

多源图像光照及边缘融合的显著性检测

宁钰豪¹, 刘昱¹, 汪少初^{1,2*}

¹天津大学微电子学院, 天津 300072;

²天津市测绘院有限公司, 天津 300381

摘要 针对 RGB 图像在光照条件较差的情况下, 难以检测其显著性目标区域, 热红外(T)图像具有红外热辐射, 难以检测其显著性目标边界的问题, 提出一种基于 RGB、T 图像信息融合互补的显著性检测算法。首先, 基于 IHS 颜色空间建立图像光照条件, 判断 RGB、T 图像的光照情况; 并提出 RGB-T 图像自适应光照融合算法, 该算法根据图像的光照情况引导 RGB-T 图像的融合, 以生成多级 RGB-T 光照融合图像, 提高显著性目标的区域检测能力。其次, 使用不同卷积核与标准差的高斯滤波器提取 RGB、T 图像中物体的边缘高频信息, 以生成不同层级的 RGB、T 高频细节图像; 并使用基于编解码器的深度学习网络, 对 RGB、T 高频细节图像进行融合, 以生成不同层级的 RGB-T 细节融合图像, 提高显著性目标的边界检测能力。最后, 根据图像光照信息进行多级 RGB-T 光照融合图像与 RGB-T 细节融合图像的融合; 并使用基于学习的算法进行显著性目标检测。实验结果表明, 所提算法在提高显著性目标检测精度的同时, 优化了显著性目标边界的检测, 与 EGNNet、PoolNet、CPDNet、DMRA 和 A2dele 等优秀的显著性检测算法在 F_{\max} 值、 F_{ave} 值和平均绝对误差等指标上相比, 具有一定的竞争力。

关键词 机器视觉; 显著性目标检测; 图像融合; 边缘检测网络; 卷积神经网络

中图分类号 TP391.41

文献标志码 A

doi: 10.3788/LOP202158.1615005

Salient Detection of Multisource Image Illumination and Edge Fusion

Ning Yuhao¹, Liu Yu¹, Wang Shaochu^{1,2*}

¹School of Microelectronics, Tianjin University, Tianjin 300072, China;

²Tianjin Institute of Surveying and Mapping Co., Ltd., Tianjin 300381, China

Abstract To resolve the problems of the difficulties in detecting the salient object area under poor lighting conditions in an RGB image and the salient object boundary because of the infrared thermal radiation in a thermal infrared (T) image, a salient detection algorithm based on the complementary fusion of RGB and T image information is proposed. First, image-lighting conditions are established based on the IHS color space to determine the lighting conditions of RGB and T images. An RGB-T image adaptive lighting fusion algorithm is proposed to guide the fusion of RGB-T images according to the lighting conditions of the images, generating a multilevel RGB-T lighting fusion image to improve the ability of salient object area detection. Second, the Gaussian filters with different convolution kernels and standard deviations are used to extract the high-frequency information of the objects' edges in the RGB and T images, generating different levels of RGB and T high-frequency detailed images. The deep learning network based on encoder-decoder is used to fuse RGB and T high-frequency detail images to generate RGB-T detail fusion images with different levels, which improves the boundary detection ability of salient objects. Finally, a multilevel fusion of RGB-T lighting fusion image and RGB-T detail fusion image is performed according to the image-lighting information, and the algorithm based on learning is used for salient object detection. Experimental results show that the proposed algorithm improves the detection accuracy of the salient area and boundary detections. The proposed algorithm is competitive compared with EGNNet, PoolNet, CPDNet, DMRA, and A2dele, which are excellent salient object detection algorithms in F_{\max} value, F_{ave} value, and average absolute error.

