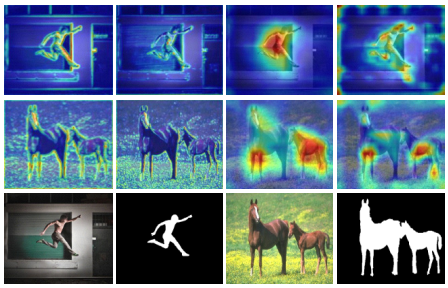
Opto-Electronic Engineering
光 电 工 程

Article

2020 年，第 47 卷，第 7 期

DOI: 10.12086/oee.2020.190627

基于子网络级联式混合信息流的显著性检测



董　波，王永雄*，周　燕，刘　涵，高远之，於嘉敏，张梦颖

上海理工大学光电信息与计算机工程学院，上海 200093

摘要：针对现有显著性检测算法在复杂场景下细节特征丢失的问题，本文提出了一种多层子网络级联式混合信息流的融合方法。首先使用 FCNs 骨干网络学习多尺度特征。然后通过多层子网络分层挖掘构建级联式网络框架，充分利用各层次特征的上下文信息，将检测与分割任务联合处理，采用混合信息流方式集成多尺度特性，逐步学习更具有辨别能力的特征信息。最后，嵌入注意力机制将显著性特征作为掩码有效地补偿深层语义信息，进一步区分前景和杂乱的背景。在 6 个公开数据集上与现有的 9 种算法进行对比分析，经实验验证，本文算法运行速度可达 20.76 帧/秒，并且实验结果在 5 个评价指标上普遍达到最优，即使对于挑战性很强的全新数据集 SOC。本文方法明显优于经典的算法，其测试结果 F-measure 提升了 1.96%，加权 F-measure 提升了 3.53%，S-measure 提升了 0.94%，E-measure 提升了 0.26%。实验结果表明，提出的模型有效提高了显著性检测的正确率，能够适用于各种复杂的环境。

关键词：显著性检测；级联式；混合信息流；注意力机制

中图分类号：TP391　　**文献标志码：**A

引用格式：董波，王永雄，周燕，等. 基于子网络级联式混合信息流的显著性检测[J]. 光电工程，2020，**47**(7): 190627

Saliency detection hybrid information flows based on sub-network cascading

Dong Bo, Wang Yongxiong*, Zhou Yan, Liu Han, Gao Yuanzhi, Yu Jiamin, Zhang Mengyin

Institute of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: In view of the detail feature loss issue existing in the complex scenario of existing saliency detection algorithms, a fusion method of multi-layer sub-network cascade hybrid information flows is proposed in this paper. We first use the FCNs backbone network to obtain multi-scale features. Through the multi-layer sub-network layering mining to build a cascading network framework, the context information of the characteristic of each level is fully used. The detection and segmentation tasks are processed jointly. Multi-scale features are integrated by hybrid information flows, and more characteristic information with discernment is learned step by step. Finally, the embedded attention mechanism effectively compensates the deep semantic information as a mask, and further distinguishes

收稿日期：2019-10-17； 收到修改稿日期：2019-12-11

基金项目：国家自然科学基金资助项目(61673276)

作者简介：董波(1998-)，男，主要从事机器视觉的研究。E-mail：535806671@qq.com

通信作者：王永雄(1970-)，男，博士，教授，主要从事智能机器人与机器视觉的研究。E-mail：wyxiong@usst.edu.cn

the foreground and the messy background. Compared with the existing 9 algorithms on the basis of the 6 public datasets, the running speed of the proposed algorithm can reach 20.76 frames and the experimental results are generally optimal on 5 evaluation indicators, even for the challenging new dataset SOC. The proposed method is obviously better than the classic algorithm. Experimental results were improved by 1.96%, 3.53%, 0.94%, and 0.26% for F-measure, weighted F-measure, S-measure, and E-measure, respectively. These experimental results show that the demonstrating the proposed model has higher accuracy and robustness and can be suitable for more complex environments, the proposed framework improves the performance significantly for state-of-the-art models on a series of datasets.

Keywords: saliency detection; cascade; hybrid information flows; attention mechanism

Citation: Dong B, Wang Y X, Zhou Y, *et al.* Saliency detection hybrid information flows based on sub-network cascading[J]. *Opto-Electronic Engineering*, 2020, **47**(7): 190627

1 引 言

显著性检测是对人类视觉注意力机制进行建模，准确定位图像中最重要的前景信息。作为计算机视觉任务的预处理过程，其类型众多，在静态图像中有RGB 图像显著性检测[1]，光场显著性检测[2]，融合深度信息的显著性检测[3]以及高分辨率的显著性检测[4]。在视频场景中有目标显著性检测[5]和注视点显著性检测[6]。

本文研究聚焦于对象级的显著性检测，其实现方法很多，可以归纳为两类：早期的计算方法和基于深度学习的方法。早期的方法主要基于各种手工特征设计显著性检测模型，由于人眼对视觉中心和周围的敏感性[7]具有一定的差异，对比度[8]成为一种广泛的研究特性，其他手工特征还包括中心先验[9]、背景先验[10]等。近年来，基于深度学习的显著性检测方法已经显示出令人印象深刻的结果。整合卷积神经网络的层次特征来实现多尺度特征融合的方法是当前的趋势。

近几年，许多研究者通过集成层次化的卷积特性，实现细粒度的显著性检测。这是因为更深层次的卷积特征倾向于对高层次知识进行编码，能够更好地定位突出的目标，而较低层次的卷积特征更有可能捕获丰富的空间信息。Liu 等人[11]将提取的多层特征输入多个子网络，预测最高分辨率的显著图，并直接融合，取得了较好的效果。但是多个子网络的方式可能会导致不同尺度的特征信息混淆，难以准确获得复杂区域边界。因此，Liu 等人[12]采用了由粗到细的特征提取方法，通过引入递归聚合方法，将各级初始特征融合在一起，逐级生成高分辨率的语义特征，并结合全局和局部的注意力机制，较好地解决了这一问题。尽管这类方法取得了良好的性能，但该方法中高层的语义信息逐层传输到浅层，所捕获的深层位置信息逐渐稀释或缺失。并且该方法中堆积大量的注意力模块，导致前景和背景区分不明显的部分的边缘模糊，某些层次的不准确信息还会导致错误检测。采用逐层融合的方式，解决了显著性目标稀释的问题。然而，由于逐级生成，随着网络深入，其特征逐渐由低级特征向高级特征发生转变，过多的低层次特征带来更多的空间细节，但这将导致高层特征无法获取准确的显著性目标，使得模型在复杂情况下可能失效。

为了有效整合多层的特征，本文借鉴多分类器级联在目标检测任务[13]上的良好性能，构建了一种用于显著性检测的多层子网络级联式全卷积神经网络框架。和逐级融合的方式不同，该模型抛弃了大量冗杂的低层特征，仅利用较高层的特征进行处理，有效避免了引入过多的低层信息，导致模型误判。为了使得子网络生成尽可能精确的显著图，引入混合信息流机制，捕捉更加有效的上下文信息，以确保子网络得到的显著图不会出现漂移。同时，利用中间层生成的显著性映射嵌入注意力机制来细化高阶特征，滤除噪声的同时改善信息流的传递方式，增强了多尺度显著性特征的有效性。与现有的显著性检测方法相比，该方法通过级联式多级细化，从而获得多层次的表征信息；利用通道组合的方式融合多分支信息流，从而获得更为有效的上下文信息，并结合注意力机制增强了显著性特征信息，从而提高了显著性检测的性能。

