

基于多尺度反卷积深度学习的显著性检测

温 静 李雨萌

山西大学计算机与信息技术学院 太原 030006



摘 要 传统的显著性检测方法大多在单一的尺度上分辨感兴趣的目标和背景,无法有效地获取多分辨率下的局部细节信息,为此提出多尺度反卷积的深度学习网络模型。首先,在多尺度下对各层特征及各层对比特征进行反卷积,充分利用反卷积层中的卷积核对输入物体的形状进行重建,在多种分辨率特征图上利用反卷积网络来学习细节特征,减少信息的丢失,以此保持不同尺寸特征图的细节信息;然后,将各尺度下的反卷积特征进行融合,形成多层次局部信息;最后,与 VGG16 网络提取的全局信息融合后,计算各个像素的显著值,从而获得最终的显著性结果。实验结果表明,多尺度反卷积结构表现出较优的性能,与传统方法相比,可以相对增强突出物体与背景之间的对比,保持细节方面的特征;与最新深度学习的方法相比,可以检测出相对清晰准确的区域,一定程度上减少了信息的损失,还原出了更多的细节,能够有效地获取各种分辨率下的显著性目标,而且各反卷积层的独立性也显著提高了本文算法的运算速度。

关键词: 显著性检测;深度学习;多尺度特征;反卷积;多分辨率

中图法分类号 TP391.41

Salient Object Detection Based on Multi-scale Deconvolution Deep Learning

WEN Jing and LI Yu-meng

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Abstract Saliency detection aims to highlight the regional objects that people pay attention to subjectively in images. However, the traditional methods mainly distinguish the objects against the background under single resolution, so it's a hard to obtain the local detailed information under various scale. In this paper, we proposed a multi-scale convolution-combined-deconvolution network model. More specifically, we applied the deconvolution on the feature layers as well as their contract features, so that more multi-scale parameters could be maintained; then the fusion of the deconvolution offsets were combined with global information to get the salient result. The experimental results show that with many uncertainty factors in the complex background, compared with traditional methods, the proposed method could get a satisfactory salient detection. Compared with the latest deep learning methods, there can be relatively clear and accurate areas, which reduces the loss of information to some extent and restores more details, at the same time, the runtime of our method has been accelerated due to the design of the independence between the deconvolution layers.

Keywords Saliency detection, Deep learning, Multi-Scale features, Deconvolution, Multiresolution

1 引言

显著性检测的目的是模仿人类的视觉系统,寻求图像中一个视觉特征不同于图像其余部分的特定区域。

现有的显著性检测方法主要分为两类,即传统的模型方法和基于学习的方法。大多数传统的方法可以追溯到特征整合理论,即选择重要的视觉特征并将其组合来模拟视觉注意力。后来 Itti^[1]等建议通过颜色、强度和方向特征的中心、环绕对比度来测量显著性。显著性目标检测方法还受到频率学

原理的启发,例如频率空间方法^[2],即基于图像的振幅或相位变换来确定频率,产生的显著性图较好地保存了图像的高层结构,之后 Fu 等^[3]利用目标唯一性和中心偏移性来计算被划分的每个聚类的显著性, Hornung 等^[4]利用颜色的超像素分割来计算每个分割区域的颜色直方图, Wei 等^[5]利用图的流形排序算法进行显著性检测。同时,背景优先^[6-7]也被许多传统模型利用,其基本假设是图像边界区域更可能是背景,然后使用边界区域作为背景种子的标签传播来识别突出区域。

传统模型的显著性检测方法既高效又有效,但在处理复

到稿日期:2019-09-02 返修日期:2020-03-27 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金青年科学基金(61703252);山西省 1331 工程项目;山西省应用基础研究计划项目(201701D121053)

This paper was supported by the Young Scientists Fund of the National Natural Science Foundation of China (61703252), 1331 Engineering Project of Shanxi Province and Shanxi Province Applied Basic Research Programs(201701D121053).

通信作者:温静(wjing@sxu.edu.cn)

杂场景时却不够稳健。最近基于学习的方法受到了更多关注,这些方法可以通过在具有标注的图像数据上训练检测器来自动学习检测显著性,其中基于深度网络的显著性模型表现得非常有竞争力。例如,Wang 等^[8]建议分别通过训练 DNN-L 和 DNN-G 的网络,来进行局部估计和全局搜索检测显著区域,但这种采用对像素进行逐个扫描的方法来获得输入图像的显著图,计算代价很高。此外,Luo 等^[9]采用简化的网格状局部与全局特征相结合的形式,有效地优化了网络结构,Zhang 等^[10]将显著性先验加入到网络推理中,利用循环全卷积网络来预测显著性,Zhang 等^[11]提出双向信息传递模型来结合不同层的特征检测显著性,但是由于在卷积池化的过程中丢失了很多接受域中的空间信息,这些空间信息对于精准定位激活域在上一层激活域中的位置至关重要,在恢复特征尺寸的过程中采用简单的上采样方法无法有效地利用基于学习的特征来构建目标的整体空间信息和局部细节信息,而 Luo 的网络结构虽然利用了反卷积网络来训练学习恢复特征,但是层层衔接的形式使得无法充分地考虑每一个分辨率下的细节特征。之后 Liu 等^[12]提出了基于 U 型结构的神经网络,来研究池化对显著性检测的改进,其中包括两种基于池化技术的模块 GGM(全局引导模块)和 FAM(特征整合模块),且这两个模块都能应用在其他金字塔模型中,具有普遍

