结合双流特征融合及对抗学习的图像显著性检测

张艺涵[1)], 张朝晖[1,2,3)*], 霍丽娜[1,2,3)], 解滨[1,2,3)], 王秀青[1,2,3)]

[1)] (河北师范大学计算机与网络空间安全学院 石家庄 050024)
[2)] (河北省供应链大数据分析与数据安全工程研究中心 石家庄 050024)
[3)] (河北师范大学河北省网络与信息安全重点实验室 石家庄 050024)
(zhangzhaohui_hbsd@163.com)

摘 要: 为实现图像显著区域或目标的低级特征与语义信息有意义的结合, 以获取结构更完整、边界更清晰的显著性检测结果, 提出一种结合双流特征融合及对抗学习的彩色图像显著性检测(SaTSAL)算法. 首先, 以 VGG-16 和 Res2Net-50 为双流异构主干网络, 实现自底向上、不同级别的特征提取; 之后, 分别针对每个流结构, 将相同级别的特征图送入卷积塔模块, 以增强级内特征图的多尺度信息; 进一步, 采用自顶向下、跨流特征图逐级侧向融合方式生成显著图; 最后, 在条件生成对抗网络的主体框架下, 利用对抗学习提升显著性检测结果与显著目标的结构相似性. 以 P-R 曲线、*F*-measure、平均绝对误差、*S*-measure 为评价指标, 在 ECSSD, PASCAL-S, DUT-OMRON 以及 DUTS-test 4 个公开数据集上与其他 10 种基于深度学习的显著性检测算法的对比实验表明, SaTSAL 算法优于其他大部分算法.

关键词: 显著性检测; 双流特征融合; 对抗学习; 卷积塔; 条件生成对抗网络
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18438

Image Saliency Detection via Two-Stream Feature Fusion and Adversarial Learning

Zhang Yihan[1)], Zhang Zhaohui[1,2,3)*], Huo Lina[1,2,3)], Xie Bin[1,2,3)], and Wang Xiuqing[1,2,3)]

[1)] (*College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang* 050024)
[2)] (*Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics & Data Security, Shijiazhuang* 050024)
[3)] (*Hebei Provincial Key Laboratory of Network & Information Security, Hebei Normal University, Shijiazhuang* 050024)

Abstract: To achieve meaningful combination of low-level features and semantic information of salient regions or targets, and to obtain saliency detection results with more complete structure and clearer boundary, an algorithm of color image saliency detection via two-stream feature fusion and adversarial learning (SaTSAL) is proposed. Firstly, different levels of image features are extracted from bottom to top by means of a two-stream heterogeneous backbone network based on VGG-16 and Res2Net-50. Secondly, in each stream, different feature maps from the same level are fetched into one convolution tower module to enrich intra-level multi-scale information. Thirdly, a predicted saliency map is generated by top-down laterally fusing of cross-stream feature maps level by level, so as to effectively make full use of high-level semantic features and low-level image features. Finally, under the mainframe of conditional generative adversarial networks (CGAN), a higher structural similarity between detected results and salient objects can be strength-

收稿日期: 2020-06-18; 修回日期: 2020-10-04. **基金项目**: 国家自然科学基金青年科学基金(61702158); 河北省自然科学基金青年科学基金(F2018205137); 河北省自然科学基金(F2018205102); 河北省教育厅重点基金(ZD2020317); 2020 年河北师范大学研究生创新资助项目(CXZZSS2020070). **张艺涵**(1996—), 女, 硕士研究生, 主要研究方向为彩色图像显著性检测; **张朝晖**(1969—), 女, 博士, 副教授, 硕士生导师, 论文通讯作者, CCF 会员, 主要研究方向为机器学习、图像识别、智能信息处理; **霍丽娜**(1982—), 女, 博士, 硕士生导师, CCF 会员, 主要研究方向为人工智能方法及应用、影像理解与解译; **解滨**(1976—), 男, 博士, 教授, 硕士生导师, CCF 会员, 主要研究方向为机器学习、粒计算与智能数据分析、人工智能的数学基础; **王秀青**(1970—), 女, 博士, 教授, 硕士生导师, 主要研究方向为先进机器人技术、人工智能、故障检测与诊断.

ened by adversarial learning. By taking P-R curve, *F*-measure, mean absolute error and *S*-measure as evaluation indexes, comparative experiments performed on four public datasets including ECSSD, PASCAL-S, DUT-OMRON and DUTS-test show that SaTSAL algorithm is superior to most of other ten saliency detection methods based on deep learning.

Key words: saliency detection; two-stream feature fusion; adversarial learning; convolutional tower; conditional generative adversarial networks

类似于人眼的注意力机制, 显著性检测旨在对输入图像中能引起人们视觉注意的目标或区域进行快速选择. 作为诸多计算机视觉应用的预处理步骤, 显著性检测已广泛用于语义分割、目标检测、图像理解等多种任务[1-3]. 然而, 显著性检测结果受图像结构、对象语义信息和上下文信息等多种因素的影响, 如何实现结构完整、边界清晰的显著性检测仍是一项具有挑战性的任务.

