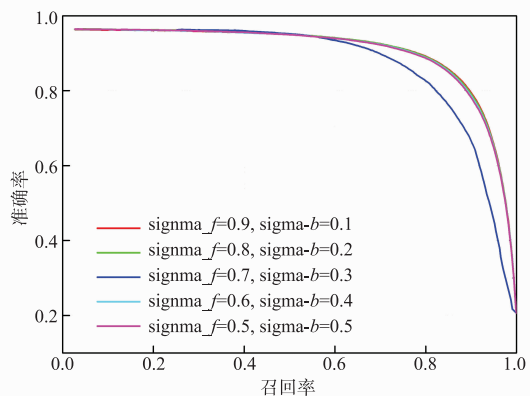


图6 参数 σ_f 和 σ_b 对P-R曲线的影响

5.4 本文算法与最新算法的比较

为进一步验证本文算法的有效性, 将本文算法与 15 种最新的显著性区域检测算法进行比较. 15 种算法分别是 GCS^[6]、FCS^[9]、RBD^[11]、LRM^[12]、DRF^[13]、HDCT^[14]、BLS^[15]、LPS^[16]、GLCS^[17]、GPD^[34]、CHS^[35]、RRWR^[36]、GBMR^[37]、MSS^[38]、MAPS^[39]. 算法的比较实验分为显著图的比较和显著性指标的定量比较.

图 7 为本文算法和其他算法在几幅有代表性图片上的显著图比较. 第 1 组图像包含大尺度显著性目标, 其不满足稀疏性、边界先验, 但在本文的算法中显著性区域是一个明显的语义显著性区域, 因此本文算法产生了最好的检测效果. 第 2、4 组的图像包含了典型的颜色非均质显著性目标, 其他算法只能检测到部分显著性区域, 如

LRM 算法只检测到了人脸. 由于语义信息的有效利用, 本文算法更完整地检测出了显著性目标区域. 第 3 组图像包含了低对比度的显著性目标, 其他算法或者难以检测出目标, 如 DRF、GCS、HDCT, 或者检测结果包含了大量的背景区域, 如 CHS、LRM、RRWR、GBMR、MSS、MAPS. 本文的算法更准确的检测出了显著性目标区域. 第 5 组图像中的显著性目标区域受到不同光照的影响, 难以完整检测. 其他算法只检出了部分显著性目标区域, 本文的算法检测出了更完整的显著性区域. 第六组图像包含了较复杂的背景区域, 本文算法更好的排除了复杂背景的干扰, 检测出了更均匀的显著性目标区域. 从以上几幅显著图的比较可以看出, 与其他基于颜色对比度、边界先验、紧凑度、中心先验、稀疏先验的算法相比, 本文算法对非均质、低对比度、大尺度、复杂背景等较难检测的图像有明显的优势, 能更完整、准确地检测出显著性目标区域, 抑制复杂背景的干扰.

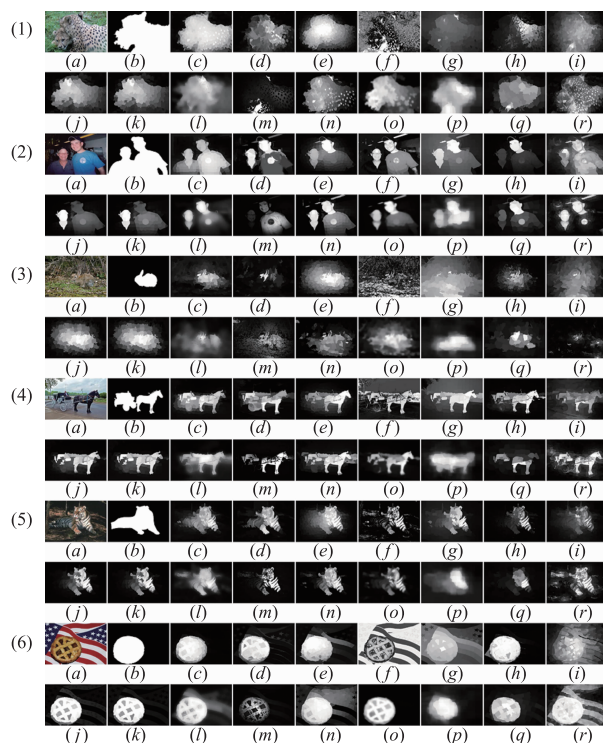


图7 本文算法和15种算法的显著图比较 (a)原图; (b)标准图; (c)本文显著图; (d)DRF; (e)GPD; (f)GCS; (g)CHS; (h)LPS; (i)LRM; (j)RRWR; (k)GBMR; (l)MSS; (m)FCS; (n)RBD; (o)MAPS; (p)BLS; (q)GLCS; (r)HDCT

图 8、9、10 分别为本文算法与其他显著性区域检测算法在 6 个测试数据集上的 F 值、MAE 值、P-R 曲线定量比较. 与其他算法相比, 本文算法在 6 个数据集上的 F 值、MAE 值、P-R 曲线都有所提升, 尤其在比较难的数据集 ICoSeg、HKU-IS、SOD、ECCSD、ROC 上, 本文算法明显优于其他算法.