适应性,但是 FAM 在整合的过程中平均中和上采样带来了混叠效应,且采用先大幅度下采样池化再上采样的方式带来的损失太大。

为此,本文提出了一种多尺度化反卷积模型,充分利用反卷积层中的卷积核在卷积过程中对物体进行学习并重构其形状的作用,在多种分辨率特征图上利用反卷积网络来学习细节特征,这样各个不同尺寸下的特征图独立学习,可以逐步还原特征图中的信息,减少信息的丢失。然后,融合各尺度下的反卷积特征形成多层次局部信息,再与 VGG16 网络提取的全局信息通过一个 SCORE 融合后,由 Softmax 函数计算各个像素的显著值,最终获得显著性结果。

2 多尺度化反卷积网络结构 (Multi-Scale Deconv Net, MSDN)

为了尽可能多地保留各个分辨率下的细节信息和上下文关系,我们构建了一个深度卷积-反卷积网络模型,如图 1 所示。本文的 MSDN 网络包含两个部分,第一部分由 VGG-16^[13]的前 13 层构成 5 个卷积块(CONV-1—CONV-5)来提取全局信息 X_G ,第二部分是通过对各个尺度下的特征和对比度特征来进行反卷积,最后融合成多分辨率的局部信息 X_L 。

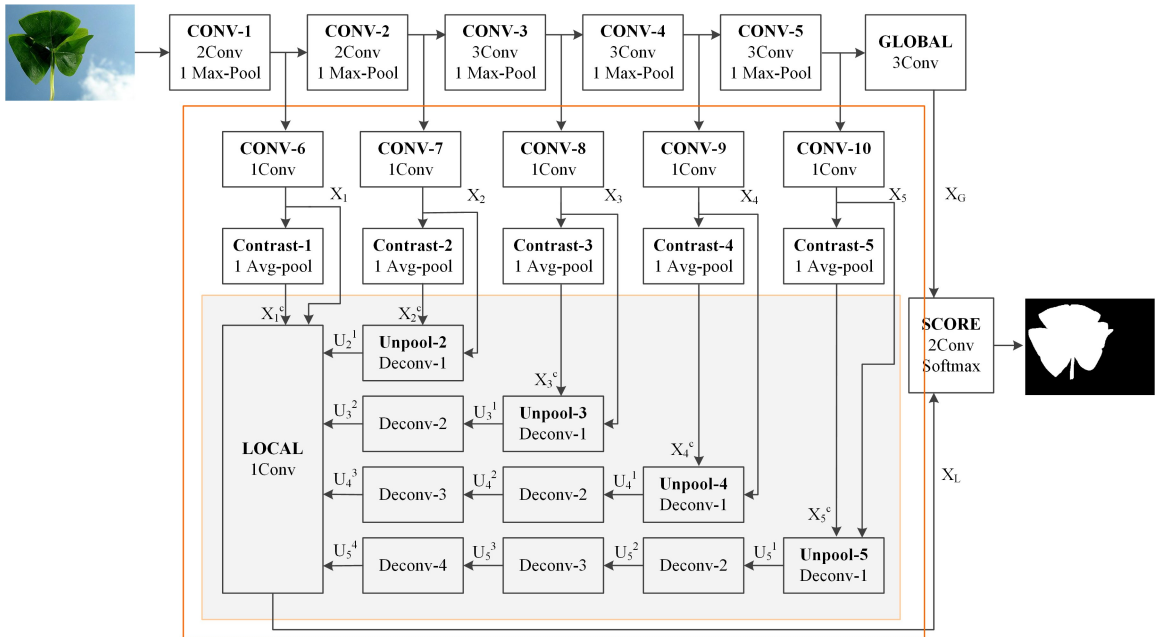


图 1 多尺度反卷积网络结构

Fig. 1 Multiscale deconvolution network

为了学习不同分辨率下的局部特征模型,MSDN 将卷积块 CONV-6—CONV-10 连接到 CONV-1—CONV-5 模块的池化层后,提取到多种尺度下的局部特征 $\{X_1, \dots, X_5\}$,如图 1 所示。

由于显著性对象是一种区别于背景区域的前景对象,因此使用对比度特征可以捕捉每个特征与其局部邻域之间的差异,找出比其邻域更亮或更暗的区域。为了捕获这种对比信息,在 MSDN 模型中添加一个与每个局部特征 X_i 相关联的