传统的图像显著性检测方法通常需要构造基于图像底层特征的显著性先验规则[4-5], 但是在面对复杂的图像结构和场景时, 基于底层特征和低阶先验的显著性检测方法, 难以捕获高层次的语义信息, 导致检测失败. 近年来, 伴随着深度学习技术的发展, 以卷积神经网络(convolutional neural networks, CNN)为代表的深度学习模型在图像高级语义特征提取方面呈现了传统方法难以比拟的优势, 在包含显著性检测在内的多种视觉任务中得到了成功的应用. 受全卷积网络在语义分割中应用的启发, 一些学者将主流的图像分类模型改造为全卷积模型, 直接用于生成输入图像的显著性预测结果. Zhang 等[6]提出一种基于编码器-解码器架构的显著性检测模型, 其在解码器中引入 R-Dropout 来学习卷积特征的不确定性, 并在解码器中借助混合上采样的平滑方案来抑制显著性预测图(简称显著图)的棋盘效应; Chen 等[7]的方法同时学习注视网络流和语义网络流, 将二者输出融合至同一模块, 以预测图像的显著性; Zhang 等[8]又利用了 CNN 固有的多尺度表示特性提取多层特征, 并结合边界细化与侧向融合生成显著图; Zhang 等[9]采用自顶向下的方式实现图像的高层语义特征和低层细节特征的有效融合, 增强了特征提取路径关于显著性目标的学习能力; 纪超等[10]提出全局上下文与局部精细检测的双分支模型, 实现深度显著特征的计算, 并利用循环结构网络对特征进行位置加权, 通过反复迭代的方式抑制噪声, 从而减少背景信息对显著性检测的影响; 方正等[11]通过主成分分析方法进行 CNN 特征的降维, 以此作为区域特征向量, 并结合多级别超像素分割和随机森林, 构建了用于显著性检测的融合模型; 项圣凯等[12]遵循编码-解码框架, 提出了密集弱注意力模块, 设计了结合全局显著性预测和基于弱注意力边缘优化的两阶段显著性检测模型, 压缩了模型的大小.

生成对抗网络(generative adversarial networks, GAN)是一种面向生成式模型建模的学习架构, 自 2014 年诞生以来, 因其特有的对抗学习机制以及在图像生成方面的出色表现, 正在不断地应用于多种视觉任务的图像生成[13-14]. 为便于有效地控制网络产生的图像类型, 以额外标签为给定条件的条件生成对抗网络(conditional GAN, CGAN)应运而生[15]. 当 CGAN 用于彩色图像显著性检测时, 通常以原始彩色图像为给定条件, 因此它有助于产生与原始彩色图像对应的显著图.

本文在 CGAN 主体学习框架下, 提出了一种结合双流特征融合和对抗学习(saliency detection via two-stream feature fusion and adversarial learning, SaTSAL)的彩色图像显著性检测算法. SaTSAL 以 VGG-16[16]和 Res2Net-50[17]双流异构网络为显著性检测的主干网络, 实现由低级到高级的多级别、细粒度、富尺度的图像特征提取; 针对每个单流结构, 引入了基于卷积塔模块的多通道特征图的处理环节, 进一步丰富了每个特征提取路径内同级特征图的多尺度信息; 采用自顶向下、跨流特征图的逐级交叉侧向融合方式生成显著图, 有效地结合目标显著性的大尺度上下文信息与显著性目标边界的小尺度特征; 基于 CGAN 主体框架下的对抗学习机制, 有助于增强显著性检测结果与显著目标的结构相似性.

1　本文算法及训练

SaTSAL 的网络结构以 CGAN 为主体框架, 由

2 个基本模块生成器网络(generator networks, GNet)和判别器网络(discriminator networks, DNet)组成. 本节首先详述 2 个基本模块的网络结构; 之后, 面向模型的学习定义损失函数; 最后, 详述 SaTSAL 的学习细节.

1.1 GNet

SaTSAL 的 GNet 如图 1 所示. GNet 由 2 个核心部分组成: (1) 基于 VGG-16 和 Res2Net-50 双流异构网络的特征提取模块(2 条特征提取路径分别对应图 1 的上蓝框标记部分和下蓝框标记部分); (2) 自顶向下、跨流特征图逐级侧向融合的显著图生成模块(见图 1 的红框标记部分).

不同图像的显著性目标不仅所处背景各有不同, 而且存在尺寸、局部结构、纹理和形状等差异, 因此显著性检测不仅需要获取反映目标显著性的大尺度上下文信息, 还需要获取用于精确定位显著性目标边界的小尺度特征. 为了实现关于输入图像更为有效的多级别、不同尺度的特征提取, SaTSAL 使用了基于 VGG-16 和 Res2Net-50 双流异构网络的特征提取模块, 图 1 中的每个蓝框标记部分即为一个单流结构的特征提取路径, 2 条路径的特征提取方式各有特色.

