

基于卷积神经网络与全局优化的协同显著性检测

吴泽民 王军* 胡磊 田畅 曾明勇 杜麟

(中国人民解放军陆军工程大学通信工程学院 南京 210007)

摘要: 针对目前协同显著性检测问题中存在的协同性较差、误匹配和复杂场景下检测效果不佳等问题, 该文提出一种基于卷积神经网络与全局优化的协同显著性检测算法。首先基于VGG16Net构建了全卷积结构的显著性检测网络, 该网络能够模拟人类视觉注意机制, 从高级语义层次提取一幅图像中的显著性区域; 然后在传统单幅图像显著性优化模型的基础上构造了全局协同显著性优化模型。该模型通过超像素匹配机制, 实现当前超像素块显著值在图像内与图像间的传播与共享, 使得优化后的显著图相对于初始显著图具有更好的协同性与一致性。最后, 该文创新性地引入图像间显著性传播约束因子来克服超像素误匹配带来的影响。在公开测试数据集上的实验结果表明, 所提算法在检测精度和检测效率上优于目前的主流算法, 并具有较强的鲁棒性。

关键词: 协同显著性; 深度学习; 卷积神经网络; 协同优化

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2018)12-2896-09

DOI: 10.11999/JEIT180241

Co-saliency Detection Based on Convolutional Neural Network and Global Optimization

WU Zemin WANG Jun HU Lei TIAN Chang ZENG Mingyong DU Lin

(College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China)

Abstract: To solve the problems in current co-saliency detection algorithms, a novel co-saliency detection algorithm is proposed which applies fully convolution neural network and global optimization model. First, a fully convolution saliency detection network is built based on VGG16Net. The network can simulate the human visual attention mechanism and extract the saliency region in an image from the semantic level. Second, based on the traditional saliency optimization model, the global co-saliency optimization model is constructed, which realizes the transmission and sharing of the current superpixel saliency value in inter-images and intra-image through superpixel matching, making the final saliency map has better co-saliency value. Third, the inter-image saliency value propagation constraint parameter is innovatively introduced to overcome the disadvantages of superpixel mismatching. Experimental results on public test datasets show that the proposed algorithm is superior over current state-of-the-art methods in terms of detection accuracy and detection efficiency, and has strong robustness.

Key words: Co-saliency; Deep Learning; Convolutional Neural Network; Global Optimization

1 引言

显著性检测旨在模拟人类的视觉注意机制(Human Visual System, HSV), 构建有效的计算模型, 以从图像视频中提取关键信息, 简洁有效地表达图像内容。传统的显著度计算往往只针对一幅静态图像, 即通过计算单幅图像内不同区域之间的相似度与对比度来描述其显著性信息。近年来, 随着多场景(图像或视频)应用的需求, 协同显著性正逐渐成为一个新兴的研究课题。所谓协同显著性, 主

要是指多场景中共同存在的相同或者相似目标的显著程度。协同显著性不仅要考虑到单幅图像的显著性计算, 还要考虑到一组图像之间的协同性信息, 从而对该组图像中的共同目标进行更加精确的标定和凸显。协同显著性对协同分割、图像匹配、多场景目标再识别、图像和视频摘要等都具有重要的应用价值。

显著性和重复性是计算协同显著性的两大要素^[1-6]。因此针对协同显著性检测, 需要考虑以下两个方面的问题: (1)如何通过有效的特征表示与计算来提取单幅图像的显著性区域; (2)如何通过有效的特征表示提取组内图像之间的协同信息对共

同显著目标区域进行强化，从而更加精准地定位一组图像中共同或相似的显著目标区域。近年来，不少学者做了大量的研究工作。如Li等人^[7]在单个图像显著性检测的基础上，通过多层图结构模型构建图像对的协同显著性。Fu等人^[8]利用颜色聚类提取多幅图像的一致性信息，然后融合颜色对比度，空间分布性和图像间一致性等线索构造协同显著图。Liu等人^[9]通过对图像的分层分割，利用颜色直方图计算分割区域的全局相似性，从而得到协同显著图。Zhang等人^[10]首先通过MR算法^[11]得到单幅图像的显著性，然后利用改进的自动元胞机模型^[12]对图像间相似区域进行显著性传播得到最终的协同显著图。以上算法在较简单的场景下均有不错的表现，但仍存在下列两个问题：一是上述算法只利用了颜色等底层特征，难以在复杂场景下取得良好的检测效果；二是上述算法难以对一组图像中的背景和前景目标进行有效区分，造成图像之间背景和显著目标的误匹配，从而影响最终检测效果。

近年来，深度学习逐渐成为机器学习领域的主流，被广泛应用于图像识别检测等任务中。深度神经网络的层级结构能够模拟人类神经系统，逐层提取图像信息，并通过海量数据的训练来学习参数，最终得到具有鲁棒性和泛化性的模型，从而大幅提升检测性能。不少学者将深度学习技术引入协同显著性检测领域并取得了良好的效果。如Zhang等人^[13]通过提取图像的深度特征，对同类图像和不同类别的图像进行综合训练，以提取共同目标的特征和背景区域的特征，但在图像间区域相似度匹配中只采用CIE-lab空间的平均颜色特征进行匹配，性能有待提升。文献^[14]通过多实例学习和自调式学习寻找图像间共同存在的目标。Wei等人^[15]通过深度学习提取输入图像组的共同语义特征来寻找共同显著目标。以上方法均需要精确标定目标训练集、训练复杂度高，难以推广。