对比特征,每个对比度特征 X_i^c 的计算方法是从 X_i 中减去局部平均池化的值。

$$X_i^c = X_i - \text{AvgPool}(X_i) \quad (1)$$

这种对比度特征在本质上类似于计算像素 RGB 颜色与图像的全局平均颜色之间的差异,与 Liu 等^[14]从高斯图像金字塔中计算对比度特征的方法相似。然而,本文方法与之不同,其特性是学习而不是预定义。

为了将多尺度下的局部特征与局部对比度特征相融合,

MSDN 采用反卷积结构的形式。反卷积网络是由 Zeiler 等^[15]于 2010 年提出的,用于无监督地学习图片的中低层特征。由于在卷积池化的过程中丢失了很多接受域中的空间信息,这些空间信息对于精准定位激活值在上一层激活域中的位置至关重要,而一旦无法精确定位激活值在上一层激活域中的对应位置,我们就无法进行精确的定位从而不能突出对象。因此,可以利用反卷积层中的卷积核对输入物体的形状进行重建。由于低层次卷积可以获得物体整体的形状信息,高层次的卷积可以获得特定类别物体的形状信息,因此,一个多尺度化的反卷积结构可以用来获取不同细节水平的物体形状,对于特定类别物体的形状信息有很高的细节还原度。因此,通过这种方法,反卷积网络就可以把特定类别的形状信息

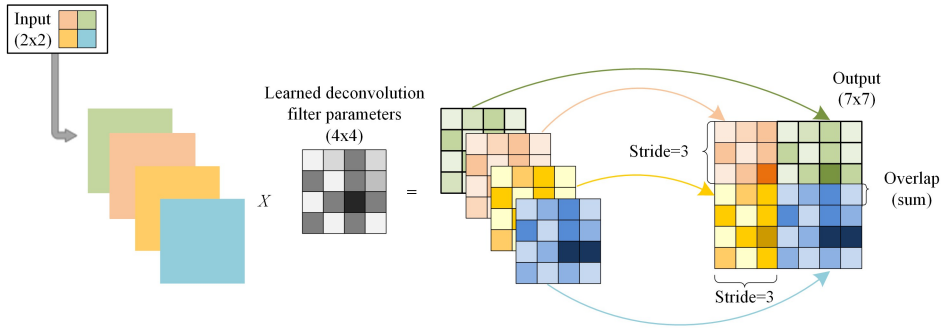


图 2 反卷积操作过程示意图

Fig. 2 Deconvolution operation schematic diagram

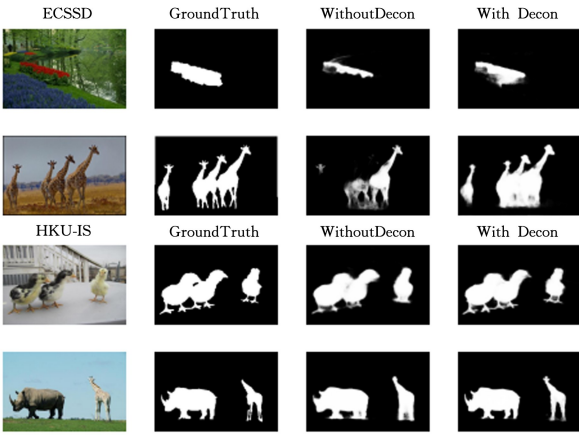


图 3 反卷积后的对比结果

Fig. 3 Comparison results of deconvolution operation

表 1 反卷积前后的评估数据

Table 1 Evaluation data of deconvolution operation

Dataset	Metric	Without Decon	With Decon
HKU-IS	max F_{β}	0.883	0.901
	MAE	0.072	0.048
ECSSD	max F_{β}	0.889	0.910
	MAE	0.063	0.062

由于我们最终的输出为固定大小 176×176 ,MSDN 分别结合不同分辨率下提取的局部特征 X_i 和局部对比度特征 X_i^c 进行反卷积式上采样。将每个不同尺度下的特征图逐步以步长为 2、卷积核为 5×5 的大小反卷积得到各自的特征图 U_i^1 ,其大小为 176×176 ,然后将各个特征图采用 concat 的融合方

纳入考量,从而对图片实现更加精准的显著性检测,而这种优势是那些仅有卷积层的网络所不具有的^[10-11]。

反卷积操作与卷积操作相反,是将特征图经过反向卷积后得到所需要的图片。假设输入图片为 x ,特征图为 y ,卷积操作可表示为 $y=Cx$,根据矩阵的运算性质可知,反卷积的过程则是 $x=C^T y$ 。即反卷积是采用卷积过程中转置后的卷积核对特征图进行卷积,可以将特征图放大。如图 2 所示,输入一个尺寸大小为 2×2 的图片,经过大小为 4×4 的卷积核进行反卷积学习的过程后,其中采用卷积的步长大小为 3,最终可以得到大小为 7×7 的示意图。图 3 为在两个不同数据集下使用了反卷积操作和没有使用反卷积操作的实验结果示意图,表 1 列出了其评估数据。