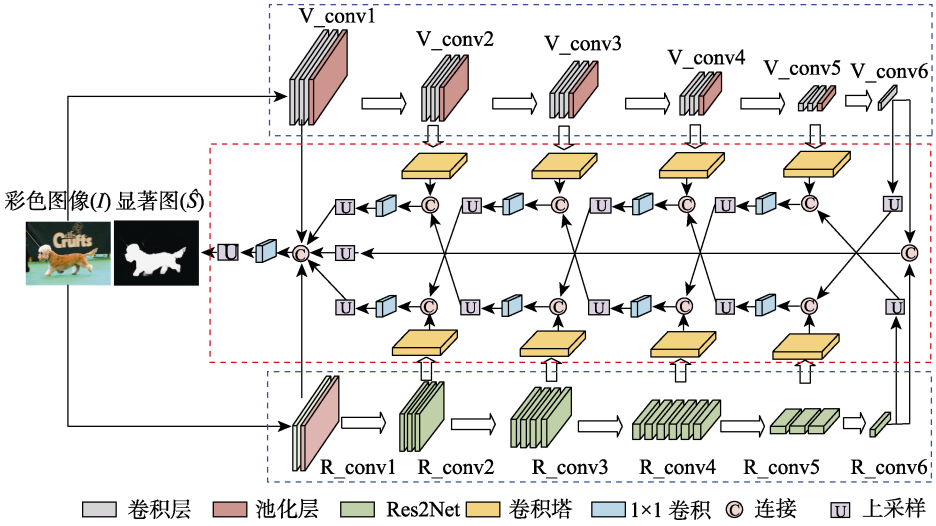


图 1 SaTSAL 网络结构的 GNet

一方面, 以 VGG-16 为代表的 VGG 网络使用固定尺寸的小卷积核, 通过持续增加网络深度而扩展感受野. 由于同一网络层的感受野相同, 因此同一网络层只能提取一种尺度的图像特征, 而更大尺度的图像特征须借助更多网络层的堆叠来获得. 另一方面, 以 Res2Net-50 为代表的 Res2Net 网络在单个残差模块中构造分层的残差类连接, 因而可以在单个网络层形成多种大小不一的感受野范围, 使 Res2Net 可以在同一网络层内以更细粒度的方式表达多尺度特性. 因此, 基于 VGG-16 和 Res2Net-50 的双流结构网络可同时使用各具特色的多级别、多尺度特征提取方式, 有助于获得关于同一彩色输入图像的多级别、富尺度、差异化的图像特征.

为进一步丰富双流结构各级特征图的多尺度信息, 本文提出可实现多尺度分层卷积的卷积塔模块. 图 2 所示为一种结合 4 个不同尺度卷积方式的卷积塔实现样例. 在该卷积塔模块的具体处理中, 使用的卷积核大小分别为 1×1, 2×2, 4×4, 6×6, 每种尺度的卷积运算都将输入的特征图通道数减少到原来的 1/4; 进一步, 将 4 种尺度卷积运算结果进行多通道连接.

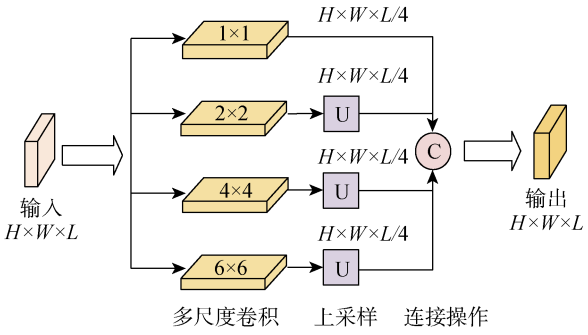


图 2 卷积塔模块

如图 1 所示, 将提出的卷积塔模块用于双流结构特征提取路径中第 2~5 级的特征图处理. 获取来自 VGG-16 路径的 V_conv2~V_conv5 级, 以及位于 Res2Net-50 路径的 R_conv2~R_conv5 级的多通道特征图, 将每个路径的同级特征图进行基于卷积塔模块的多尺度分层卷积, 从而丰富了该级特

征图的多尺度信息.

在上述特征提取的基础上, 构建自顶向下、跨流特征图交叉侧向融合的显著图生成模块(见图 1 的红框标记部分), 以实现不同路径、不同级别的多尺度特征的优势融合. 如图 1 所示, 来自 Res2Net-50 特征提取路径的第 5 级特征图经由卷积塔模块处理之后, 与 VGG-16 中经过双线性插值上采样处理的第 6 级特征图进行跨路径的交叉融合; 而来自 VGG-16 特征提取路径的第 5 级特征图经由卷积塔操作后, 与 Res2Net-50 中经过双线性插值上采样处理的第 6 级特征图进行跨路径的交叉融合; 2 条特征提取路径的其他级别特征图的融合方式类似. 针对双流结构中来自 2 个特征提取路径的特征图, 按照由高级到低级的处理顺序, 以卷积压缩的方式逐级、顺次跨流交叉融合, 完成了 2 条融合路线中的由第 6 级至第 2 级的特征图融合.

考虑输入图像的第 1 级特征为反映图像空间局部细节的底层特征, 故未对该级特征进行卷积塔处理, 而是直接将其用于最终显著图生成时的空间局部细节补充. 此外, 为了强调目标显著性的语义信息, 来自 2 个特征提取路径的第 6 级特征图经过融合后, 也将直接参与最终显著图的生成过程, 以补充高级语义信息.

最终, 将 2 条融合路线的第 6 级至第 2 级的特征图融合结果、双流特征提取路径的第 1 级特征图, 以及来自 2 个特征提取路径的第 6 级特征图融合后的结果, 顺次经过连接、卷积压缩、基于 Sigmoid 激活函数的非线性运算和空间上采样, 生成与原图像大小一致的显著图.

1.2　DNet

DNet 的主要功能是对 GNet 生成的显著图进行判断. 显著图的得分越接近 1, 则认为它越接近真值图. 如图 3 所示, 本文遵从 CGAN 的学习策略, 构造了基于二分类模型的 DNet.