Key words machine vision; salient object detection; image fusion; edge detection network; convolution neural

收稿日期: 2020-10-26; 修回日期: 2020-11-26; 录用日期: 2020-12-14

基金项目: 天津市科技计划项目重大专项(18ZXZRHSHY00190)

通信作者: *wangshaochu@tju.edu.cn

network

OCIS codes 150.0155; 330.7326; 110.2970

1 引言

显著性目标检测任务旨在检测图像中的突出目标,准确分割输入图像中显著性物体的像素。显著性目标检测算法可以分为传统方法^[1-2]和基于学习的方法^[3-7]。传统方法大多依靠手工制作的特征或同时捕获的局部细节和全局信息来完成图像中突出目标的检测。缺少高级语义信息限制了传统方法在复杂场景中的应用。受益于卷积神经网络强大的特征提取能力,基于学习的显著性目标检测算法的精度逐渐超过了传统算法的精度^[2]。

显著性目标检测任务包括对显著性目标的区域检测和边界检测。在光照条件较差、显著性物体与背景相似等复杂环境中,使用 RGB 图像进行显著性检测时,难以检测显著性目标区域。由于缺乏物体纹理信息,使用热红外(T)图像进行显著性检测时,难以检测显著性目标边界^[8]。融合互补 RGB 图像信息与 T 图像信息的算法可以在提高显著性目标边界检测能力的同时提高显著性目标区域检测能力,已经被证明是一种有效的显著性检测方法。文献[8]提出了一种基于多任务流形排序的 RGB-T 图像显著性检测算法。文献[9]构建了一种新的协同图算法。该算法将超像素作为图节点,在一个统一的优化框架中协同使用深层次特征,提高了 RGB-T 显著性检测算法的精度。但是由于缺少融合的 RGB、T 图像高级语义信息,在复杂场景中,上

述算法的显著性目标检测精度较低。

不同于目标检测、行人追踪等任务旨在对目标的区域范围进行检测,图像语义分割、显著性目标检测等任务对物体边界的检测精度要求较高,使用基于学习的算法提取 RGB-T 图像特征进行融合,并进行显著性目标检测,仍然是计算机视觉领域相对较新的任务^[9]。

近些年来许多信号处理的方法被应用于图像融合领域来提取图像特征,例如小波变换^[10]、轮廓波变换^[11]等。文献[12]首先在红外与可见光融合的任务中提出了一种通过均值滤波器将图像分解为基础层与细节层,之后通过视觉显著性获得的权重图引导基础层与细节层融合的方法。文献[13]通过均值滤波器将图像分解为基础层与细节层,通过等比例加权的方式对 RGB-T 基础层进行融合,通过预训练的 VGG-19 网络对 RGB-T 细节层进行融合,最后叠加基础层与细节层获得融合图像。由于缺少对图像光照条件的考虑,上述算法生成的 RGB-T 图像的显著性检测精度较低。