式,其中卷积核的大小为 1×1 ,最终结合形成局部特征 X_L 。

我们在 conv-5 模块之后添加了 3 个卷积层来计算全局特征 X_G ,前两个卷积层的内核大小为 5,最后一个卷积层的内核大小为 3,每个卷积层都有 128 个特征通道。最后分别用 1×1 的卷积核对局部特征 X_L 和全局特征 X_G 进行卷积得到局部分数和全局分数,将两部分分数相加得到总体分数,最后通过 *Softmax* 函数预测各个像素的显著性,从而得到最终显著性特征图。

$$U_i^1 = \text{UNPOOL}(X_i, X_i^c) \quad (2)$$

$$U_i^{i-1} = \text{UNPOOL}(U_i^{i-2} \dots U_i^1) \quad (3)$$

$$X_L = \text{CONV}(X_1, X_1^1, U_2^1, U_3^1, U_4^1, U_5^1) \quad (4)$$

3 计算显著性特征图

MSDN 模型使用监督的深度卷积网络,损失近似于 *MS* 函数^[16],由 GroundTruth 和 Estimated Saliency 之间的交叉熵数据保真项和边界损失项之和构成:

$$F^{MS} \approx \underbrace{\sum_j \lambda_j \int_{v \in \Omega_j} H_j(y(v), \hat{y}(v))}_{\text{crossEntropy}} + \underbrace{\sum_j \gamma_j (1 - \text{IoU}(c_j, \hat{c}_j))}_{\text{boundaryIoULoss}} \quad (5)$$

H_j 是图像区域 Ω_j 内部的所有像素 (v) 的 GroundTruth (y) 和 Estimated Saliency (\hat{y}) 之间的总交叉熵, $\text{IoU}(c_j, \hat{c}_j)$ 是在真正的边界上的像素 c_j 和估计边界上的像素 \hat{c}_j 之间的交并比。

基于 *Dice* 损失和 *IoU* 边界损失在医学图像分割中的重要应用,本文利用 *IoU* 边界损失项对式(5)的边界长度进行

惩罚。这种损失函数是分别检查每个像素,将类预测(深度方向的像素向量)和目标向量进行比较,来单独评估每个向量的类别预测,然后对所有像素求平均值,这样每个像素就能被平等学习,而显著性目标检测就是需要在全局中公平地学习预测所需的类别,最终寻找在目标物体。

为了计算边界损失,我们还使用 Sobel 操作符和激活函数 \tanh 来近似显著性映射梯度大小, \tanh 将显著性梯度大小控制在概率范围 $[0, 1]$ 。对于显著性特征图 \hat{c}_j 的梯度幅度和真正的显著性图 c 的每个区域 j 的梯度幅度,计算 Dice 和 IoU 的边界损失为:

$$IoU\ Loss = 1 - \frac{2|c_j \cap \hat{c}_j|}{|c_j| + |\hat{c}_j|} \quad (6)$$

显著性特征图由局部特征 X_L 和全局特征 X_G 经过两个线性操作 (W_L, b_L) 和 (W_G, b_G) 计算得出, softmax 函数用来计算每个像素是否是显著性的概率。

$$\hat{y}(v) = p(y(v) = c) = \frac{e^{W_L^c X_L(v) + b_L^c + W_G^c X_G + b_G^c}}{\sum_{c' \in \{0,1\}} e^{W_L^{c'} X_L(v) + b_L^{c'} + W_G^{c'} X_G + b_G^{c'}} \quad (7)$$

4 实验与结果分析

4.1 实验数据集与配置

为了验证上述算法的可行性,我们在 5 个数据集上进行测试,其中包括 3 个公共的标准数据集 MSRA-B, HKU-IS, ECSSD 和两个交通标志数据集 CCTSDB, GTSDDB。

MSRA-B 数据集包括 5 000 张图片,在视觉显著性检测方面应用最广泛,大部分图像都有一个突出的对象。HKU-IS 数据集包含 4 447 张图片,其中大部分具有低对比度以及多个突出对象。ECSSD 数据集由 1 000 张结构复杂的图片组成,它的 GroundTruth 是由 5 位受试者进行标记的。中国自然场景中交通标志数据集 CCTSDB^[17] 用来针对交通标志进行目标检测,此数据库是街拍的真实场景,我们从中选取了 3 500 张图片,让项目组的很多成员进行手工标注 Ground-Truth,该数据集图片样本多,样本样式丰富,具有不同的背景、不同的拍摄角度、不同的时间、不同逆光程度以及一张图片中出现多个交通标志等特征。为了公平比较,我们在公开的德国交通标志数据集 GTSDDB^[17] 上验证本文方法是否可以高效地应用于真实的场景。该数据集包含了 900 张涵盖各种距离、照度、天气条件以及部分遮挡和旋转变化的图片。本文的对比方法有 Cluster-Based Co-Saliency (CS)^[33], Contrast based filtering for salient region detection (SF)^[4], Geodesic Saliency Using Background Priors (GS)^[5], Graph-Based Manifold Ranking (MR)^[6], Robust Background Detection (wCtr)^[7], Non-local Deep Features (NF)^[9], Progressive attention guided recurrent network (PAG)^[10]。