2 算法理论推导

在显著性检测中，信息的提取和流动方式决定了最终特征融合的效果。本文提出的多层子网络级联式混合信息流框架如图 1 所示。输入的图像经骨干网络处理后，获取的多层次特征分别进入不同的子网络，每一个子网络由混合信息流模块获取显著性特征信

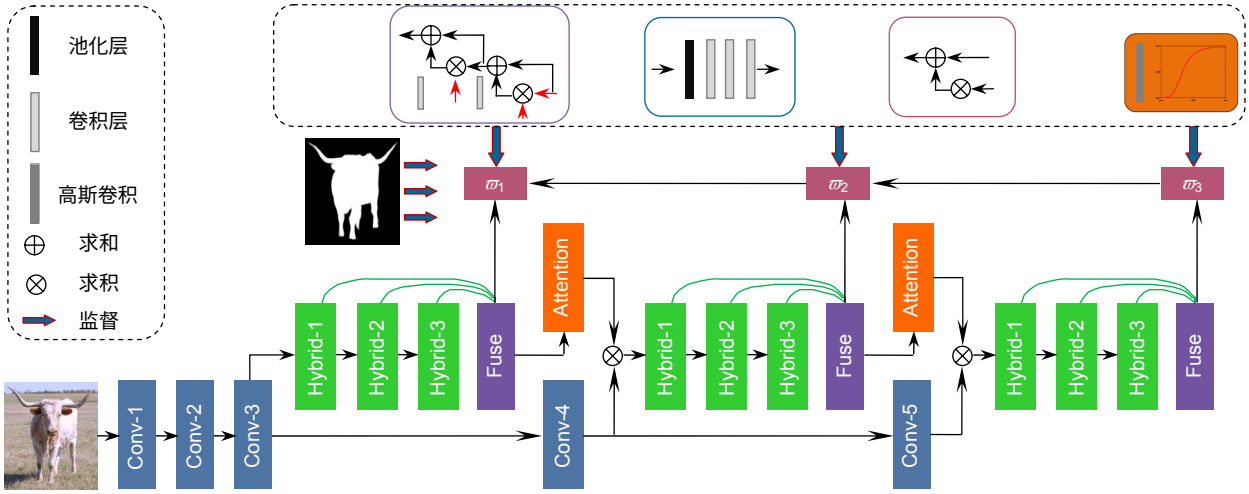


图 1 基于多层子网络级联式混合信息流的显著性检测模型

Fig. 1 The saliency detection hybrid information flows based on the multi-layer sub-network cascading model proposed in this paper

息，并将获得的三个尺度的特征经通道融合的方式整合。由于直接利用各阶段深层次的卷积特征提取显著性对象仍然存在一定的不足，因此本文使用注意力机制，将该层次的信息过滤后流入更深的层次中，从而辅助深层提取显著性信息。最后经非线性加权融合后即可获得准确的显著图。

2.1 级联式主框架

全卷积神经网络[14](FCNs)是显著性检测模型中应用最为广泛的网络，该网络模型的较浅层能提取到低层次特征，较深层能提取到更有效的高层次特征。本文将 FCNs[14]作为骨干网络获得多层次特征，并采用级联式混合信息流方式进行多尺度特征融合，有效避免了 FCNs[14]对应点聚合的弊端，形成更加丰富的特征信息。

本文模型网络结构如图 1 所示，首先将图像 $I\in\mathbb{R}^{H\times W\times C}$ 输入网络，其中 C 为输入图像的通道数，H 和 W 分别为输入图像的高和宽。利用骨干网络获取特征，通过级联的方式对特征进行处理，特征传递方式如下：

$$\chi_t^{\text{key}}=P_t(x,\chi_t^{\text{fuse}}) \quad , \tag{1}$$

$$\chi_t^{\text{fork}}=B(\chi_t^{\text{key}}) \quad , \tag{2}$$

$$\chi_t^{\text{fuse}}=F(\chi_1^{\text{fork}},\chi_2^{\text{fork}},\chi_3^{\text{fork}}) \quad , \tag{3}$$

其中：x 表示 FCNs[14]骨干网络获取的特征，χ_t^{key} 是由融合混合信息流的特征掩膜 χ_t^{fuse} 和骨干特征 x 处理后的信息流，χ_t^{fork} 是混合信息流处理后特征映射，$B(\cdot)$ 表示分层式结构处理，每一个阶段由三个分支组成，每一分支通过分层式结构处理后，经过 $F(\cdot)$上采样-卷积模块实现高效融合，从而实现混合信息流的目的，扰乱了模型重复的卷积、采样带来的重叠效应，通过 $P_t(\cdot)$注意力机制改进信息流的传递方式，利于将显著性特征掩码信息传递到深层次的特征。通过多层子网络级联强调细微信息，结合混合信息流辅助处理，增强了多尺度显著性特征的有效性；利用注意力机制将有效的显著性特征作为掩码信息传递到深层特征，实现信息流的高效传递；最后将多阶段的显著性映射非线性加权融合，互补滤除冗余特征。

2.2 混合信息流

当显著性目标结构复杂时，单个层次的特征往往无法提供足够的细节辨别能力，为了区分前景和复杂的背景信息，需要准确获取空间上下文信息。本文增加了一组混合任务分支，用于建立整个图像像素的上下文语义关联信息。具体是将前一分支的掩码特征反馈到当前掩码特征，采用直连方式加强掩码分支之间的信息流，利用分层式结构和新增的特征融合层实现多尺度最优融合。上下文信息是对多尺度特征的有效补充，因此组合他们可以获得更好的预测结果。

该模块首先通过 1×1 的卷积层将特征投影到一个不同的特征空间，接着，为了充分构建多个层次之间的上下文信息，利用 $n\times1$ 和 $1\times n$ 对流层来处理这些特征，并将结果反馈到每一个分支进行处理，其中 n=(3, 3, 5, 7)，以得到多个层次的特性并弥补不同特征在空间语义上的不足，如图 2 所示。此外，为了扩大各分支的感知区域，采用相对应大小的扩张卷积进行上采样解码。最后将转换后的不同层次特征图通过连接进行融合，使得模型捕获到任意空间位置在不同尺度下的上下文语义信息。结合本地特征对融合后的特征图