针对算法存在的问题,本文提出了一种多源图像光照及边缘融合的显著性检测算法。所提算法在提高显著性目标区域检测能力的同时增强了显著性目标边界检测能力。

2 基本原理

所提算法的整体框架如图 1 所示,其中 I_L 为 RGB-T 光照图像, I_D 为 RGB-T 细节图像, I_F 为最

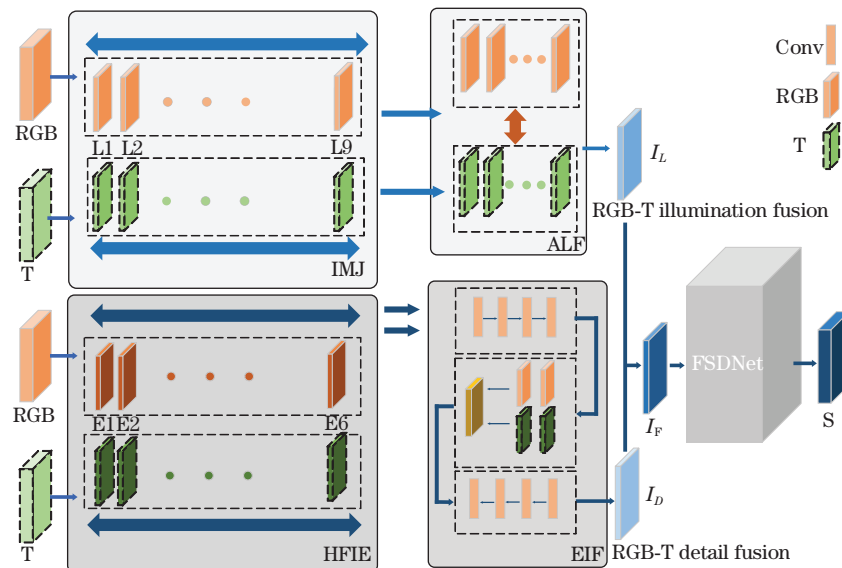


图 1 所提算法的整体框架

Fig.1 Architecture of proposed algorithm

终的 RGB-T 融合图像, L 与 D 分别表示 RGB-T 光照图像和 RGB-T 细节图像的等级, $L=9, D=6$ 。首先, 将 RGB、T 图像分别输入 RGB-T 图像光照信息判别(IMJ)算法, 生成 RGB、T 图像的光照评分, 根据评分确定 RGB、T 图像的光照等级; 并通过 RGB-T 自适应光照融合(ALF)算法, 根据 RGB、T 图像的光照评分和等级确定 RGB-T 图像的融合比例, 生成 RGB-T 光照融合图像。其次, 通过图像高频信息提取(HFIE)算法提取不同层级的显著性物体边缘信息; 并利用细节信息融合(EIF)网络生成 RGB-T 细节融合图像。最后, 根据 RGB、T 图像的光照等级, 对同等级的 RGB-T 光照融合图像与 RGB-T 细节融合图像进行融合, 生成最终的 RGB-T 融合图像; 并将其输入到显著性检测模块(FSDNet), 从而最终预测生成显著图。

2.1 RGB-T 图像光照信息检测模块

IHS 色彩表征系统(I: 明度, H: 色调, S: 饱和度)是计算机定量处理色彩时常用的 RGB 表色系统, 可以有效地将图像分为明度、色调和饱和度信息。IHS 色彩表征系统对图片的颜色信息与亮度信息进行了很好的区别, 更容易识别与量化 RGB 图像的亮度信息^[14]。利用 IHS 空间坐标系中图像亮度的计算方式^[14]来表征图片的光照情况, 图像亮度的计算公式为

$$I = \frac{R + G + B}{3}, \quad R, G, B \in [0, 255], \quad (1)$$

$$I' = \frac{\sum_{i=0}^m \sum_{j=0}^n I}{m \times n}, \quad (2)$$

式中: R, G, B 分别表示 RGB 图像中某像素点在 R, G, B 空间的像素值; I 表示 RGB 图像中某像素点的亮度值; I' 表示图像的整体光照条件通过归一化计算后的光照评分; m, n 分别表示 RGB 图像的长度、宽度。通过 IMJ 模块将 VT1000 数据集^[8]中的 RGB、T 图像分为 9 个不同的光照等级(L1~L9), 具体如表 1 所示, 其中 U 和 Q 分别为光照评分的上界和下界。由于很暗或者曝光严重的图像较少, 在 RGB、T 图像依据光照评分等比例分级的基础上, 稍微提高了 L1 级和 L9 级所包含的评分范围。

表 1 RGB、T 图像光照分级

Table 1 RGB, T image illumination classification

Condition	L1	L2	L3	L4	L5	L6	L7	L8	L9
U	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	1
Q	0	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85

2.2 RGB-T 光照融合模块

ALF 模块使用自适应加权的方法引导 RGB 图像与 T 图像融合。当 RGB 图像的光照评分较低时, 提高 T 图像的融合比例; 反之亦然。提出的 RGB-T 光照融合算法的核心在于通过提高融合图像的光照评分, 来提高 RGB-T 融合图像的显著性目标区域检测能力。RGB、T 图像融合比例的计算方程为

$$\alpha \times \epsilon T + \beta \times \lambda R = (1 + \eta)R, \quad (3)$$

$$\alpha = \frac{\eta R}{2T - R}, \quad \alpha, \beta \in (0, 1), \alpha + \beta = 1, \quad (4)$$

$$\beta' = R / (R + T), \quad (5)$$

式中: T 与 R 分别表示 T 图像与 RGB 图像的光照评分; α 与 β 分别表示 T 图像与 RGB 图像融合时的比例系数; ϵ 与 λ 分别代表 T 图像与 RGB 图像光照评分的权重; η 表示融合图像的光照评分相比 RGB 图像的光照评分预计提高的百分数; β' 表示 RGB 图像与 T 图像光照评分的加权值; $(1 + \eta)R$ 表示 RGB-T 光照融合图像的光照评分。