本文的网络模型是在 TensorFlow^[18] 中实现的。CONV-1—CONV-5 块中的权值用 VGG-16^[18] 的预训练权值进行初始化,所有新添加的卷积和反卷积层的权重随机初始化 ($\delta =$

0.01), 偏差初始化为 0。我们用 Adam 优化器^[19] 来训练模型,初始学习率为 10^{-6} , $\beta_1 = 0.9$, $\beta_2 = 0.999$ 。式(5)中, λ_j 和 γ_j 都设置为 1。为了与其他方法进行公平比较,我们将公共数据集分为 3 部分: 2 500 张用于训练, 500 张用于验证, 剩下的 1 000 张用于测试。我们使用 NVIDIA 1080Ti GPU 的设备,在一个图像批处理大小完成 20 个 epoch 的情况下,整个训练过程需要约 6h,同时为保证实验的可信度,在相同的实验条件下训练测试基于学习的其他方法,然后统一根据我们的评估方法得出表 2 所列的评估结果。

表 2 3 个基准数据集的评估数据

Method	Metric	MSRA-B	HKU-IS	ECSSD
GS	max F_β	0.777	0.682	0.661
	MAE	0.144	0.167	0.206
MR	max F_β	0.824	0.715	0.736
	MAE	0.127	0.174	0.189
wCtr	max F_β	0.820	0.726	0.716
	MAE	0.110	0.141	0.171
NF	max F_β	0.910	0.888	0.889
	MAE	0.048	0.048	0.063
PAG	max F_β	0.912	0.897	0.904
	MAE	0.048	0.048	0.061
Our	max F_β	0.911	0.901	0.910
	MAE	0.047	0.048	0.062

4.2 实验评估

在性能上我们采用标准显著性评价指标 F 度量值 (F-measure) 和平均绝对误差 (Mean Absolute Error, MAE)。F-measure 的定义为式(8), 其中采用 $\beta = 0.3$ ^[20] 来提高精确度的权重。

$$F_\beta = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad (8)$$

MAE 是直接计算模型输出的显著性图 S 与 L (Ground-truth) 之间的平均绝对误差, W 和 H 是给定图像的宽度和高度。

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - L(x, y)| \quad (9)$$

表 2 列出了运用两个指标对算法进行评估的结果, F 值是指精确度, 越高越好, MAE 是指与 GroundTruth 的误差, 越小越好。我们可以从表 2 看出本文方法在下度量值以及 MAE 评估方面有了较大的提升。表 3 列出了在 NVIDIA 1080Ti GPU 设备下对应每种方法测试一张图片的时间, 可以看出本文方法具有较好的运算速度。

表 3 各种方法平均运行时间的比较

Method	GS	MR	wCtr	NF	PAG	Our
Time/s	0.214	0.256	0.249	0.08	0.04	0.03

4.3 实验结果对比

实验 1 本文算法与 6 种算法在 3 个数据集上进行对比, 结果如图 4 所示, 依次为 GroundTruth, CS^[33], GS^[5], MR^[6], wCtr^[7], NF^[9], PAG^[10], Our Method。