本文针对目前协同显著性检测中存在的上述问题，结合显著性和重复性两大要素，首先基于VGG16Net^[16]深度学习网络框架，构建了全卷积结构的显著性检测网络。采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，从而更好地模拟人类的视觉注意机制，提取图像中显著目标区域。为了更好地对一组图像之间的协同性进行建模，本文基于传统单幅图像的优化模型构造了全局协同显著性优化模型。该优化模型充分利用图像内与图像间的协同性信息，实现了局部超像素块显著值在图像间的传播与共享，使得优化后的显著图与真值更加接近。最后，为了克服图像间相似目标区域误匹配带来的影响，本文创新性地引入图像间显著性传播约束因子。在数据集上的实验表明本文算法优于目前主流的协同显著性检测算法，具有较强的鲁棒性。

2 基于全卷积结构的显著性检测网络

2.1 网络结构

网络的整体结构如图1所示，本文显著性检测网络采用端到端的方式将整个图像作为处理对象进行训练和预测，与基于图像块的训练方法相比^[13-15]，大大提高了训练与测试效率。其中第1部分(前13个卷积层与4个池化层)基于VGG16卷积网络。对于输入的每幅图片，该网络首先将其尺寸调整为 224×224 ，然后通过一系列卷积与池化操作得到不同层次的深度特征。对于卷积神经网络而言，比较浅的卷积层提取的往往是图像颜色、边缘和强度等底层特征，而比较深的卷积层提取的往往是高级语义特征。即较深的卷积层能够强化目标区域的整体特征，弱化大量细节性信息。因此，高层卷积特征更能够从整体上标注出一幅图像中的显著性目标区域。

为了使网络模型更加适合进行显著性检测，在本

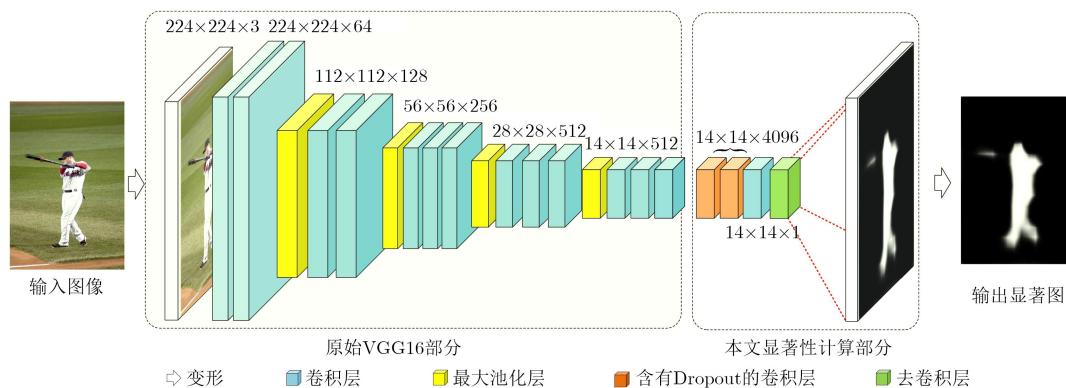


图1 显著性检测网络结构示意图

文显著性检测网络的第2部分中, 本文借鉴文献[17]的做法, 去掉了VGG16中的全连接层, 并额外添加两个卷积层对第1部分得到的深度特征做进一步抽象, 以提高其泛化能力, 同时引入Dropout。最后通过一个卷积层与去卷积层输出显著图(显著值范围为[0, 1])。每个卷积层的特征都通过激活函数后送入下一个卷积层。激活函数为

$$y = \max \{x, 0\} \quad (1)$$

2.2 网络训练

给定训练样本集 $\mathbf{X} = \{X_i\}_{i=1}^N$ 以及对应的显著图真值标注 $\{\text{GT}_i\}_{i=1}^N$, N 表示数据集中的训练样本个数。该卷积网络输出的显著图可以表示为

$$S_{\text{cnn}}^i = f_s(X_i; \Theta) \quad (2)$$

其中, $f_s(\cdot)$ 表示整个网络所构建的从输入图像 X_i 到输出显著图 S_{cnn}^i 的映射模型。 Θ 表示网络中所有参数的集合, 其主要由两部分构成: 第1部分为VGG16Net中的参数, 在训练中保持不变; 第2部分为新添加的相关卷积与去卷积层参数。训练的目的就是利用给定的训练集对后边几层参数进行微调, 使得损失函数最小, 求出最优模型:

$$L(\mathbf{X}; \Theta) = \frac{1}{N} \sum_{i=1}^N \|\text{GT}_i - f_s(X_i; \Theta)\|_F^2 \quad (3)$$

$$\Theta = \arg \min_{\Theta} L(\mathbf{X}; \Theta) \quad (4)$$

本文深度显著性检测网络的搭建基于Caffe toolbox^[18]。实验中, 通过随机梯度下降法(SGD)进行网络训练, 采用最常用的MSRA10K^[19]数据库作为训练集。MSRA10K数据库包含丰富的图像样本(10000张图像), 每张图像都有着像素级别的真值标注, 广泛应用于显著性检测网络的训练。训练过程中学习率为0.001, 批处理数量设置为20, 权重衰减系数设置为0.005, 训练迭代次数为50。

3 基于广义匹配邻居的协同优化模型

基于训练好的深度卷积神经网络, 可以得到初始显著图 S_{cnn} , 如图2(c)所示。该显著图从整体上标注了每幅图片中显著性目标区域, 但其对于目标的边缘等细节部分检测效果不够理想, 还存在一定程度的背景噪声。而且, 对于协同显著性检测而言, 其没有考虑到一组图像之间的协同性。