在相同的光照条件下(特别在黑暗环境下), 相比 RGB 图像, 使用 T 图像可以更好地检测显著性目标区域, 直接采用 β' 作为 RGB-T 光照融合图像中 RGB 图像所占比例并不是最好的选择, 因此对 RGB 图像和 T 图像的光照评分设置了不同的权重系数来区分它们。首先, 通过定性观察, 当 T 图像的光照评分是 RGB 图像的一半时, 仍可以很好地区分两者的显著性目标; 另外, 通过改变图像的光照系数(ϵ, λ), 进行多次定量的实验后, 发现将 ϵ 与 λ 设置为 $\epsilon=2, \lambda=1$ 时, 模型具有不错的性能。为了避免出现过拟合的问题, 将 $\epsilon=2, \lambda=1$ 作为模型的最终参数。(4)式由 $\epsilon=2, \lambda=1$ 与 $\alpha+\beta=1$ 代入(3)式所得, 由(4)式可知: η 值可以直接影响 T 图像的融合比例(α), 当 RGB、T 图像的光照质量近似相同($T=R$)时, η 等于 α ; 当参数 η 提高时, α 增大, 融合图像中 T 图像的比例增大。因此在图像光照质量较好时, 为了避免引入过多的物体红外热辐射信息, 需要降低 η 值, 减少 T 图像所占比例; 反之亦然。

使用 ALF 模块, 可以通过调整 η 值的大小调整不同光照等级下 RGB、T 图像的融合比例, 获得光照增强程度不同的 RGB-T 融合图像。表 2 是根据 RGB 图像的光照等级生成融合图像时对应的 η 值。此外, 针对(4)式中 $\epsilon T - R < 0$ 使得 $\alpha < 0$ 的情况, 使用(5)式中的 β' 值作为 RGB-T 光照融合图像中

RGB 图像所占比例。

表 2 RGB、T 图像光照融合比例

Table 2 RGB, T image light fusion ratio

Parameter	L1	L2	L3	L4	L5	L6	L7	L8	L9
η	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

2.3 图像细节信息提取模块

不同于文献[13]使用均值滤波器将图像分解为基础层和细节层,为了降低图像噪声,HFIE 模块选用高斯卷积滤波器提取图像边缘细节信息,并且对不同光照等级的图像设置了不同的高斯卷积滤波器来提取图像细节信息。由 ALF 模块可知,在生成 RGB-T 光照融合图像时,光照条件越差,T 图像所占的比例越高,导致融合图像中的细节信息减少。因此在图像光照条件较差时,HFIE 模块采用卷积核与方差更大的高斯卷积滤波器来提取图像细节层的信息,用以突出图像中物体的边缘细节。卷积核大小为 $(2K+1)(2K+1)$ 时卷积核中各个元素值为

$$H_{(i,j)} = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(i-K-1)^2 + (j-K-1)^2}{2\sigma^2}\right], \quad (6)$$

式中: σ 表示高斯卷积核的标准差; (i,j) 表示高斯

卷积核中像素点的坐标。经过高斯卷积滤波之后的新图像中某点像素值(Q)为

$$Q_{(x,y)} = \sum_{i=1}^{2K+1} \sum_{j=1}^{2K+1} P_{(x+i,y+j)} H_{(x+i,y+j)}, \quad (7)$$

式中: P 表示原图像中某点的像素值; (x,y) 表示新图像的像素坐标。为了提取绝大部分的图像信息,需要将高斯卷积核的大小设置为标准差的 3 倍。使用不同卷积核与方差大小的高斯卷积滤波器生成不同等级的 RGB、T 细节图像。表 3 中,R1 到 R6 代表不同等级高斯滤波器中卷积核与标准差的大小。

表 3 RGB、T 图像细节信息的提取

Table 3 Detail information extraction of RGB、T image

Parameter	R1	R2	R3	R4	R5	R6
σ	2	3	4	5	6	7
K	2	4	5	7	8	10

2.4 RGB-T 图像细节信息的融合模块

EIF 模块将 HFIE 模块生成的 RGB、T 细节图像融合,生成 RGB-T 细节融合图像,从而提高显著性目标边界检测的能力。EIF 模块包含 RGB、T 细节图像特征提取与 RGB、T 细节图像融合两个部分,如图 2 所示。