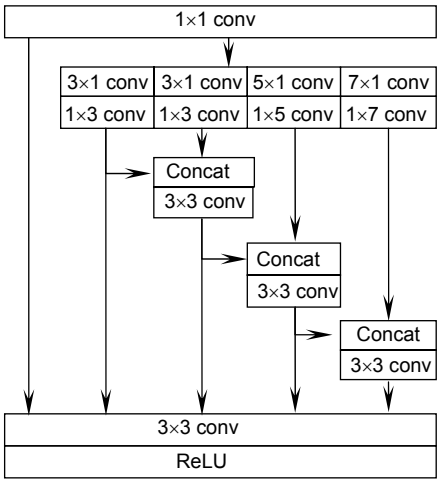


图 2 混合信息流

Fig. 2 Hybrid information flows

进行补充，利用 ReLU 消除冗杂信息，从而建立显著性对象区域的上下文信息。

密集连接有利于改善特征的表示方式，提升显著性检测性能。本文将深层次的语义特征与各层次的上下文信息结合，从而建立不同层之间的连接关系，实现多分支之间的语义信息交互。为了解决语义信息的融合问题，采用通道组合的方式，逐步融合高层的语义信息以获取更有效的上下文信息，并将上采样处理后采用卷积层对融合后的特征图进行平滑处理，使得下一步融合更加有效。该机制如下所示：

$$\tilde{S}_{\alpha}=conv_{\rho}\{(\xi^{1}(S_{1})\otimes S_{2})\oplus\xi^{2}(S_{1})\}\ , \tag{4}$$

$$\tilde{S}_{\beta}=conv_{\theta}\{(\xi^{3}(S_{1})\otimes\xi^{4}(S_{2})\otimes S_{3})\oplus\tilde{S}_{\alpha}\}\ , \tag{5}$$

$$\xi^{j}(S_{i})=conv_{3\times3}(up_{k=2}(S_{i}))\ , \tag{6}$$

其中：⊕表示特征矩阵元素求和，$\xi^{j}\in\{\xi^{1},\xi^{2},\xi^{3},\xi^{4}\}$ 是双线性插值与卷积核大小为 3 的组合，$S_{i}\in\{S_{1},S_{2},S_{3}\}$ 表示经分层式结构处理后大小为 $[H/2^{1+i},W/2^{1+i}]$ 的显著性特征映射。

为了验证该机制的有效性，本文对其附近的特征图进行了可视化处理。如图 3 所示，其中图 3(a)、3(b)列表示较浅层的特征信息，能够凸出显著性目标的空间细节信息，图 3(c)、3(d)列表示深层次的特征映射，能够很好地定位显著性目标。经混合信息流机制处理后特征图 3(a)、3(c)相对于未处理的特征图 3(b)、3(d)，能够更好地捕获显著性目标的空间细节以及定位信息，有助于模型持续关注图像中的显著性对象。

2.3 注意力机制

混合信息流机制能够获得不同层次特征之间的上下文语义信息，但是不同层次的语义信息比较独立，难以在全局范围内构建信息网络流。本文提出了一种注意力机制，弥补了混合信息流机制的不足，实现了不同层次特征的语义信息传递方式。该机制引入高斯滤波方式，降低噪声特征，提高模型对边界区域的感知性能。再将特征图映射到[0,1]，利用$\xi(\cdot)$函数的分类机制，提高深层次中显著性目标区域的权值，并且减小了非显著性区域的权值，从而增强了全局注意力图。最后将空间注意力图作为权重与骨干特征 x 的各个通道相乘，得到带有空间注意力的特征 x'。计算方式：

$$P_{t}(x,\chi_{t}^{\text{fuse}})=mul\{x,\xi(G_{\text{norm}}(conv(\chi_{t}^{\text{fuse}},k)))\}\ , \tag{7}$$

$$\xi(x)=1-\frac{1}{1+(\alpha x)^{\beta}}\ , \tag{8}$$

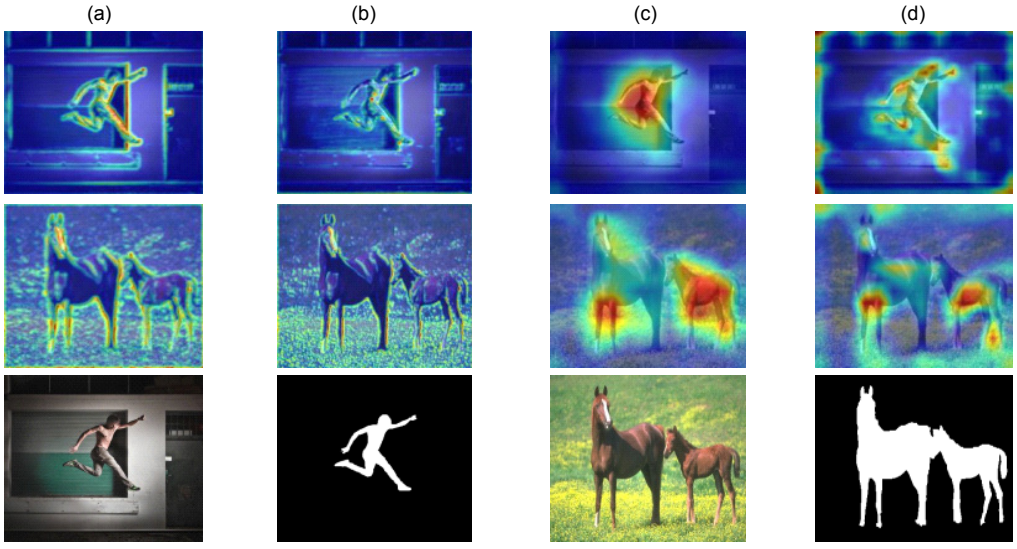


图 3 混合信息流前后的可视化对比

Fig. 3 Visual comparison before and after hybrid information flow

其中：χ_t^{fuse} 表示经混合信息流处理后的特征，$mul\{\cdot\}$ 函数表示各元素相乘，$conv(\cdot)$表示高斯卷积操作，初始化高斯核 k 为 16，标准差为 0，该卷积操作能够突出对象边缘，$G_{\text{norm}}(\cdot)$为归一化函数，将变量映射到[0,1]。为了利于$\xi(\cdot)$函数增大显著性区域的权值，如表 1 所示，参数α置为 2.5，β置为 5 时，在 ECSSD[15]和 HKUIS[16]两个数据集上平均绝对误差最佳(红色加粗标记)。该注意力机制有效改进了信息流，实现了深层特征的精细化处理。

表 1 $\xi(\cdot)$函数参数选择

Table 1 Parameter selection of function $\xi(\cdot)$

α	β	ECSSD[15]	HKUIS[16]
3.5	3.0	0.047	0.036
3.0	4.0	0.042	0.033
2.5	6.0	0.041	0.032
2.5	**5.0**	**0.040**	**0.033**
2.0	6.0	0.043	0.034
2.0	5.0	0.045	0.035

3 实验步骤

3.1 实验细节

本文实验基于 PYTORCH 框架实现，使用 NVIDIA GeForce GTX 1080Ti GPU 进行加速训练。训练阶段采用 VGG-16 预训练模型初始化骨干网络参数。利用数据集作为训练集，包含 10553 幅图像，将输入图像大小调整为 350×350，仅使用简单的随机水平翻转来增强数据集。模型采用 Adam 优化器，初始化学习率为 $5e^{-5}$，衰减权值为 $5e^{-4}$，共进行 25 个 epoch。采用的目标函数[17]为

$$L_{\text{Loss}}=(1-\lambda)\frac{1}{N}\sum_{i=1}^{N}\Psi[F(x_i)-y_i]+\lambda\times\frac{N^-}{N}\sum_{j=1}^{N^+}\Psi[F(x_j)-y_j]\ , \tag{9}$$