DNet 以训练样本集的每一幅彩色图像为给定

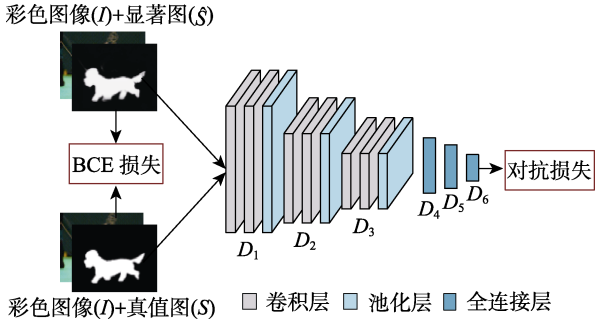


图 3　SaTSAL 网络结构的 DNet

条件, 在模型的学习阶段扮演 2 个角色: 一是对真值图(即“真图”)进行类别的有效建模; 另一个是对 GNet 生成的显著图(即“假图”)进行质量鉴别, 判断是否能够以假乱真, 并将判别损失反馈至 GNet, 以借助对抗学习改善 GNet 关于显著图的预测能力.

1.3　损失函数设计

1.3.1　DNet 的损失函数

DNet 是一种基于 CNN 的二分类模型. 给定彩色输入图像 I 以及对应的二值真值图 S, DNet 以该图像对(I, S)为输入, 在输出端产生单值预测输出 $D(I, S)$, 它是 DNet 将图像 I 的真值图 S 正确预测为真图的概率. DNet 的学习目的在于改善其关于真值图的正确预测能力. 为此, 定义 DNet 对真值图 S 的真图类别进行正确预测的二值交叉熵(binary cross entropy, BCE)损失为

$$L(D)=-[1\cdot\mathrm{lb}\,D(I,S)+0\cdot(1-\mathrm{lb}\,D(I,S))]=-\mathrm{lb}\,D(I,S).$$

$L(D)$取值越小, DNet 关于 S 的真图类别正确预测能力越好.

1.3.2　GNet 的损失函数

SaTSAL 的学习是一种基于 GNet 和 DNet 的博弈式对抗学习过程, 其学习的目的在于使 GNet 较好地获取显著性目标的结构信息, 从而针对彩色输入图像可以生成高质量的显著图.

首先, GNet 生成的显著图与给定的真值图在内容上尽可能一致. 为此, 引入 GNet 的内容损失函数. 给定一幅由 N 个像素组成的彩色输入图像 I 以及真值图 S, 由 GNet 生成同分辨率的显著图 $G(I)$ 记为 $\hat{S}$. 对于任意像素 $j\in\{1,2,\cdots,N\}$, 有 $\hat{S}(j)\in[0,1]$, $S(j)\in\{0,1\}$; $\hat{S}(j)$ 表示 GNet 将图像 I 的像素 j 预测为显著性目标像素的概率. 估计任意像素 j 处 $\hat{S}(j)$ 与 $S(j)$ 之间的 BCE, 将图像中所有像素的 BCE 取均值, 得到 GNet 针对单幅输入图像 I 进行显著性预测的内容损失函数, 记为

$$L_{\mathrm{BCE}}(G)=-\frac{1}{N}\sum_{j=1}^{N}[S(j)\,\mathrm{lb}\,\hat{S}(j)+(1-S(j))\,\mathrm{lb}(1-\hat{S}(j))] \tag{1}$$

其中, $\hat{S}=G(I)$. $L_{\mathrm{BCE}}(G)$ 值越小, 显著图 $\hat{S}$ 与真值图 S 的内容越接近.

此外, GNet 生成的显著图对于 DNet 的欺骗能力应尽可能高, 即 DNet 将该假图错误预测为真图的概率尽可能高. 当 DNet 的输入信号为图像对 $(I,G(I))$ 时, 相应的单值预测输出为 $D(I,G(I))$, 它是 DNet 将图像 I 的显著图 $G(I)$ 错误预测为真图的概率, 该值越大, 表明显著图 $G(I)$ 对于 DNet 的

欺骗程度越强.

为体现由 GNet 关于彩色输入图像 I 的显著图 $G(I)$对 DNet 的欺骗能力, 进一步构造对抗损失函数

$$L_{\mathrm{ADV}}(G)=-\mathrm{lb}\,D(I,G(I)) \tag{2}$$

$L_{\mathrm{ADV}}(G)$值越小, 表明显著图 $G(I)$对 DNet 的欺骗能力越强.

最终, 面向 GNet 的学习, 将式(1)所示的内容损失函数 $L_{\mathrm{BCE}}(G)$ 与式(2)所示的对抗损失函数 $L_{\mathrm{ADV}}(G)$结合, 得到 GNet 的综合损失函数, 即

$$L(G)=\alpha\cdot L_{\mathrm{BCE}}(G)+L_{\mathrm{ADV}}(G).$$

其中, α取经验值 0.05. $L(G)$值越小, 由 GNet 预测的显著图不仅与真值图的内容在像素级上更接近, 而且其对 DNet 的欺骗能力更强, 因而由 GNet 生成的显著图更接近真值图.

1.4 SaTSAL 的学习

不同于一般意义的 GAN, SaTSAL 是一种基于 CGAN 学习框架的算法. 这主要表现为: 一方面, GNet 以彩色图像 I 为输入, 在输出端生成该图像的显著图 $G(I)$; 另一方面, 无论是 DNet 对真值图的真图类别学习, 还是 DNet 对显著图 $G(I)$的"以假乱真"程度的预测, 都以相应的彩色图像 I 为输入条件.