为了对初始显著图做进一步优化, 同时便于后期图像间相似目标区域的匹配以构建有效的协同优化模型, 本文使用SLIC^[20]算法(该算法在对图像做分割时能够保留目标边缘信息)对原始图像 I_i 做超像素分割, 形成一系列超像素集 $R_i = \{r_i\}_{i=1}^{N_i}$, N_i 表示分割形成的超像素的个数, 如图2(b)所示。然后定义超像素 r_i 的初始显著值为

$$f_i = \frac{1}{M} \sum_{p \in \Omega} S_{\text{cnn}}(p) \quad (5)$$

其中, Ω 表示位于超像素 r_i 内的所有像素点, M 表示超像素 r_i 中所包含总的像素点个数, $S_{\text{cnn}}(p)$ 表示初始显著图中像素 p 处的显著值。由此可以得到超像素层次的显著图 $\mathbf{F}_i = [f_1, f_2, \dots, f_{N_i}]$, 如图2(d)。

为了保证空间相邻的相似区域显著性一致, 很多单幅图像的检测算法都采用一定的优化模型来得到更加精确的显著图。典型的优化模型^[21]如式(6)所示。

$$\arg \min_{\mathbf{s}} \frac{1}{2} \sum_i (s_i - f_i)^2 + \sum_{i,j \in \mathfrak{R}} \omega_{i,j} (s_i - s_j)^2 \quad (6)$$

其中, $\omega_{i,j}$ 为超像素 r_i, r_j 的相似度权重, \mathfrak{R} 表示超像素 r_i 的“邻居”(与其空间上相邻的超像素), f_i 为超像素 r_i 的初始显著值, $\mathbf{s} = [s_1 \ s_2 \ \dots \ s_j]$ 为优化后各超像素的显著值。式(6)的第1项融入先验显著性信息, 使得优化后的显著图与初始显著图尽可能保持一致。第2项融入空间一致性信息, 使得邻居中

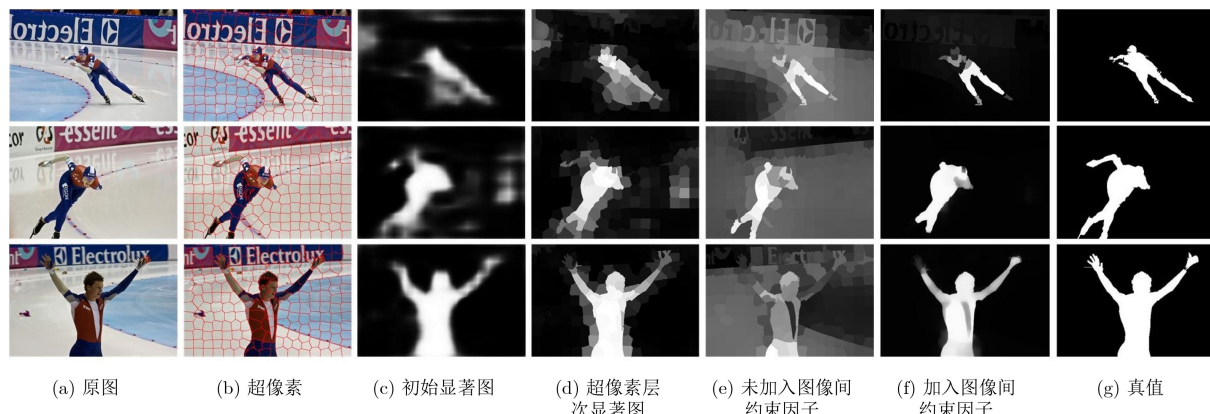


图2 本文算法每个步骤所生成的显著图对比

相似度权重较大的超像素块之间的显著值尽可能一致，因此实现了邻居之间显著值的传播。

对于协同显著性而言，本文对“邻居”的定义做了如下扩展：即超像素块 $r_{i,s}$ 的邻居不仅包括其在单幅图像上相邻的超像素块，还包括该组图像其他图像中与其特征相匹配的超像素块(广义邻居)，如图3所示。因此，对于协同显著性检测而言，新的优化模型为

$$\arg \min_s \frac{1}{2} \sum_i (s_i - f_i)^2 + \underbrace{\alpha \sum_{i,j \in \mathfrak{R}_1} \omega_{i,j}^* (s_i - s_j)^2}_{\text{图像内一致性}} + \underbrace{\beta \sum_{i,j \in \mathfrak{R}_2} \omega_{i,j}^* \psi_{i,j} (s_i - s_j)^2}_{\text{图像间一致性}} \quad (7)$$

其中， \mathfrak{R}_1 表示超像素 r_i 的图像内邻居(图3中间图像小圆点标注部分)， \mathfrak{R}_2 表示超像素 r_i 的图像间邻居(图3四周图像圆点和方块标注部分，四周图像中圆点标注部分为正确匹配的超像素，方块标注部分为误匹配的超像素)， $\omega_{i,j}^*$ 为协同优化模型中的相似度权重， $\psi_{i,j}$ 为图像间显著性传播约束因子。协同优化示意图如图3所示，可以看出，相对原始优化模型，协同优化模型充分利用了图像间相似目标区域的一致性，实现了局部超像素的显著值在图像内与图像间相似区域的传播。下面，首先介绍本文对相似度权重 $\omega_{i,j}^*$ 的定义以及图像间广义邻居的匹配方法，随后介绍图像间超像素显著性传播约束因子 $\psi_{i,j}$ 。