其中：平衡因子λ=0.5，N^+表示显著性区域的像素数，N^-表示非显著性区域的像素数，N 表示总像素数，x_i 表示输入图像中的每一个像素值 $X=\{x_i\,|\,i=1,2,\ldots,N\}$，$y_i$代表的是真值图中的每一个像素值 $Y=\{y_i\,|\,i=1,2,\ldots,N\}$，$y_i\in[0,1]$，$F(\cdot)$表示网络处理的过程，$\Psi(\cdot)$表示 Smooth-L1 函数。

3.2 数据集

本文使用 5 个显著性基准数据集评估本文的模型，分别为 ECSSD[15]、PASCAL[18]、DUT-OMRON[19]、HKUIS[15]和 DUTS[20]，其中 ECSSD[15]有 1000 个语义上有意义且复杂的图像，包含各种复杂的场景。PASCAL[18]数据集由 850 幅图像组成，均是带有像素级注释的自然图像。DUT-OMRON[19]包括 5168 张具有挑战性的图片，每张图片通常都有复杂的背景。HKUIS[15]包含 4447 张低对比度的图片，每张图片中都有多个前景对象。DUTS[19]数据集是目前数量最多的显著性检测基准数据集，包含了用于训练的 10553 张图像(DUTS-TR)和测试评估的 5019 张图像(DUTS-TE)。该数据集大多数显著性目标位置和规模不同，并具有复杂的背景。除此之外，本文对一个全新的数据集 SOC[21]进行了探索研究，它包含日常物体类别中显著和非显著物体的图像，并且显著图像含有单个或多个显著性目标。除了对象类别注释之外，每个显著的图像都伴有反映现实场景中常见挑战的属性，极具挑战性。

3.3 评估指标

本文使用 2 个常用指标(F-measure[22]，平均绝对误差(MAE)[23])以及最近提出的 3 个新的指标(加权 F-measure[24]，S-measure[25]，E-measure[26])进行评估。

1) F-measure[22] (F_β)：将预测的显著图与其对应的真值图进行对比，通常使用一个阈值来将一个显著性映射二值化成一个前景掩码映射，计算出平均准确率(Precision，用η_{P}表示)和召回率(Recall，用η_{R}表示)：

$$\eta_{\text{P}}=\frac{\sum_{(x,y)}G(x,y)S(x,y)}{\sum_{(x,y)}S(x,y)},\eta_{\text{R}}=\frac{\sum_{(x,y)}G(x,y)S(x,y)}{\sum_{(x,y)}G(x,y)}\ , \tag{10}$$

$$F_\beta=\frac{(1+\beta^2)\eta_{\text{P}}\times\eta_{\text{R}}}{\beta^2\times\eta_{\text{P}}+\eta_{\text{R}}}\ , \tag{11}$$

其中：阈值设置从 0 到 255 变化。由于准确率通常比召回率更重要，因此将 β^2 设置为 0.3。

2) MAE[23]：计算显著图与真值图的平均绝对误差(用σ_{MAE}表示)，广泛应用于显著性检测评估任务中。

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

其中：W 和 H 表示显著图 S 的宽度和高度，$S(x, y)$和 $G(x, y)$表示像素(x, y)处的显著性值和二元真值。MAE 分数越小，显著图与真值图之间相似程度高。

3) 加权 F-measure[24] (F_β^{w})：该指标由精度和查全率的加权均值计算：

$$F_\beta^{\text{w}}=\frac{(1+\beta^2)\eta_{\text{P}}^{\text{w}}\times\eta_{\text{R}}^{\text{w}}}{\beta^2\times\eta_{\text{P}}^{\text{w}}+\eta_{\text{R}}^{\text{w}}}\ , \tag{13}$$

其中：精度($\eta_{\text{P}}^{\text{w}}$)和查全率($\eta_{\text{R}}^{\text{w}}$)为加权精度和加权查全率，可以直接将显著图与真值图进行比较而不需要阈

值化，从而避免了 F_β 插值缺陷。同样 β^2 置 0.3，用于强调模型精度。

4) S-measure[25](S_α)：在文献[19]提出了 S-measure 指标，用于测量显著图的空间结构相似性：

$$S_\alpha = \alpha S_0 + (1-\alpha) S_r \quad , \tag{14}$$

其中 α 是一个平衡目标结构相似性 S_0 与区域相似性 S_r 的参数，在文献[19]中建议设置 α =0.5。

5) E-measure[26] (E_m)：最近 Fan 等人提出了一种增强定位度量的方法，该指标能够主动适应度量整体与局部的显著性差异。为了比较非二值显著性映射和二值映射，我们采用了一种类似于上述最大 F-measure 的方法，首先通过运行所有可能的阈值将显著性映射二值化，再将两个二进制映射的全局平均值对其进行对齐，然后计算局部像素相关性，最后得出最大 E_m，如下所示：

$$Q_{FM} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \varphi_{FM}(x, y) \quad , \tag{15}$$

其中 $\varphi_{FM}(\cdot)$ 表示增强的一致性矩阵操作用于捕获二值图中像素级匹配和图像级统计的属性。

在上面的 5 个指标中 F_β，F_β^w，S_α，E_m 越高，MAE 越低表示更好的性能。

4 实验结果与分析

4.1 定量分析

对现有的 8 个模型进行了比较，其中包括 7 个基于深度学习的模型与 1 个传统的显著性模型，基于深度学习的模型包括：DCL[27](深度对比学习)，DSS[28](具有短连接的深度监督模型)，DHS[29](深度分层学习)，Amulet[30](聚合多层次特征模型)，DLS[31](深层次聚合)，NLDF[32](非线性深度连接)，SRM[33](逐步修正模型)，同时本文对比了一种传统方法：DRFI[34](区域特征融合)。DCL[27]，DSS[28]，DHS[29]，Amulet[30]，DLS[31]，NLDF[32]利用 VGG-16 作为骨干网络，而 SRM[33]是在 ResNet-50 上实践。实验结果如表 2~5 所示，其中，-TR-表示该数据集作为该方法的训练集，红色表示方法的最优结果，蓝色表示方法的次优结果。

表 2 在 5 个基准数据集的 F_β 评分结果(越高越好)

Table 2 F_β score of five benchmark datasets (the higher the better)

Methods/Datasets	ECSSD[15]	DUT-OMRON[19]	PASCAL-S[18]	HKU-IS[15]	DUTS[20]
DRFI[34]	0.6899	0.6237	0.6382	0.7177	0.5857
DCL[27]	0.8820	0.6993	0.8220	0.8849	0.7820
DSS[28]	**0.9062**	**0.7369**	0.8111	**0.9011**	0.7773
DHS[29]	0.8937	-TR-	0.7984	0.8772	0.7813
Amulet[30]	0.9050	0.7154	0.8165	0.8888	0.7504
DLS[31]	0.8257	0.6448	0.7200	0.8074	-TR-
NLDF[32]	0.8887	0.6993	0.8027	0.8876	**0.8120**
SRM[33]	0.9048	0.7253	**0.8250**	0.8915	0.7976
Proposed	**0.9332**	**0.7892**	**0.8692**	**0.9236**	**0.8643**