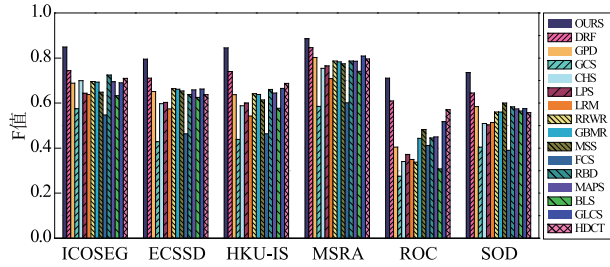


图8 本文算法与15种算法的F值比较

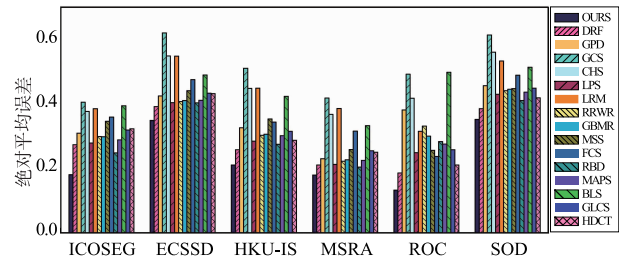


图9 本文算法与15种算法的MAE值比较

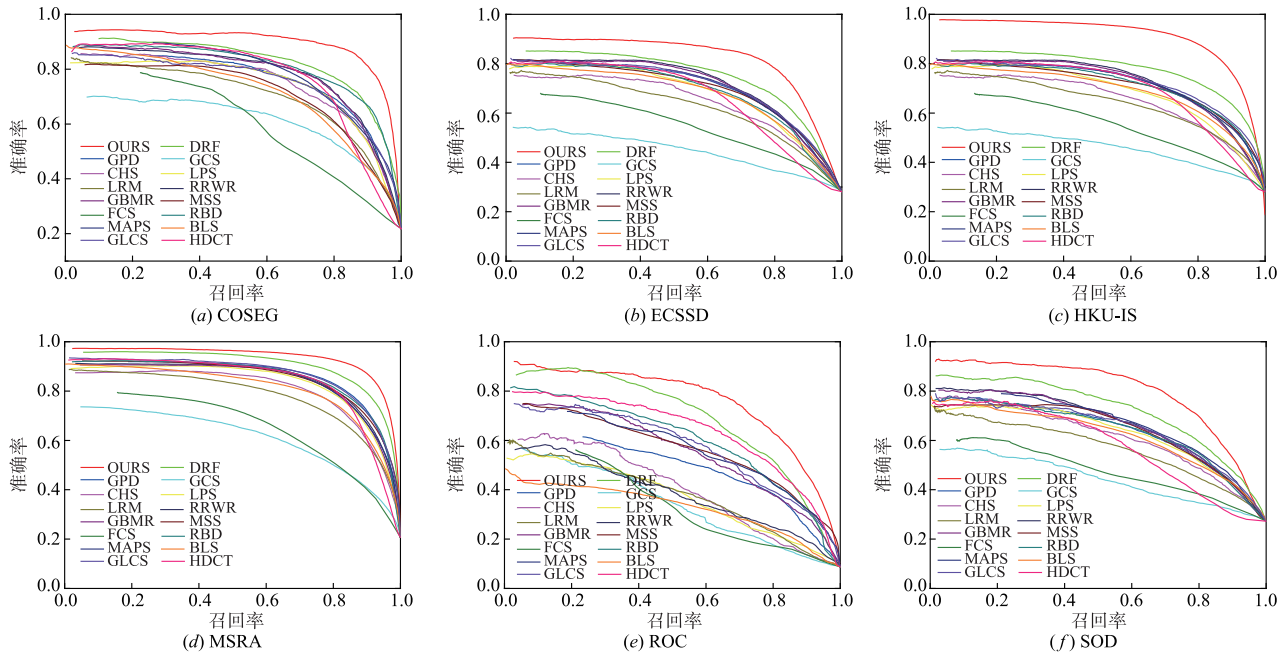


图10 本文模型与15种算法的P-R曲线比较

6 结论

本文针对传统显著性检测算法中语义信息难以获取和利用的难题,构建了基于全卷积结构的语义显著性区域检测网络,经过训练后,能有效的提取图像中的语义显著性区域.针对语义显著性区域边界不准确、区域不完整的缺陷,提出一个优化模型有效融合基于颜色的前景和背景概率信息、语义信息、空间一致信息得到更准确的显著性区域图.通过定性与定量实验比较,验证了本文提出的全卷积结构的语义显著性区域检测网络能有效的提取图像中的语义显著性区域,证明了本文算法的有效性和鲁棒性.

在以后的工作中,将进一步探索并优化网络结构,直接得到更准确的语义显著性区域图,提高显著性区域检测的效率.

参考文献

[1] Koch K, McLean J, Segev R, Freed M A, Berry M J, Balasubramanian V, Sterling P. How much the eye tells the

brain[J]. *Current Biology*, 2006, 25(16):1428 - 1434.

[2] 姜维, 卢朝阳, 李静, 刘晓佩, 姚超. 基于视觉显著性与文字置信图的场景文字的背景抑制方法[J]. *电子学报*, 2015, 42(9):62 - 68.

JIANG Wei, LU Zhao-yang, LI Jing, LIU Xiao-pei, YAO Chao. Visual saliency and text confidence map based background suppression for scene text[J]. *Acta Electronica Sinica*, 2015, 42(9):62 - 68. (in Chinese)

[3] 赵宏伟, 李清亮, 刘萍萍. 基于分级显著信息的空间编码方法[J]. *电子学报*, 2015, 43(1):1863 - 1867.

ZHAO Hong-wei, LI Qing-liang, LIU Ping-ping. Spatial encoding based on hierarchical salient information[J]. *Acta Electronica Sinica*, 2015, 43(1):1863 - 1867. (in Chinese)

[4] Borji A, Itti L. State-of-the-art in visual attention modeling [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2013, 35(1):185 - 207.

[5] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1998, 20(11):1254 - 1259.

- [6] Cheng M M, Mitra N J, Huang X L, Torr P H S, Hu S M. Global contrast based salient region detection [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 569 – 582.
- [7] Bao L, Lu J J, Li Y, Shi Y W. A saliency detection model using shearlet transform [J]. *Multimedia Tools and Applications*, 2015, 74(11): 4045 – 4058.
- [8] Bao L, Zhang X W, Zheng Y F, Y Li. Video saliency detection using 3D shearlet transform [J]. *Multimedia Tools and Applications*, 2015, 75(13): 7716 – 7778.
- [9] Perazzi F, Krahenbuhl P, Pritch Y, Hornung A. Saliency filters; contrast based filtering for salient region detection [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Providence: IEEE Press, 2012. 733 – 740.
- [10] 钱晓亮, 郭雷, 韩军伟, 等. 一种基于加权稀疏编码的频域视觉显著性检测算法 [J]. *电子学报*, 2013, 41(6): 1159 – 1165.
QIAN Xiao-liang, GUO Lei, HAN Jun-wei, CHENG Gong, YAO Xi-wen. A spectral algorithm based on weighted sparse Coding for visual saliency detection [J]. *Acta Electronica Sinica*, 2013, 41(6): 1159 – 1165. (in Chinese)
- [11] Zhu W J, Liang S, Wei Y C, J Sun. Saliency optimization from robust background detection [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Columbus: IEEE Press, 2014. 2814 – 2821.
- [12] Shen X H, Wu Y. A unified approach to salient object detection via low rank matrix recovery [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Providence: IEEE Press, 2012. 853 – 860.