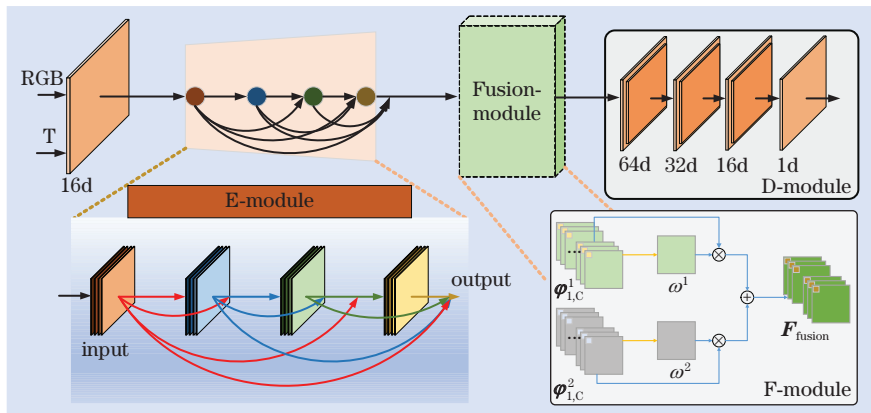


图 2 RGB-T 图像细节信息融合框图

Fig. 2 Block diagram of RGB-T image detail information fusion

2.4.1 RGB、T 细节图像的特征提取

在特征提取阶段,EIF 网络中的 E-module 参考文献[15]中跳跃连接的方式建立不同层级特征的联系。EIF 网络通过 E-module 提取图像的深层语义信息,之后使用 D-module 逐级减少输出特征维度,重建输入图像。在训练数据获取方面,DUTS 数据集^[16]的训练集中的 10553 张 RGB 图片通过 HFIE 模块后生成对应的细节图像,将细节图像作为 EIF 网络的训练数据。EIF 网络使用结构相似度(SSIM)^[17]与平均绝对误差(MAE)作为损失函数,并使用 Adam 优化器(设置条件:学习率为 0.0001,

训练迭代 20 次)。值得注意的是,EIF 网络在训练过程中不使用 Fusion-module。

2.4.2 RGB、T 细节图像的融合

图 2 中,首先,E-module 分别以 RGB、T 细节图像为输入,提取 RGB、T 细节图像的深度特征,生成 64 通道的 RGB、T 深度特征图像 $\phi_{1,C}^r$,其中 $r=1$ 表示 RGB 图像, $r=2$ 表示 T 图像, ϕ_f^r 为第 f 个通道的特征, C 为总通道数;其次,通过 Fusion-module,使用 L1-norm 算法^[13]生成 RGB、T 细节图像的融合特征 F_{fusion} ;最后,通过解码器网络生成 RGB-T 细节融合图像。Fusion-module 中 RGB、T

细节图像深度特征融合的计算公式为

$$C^r(x, y) = \|\varphi_{1,c}^r(x, y)\|_1, \quad (8)$$

$$\omega^r(x, y) = C^r(x, y) / \sum_{r=1}^2 C^r(x, y), \quad (9)$$

$$F_{\text{fusion}}(x, y) = \sum_{r=1}^2 \omega^r(x, y) \times \varphi_{1,c}^r(x, y), \quad (10)$$

式中: $C^r(x, y)$ 表示经过 1-范数处理后各坐标值; $\omega^r(x, y)$ 表示融合特征在各个坐标位置上, RGB 通道或者 T 通道信息所占的权重。

2.5 RGB-T 光照图像与 RGB-T 细节图像的融合

通过光照图像与细节图像的融合,生成最终的 RGB-T 融合图像 (I_F), 计算公式为

$$I_F(x, y) = \alpha_1 I_L(x, y) + \alpha_2 I_D(x, y). \quad (11)$$

由于 I_D 图像中所含信息量较少, 而 I_L 图像包含大部分的图像信息, 可以视为融合图像的基础部分, 借鉴文献[13]融合图像基础层与图像细节层的思路, 将 I_L 与 I_D 融合时某点像素的权重取值为 $\alpha_1 = 0.5, \alpha_2 = 0.5$ 。为了确保最终的 RGB-T 融合图像具有良好的物体边界信息且不引入过多噪声, 融合时 RGB-T 细节图像的等级(D)依据对应 RGB-T 光照图像的等级(L)来确定。具体融合规则如表 4 所示。