3.1 基于颜色与纹理特征的超像素匹配

颜色与纹理特征是描述图像区域属性的重要特征，在模式匹配中，这种区域性的特征具有较大的

优越性，不会由于局部的偏差而无法匹配成功，而且实现起来简单高效。因此，为了有效地描述超像素块之间的相似性，本文结合纹理与颜色特征来定义超像素块之间的相似度：

$$\omega_{i,j}^* = (1 - \varepsilon) \cdot f_{\text{Lab}}(r_i, r_j) + \varepsilon \cdot f_{\text{Lbp}}(r_i, r_j) \quad (8)$$

其中， ε 为控制权重大小，平衡纹理特征与颜色特征的影响程度大小。 f_{Lab} 为描述超像素块之间颜色相似度的函数，其被定义为： $f_{\text{Lab}}(r_i, r_j) = \exp(-\|c_i - c_j\|^2 / (2\delta_c^2))$ ， c_i, c_j 表示超像素 r_i, r_j 在Lab空间的平均颜色特征向量。 f_{Lbp} 为描述超像素块之间纹理相似度的函数，其被定义为： $f_{\text{Lbp}}(r_i, r_j) = \exp(-\|b_i - b_j\|^2 / (2\delta_b^2))$ ， b_i, b_j 为超像素 r_i, r_j 的Lbp纹理特征向量。 δ_c, δ_b 分别为描述颜色特征相似性与纹理特征相似性的控制参数。本文取图像组中满足匹配条件 $\{r_j | \omega_{i,j}^* \geq \xi\}$ 的超像素构成超像素 r_i 的图像间邻居， ξ 为匹配阈值。

3.2 图像间显著性传播约束因子

需要说明的是，在一组图像中，对超像素 r_i 而言，满足与其匹配条件的“邻居”有时候可能既包含显著性目标自身，也包含其他部分(误匹配)。如图3所示，中心图像中绿色圆点标注区域为当前待匹配的超像素块，周围白色圆点标注区域为其图像内的“邻居”，四周图像中圆点和方块标注部分均为满足匹配条件的图像间邻居。对于协同显著性检测而言，我们更希望超像素 r_i 的显著值与圆点标注区域具有一致性。为了克服误匹配带来的影响，本文引入图像间显著性传播约束因子 $\psi_{i,j}$ ，定义为

$$\psi_{i,j} = f_i \cdot f_j \quad (9)$$

可以看出，约束因子 $\psi_{i,j}$ 综合考虑了超像素块 r_i, r_j

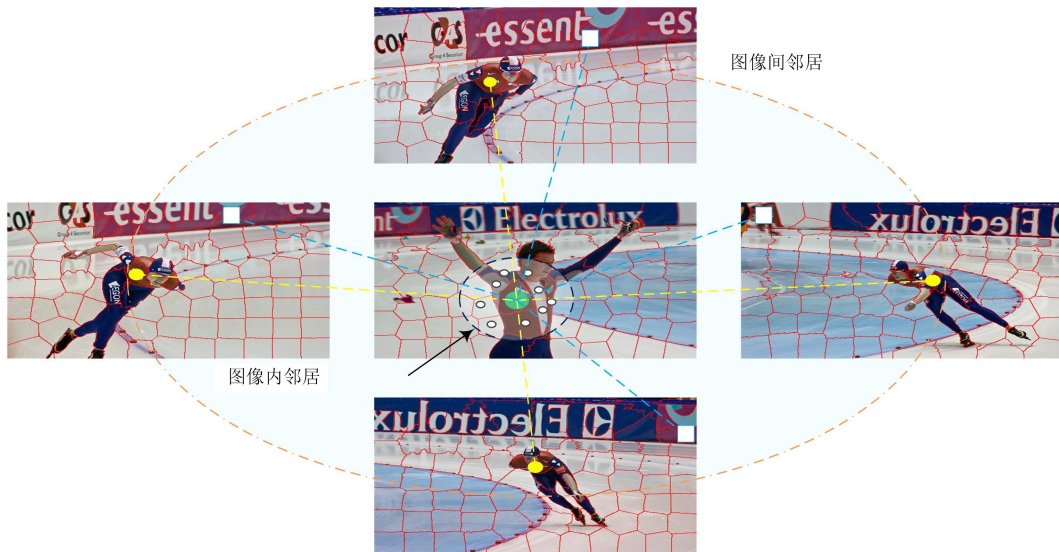


图3 协同显著性优化示意图

各自的初始显著性大小。由于 $f_i, f_j \in [0, 1]$ ，如果满足匹配条件的两个超像素的初始显著性均比较大(接近于1)，则在协同优化时，对两者之间的相似度权重 $\omega_{i,j}^*$ 的约束较小。如果其中任意一个超像素块的初始显著值较小(接近于0)，在进行优化时对相似度权重 $\omega_{i,j}^*$ 的“约束”较大，这样的两个超像素虽然特征较为相似，但由于相似度权重 $\omega_{i,j}^*$ 的约束，仍然能够阻碍显著值在这两个超像素块之间传播，从而降低了误匹配的影响。

为直观地说明其有效性，图2(e)，图2(f)给出了几个示例。结合前文相关论述可以看出，图2(e)未加入约束因子时，由于目标与背景之间的误匹配，造成最后的显著图中包含大量背景噪声，显著图质量较差。图2(f)通过引入传播约束因子，保证了超像素块之间的显著性能够在目标区域内部之间进行传递，降低了误匹配带来的影响，因此所得到的显著图与真值也更加接近。