- [13] Jiang H Zu, Wang J D, Yuan Z J, Wu Y, N N Zheng, S P Li. Salient object detection; a discriminative regional feature integration approach [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Portland: IEEE Press, 2013. 2083 – 2090.
- [14] Kim J H, Han D, Tai Y W, J Kim. Salient region detection via high dimensional color transform [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Columbus: IEEE Press, 2014. 883 – 890.
- [15] Tong N, Lu H C, Ruan X, M H Yang. Salient object detection via bootstrap learning [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Boston: IEEE Press, 2015. 1884 – 1892.
- [16] Li H Y, Lu H C, Lin Z, Shen X H, Price B. Inner and inter label propagation; salient object detection in the wild [J]. *IEEE Transactions on Image Processing*, 2015, 10(24): 3176 – 3186.
- [17] Tong N, Lu H C, Zhang Y, Ruan X. Salient object detection via global and local cues [J]. *Pattern Recognition*, 2015, 48: 3258 – 3267.
- [18] Einhäuser W G, Spain M, Perona P. Objects predict fixations better than early saliency [J]. *Journal of Vision*, 2008, 8(14): 1 – 26.
- [19] Henderson J M, Hollingworth A. High-level scene perception [J]. *Annu Rev Psychol*, 1999, 50: 243 – 71.
- [20] Athanasiadis T, Mylonas P, Avrithis Y, Kollias S. Semantic image segmentation and object labeling [J]. *IEEE Transaction on Circuits and Systems for Video Technology*, 2007, 17(3): 298 – 312.
- [21] Long J, Shelhamer E, T Darrell. Fully convolutional networks for semantic segmentation [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Boston: IEEE Press, 2015. 3431 – 3440.
- [22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. <http://arxiv.org/pdf/1409.1556>.
- [23] Cheng M M, Prisacariu V A, Zheng S, Torry P H S, Roth C. Densecut; densely connected CRFs for realtime grabcut [J]. *Computer Graphics Forum*, 2015, 34(7): 193 – 201.
- [24] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274 – 2281.
- [25] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2001, 23(11): 1222 – 1239.
- [26] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2004, 26(9): 1124 – 1137.
- [27] Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H Y. Learning to detect a salient object [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2011, 33(2): 353 – 367.
- [28] Batra D, Kowdle A, Parikh D, Luo J, Chen T. Icoseg; interactive co-segmentation with intelligent scribble guidance [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. San Francisco: IEEE Press, 2010. 3169 – 3176.
- [29] Li G B, Yu Y Z. Visual saliency based on multiscale deep features [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. Boston: IEEE Press, 2015. 5455 – 5463.
- [30] Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to eval-