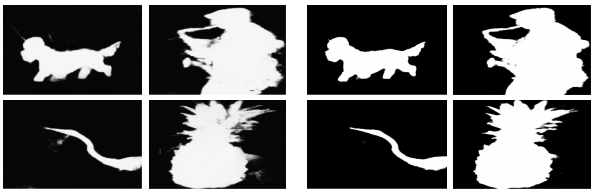
在完成了损失函数的构造之后, 即可进入 SaTSAL 的学习阶段. 首先, 分别基于 $L_{\mathrm{BCE}}(G)$ 和 $L(D)$ 损失函数的优化进行 GNet 和 DNet 的初步学习; 之后, 针对 GNet 的学习, 在内容损失 $L_{\mathrm{BCE}}(G)$ 的基础上引入对抗损失 $L_{\mathrm{ADV}}(G)$, 将 SaTSAL 的学习转化为 $\min\limits_{D}L(D)$ 和 $\min\limits_{G}L(G)$ 交替寻优的对抗学习问题.

本文基于 PyTorch 框架, 采用 PyCharm 2019 完成了 SaTSAL 算法的代码实现. 在学习阶段, 统一将彩色样本图像及其对应的真值图调整至 256 行×256 列. GNet 的输入为 RGB 彩色图像 I, 而 DNet 的输入是由 RGB 彩色图像 I 及其相应的真值图 S (或由 GNet 产生的显著图 $\hat{S}$)构成的 $256\times256\times4$ 的 3 阶张量.

SaTSAL 的学习所使用的训练样本集为 DUTS[18]中的 DUTS-train 数据集, 它内容丰富、场景复杂且具有大样本容量, 由 10553 幅彩色图像及其对应的真值图组成. 基于该训练样本集, 采用 mini-batch 的梯度下降法, 结合 AdaGrad 优化方式, 在 R5-3600X 3.80 GHz CPU, NVIDIA RTX2070Super 显卡的台式机上完成了 SaTSAL 的学习. 设定每个 mini-batch 的样本数目为 4, 经过 1380 轮的参数更新, 得到最终学习结果, 学习过程耗时约 105.4 h. 在完成了 SaTSAL 的学习后, 即可使用 GNet 产生关于彩色输入图像的显著图.

为充分地利用显著目标局部结构的空间一致性, 在 SaTSAL 算法中还引入了基于全连接条件随机场模型(conditional random field, CRF)[19]的显著性检测后处理环节.

图 4 所示为部分样本图像在引入 CRF 后处理环节前后的显著性检测效果对比图. 其中, 图 4a 所示为由 GNet 直接生成的显著图, 图 4b 所示为基于 CRF 后处理的二值显著图.



a. 处理前 b. 处理后

图 4 基于 CRF 后处理环节前后视觉效果比较

由图 4 的视觉比较可知, 基于 CRF 的后处理有助于获得边界更为清晰、目标区域更为致密平滑的显著性检测结果. 下面将结合 CRF 的显著性检测后处理的结果对 SaTSAL 算法进行评价.

2 实验与结果分析

本文以 4 个典型数据集 ECSSD[20], PASCAL-S[21], DUT-OMRON[22]以及 DUTS[18]中的 DUTS-test 为测试集, 对 SaTSAL 算法进行实验测试与结果分析. 其中, ECSSD 数据集包含 1000 幅背景复杂的测试图像及真值图; PASCAL-S 数据集包含 850 幅具有杂乱背景与复杂目标的测试图像及真值图; DUT-OMRON 数据集包含 5168 幅背景更复杂、更具挑战的测试图像及真值图; DUTS-test 数据集包含 5019 幅场景复杂的测试图像及真值图. 首先给出用于算法评价的 4 个指标, 然后是具体实验与性能比较.

2.1 评价指标

下面给出本文用于算法评价的 4 个指标.

(1) P-R(precision-recall)曲线[23]. 首先针对测试集的每一幅样本图像, 将实验得到的显著图 $\hat{S}$ 基于 0~255 的 256 个不同阈值分别进行二值化, 并进行显著图 $\hat{S}$ 的二值化结果与二值真值图 S 的逐像素比较, 按照

$$P=\frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FP}}},\quad R=\frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FN}}}$$

计算得到每个阈值下该样本图像显著性检测结果的精度 P 和召回率 R. 其中, $n_{\mathrm{TP}}, n_{\mathrm{TN}}, n_{\mathrm{FP}}, n_{\mathrm{FN}}$ 分别代表给定的样本图像中显著性检测结果为真阳性、真阴性、假阳性、假阴性的像素数目. 对于 256 个不同的阈值, 分别计算每个阈值下同一样本集内不同样本图像的 P 均值和 R 均值, 得到该样本集的 256 对 P 均值和 R 均值, 绘制得 P-R 曲线. P-R 曲线越靠近坐标系右上方, 表示算法性能越优越.

(2) F-measure[23]. 当显著性检测算法用于给定的测试集时, 还可结合不同阈值下每个测试集的 P 均值和 R 均值, 计算二者的加权调和平均, 得到该阈值下相应测试集的 F_β 值, 即

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

其中, β^2 值越接近 0, 越强调 P 相对于 R 的重要性; $\beta^2=1$时, 二者同等重要. 本文实验中更强调 P, 取经验值为 $\beta^2=0.3$, 并取其中的 F_β 最大值(记为 $F_{\max}$)作为该数据集的评价结果; $F_{\max}$ 越大, 表示算法性能越好.