表 3 在 5 个基准数据集的 MAE 评分结果(越低越好)

Table 3 MAE score of five benchmark datasets (the lower the better)

Methods/Datasets	ECSSD[15]	DUT-OMRON[19]	PASCAL-S[18]	HKU-IS[15]	DUTS[20]
DRFI[34]	0.1639	0.1554	0.2034	0.1394	0.1453
DCL[27]	0.0679	0.0797	0.1080	0.0481	0.0880
DSS[28]	**0.0517**	**0.0628**	0.0977	**0.0401**	0.0618
DHS[29]	0.0588	-TR-	0.0959	0.0519	0.0651
Amulet[30]	0.0589	0.0976	0.0992	0.0501	0.0841
DLS[31]	0.0859	0.0894	0.1328	0.0696	-TR-
NLDF[32]	0.0626	0.0796	0.1007	0.0480	0.0660
SRM[33]	0.0543	0.0694	**0.0867**	0.0457	**0.0583**
Proposed	**0.0402**	**0.0572**	**0.0715**	**0.0327**	**0.0431**

表 4 在 5 个基准数据集的 F^w_β 评分结果(越高越好)

Table 4 F^w_β score of five benchmark datasets (the higher the better)

Methods/Datasets	ECSSD[15]	DUT-OMRON[19]	PASCAL-S[18]	HKU-IS[15]	DUTS[20]
DRFI[34]	0.4629	0.3720	0.4695	0.5284	0.3261
DCL[27]	0.8387	0.6392	0.7332	0.8411	0.6927
DSS[28]	**0.8928**	**0.7118**	**0.7815**	**0.8891**	**0.7377**
DHS[29]	0.8422	-TR-	0.7397	0.8411	0.6965
Amulet[30]	0.8321	0.5928	0.7174	0.8022	0.6306
DLS[31]	0.7933	0.5966	0.6734	0.7750	-TR-
NLDF[32]	0.8547	0.6407	0.7484	0.8502	0.7103
SRM[33]	0.8628	0.6622	0.7617	0.8455	0.7252
Proposed	**0.8839**	**0.7112**	**0.7987**	**0.8815**	**0.7957**

表 5 在 5 个基准数据集的 S_α 和 E_m 评分结果(越高越好)

Table 5 S_α and E_m score of five benchmark datasets (the higher the better)

Methods/Datasets	ECSSD[15]		DUT-OMRON[19]		PASCAL-S[18]		HKU-IS[15]		DUTS[20]	
	S_α	E_m	S_α	E_m	S_α	E_m	S_α	E_m	S_α	E_m
DRFI[34]	0.7202	0.7631	0.6978	0.7938	0.6487	0.7450	0.7277	0.8329	0.6725	0.7620
DCL[27]	0.8684	0.9163	0.7710	0.8261	0.7855	0.8490	0.8770	0.9318	0.7891	0.8448
DSS[28]	0.8821	0.9306	0.7899	**0.8450**	0.7926	0.8560	0.8783	0.9414	0.8106	0.8717
DHS[29]	0.8842	0.9279	-TR-	-TR-	0.8045	0.8592	0.8698	0.9311	0.8201	0.8802
Amulet[30]	0.8941	0.9315	0.7805	0.8339	0.8193	0.8653	0.8860	0.9344	0.8039	0.8507
DLS[31]	0.8066	0.8726	0.7250	0.7978	0.7198	0.7941	0.7986	0.8769	-TR-	-TR-
NLDF[32]	0.8747	0.9221	0.7704	0.8200	0.8012	0.8591	0.8782	0.9344	0.8163	0.8716
SRM[33]	**0.8952**	**0.9371**	**0.7977**	0.8438	**0.8306**	**0.8787**	**0.8871**	**0.9442**	**0.8356**	**0.8910**
Proposed	**0.9125**	**0.9512**	**0.8188**	**0.8566**	**0.8562**	**0.8876**	**0.9089**	**0.9498**	**0.8707**	**0.9110**

由表 2~表 5 可以看出，基于深度学习方法 DCL[27]、DSS[28]、DHS[29]、Amulet[30]、DLS[31]、NLDF[32]、SRM[33] 在各个数据集下的平均绝对误差都有所降低，并且 F_β，F^w_β，S_α，E_m 结果都明显优于早期的 DRFI 方法[34]。

本文首先与次优算法 SRM 在多个数据集下的测试结果进行了横向对比。其中，对于平均绝对误差 MAE 指标，ECSSD 下降了 1.41%，DUT-OMRON 下降了 1.22%，PASCAL-S 下降了 1.52%，HKU-IS 下降了 1.3%，DUTS 下降了 1.52%。其次，在 DUTS 数据集下的结果进行纵向对比，其中 F_β 有 6.03%的提升，F^w_β 提升了 7.05%，S_α 提升了 3.48%，E_m 提升了 2%，并且 MAE 下降了 0.64%。因此，本文模型的错误预测数明显少于其他方法，能够适应各种复杂场景，体现了本文算法的准确性。

此外，本文对 SOC 数据集进行了评估，这是最近提出用于显著性检测的一个全新数据集，具有很强的

表 6 在 SOC 基准数据集的测试结果

Table 6 SOC benchmark data set scoring

Methods	F_β	MAE	F^w_β	S_α	E_m
DCL[28]	0.6440	0.1373	0.5570	0.6960	0.7712
DSS[28]	0.6284	0.1411	0.5625	0.6726	0.7593
DHS[29]	0.6844	0.1123	0.6103	0.7354	0.8005
RFCN[35]	0.6581	0.1276	0.5797	0.7180	0.7811
NLDF[32]	0.6663	0.1285	0.5774	0.7234	0.7843
SRM[33]	**0.7071**	**0.1074**	**0.6147**	**0.7632**	**0.8184**
Proposed	**0.7268**	**0.0975**	**0.6500**	**0.7726**	**0.8210**

挑战性。与 DCL[27]、DSS[28]、DHS[29]、RFCN[35]、NLDF[32]、SRM[33]这 6 种先进的算法进行比较，如表 6 所示(最优结果红色标记)。所提方法能够很好地适应这样一个新的数据集，在各个指标下均达到最优结果，进一步说明本文模型的准确性与鲁棒性。

为了更加直观地表明评价之间的关系，本文对 MAE 与 S_α，E_m 的分布趋势进行关联分析，如图 4 所示。本文的方法基本位于最左上角，在平均绝对误差 MAE 减小的前提下，S_α和 E_m 均有提升，表明所提方法能够更加突出前景区域和结构。这主要得益于混合信息流方式集成各个阶段互补特性，逐步学习更多具有辨别的特征信息，同时结合注意力机制使得信息之间可以实现非线性流动，从而使得最终的显著性映射集成了更多有效的多尺度的上下文信息。