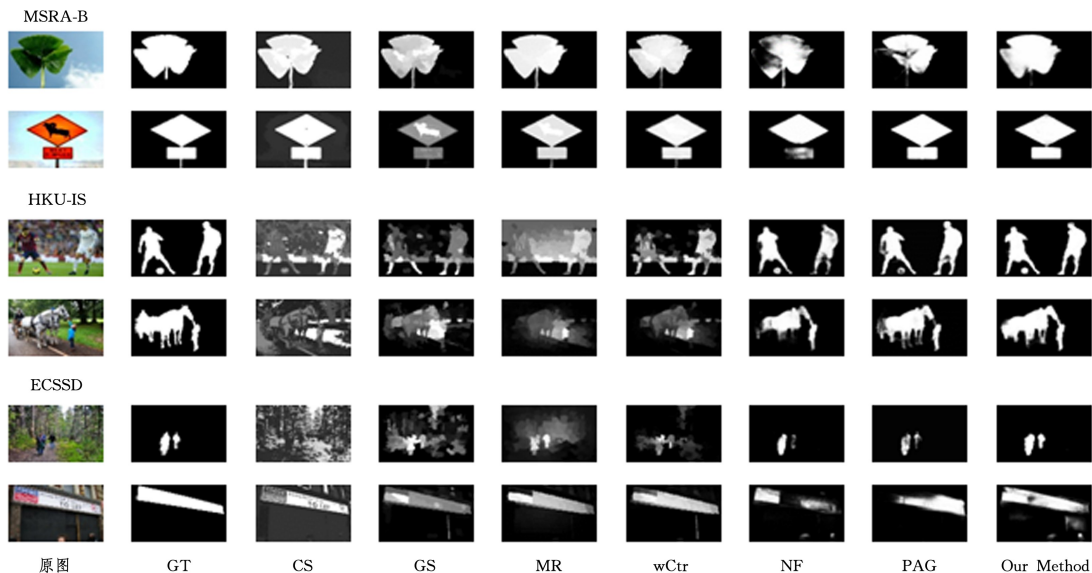


图4 3种基准数据集下7种方法的对比结果

Fig. 4 Comparison results of seven methods under three benchmark datasets

CS^[3]算法是利用聚类的方法对图像中的像素进行分组,然后利用唯一性测度和中心偏移性测度计算每个聚类的显著性。对于MSRA-B数据集的特点,图片相对单一,背景简单,大部分图片在中心位置有一个突出的显著对象,CS可以利用目标唯一性以及中心位置原理将显著目标提取出来。但当在HKU-IS和ECSSD数据集上进行测试时,由于图片的场景复杂,并且物体间对比度低,大多是针对森林、城市街景、道路、建筑物等室外场景,其中有多个显著性对象,此时唯一性的信息就会相应减弱,而物体与图像中心距离增加,显著性注意力增益也会降低,导致无法正确地辨别出突出物体。

GS^[5]算法是一种测地线距离度量的显著性检测方法,该方法利用的是背景的先验知识,但其提取的绝对背景容易把轻微接触边缘的显著区域错判为背景,当显著性区域不止一个时,显著性区域会与复杂背景交错相间,不能清晰地辨别出突出物体。例如图4所示的第四张图最左边的人接触到图片边界,就会有很大概率被当作背景来处理,且结果表明确实在处理时被误认为是背景。

MR^[6]算法是一种基于图的流形排序的显著性方法,该方法需要选取背景种子点后进行相似度流形排序,这样就使得背景假设过于理想化。当较大面积的真正显著性目标接触到图像边界时算法容易有缺陷,例如图4所示的第四张图片。

wCtr^[7]算法利用的是背景检测边界连续性方法,即当某一区域和图像边缘连接紧密时,称此区域属于背景信息。该方法的优点是当图像内容发生变化时,它的性能也应该是相对稳定的。但是在实际的图像中例如HKU-IS和ECSSD数据集,目标区域的边界并不是很清晰,这就导致显著区域的计算比较困难从而产生偏差,特别是有多个形状的突出对象时偏差效果更明显。

NF^[9]算法和PAG^[10]算法利用深度学习来提取特征。但是由于在卷积池化的过程中丢失了很多接受域中的空间信息,这些空间信息对于精准定位激活值在上一层激活域中的

位置至关重要。而本文算法利用多个尺度下的独立反卷积模型,可以通过学习来重建物体形状,相对减少了特征信息的丢失,具有更为清晰的显著性区域。

显著性检测的目的是模仿人类的视觉系统,而人类视觉系统的选择性注意可以分为两种机制^[21-24],一种是基于数据驱动自底向上(bottom-up)的注意机制,它是一种基于显著性驱动的与任务无关的方式,这种方式是相对快速的,另一种是基于任务驱动自上而下(top-down)的注意机制,它是一种受到我们意志力控制的,与任务相关的方式,这种方式是相对较慢的。

我们一般采用基于显著性驱动的与任务无关的方式进行检测,因为这种方式是快速的。但是在现实场景中,例如交通需求中^[25]自动驾驶等的应用需要检测特定的目标,如驾驶人员需要关注的交通标志,这种情况下就需要采用基于任务驱动的方式来检测所关注的交通标志,于是本文根据所需要的任务让本文网络进行训练学习,然后在两个数据集下分别进行测试,如实验2和实验3所示。

实验2 CCTSDB^[17]是自然场景下的数据集,该数据集展示了真实场景下的图片且样本样式丰富,我们利用此数据集来验证本文算法在处理复杂背景下图片时的相对高效性。大多数的图片是在人群密集以及车流量大的情况下自然拍摄的,数据特点多样且环境多变,挑战性非常大,在这种环境下找出我们所关注的交通标志相对不易。本文方法弥补了CS^[3]算法基于唯一性的缺点,如图5所示,当背景复杂时CS算法会认为很多唯一性的对象为显著区域,同时也避免了GS^[5],MR^[6],wCtr^[7]算法利用背景先验的知识,过分依赖于背景与显著区域的关系,这样就容易产生非背景边缘的复杂区域被误判为显著性区域的问题。本文方法消除了SF^[4]算法只对于颜色的敏感性,利用多尺度的反卷积来保持细节特征,可以尽量减少信息的丢失。如图5所示,实验结果表明在各种复杂多变的场景下,本文方法相对具有良好的性能。