- uating segmentation algorithms and measuring ecological statistics [A]. International Conference on Computer Vision [C]. Vancouver: Oxford University Press, 2001. 416 – 423.
- [31] Yan Q, Xu L, Shi J, Jia J. Hierarchical saliency detection [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland: IEEE Press, 2013. 1155 – 1162.
- [32] Li J, Martin D L, An X J, He H. Saliency detection based on frequency and spatial domain analysis [A]. British Machine Vision Conference [C]. Dundee: BMVA Press, 2011.
- [33] Radhakrishna A, Sheila, Francisco E, Sabine S. Frequency-tuned salient region detection [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Miami: IEEE Press, 2009. 1597 – 1604.
- [34] Jiang P, Vasconcelos N, Peng J L. Generic promotion of diffusion-based salient object detection [A]. International Conference on Computer Vision [C]. Santiago: Oxford University Press, 2015. 217 – 225.
- [35] Yang C, Zhang L H, Lu H C. Graph-regularized saliency detection with convex-hull-based center prior [J]. IEEE Signal Processing Letters. 2013, 7(20): 637 – 640.
- [36] Li C Y, Yuan Y C, Cai W D, Xia Y, Feng D D. Robust saliency detection via regularized random walks ranking [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston: IEEE Press, 2015. 2710 – 2717.
- [37] Yang C, Zhang L H, Lu H C, Ruan X, Yang M H. Saliency detection via graph-based manifold ranking [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland: IEEE Press, 2013. 3166 – 3173.
- [38] Tong N, Lu H C, Zhang L H, Ruan X. Saliency detection with multi-scale superpixels [J]. IEEE Signal Processing Letters. 2014, 9(21): 1035 – 1039.
- [39] Sun J, Lu H C, Liu X P. Saliency region detection based on markov absorption probabilities [J]. IEEE Transactions on Image Processing, 2015, 5(24): 1639 – 1649.

作者简介



郑云飞 男, 1983 年生, 解放军陆军工程大学博士研究生, 解放军炮兵防空兵学院讲师. 研究方向为图像与视频的显著性检测.
E-mail: yfzheng83@163.com



张雄伟 男, 解放军陆军工程大学教授, 研究方向为多媒体信息处理.



曹铁勇(通信作者) 男, 解放军陆军工程大学教授, 研究方向为多媒体信息处理.
E-mail: cty_ice@sina.com



孙蒙 男, 比利时鲁汶大学博士, 解放军陆军工程大学讲师, 研究方向为多媒体信息处理, 机器学习.