(3) 平均绝对误差(mean absolute error, MAE)[23]. 将一幅 $M\times N$ 样本图像的显著图 $\hat{S}$ 与真值图 S 进行逐像素的取值比较, 可得到该图像显著性检测的平均绝对误差为

$$E_{\mathrm{MAE}}=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}\left|\hat{S}(x,y)-S(x,y)\right|.$$

取测试集内所有样本图像显著性检测的 E_{MAE} 值的算术均值作为该数据集的 E_{MAE} 值, 以此评价测试集在统计意义上显著性检测结果与真值图的内容一致性. 对于给定的测试集, 其 E_{MAE} 值越接近 0, 意味着显著性检测结果与真值图的内容越接近.

(4) S-measure[23]. 除了上述基于逐像素计算的评价方式外, 本文还针对显著性检测结果引入了基于 S-measure 的结构相似性度量, 即

$$S_{\mathrm{S}}=\alpha\cdot S_{\mathrm{object}}+(1-\alpha)\cdot S_{\mathrm{region}}\qquad(3)$$

其中, S_{object} 和 S_{region} 分别度量显著性检测结果与真值图之间面向目标和面向区域的结构相似性; 控制参数 $\alpha\in[0,1]$, 用于平衡 S_{object} 和 S_{region} 的相对重要性, 通常设置 α=0.5; S_{S} 值最大为 1. 利用式(3)得到测试集内每个样本图像的 S_{S} 值, 最后将测试集内各样本图像的 S_{S} 值进行算术平均运算, 得到该数据集的 S_{S} 值, 以此评价测试集在统计意义上显著性目标区域与真值图的结构一致性. 一个测试集的 S_{S} 值越接近 1, 意味着该测试集的显著性检测结果与真值图的结构一致性越好.

2.2　消融实验

为了验证 SaTSAL 算法中使用多尺度分层卷积的卷积塔模块和跨流交叉特征融合方式的有效性, 本节结合 4 个基准数据集, 分别采用 2 种方法进行了消融实验. 方法 1 是在保留逐级、顺次跨流交叉特征融合的前提下, 考查去掉 SaTSAL 算法中的卷积塔模块进行图像显著性检测; 方法 2 是在保留卷积塔模块的基础上, 考查去掉 SaTSAL 算法中的逐级、顺次跨流交叉特征融合部分, 而在 2 条特征提取路径中采用从高级到低级顺次进行路径内特征结合, 并最终在最低层融合 2 条路径特征的方式进行的图像显著性检测.

表 1 列出了 SaTSAL 算法及其上述 2 种方法关于 4 个数据集显著性检测结果的各评价指标对比. 由表 1 各指标评价数据可知, 同时结合了卷积塔模块并对 2 条路径的各级特征进行顺次跨流交叉融合的 SaTSAL 算法总体表现最好.

2.3　性能比较

本节结合 4 个基准数据集, 将 SaTSAL 算法与 MWS[24], RSDNet[25], PAGRN[9], C2S-Net[26], SBF[27], NLDF[28], UCF[6], Amulet[8], DCL[29]和 MDF[30]这 10 种性能优良的显著性检测算法进行比较, 以验证 SaTSAL 算法的有效性.

2.3.1　显著性检测的视觉效果比较

为便于视觉比较, 首先从上述测试集内选择了 12 幅代表性的样本图像. 在表 2 给出了 11 种算法显著性检测结果的对比.

表 1　3 种方法关于 4 个数据集的 $F_{\max}$, S_{S} 和 E_{MAE} 值

方法	ECSSD			PASCAL-S			DUT-OMRON			DUTS-test		
	$F_{\max}$	S_{S}	E_{MAE}	$F_{\max}$	S_{S}	E_{MAE}	$F_{\max}$	S_{S}	E_{MAE}	$F_{\max}$	S_{S}	E_{MAE}
SaTSAL	**0.921**	**0.904**	**0.038**	**0.831**	**0.821**	**0.079**	**0.744**	**0.802**	**0.064**	**0.839**	0.857	**0.042**
方法 1	0.909	0.896	0.047	0.825	0.820	0.085	0.735	**0.802**	0.069	**0.839**	**0.864**	0.045
方法 2	0.904	0.891	0.055	0.826	**0.821**	0.088	0.724	0.793	0.077	**0.839**	0.858	0.050

注. 粗体表示评价结果最好的值.

表 2　11 种显著性检测算法的显著性预测图视觉比较

样本	彩色图像	真值图	算法										
			SaTSAL	MWS[24]	RSDNet[25]	PAGRN[9]	C2S-Net[26]	SBF[27]	NLDF[28]	UCF[6]	Amulet[8]	DCL[29]	MDF[30]
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													

由表 2 的视觉比较可知，SaTSAL 算法在不同场景下的显著性检测结果更接近真值图，在视觉上取得了综合检测效果最好的结果，具体体现为：(1) SaTSAL 算法能够针对多种类型目标的显著性检测，产生边界更为清晰(即使输入图像中只含小目标，如样本 3, 5)、结构更为完整(如样本 1, 2, 6, 8, 10, 11, 12)、目标区域更为干净、平滑的显著图；(2) 与其他算法相比，即使场景复杂(如样本 4, 9)，甚至包含视觉上难以发现的小目标(如样本 3, 5)，或者前景目标与背景相似(如样本 7)，本文算法仍能有效检测.