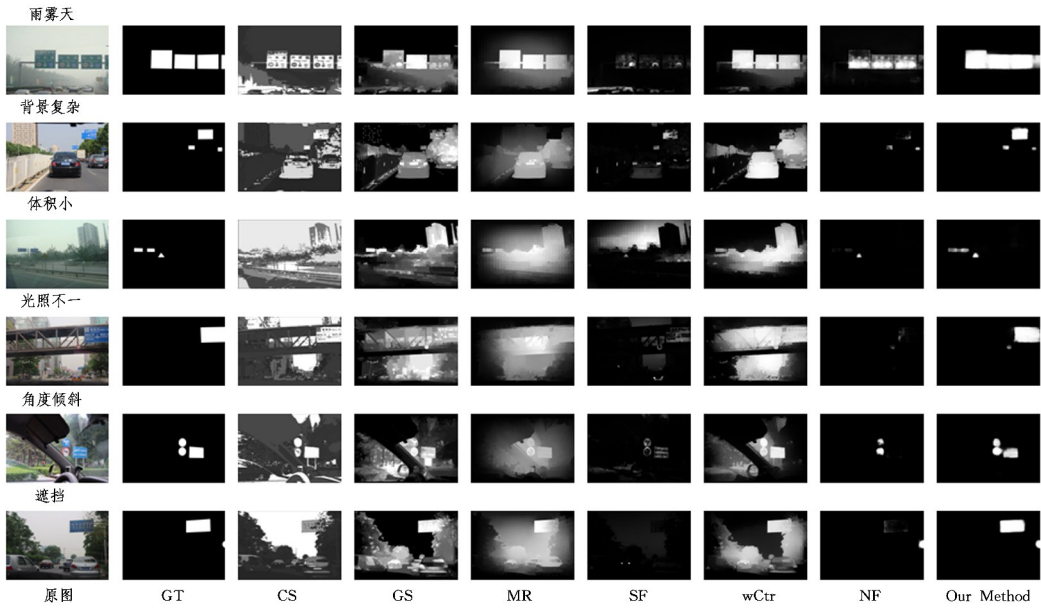


图5 不同自然场景下测试7种方法的结果

Fig. 5 Results of seven methods in different natural scenarios datasets

实验3 为了验证该方法适用于各种真实场景,我们利用德国交通标志检测基准 GTSD^[17]来进行测试。如图6所示,相比传统方法,本文方法通过多尺度反卷积网络的学习后,可以相对高效地处理复杂背景下的较为显著突出的物体。

总体上来看,本文方法可以表现出较优的性能,与传统方法相比,可以相对增强突出物体与背景之间的对比,保持细节

方面的特征,与最新的深度学习方法相比,本文方法可以有相对清晰准确的区域,在一定程度上尽可能地减少了信息的损失,还原出了更多的细节。但是网络还是会不同程度地丢失一些重要信息,因此检测一些被关注的小尺寸物体仍然有较大的困难,例如图5所示的第三张和第四张图片以及图6所示的第五张和第六张图片,因此我们还需要进一步地针对小型物体提高其检测性能。

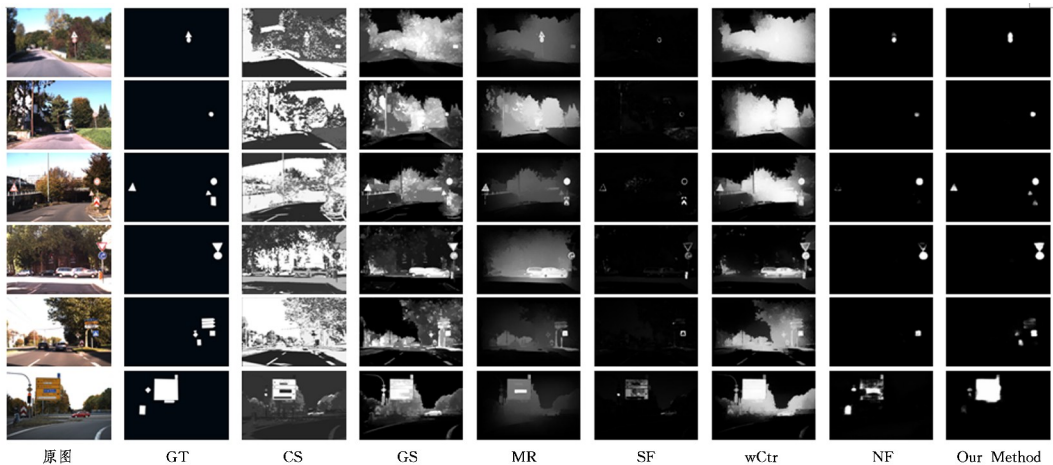


图6 德国交通标志测试与验证结果

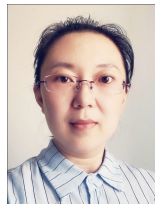
Fig. 6 Results of German traffic signs datasets