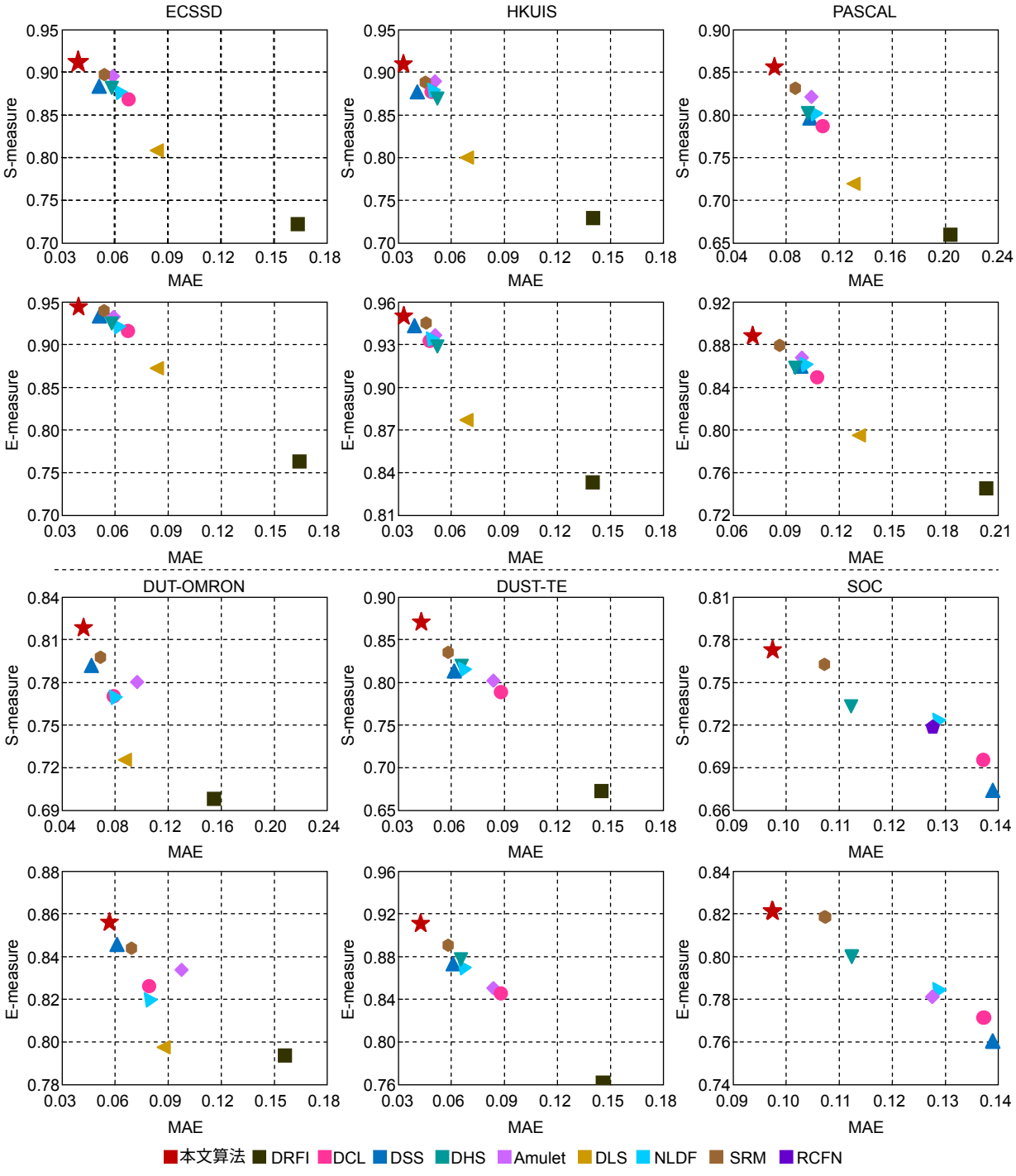


图 4 MAE 与 S-measure，E-measure 进行关联分析

Fig. 4 Association analysis between MAE and S-measure, E-measure

4.2 定性对比分析

为了进一步说明本文方法的优点，在图 5 中给出了部分数据的可视化结果。红色方框内是本文结果与真值图，可以得出该方法能够准确地识别图像中最显著的目标对象，并且几乎在所有情况下都能保持其尖锐的边界分割，目标区域高亮均匀，在特征融合以及抗噪性能方面达到了最优效果。

4.3 时间复杂度分析

本文利用 ECSSD[15]数据集作为测试，比较不同算法的运行时间与精度如表 7。从表 7 可以看出，与 DHS[29]、DSS[28]、NLDF[32]、Amulet[30]、SRM[33]算法相比，所提算法在保证处理效率的同时，精度大有提高，体现了本文算法的高效性。

4.4 有效性验证

为了验证本文提出的混合信息流、注意力机制以及非线性加权融合方式的有效性，在本文提出的级联框架上对各模块进行实验分析。分别用两层卷积替换对比，采用 DUTS-TR[19] 数据集作为训练集，DUTS-TE[19]数据集作为测试集，结果如表 8 所示，最佳结果用红色突出显示。实验结果说明，各个模块对模型精度都有一定的提升，缺失任何模块都会对模型精度造成影响。