4 实验与分析

4.1 实验设置

测试数据集：本文在两大数据集上测试各算法：ImgPair^[22]和iCoSeg^[23]。其中，ImgPair包含105组图像，每组只有两幅图像，相对简单；iCoSeg包含38组图像，每组均包含多幅图像，背景复杂，具有一定挑战性。比较算法：ACM^[7]，CB-C^[8]，CB-S^[8]，HS^[9]，SM^[10]，LDW^[13]，SA^[24]，以及IPIM^[25]。

评价指标：本文采用P-R(精度-召回率)曲线，F-measure(下)曲线，ROC曲线面积(Area Under ROC Curve, AUC)，平均F-measure值(Average F-measure, AF)和平均绝对误差(Mean Absolute Error, MAE)进行算法评估。P-R曲线通过[0, 255]的滑动阈值对每个显著图进行二值化，并和真值进行对比求得精度(Precision, P)和召回率(Recall, R)：

$$P = \frac{|M \cap G|}{|M|}, R = \frac{|M \cap G|}{|G|} \quad (10)$$

平均F-measure(AF)值一般采用一个自适应阈值(平均显著值的2倍)对显著图进行二值化，然后求得精度和召回率：

$$AF = \frac{(1 + \beta^2) P \cdot R}{\beta^2 \cdot P + R} \quad (11)$$

这里一般取 $\beta^2 = 0.3$ 以强调精度。F-measure曲线图通过一个[0, 255]的滑动阈值对显著图进行二值化，和真值对比求得每个阈值下的F-measure值。MAE被定义为显著图S和真值G(两者归一化到[0, 1])绝对误差的平均值：

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)| \quad (12)$$

在实验中，协同优化模型中的控制参数 α, β 均取1。控制参数 $\varepsilon, \delta_b, \delta_c$ 分别取0.5, 10, 10，匹配阈值 ξ 取0.8。iCoSeg和MSRC两个数据集的超像素分割参数为[300 30]，ImgPair数据集上的图片分辨率较小，因此超像素分割参数设置为[100 10]。

4.2 定性分析

图4，图5给出了本文算法与其他算法在不同数据库上的检测结果示例。其中，图4中所包含的检测样本来自于ImgPair数据集。由于每组图像只包含两个样本，检测难度相对较小，因此大部分算法均能够检测到共同目标区域，但都存在不同程度的背景噪声，其中ACM算法检测结果不完整，CB-C算法对共同目标区域的凸显程度不够。图5所展示的一组图像来自iCoSeg数据集，背景较为复杂而且背景中包含颜色外观与目标比较相似的区域，具有一定的挑战难度。LDW，IPIM和SM等算法均出现不同程度的误匹配问题，把背景部分误检测为显著区域。SA算法整体检测结果相对较好，但其算

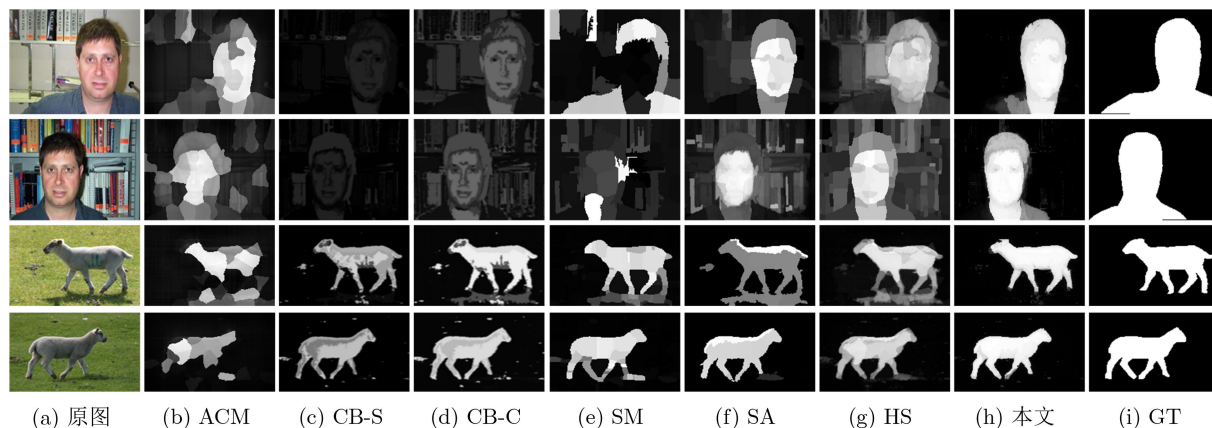


图4 ImgPair数据集上部分实验结果对比示例(GT表示真值图)