表 4 RGB、T 图像融合规则

Table 4 RGB、T Image fusion rules

L	L1	L2	L3	L4	L5	L6	L7	L8	L9
D	R6	R6	R5	R4	R3	R2	R2	R1	R1

当光照条件较差时, 由于 RGB-T 光照融合图像中 T 图像所占比例较高, RGB-T 光照融合图像丢失了很多的细节信息, 因此所提算法使用细节信息更加丰富的 RGB-T 细节图像与 RGB-T 光照图像进行融合。

2.6 显著性检测模块

FSDNet 模块参考 CPDNet 模型^[3]进行 RGB-T 图像显著性目标检测, 唯一的不同是 FSDNet 参考文献[18]模型的训练算法, 使用 MAE 与 SSIM^[17]二种损失函数进行联合训练, 分别从像素和结构两个方面准确地预测显著性目标, 这与本文强调显著性区域与显著性边界共同作用的算法相适应。FSDNet 模型的整体架构如图 3 所示, FSDNet 由编码器、解码器和细化网络三个部分组成。编码器部分采用 VGG16 网络模型, 由 Conv1_2 至 Conv5_4 五个卷积块组成, 用于提取 RGB-T 融合图像 (I_F) 的深层特征。解码器部分和细化网络部分的结构基本一致。其中解码器输入编码器部分提取的深层特

征 (F_1^D, F_2^D, F_3^D), 经过感受野区域 (RFB) 模块扩大多级特征感受野和 Aggregation 模块整合多级上下文特征后输出聚合特征 (F^A), 并生成粗糙的显著图 (S_1)。细化网络部分首先输入聚合特征 (F^A), 通过整体注意力 (HA) 模块获取特征图中的显著性目标边缘信息, 同时扩大显著图的覆盖范围, 随后将特征输入到 Conv4_4 卷积块中, 之后采用与编解码器网络相同的方式将特征输入到 RFB、Decoder 模块, 生成最终的显著图 (S_2)。参考 CPDNet^[3], 使用标注值 (GT) 同时监督 S_1 和 S_2 进行模型的训练。为了更好地对细化网络部分和解码器部分的数据流向进行区分, 图 3 中使用实线箭头表示编码器与解码器的数据流向, 使用虚线箭头表示细化网络的数据流向。

以编码器模块输出的特征 F_1^D 在 RFB 模块的数据流向为例, 对 RFB 模块的作用进行说明。RFB 模块借鉴了 InceptionNet 模型的思想, 在多个分支结构中使用不同感受野大小的卷积核提取特征, 增强模块对局部细节信息和高级语义信息的提取能力。 F_1^D 特征大小为 64, 通道数为 256。 F_1^D 特征大小输入到 RFB 模块之后, 首先在 RFB 模块的每一个分支中, 均通过一个卷积层 (虚线方块表示) 将输入特征通道数由 256 改变为 32。RFB 模块在三个分支结构中分别使用了卷积核大小为 3、5、7 的非对称卷积模型代替常规的卷积操作, 减少了模型的计算量。紧接着分别使用了空洞率为 3、5、7 的空洞卷积进一步增加了模型的感受野。InceptionNet 模型将各个分支网络进行拼接, 然后进行输出, 所以其输出特征的通道数。与 InceptionNet 模型输出的方式不同, RFB 模块对 4 个分支结构获得的不同感受野的特征进行拼接, 之后通过 ConvR_1 卷积层整合各个分支特征, 使输出通道数转变为 32, 最后与原始的特征图相加作为 RFB 模块的输出特征, 大大减少了之后解码器模块的计算量。

以特征 F_4^D, F_5^D 和 F_6^D 在 Aggregation 模块的流向为例, 对 Aggregation 模块的作用进行说明。特征 F_4^D, F_5^D 和 F_6^D 输入到 Decoder 模块, 首先低级特征 F_4^D 通过卷积与上采样操作后与高级特征 F_5^D 和 F_6^D 相乘, 使 F_4^D 特征信息输入至高级特征 F_5^D 和 F_6^D 中, 实现了多级特征的跨越连接; 之后采用 F_6^D 特征信息逐级传递的方法, 低级特征 F_4^D 通过卷积与上采样操作后逐级累加至高级特征中, 输出特征 F^H 。通过卷积与上采样操作, Aggregation 模型输出预测显著图, 实现了多级特征之间的联系, 进一步