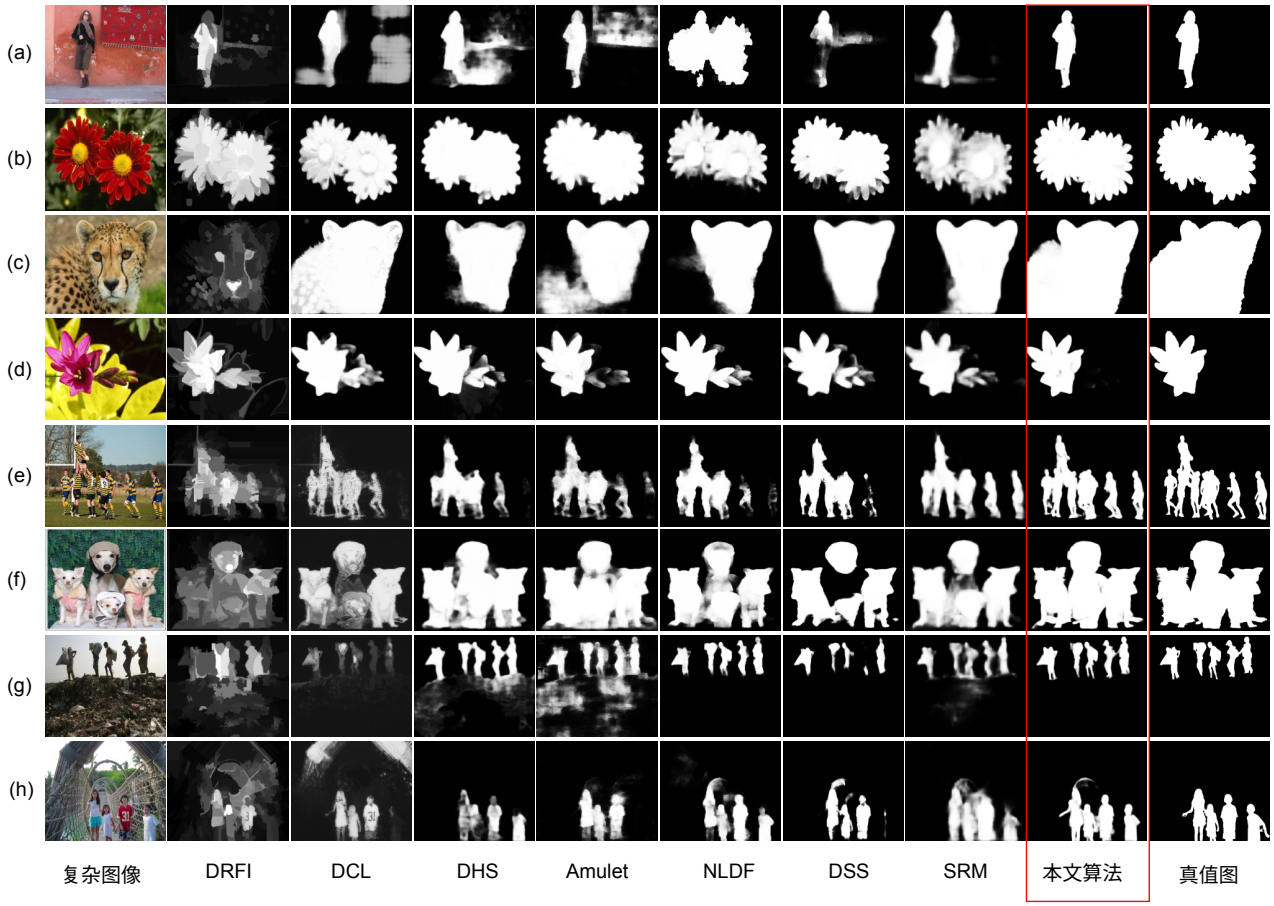


图 5 本文算法与其他模型定性比较结果

Fig. 5 Qualitative comparison between the proposed algorithm and other models

表 7 时间复杂度对比(ECSSD[15]-MAE)

Table 7 Time complexity comparison (ECSSD[15]-MAE)

Method	DHS[29]	DSS[28]	NLDF[32]	Amulet[30]	SRM[33]	Proposed
Input size	224×224	400×300	400×300	256×256	353×353	350×350
MAE	0.059	0.052	0.0.63	0.059	0.054	0.040
FPS	23	12	12	16	14	20

表 8 基于 DUTS-TE[19]数据集的有效性分析

Table 8 Validity analysis based on DUTS-TE[19] dataset

混合信息流	注意力机制	非线性加权融合	MAE
✓			0.050
	✓		0.126
✓	✓		0.047
✓	✓	✓	**0.043**

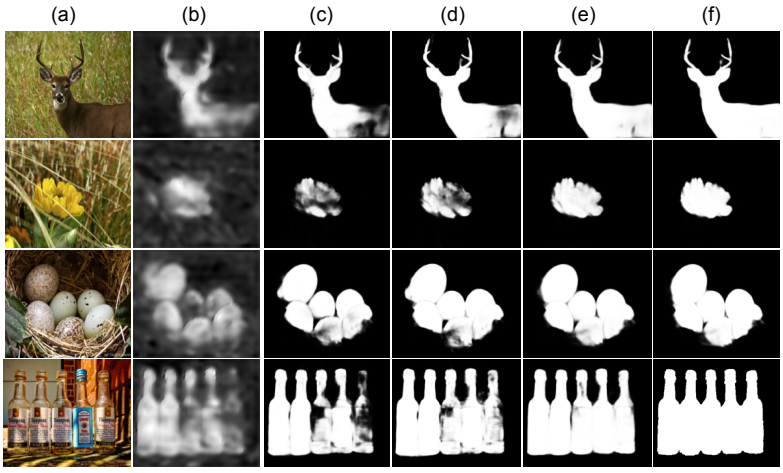


图 6 (a) 原图; (b) FCNs 网络; (c) FCNs+级联方式;
(d) 加入混合信息流机制后的效果; (e) 引入注意力机制的效果; (f) 非线性融合

Fig. 6 (a) Original image; (b) FCNs network; (c) FCNs+ cascade mode;
(d) Performance after adding hybrid information flow mechanism;
(e) Performance of introducing attention mechanism; (f) Nonlinear fusion

由图 6 中的显著性映射可以看出，在一些复杂的场景中，仅使用骨干网络很难定位出显著的目标，结果较为模糊。引入级联式框架后，得到的显著图的质量有较大的改善。结合混合信息流以及注意力机制，显著的区域可以被精确地分割，从而说明了本文模型的有效性。

5 结 论

本文提出的子网络级联式混合信息流框架，有效解决了复杂场景下的显著性检测方法存在多尺度融合问题。该方法结合了级联式的优势，提出的混合信息流机制可充分提取各层次特征，并利用注意力机制增强特征提取。最后通过非线性方式融合不同尺度的特征信息，取得了很好的效果。在 5 个广泛使用的基准数据集上进行定量分析，5 个评估指标均达到最优，同时在一个全新的 SOC 数据集上测试结果达到最佳，有效地验证了本文模型的准确性与鲁棒性。通过可视化定性分析，同时对不同指标结果进行关联分析和运行速度分析，验证了所提模型的性能都有较大提升。本文还对级联式框架、混合信息流机制、注意力机制进行了有效性验证，进一步说明了所提方法的良好效果。

参考文献

[1] Zhang X D, Wang H, Jiang M S, *et al*. Applications of saliency analysis in focus image fusion[J]. *Opto-Electronic Engineering*, 2017, **44**(4): 435–441.
张学典, 汪泓, 江旻珊, 等. 显著性分析在对焦图像融合方面的应用[J]. 光电工程, 2017, **44**(4): 435–441.

[2] Piao Y R, Rong Z K, Zhang M, *et al*. Deep light-field-driven saliency detection from a single view[C]//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019: 904–911.

[3] Zhao J X, Cao Y, Fan D P, *et al*. Contrast prior and fluid pyramid integration for RGBD salient object detection[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 3927–3936.

[4] Zeng Y, Zhang P P, Lin Z, *et al*. Towards high-resolution salient object detection[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 7233–7242.

[5] Fan D P, Wang W G, Cheng M M, *et al*. Shifting more attention to video salient object detection[C]//*Proceedings of 2019*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8554–8564.

[6] Shen C, Huang X, Zhao Q. Predicting eye fixations on webpage with an ensemble of early features and high-level representations from deep network[J]. *IEEE Trans. Multimedia*, 2015, **17**(11): 2084–2093.

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

[8] Perazzi F, Krähenbühl P, Pritch Y, *et al*. Saliency filters: Contrast based filtering for salient region detection[C]//*2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 733–740.

[9] Zhao H W, He J S. Saliency detection method fused depth information based on Bayesian framework[J]. *Opto-Electronic Engineering*, 2018, **45**(2): 170341.
赵宏伟, 何劲松. 基于贝叶斯框架融合深度信息的显著性检测[J]. 光电工程, 2018, **45**(2): 170341.