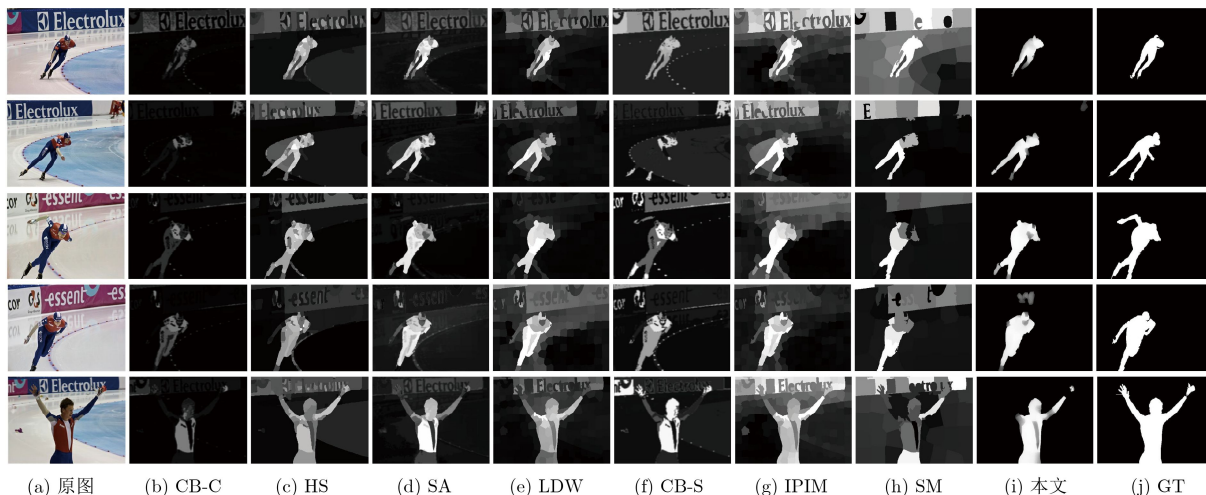


图5 iCoSeg数据集上部分实验结果对比示例

法的核心是对多个单幅图像显著性算法的融合，其性能对各算法检测的效果依赖性很强。本文首先运用深度网络模拟人眼的视觉注意机制，能够在不同背景下准确判定显著目标区域，提取语义显著图；然后综合颜色与纹理特征完成相似目标区域的有效匹配，最后运行改进的全局优化模型进行显著性的图像内与图像间传播；为克服误匹配带来的影响，本文还引入了图像间显著性传播约束因子。因此，本文算法在上述场景下都能够更加精确地检测出共同目标区域，而且几乎不存在误检测问题，体现了较强的鲁棒性。

4.3 定量分析

本文算法涉及到由卷积神经网络输出初始显著图以及协同优化两个步骤，为了说明每个步骤对最终算法的贡献大小，图6给出了在iCoSeg数据集上每个步骤对算法性能贡献程度的定量分析。其中，本文CNN表示本文构建的深度显著性检测网络的性能，本文CNN+COOPT表示协同优化后的整体算法性能。从F-measure曲线上可以明显看出，初始显著性的计算由于没有考虑到一组图像之间的协同性与一致性信息，其性能明显低于协同优化后的结果。协同优化后的算法性能好于仅使用CNN的

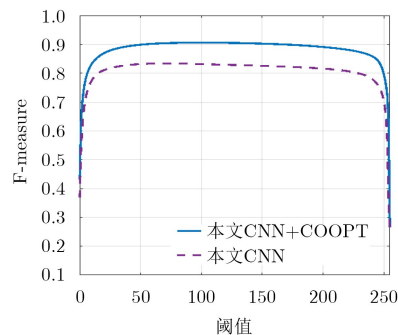


图6 本文算法在iCoSeg数据集上的量化分析(F-measure曲线图)

方法，说明了本文构建的协同优化模型的有效性。

图7和图8给出了本文算法和部分其他算法在两大数据集上的P-R曲线，F-measure曲线图对比；表1给出以上算法在ImgPair和iCoseg两大数据库上的AUC值，AF值与MAE值。可以看出，在两大数据集上，本文算法均有着更加饱和的P-R曲线以及更宽范围阈值的高F-measure值，说明本文算法在精度与召回率方面相对于以往算法均有较大提升，同时，本文算法在两大数据集上的MAE值均最低。以上实验数据说明本文算法所产生的显著图与真值更加接近。