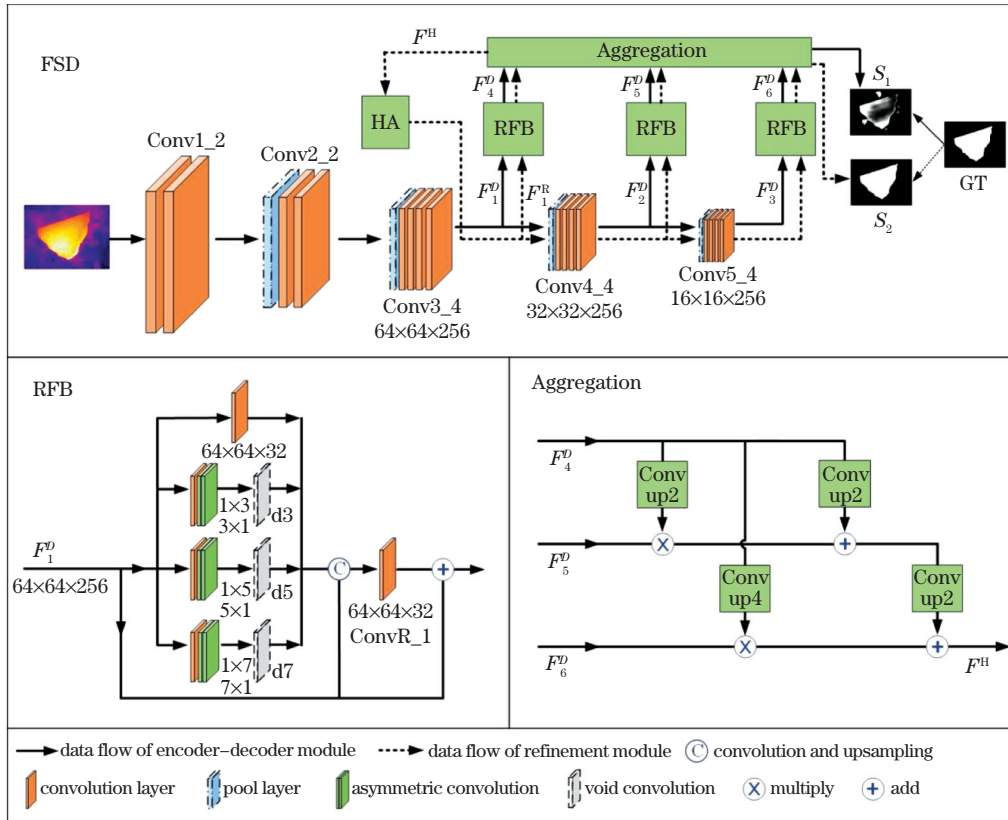


图 3 所提 FSDNet 模块的架构

Fig.3 Architecture of the proposed FSDNet module

降低了模型的推理时间。

为了使显著性物体的边界信息更加清晰, FSDNet 在解码器部分生成粗糙的预测显著图 (S_1) 之后, 使用细化网络进一步优化预测显著图的细节。由于解码器部分预测的粗糙显著图 (S_1) 会影响细化网络生成的显著图 (S_2) 的精度, 因此使用 HA 模块。HA 模块在获取特征 F^H 中的显著性目标边缘信息的同时扩大显著图的覆盖范围, 降低 S_1 精度较低时对细化网络的影响。HA 模块的计算公式为

$$F_1^R = \text{MAX} [f_N(\text{Conv}(F^H, K^H, \sigma^H), S_1^D)] \quad (12)$$

式中: F_1^R 表示经过 HA 模块之后输出的特征; $\text{Conv}(\cdot)$ 表示对输入特征 F^H 进行高斯核大小为 $K^H \times K^H$ 、方差为 σ^H 的卷积运算, 过滤出显著性特征 F_1^R 的边缘信息, K^H 与 σ^H 分别设置为 32 与 4, 并可以随神经网络的反向传播进行更新; $f_N(\cdot)$ 表示归一化函数, 使输出特征映射至 $[0, 1]$; $\text{MAX}[\cdot]$ 为最大值函数, 使归一化函数 $f_N(\cdot)$ 的输出特征与特征 S_1^D 进行比较, 扩大初始化特征的边界范围。最后将 HA 模块的输出特征依次输入到 Conv4_4、Conv5_4 卷积块, 输出特征 F_2^R 与 F_3^R , 并使用 RFB、Decoder 模块输出最终预测的显著图。