[10] Wei Y C, Wen F, Zhu W J, *et al*. Geodesic saliency using background priors[C]//*Proceedings of the 12th European Conference on Computer Vision*, 2012: 29–42.

[11] Liu J J, Hou Q, Cheng M M, *et al*. A Simple Pooling-Based Design for Real-Time Salient Object Detection[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3917–3926.

[12] Liu N, Han J W, Yang M H. PiCANet: Learning pixel-wise contextual attention for saliency detection[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 3089–3098.

[13] Chen K, Pang J M, Wang J Q, *et al*. Hybrid task cascade for instance segmentation[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 4974–4983.

[14] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431–3440.

[15] Yan Q, Xu L, Shi J P, *et al*. Hierarchical saliency detection[C]//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 1155–1162.

[16] Li G B, Yu Y Z. Visual saliency based on multiscale deep features[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 5455–5463.

[17] Xi X Y, Luo Y K, Wang P, *et al*. Salient object detection based on an efficient end-to-end saliency regression network[J]. *Neurocomputing*, 2019, **323**: 265–276.

[18] Li Y, Hou X D, Koch C, *et al*. The secrets of salient object segmentation[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 280–287.

[19] He X M, Zemel R S, Carreira-Perpinan M A. Multiscale conditional random fields for image labeling[C]//*Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004: II.

[20] Wang L J, Lu H C, Wang Y F, *et al*. Learning to detect salient objects with image-level supervision[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 136–145.

[21] Fan D P, Cheng M M, Liu J J, *et al*. Salient objects in clutter: bringing salient object detection to the foreground[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 186–202.

[22] Cheng M M, Mitra N J, Huang X L, *et al*. Global contrast based salient region detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(3): 569–582.

[23] Cheng M M, Warrell J, Lin W Y, *et al*. Efficient salient region detection with soft image abstraction[C]//*Proceedings of 2013 IEEE International Conference on Computer vision*, 2013: 1529–1536.

[24] Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 248–255.

[25] Fan D P, Cheng M M, Liu Y, *et al*. Structure-measure: a new way to evaluate foreground maps[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 4548–4557.

[26] Fan D P, Gong C, Cao Y, *et al*. Enhanced-alignment measure for binary foreground map evaluation[C]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018: 698–704.

[27] Li G B, Yu Y Z. Deep contrast learning for salient object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 478–487.

[28] Hou Q B, Cheng M M, Hu X W, *et al*. Deeply supervised salient object detection with short connections[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 3203–3212.

[29] Liu N, Han J W. DHSNet: deep hierarchical saliency network for salient object detection[C]//*Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 678–686.

[30] Zhang P P, Wang D, Lu H C, *et al*. Amulet: aggregating multi-level convolutional features for salient object detection[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 202–211.

[31] Hu P, Shuai B, Liu J, *et al*. Deep level sets for salient object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2300–2309.

[32] Luo Z M, Mishra A, Achkar A, *et al*. Non-local deep features for salient object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6609–6617.

[33] Wang T T, Borji A, Zhang L H, *et al*. A stagewise refinement model for detecting salient objects in images[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 4039–4048.

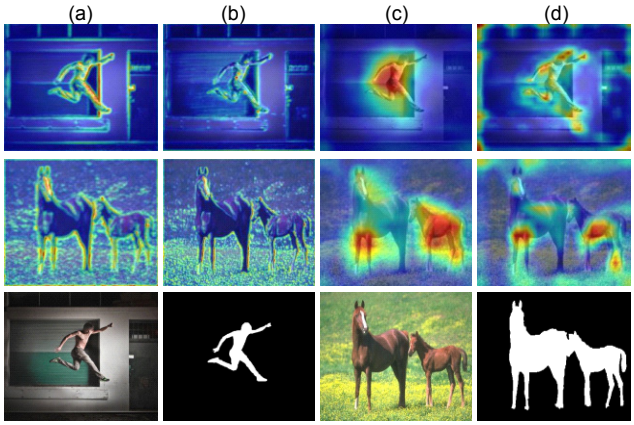
[34] Jiang H D, Wang J D, Yuan Z J, *et al*. Salient object detection: a discriminative regional feature integration approach[C]//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 2083–2090.

[35] Wang L Z, Wang L J, Lu H C, *et al*. Saliency detection with recurrent fully convolutional networks[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 825–841.

Saliency detection hybrid information flows based on sub-network cascading

Dong Bo, Wang Yongxiong*, Zhou Yan, Liu Han, Gao Yuanzhi, Yu Jiamin, Zhang Mengyin

Institute of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China



Visual comparison before and after hybrid information flows

Overview: Saliency detection (SOD) is to detect and segment most important foreground objects that are modeled to accurately locate the mechanism of human visual attention. It has many types, including RGB SOD, light field SOD, RGB-D SOD, and high-resolution SOD. In the video scene, there are object SOD and fixation SOD, while the specific task is broken down into object-level saliency detection and instance-level significance detection. In view of the multi-scale feature fusion problem existing in the complex scenario of the existing saliency object detection algorithms, a fusion method of multi-layer sub-network cascade hybrid information flows is proposed in this paper. First of all, the FCNs backbone network and feature pyramid structure are used to learn multi-scale features. Then, through the multi-layer sub-network layering mining to build a cascading network framework, the context information of the characteristic of each level is fully used. The method of information extraction and flows determines the effect of final feature fusion, so we use the hybrid information flows to integrate multi-scale characteristics and learn more characteristic information with discernment. In order to solve the problem of semantic information fusion, high-level semantic information is used to guide the bottom layer, obtaining more effective context information. In this paper, we adopt the way of channel combination fusion, and the sampling processing is accompanied by the convolution layer smoothing the fusion feature map, making the next fusion more effective. Finally, the effective saliency feature is transmitted as mask information, which realizes the efficient transmission of information flows and further distinguishes the foreground and messy background. Finally, the multi-stage saliency mapping nonlinear weighted fusion is combined to complement the redundant features. Compared with the existing 9 algorithms on the basis of the 6 public datasets, the run speed of the proposed algorithm can reach 20.76 frames and the experimental results are generally optimal on 5 evaluation indicators, even for the challenging new dataset SOC. The proposed method is obviously better than the classic algorithm. Experimental results were improved by 1.96%, 3.53%, 0.94%, and 0.26% for F-measure, weighted F-measure, S-measure, and E-measure, respectively, effectively demonstrating the accuracy and robustness of the proposed model. Through the visual qualitative analysis verification, the correlation analysis and running speed analysis of different indicators are carried out, which further highlights the superior performance of the proposed model. In addition, this paper verifies the effectiveness of each module, which further explains the efficiency of the proposed cascading framework that mixes information flow and attention mechanisms. This model may provide a new way for multi-scale integration, which is conducive to further study.

Citation: Dong B, Wang Y X, Zhou Y, *et al*. Saliency detection hybrid information flows based on sub-network cascading[J]. *Opto-Electronic Engineering*, 2020, **47**(7): 190627

Supported by National Natural Science Foundation of China (61673276)

* E-mail: wyxiong@usst.edu.cn