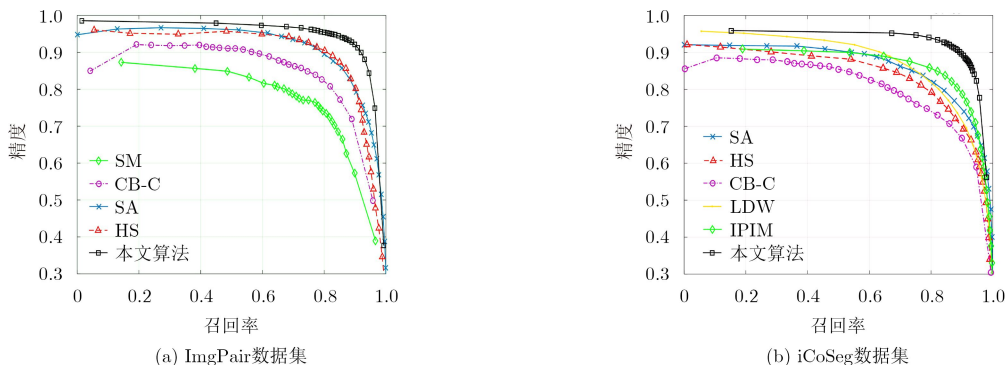


图7 本文算法与其他算法在两大数据集上的P-R曲线对比

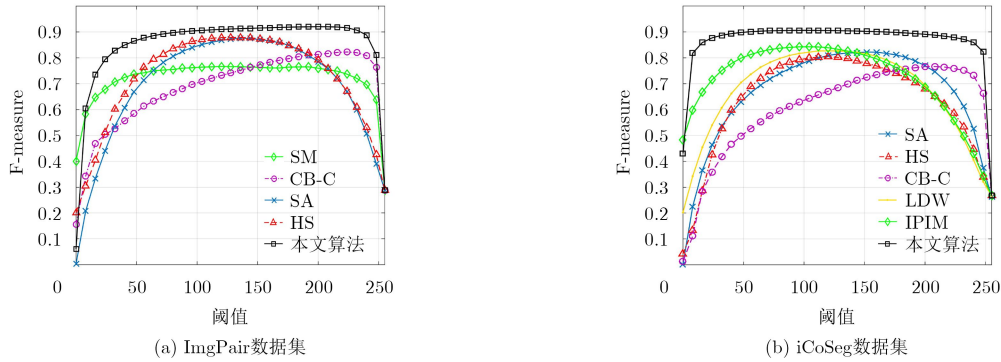


图8 本文算法与其他算法在两大数据集上的F-measure曲线对比

表1 不同算法在两大数据库上的测试结果对比

算法	ImgPair			iCoSeg		
	AUC	AF	MAE	AUC	AF	MAE
SA	0.967	0.826	0.160	0.965	0.720	0.160
HS	0.954	0.821	0.147	0.954	0.640	0.180
CB-C	0.931	0.782	0.178	0.913	0.647	0.198
CB-S	0.927	0.749	0.181	0.935	0.688	0.173
ACM	0.880	0.719	0.197	-	-	-
SM	0.879	0.724	0.166	0.621	0.580	0.234
LDW	-	-	-	0.957	0.699	0.178
IPIM	-	-	-	0.964	0.703	0.159
本文CNN	0.958	0.811	0.098	0.932	0.761	0.081
本文CNN+COOPT	0.981	0.904	0.075	0.962	0.848	0.056

4.4 运行时间分析

表2给出了本文算法和其他几种已公开代码的协同显著性算法的平均运行时间比较。所有代码均基于Ubuntu 14.04 系统下的MATLAB R2014b平台进行测试。测试环境为4核心的PC(CPU: i7-4790, RAM16GB.GPU: NVIDIA Titan X, RAM12GB)。测试数据集为iCoSeg, 测试图片大小一般为 400×500 。需要说明的是, 本文算法在调用卷积神经网络处理图像数据时启用了GPU图形处理器加速功能, 其余部分代码均只用CPU处理器进行处理。SA算法由于是融合多种单幅图像显著性检测算法的结果, 所以其整体运行时间取决于各算法的时间, 这里只给出融合算法的时间。可以看出, 在GPU加持下, 卷积神经网络处理每幅图像仅需0.12 s, 本文协同优化算法也仅需2.58 s。

本文整个算法的时间复杂度和SA融合算法相当, 比SM算法和CB-C算法大大降低。

5 结论

本文针对目前协同显著性检测算法中存在的问题, 基于VGG16Net构造全卷积显著性检测网络, 采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, 从而更好地模拟人类的视觉注意机制, 提取图像中显著区域。为了充分利用图像间的辅助信息对初始显著图做进一步优化, 本文在传统优化单幅图像显著性优化函数的基础上, 扩展了对超像素邻居的定义, 构造了全局优化模型, 同时引入图像间显著性传播约束因子来克服超像素误匹配带来的影响。在数据集上的实验表明, 本文算法在检测精度和检测效率上优于目前的主流算法, 具有较强的鲁棒性。

表2 不同协同显著性算法平均运算时间比较

算法	CB-C	SA	SM	本文CNN	本文CNN+COOPT
时间(s)	5.40	2.10	6.60	0.12	2.70
处理器	CPU	CPU	CPU	GPU	CPU+GPU