结束语 本文提出的多尺度化反卷积网络模型,在独立不同分辨率的反卷积层上进行逐步上采样,可以相对全面细节地提取所需特征,还原特征图中的重点信息,减少信息的丢失。我们在多种公共基准数据集下对所提方法进行验证,使用显著性检测统一评价标准进行评估,并在多种真实场景下验证得出本文方法比传统方法更能表现出较为清晰准确的显著区域。但是,在突出所关注的小型物体上仍然有较大的困难,因此本文模型还需要进一步的优化和提升来达到更高效的性能。

参考文献

- [1] ITTI L, KOCH C, NIEBUR E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20 (11): 1254-1259.
- [2] HOU X, ZHANG L. Saliency Detection: A Spectral Residual Approach [C] // IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2007: 1-8.

- [3] FU H, CAO X, TU Z. Cluster-Based Co-Saliency Detection[J]. IEEE Transactions Image Processing, 2013, 22(10): 3766-3778.
- [4] HORNUNG A, PRITCH Y, KRAHENBUHL P, et al. Saliency filters: Contrast based filtering for salient region detection[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2012: 733-740.
- [5] WEI Y, WEN F, ZHU W, et al. Geodesic Saliency Using Background Priors[C]// European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012: 29-42.
- [6] YANG C, ZHANG L, LU H, et al. Saliency Detection via Graph-Based Manifold Ranking[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2013: 3166-3173.
- [7] ZHU W, LIANG S, WEI Y, et al. Saliency Optimization from Robust Background Detection[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2014: 2814-2821.
- [8] WANG L, LU H, RUAN X, et al. Deep networks for saliency detection via local estimation and global search[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015: 3183-3192.
- [9] LUO Z, MISHRA A, ACHKAR A, et al. Non-local Deep Features for Salient Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017: 6593-6601.
- [10] ZHANG X, WANG T, QI J, LU H, et al. Progressive attention guided recurrent network for salient object detection[C]// IEEE Conf. Comput. Vis. Pattern Recognit, 2018: 714-722.
- [11] ZHANG L, DAI J, LU H, et al. A bi-directional message passing model for salient object detection[C]// IEEE Conf. Comput. Vis. Pattern Recognit, 2018: 1741-1750.
- [12] LIU J J, HOU Q, CHENG M M, et al. A Simple Pooling-Based Design for Real-Time Salient Object Detection[C]// Proceedings of International Conference on Computer Vision and Pattern Recognition, 2019.
- [13] SIMONYAN K, ZISSERMAN A, et al. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]// Proceedings of the International Conference on Learning Representations, 2015.
- [14] LIU F, GLEICHER M, et al. Region Enhanced Scale-Invariant Saliency Detection[C]// Proceedings of the 2006 IEEE International Conference on Multimedia and Expo (ICME 2006). Toronto, Ontario, Canada. IEEE, 2006: 1477-1480.
- [15] ZEILER M, KRISHNAN D, TAYLOR W, et al. Deconvolutional net-works[C]// Computer Vision & Pattern Recognition, 2010: 2528-2535.
- [16] MUMFORD D, SHAH J, et al. Optimal approximations by piecewise smooth functions and associated variational problems [J]. Communication Pure & Applied Mathematics, 1989, 42(5): 577-685.
- [17] ZHANG J, HUANG M, JIN X, et al. A Real-Time Chinese Traffic Sign Detection Algorithm Based on Modified YOLOv2 [J]. Algorithms, 2017, 10(4): 127.
- [18] Abadi M, Barham P, et al. TensorFlow: a system for large-scale machine learning[C]// Usenix Symposium on Operating Systems Design and Implementation, 2016: 265-283.
- [19] KINGMA P, BA J. Adam: A Method for Stochastic Optimization [C]// Proceedings of the International Conference on Learning Representations, 2015.
- [20] RADHAKRISHNA A, SHEILA H, Francisco E, et al. Frequency-tuned Salient Region Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009.
- [21] BUSCHMAN T J, MILLER E K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices[J]. Science, 2007, 315(30): 1860-1862.
- [22] ITTI L. Models of bottomup and top-down visual attention[D]. California Institute of Technology, 2000.
- [23] ITTI L, KOCH C. Computational modeling of visual attention [J]. Nature Reviews Neuroscience, 2001, 2(3): 194-203.
- [24] MIRPOUR K, ARCIZET F, ONG W S, et al. Been There, Seen That: A Neural Mechanism for Performing Efficient Visual Search[J]. Journal of Neurophysiology, 2009, 102(6): 3481-3491.
- [25] DENG T, LUO E Q, ZHANG Y S, et al. Selective Attention-Based Saliency of Traffic Images and Characteristics of Eye Movement[J]. Journal of University of Electronic Science and Technology of China, 2014, 43(4): 624-628.



WEN Jing, born in 1982, Ph.D, associate professor, M. S. supervisor, is a member of China Computer Federation. Her main research interests include computer vision, image processing and pattern recognition.