2.3.2　显著性检测的客观评价

图 5 所示为 11 种算法关于 4 个数据集的 P-R 曲线. 由图 5 发现，SaTSAL 算法关于 ECSSD, DUT- OMRON, DUTS-test 这 3 个数据集的 P-R 曲线更靠近坐标系的右上角，是 11 种算法中显著性检测综合性能最好的；而在 PASCAL-S 数据集的显著性检测中，RSDNet 算法的 P-R 曲线最靠近坐标系右上角，其显著性检测的综合表现更突出.

表 3 列出了 11 种算法关于 4 个数据集显著性检测结果的评价指标对比，并且针对每个数据集，分别以红(最佳)、蓝(次佳)、绿(第 3)3 种颜色标记

图 5　11 种算法在 4 个数据集的 P-R 曲线

表 3　11 种算法在 4 个数据集的 F_{max}, S_S 和 E_{MAE} 值

算法	ECSSD			PASCAL-S			DUT-OMRON			DUTS-test		
	F_{max}	S_S	E_{MAE}	F_{max}	S_S	E_{MAE}	F_{max}	S_S	E_{MAE}	F_{max}	S_S	E_{MAE}
SaTSAL	0.921	0.904	0.038	0.831	0.821	0.079	0.744	0.802	0.064	0.839	0.857	0.042
MWS[24]	0.859	0.827	0.099	0.761	0.766	0.137	0.677	0.756	0.108	0.720	0.759	0.092
RSDNet[25]	0.880	0.788	0.173	0.849	0.803	0.160	0.715	0.644	0.178	0.798	0.720	0.161
PAGRN[9]	0.904	0.889	0.061	0.814	0.822	0.089	0.707	0.775	0.071	0.817	0.838	0.056
C2S-Net[26]	0.902	0.896	0.053	0.827	0.839	0.046	0.722	0.799	0.072	0.784	0.831	0.062
SBF[27]	0.833	0.832	0.091	0.726	0.758	0.133	0.649	0.748	0.110	0.657	0.743	0.109
NLDF[28]	0.889	0.875	0.063	0.795	0.805	0.098	0.699	0.770	0.080	0.777	0.816	0.065
UCF[6]	0.890	0.883	0.069	0.787	0.805	0.115	0.698	0.760	0.120	0.742	0.782	0.112
Amulet[8]	0.905	0.894	0.059	0.805	0.818	0.100	0.715	0.780	0.098	0.750	0.804	0.085
DCL[29]	0.882	0.868	0.075	0.787	0.796	0.113	0.699	0.771	0.086	0.742	0.796	0.149
MDF[30]	0.797	0.776	0.105	0.704	0.696	0.142	0.643	0.721	0.092	0.657	0.728	0.114

了评价结果最好的前 3 种算法. 由表 3 中各指标的评价结果可以观察到, SaTSAL 算法在 ECSSD, DUT-OMRON 和 DUTS-test 这 3 个数据集上的 F_{max}, S_S 以及 E_{MAE} 值均取得了最佳结果; SaTSAL 算法在 PASCAL-S 数据集上的 F_{max} 和 E_{MAE} 值表现次佳, S_S 值则排名第 3. 总体来说, 结合表 3 评价指标, 相对于其他 10 种算法, 本文 SaTSAL 算法在 4 个数据集上的综合表现最好.

2.4　显著性检测的时间性能

为考查算法的时间性能, 本文基于 4 个数据集, 对 SaTSAL 算法进行了显著性检测的运行时间测试, 实验结果表明, 分辨率为 256 像素×256 像素的单幅彩色图像显著性检测的平均耗时为 0.076 s; 而要完成原始分辨率下 4 个数据集共计 12 037 幅彩色图像的显著性检测, 总体耗时约 1 586.4 s. 这表明本文算法可以满足彩色图像显著性检测实时性的需要.

2.5　显著性检测的失败案例

尽管 SaTSAL 算法在大多数情况下表现突出, 但也存在显著性检测失败的情况. 图 6 给出几个典型的失败案例, 失败情况主要体现为: (1) 因输入图像中存在镜面目标或者图像相邻区域之间的高对比度, 导致显著性检测结果存在误报情况(如样本 1, 2 的显著图内以蓝框标记的误报区域); (2) 因人工标注的局限性, 导致显著性检测结果存在争议或误报(如样本 5 的显著图内以蓝框标记的争议区域), 以及显著性检测结果与真值图存在结构上的差异(如样本 2 的显著图内黄框标记部分, 以及样本 5 的显著图内黄框标记部分); (3) 因图像前景背景高度相似、语义结构过于复杂, 导致显著性检测结果结构不完整(如样本 4 的显著图内黄框标记部分)、漏检(如样本 4 的真值图内红框标记部分、样本 3 的真值图内红框标记的小目标均为漏检区域), 以及误报(如样本 4 的显著图内蓝框标记部分为误报区域).