参考文献

- [1] CHANG Kaiyueh, LIU Tyngluh, and LAI Shanghong. From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado, USA, 2011: 2129–2136. doi: [10.1109/CVPR.2011.5995415](https://doi.org/10.1109/CVPR.2011.5995415).
- [2] JACOBS D E, DAN B G, and SHECHTMAN E. Co-saliency: Where people look when comparing images[C]. *ACM Symposium on User Interface Software and Technology*, New York, USA, 2010: 219–228. doi: [10.1145/1866029.1866066](https://doi.org/10.1145/1866029.1866066).
- [3] YE Linwei, LIU Zhi, ZHOU Xiaofeng, *et al.* Saliency detection via similar image retrieval[J]. *IEEE Signal Processing Letters*, 2016, 23(6): 838–842. doi: [10.1109/LSP.2016.2558489](https://doi.org/10.1109/LSP.2016.2558489).
- [4] YE Linwei, LIU Zhi, LI Junhao, *et al.* Co-saliency detection via co-salient object discovery and recovery[J]. *IEEE Signal Processing Letters*, 2015, 22(11): 2073–2077. doi: [10.1109/LSP.2015.2458434](https://doi.org/10.1109/LSP.2015.2458434).
- [5] LIU Zhi, ZOU Wenbin, and OLIVIER L M. Saliency tree: A novel saliency detection framework[J]. *IEEE Transactions on Image Processing*, 2014, 23(5): 1937–1952. doi: [10.1109/TIP.2014.2307434](https://doi.org/10.1109/TIP.2014.2307434).
- [6] REN Jingru, ZHOU Xiaofei, LIU Zhi, *et al.* Saliency integration driven by similar images[J]. *Journal of Visual Communication and Image Representation*, 2018, 50: 227–236. doi: [10.1016/j.jvcir.2017.12.002](https://doi.org/10.1016/j.jvcir.2017.12.002).
- [7] LI Yijun, FU Keren, LIU Zhi, *et al.* Efficient saliency-model-guided visual co-saliency detection[J]. *IEEE Signal Processing Letters*, 2015, 22(5): 588–592. doi: [10.1109/LSP.2014.2364896](https://doi.org/10.1109/LSP.2014.2364896).
- [8] FU Huazhu, CAO Xiaochun, and TU Zhuowen. Cluster-based co-saliency detection[J]. *IEEE Transactions on Image Processing*, 2013, 22(10): 3766–3778. doi: [10.1109/TIP.2013.2260166](https://doi.org/10.1109/TIP.2013.2260166).
- [9] LIU Zhi, ZOU Wenbin, LI Lina, *et al.* Co-saliency detection based on hierarchical segmentation[J]. *IEEE Signal Processing Letters*, 2014, 21(1): 88–92. doi: [10.1109/LSP.2013.2292873](https://doi.org/10.1109/LSP.2013.2292873).
- [10] ZHANG Zhaofeng, WU Zemin, JIANG Qingzhu, *et al.* Co-saliency detection based on superpixel matching and cellular automata[J]. *KSII Transactions on Internet and Information Systems*, 2017, 11(5): 2576–2589. doi: [10.3837/tiis.2017.05.015](https://doi.org/10.3837/tiis.2017.05.015).
- [11] YANG Chuan, ZHANG Lihe, LU Huchuan, *et al.* Saliency detection via graph-based manifold ranking[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, USA, 2013: 3166–3173. doi: [10.1109/CVPR.2013.407](https://doi.org/10.1109/CVPR.2013.407).
- [12] QIN Yao, LU Huchuan, XU Yiqun, *et al.* Saliency detection via cellular automata[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, 2015: 110–119. doi: [10.1109/CVPR.2015.7298606](https://doi.org/10.1109/CVPR.2015.7298606).
- [13] ZHANG Dingwen, HAN Junwei, LI Chao, *et al.* Co-saliency detection via looking deep and wide[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, 2015: 2994–3002. doi: [10.1109/CVPR.2015.7298918](https://doi.org/10.1109/CVPR.2015.7298918).
- [14] ZHANG Dingwen, MENG Deyu, HAN Junwei, *et al.* Co-saliency detection via a self-paced multiple-instance learning framework[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(5): 865–878. doi: [10.1109/TPAMI.2016.2567393](https://doi.org/10.1109/TPAMI.2016.2567393).
- [15] WEI Lina, ZHAO Shanshan, BOURAHLA O E F, *et al.* Group-wise deep co-saliency detection[C]. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia, 2017: 3041–3047. doi: [10.24963/ijcai.2017/424](https://doi.org/10.24963/ijcai.2017/424).
- [16] SIMONYAN K, and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[OL]. *Computer Science*, 2014. arXiv: 1409.1556.
- [17] LI Xi, ZHAO Liming, WEI Lina, *et al.* Deepsaliency: Multi-task deep neural network model for salient object detection[J]. *IEEE Transactions on Image Processing*, 2016, 25(8): 3919–3930. doi: [10.1109/TIP.2016.2579306](https://doi.org/10.1109/TIP.2016.2579306).
- [18] JIA Yangqing, SHELHAMER E, DONAHUE J, *et al.* Caffe: Convolutional architecture for fast feature embedding [C]. *Proceedings of the ACM International Conference on Multimedia*, 2014: 675–678. doi: [10.1145/2647868.2654889](https://doi.org/10.1145/2647868.2654889).
- [19] CHENG Mingming, ZHANG Guoxin, NILOY J, *et al.* Global contrast based salient region detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Spring, USA, 2011: 409–416. doi: [10.1109/CVPR.2011.5995344](https://doi.org/10.1109/CVPR.2011.5995344).
- [20] ACHANTA R, SHAJI A, SMITH K, *et al.* Slic superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274–2282. doi: [10.1109/TPAMI.2012.120](https://doi.org/10.1109/TPAMI.2012.120).
- [21] LU Song, MAHADEVAN V, and VASCONCELOS N. Learning optimal seeds for diffusion-based salient object detection[C]. *IEEE Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA, 2014: 2790–2797. doi: [10.1109/CVPR.2014.357](https://doi.org/10.1109/CVPR.2014.357).
- [22] LI Hongliang and NGAN K N. A co-saliency model of image

- pairs[J]. *IEEE Transactions on Image Processing*, 2011, 20(12): 3365–3375. doi: [10.1109/TIP.2011.2156803](https://doi.org/10.1109/TIP.2011.2156803).
- [23] BATRA D, KOWDLE A, PARIKH D, *et al.* Interactively co-segmentating topically related images with intelligent scribble guidance[J]. *International Journal of Computer Vision*, 2011, 93(3): 273–292. doi: [10.1007/s11263-010-0415-x](https://doi.org/10.1007/s11263-010-0415-x).
- [24] CAO Xiaochun, TAO Zhiqiang, ZHANG Bao, *et al.* Self-adaptively weighted co-saliency detection via rank constraint[J]. *IEEE Transactions on Image Processing*, 2014, 23(9): 4175–4186. doi: [10.1109/TIP.2014.2332399](https://doi.org/10.1109/TIP.2014.2332399).
- [25] ZHANG Dingwen, HAN Junwei, HAN Jungong, *et al.* Co-saliency detection based on intra-saliency prior transfer and deep inter-saliency mining[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(6): 1163–1176. doi: [10.1109/TNNLS.2015.2495161](https://doi.org/10.1109/TNNLS.2015.2495161).
- 吴泽民: 男, 1973年生, 副教授, 硕士生导师, 研究方向为图像分析、数据融合.
- 王 军: 男, 1995年生, 硕士生, 研究方向为深度学习、图像与视频的显著度研究.
- 胡 磊: 男, 1987年生, 博士, 研究方向为目标跟踪与识别、数据融合.
- 田 畅: 男, 1963年生, 教授, 博士生导师, 研究方向为数据链技术、图像视频处理.
- 曾明勇: 男, 1988年生, 博士生, 研究方向为目标检测与识别.
- 杜 麟: 男, 1990年生, 博士生, 研究方向为视频编码与视频传输保障.