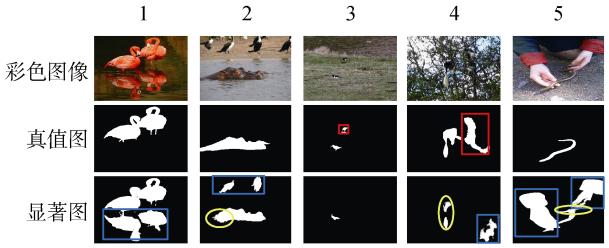


图 6　SaTSAL 算法显著性检测失败示例

3　结　语

本文提出了 SaTSAL 彩色图像显著性检测算法. 在对抗学习框架下, 将双流异构主干网络的特征提取与自顶向下、跨流特征图交叉融合方式有机结合, 以实现具有复杂背景及语义结构的图像显著性检测. 本文算法与其他 10 种显著性检测算法在 4 个公开数据集的比较实验表明, 本文算法可生成边界更清晰且结构更完整的显著图. 在运行效率上, 本文算法可以满足实时进行显著性检测的要求. 但是, 当图像中存在镜面目标、前景与背景高度相似等情况时, 本文算法仍有检测失败的可能.

生成器内多级别特征的结合方式的设计、多尺度边缘信息的结合, 以及损失函数的设计是后续显著性检测研究进一步努力的方向.

参考文献(References):

[1] Wang X, You S D, Li X, *et al*. Weakly-supervised semantic

segmentation by iteratively mining common object features[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 1354-1362

[2] Zhang D W, Meng D Y, Zhao L, *et al*. Bridging saliency detection to weakly supervised object detection based on self-paced curriculum learning[C] //Proceedings of the 25th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2016: 3538-3544

[3] Zhang F, Du B, Zhang L P. Saliency-guided unsupervised feature learning for scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(4): 2175-2184

[4] 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

[5] Achanta R, Hemami S, Estrada F, *et al*. Frequency-tuned salient region detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2009: 1597-1604

[6] Zhang P P, Wang D, Lu H C, *et al*. Learning uncertain convolutional features for accurate saliency detection[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 212-221

[7] Chen X W, Zheng A L, Li J, *et al*. Look, perceive and segment: finding the salient objects in images via two-stream fixation-semantic CNNs[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1050-1058

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

[9] Zhang X N, Wang T T, Qi J Q, *et al*. Progressive attention guided recurrent network for salient object detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 714-722

[10] Ji Chao, Huang Xinbo, Cao Wen, *et al*. Fusion of deep learning and global-local features of the image salient region calculation[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(10): 1838-1846(in Chinese)
(纪超, 黄新波, 曹雯, 等. 结合深度学习和全局-局部特征的图像显著区域计算[J]. 计算机辅助设计与图形学学报, 2019, 31(10): 1838-1846)

[11] Fang Zheng, Cao Tieyong, Zheng Yunfei, *et al*. Extraction of refined deep feature and its application in saliency detection[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(2): 324-331(in Chinese)
(方正, 曹铁勇, 郑云飞, 等. 高效深度特征提取及其在显著性检测中的应用[J]. 计算机辅助设计与图形学学报, 2019, 31(2): 324-331)

[12] Xiang Shengkai, Cao Tieyong, Fang Zheng, *et al*. Dense weak attention model for salient object detection[J]. Journal of Image and Graphics, 2020, 25(1): 136-147(in Chinese)
(项圣凯, 曹铁勇, 方正, 等. 使用密集弱注意力机制的图像显著性检测[J]. 中国图象图形学报, 2020, 25(1): 136-147)

[13] Reed S, Akata Z, Yan X C, *et al*. Generative adversarial text to image synthesis[C] //Proceedings of the 33rd International Conference on Machine Learning. New York: ACM Press, 2016: 1060-1069

[14] Zhu J Y, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2242-2251

[15] Mirza M, Osindero S. Conditional generative adversarial nets[OL]. [2020-06-18]. https://arxiv.org/abs/1411.1784

[16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2020-06-18]. https://arxiv.org/abs/1409.1556

[17] Gao S H, Cheng M M, Zhao K, *et al*. Res2Net: a new multi-scale backbone architecture[OL]. [2020-06-18]. https://arxiv.org/abs/1904.01169

[18] Wang L J, Lu H C, Wang Y F, *et al*. Learning to detect salient objects with image-level supervision[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 3796-3805

[19] Krahenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[OL]. [2020-06-18]. https://arxiv.org/abs/1210.5644

[20] Yan Q, Xu L, Shi J P, *et al*. Hierarchical saliency detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 1155-1162

[21] Li Y, Hou X D, Koch C, *et al*. The secrets of salient object segmentation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 280-287

[22] Yang C, Zhang L H, Lu H C, *et al*. Saliency detection via graph-based manifold ranking[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 3166-3173

[23] Wang W G, Lai Q X, Fu H Z, *et al*. Salient object detection in the deep learning era: an in-depth survey[OL]. [2020-06-18]. https://arxiv.org/abs/1904.09146

[24] Zeng Y, Zhuge Y Z, Lu H C, *et al*. Multi-source weak supervision for saliency detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 6067-6076

[25] Islam M A, Kalash M, Bruce N D B. Revisiting salient object detection: simultaneous detection, ranking, and subitizing of multiple salient objects[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 7142-7150

[26] Li X, Yang F, Cheng H, *et al*. Contour knowledge transfer for salient object detection[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2018: 370-385

[27] Zhang D W, Han J W, Zhang Y. Supervision by fusion: towards unsupervised learning of deep salient object detector[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 4068-4076

[28] Luo Z M, Mishra A, Achkar A, *et al*. Non-local deep features for salient object detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 6593-6601

[29] Li G B, Yu Y Z. Deep contrast learning for salient object detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 478-487

[30] Li G B, Yu Y Z. Visual saliency based on multiscale deep features[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 5455-5463