由上可知: FSDNet 模块在经过编码器、解码器部分之后, 通过 HA 模块提取了初始显著图 (S_1) 的边界范围并通过高斯卷积操作进一步过滤多余的噪声, 有利于提取 RGB-T 融合图像 (I_F) 的显著性物体边界信息; 并且通过 RFB 模块、Aggregation 模块, 可以快速地预测显著图, 减少了推理时间。

3 分析与讨论

3.1 实验配置

3.1.1 数据集

由于用于 RGB-T 显著性检测任务的数据有限, 因此使用目前所知的两个公开的数据集: VT821^[19] 和 VT1000^[5]。其中, VT821 数据集包含 821 对 RGB-T 图像及对应的真值, 但由于很多的 T 图像未与 RGB 图像校准对齐, 故从 VT821 数据集中人工挑选 223 张校准情况相对较好的图像作为算法的测试数据。VT1000 数据集较好地解决了 RGB、T 图像的校准问题, 包含 1000 对 RGB-T 图像及对应的真值, 并包含低光照环境、多个显著性物体、显著性物体与背景颜色相似等多个子集, 可以针对不同条件测试显著性检测模型的精度。使用整个 VT1000 数据集对所提算法进行测试。

3.1.2 评价指标

采用显著性检测任务常用的评价指标进行评价:查准率(P)和查全率(R)。使用不同的阈值对测试结果进行二值化处理,之后将结果与二值化的 GT 比较,即可计算出相应阈值对应的 P - R 值。阈值按照图像像素值 $[0, 255]$ 中以整数为间隔进行选取,即可得到 256 组 P - R 值,绘制 P - R 曲线。F-measure 指标是对显著性图像查准率与查全率水平的综合评价,通过改变阈值 $[0, 255]$ 的大小可以得到该指标的最大值 F_{\max} 和平均值 F_{ave} 。F-measure 的计算公式为

$$F_{\beta} = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \quad (13)$$

依据惯例^[4-5], (13)式中的 β^2 设置为 0.3。生成的显著性图像与真实值的像素平均绝对误差 (MAE)的计算公式为

$$E_{\text{MA}} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)|, \quad (14)$$

式中: W 与 H 分别表示图像的宽和高; S 表示算法生成的显著性图像; G 表示人工标注的真值图。

3.2 实验评估

为了证明所提算法的有效性,对所提算法与 7 种现有的显著性检测算法进行了比较。其中包括 2 种基于 RGB-T 的显著性检测算法,即 CGL^[9] 和

CRA^[8],两者均使用传统算法提取图像中的显著性特征,之后借助图论的算法来得到显著性检测结果;3 种基于 RGB 图像的显著性检测算法,即 PoolNet^[4]、CPDNet^[3]和 EGNet^[5];2 种基于 RGB-D 图像的显著性检测算法,即 DMRA^[20]和 A2dele^[21]。由于现阶段基于 RGB-T 图像的显著性检测算法大多使用传统的算法,没有提供公开代码且精度不如基于深度学习的算法,因此仅给出了两种 RGB-T 算法的定量比较结果。上述基于 RGB 的算法均使用提出的显著性检测模型,在 VT1000 数据集中使用 1000 张图片进行测试。由于 CPDNet、PoolNet 和 EGNet 均为以 RGB 图像为输入的深度学习模型,为了公平比较,对 RGB 图像与 T 图像进行拼接组成 RGB-T 特征,然后通过一个 3×3 的卷积层,将拼接后的特征转变为 3 通道,作为 CPDNet、PoolNet 和 EGNet 的 RGB-T 输入。从 VT1000 数据集中随机挑选 900 对图像作为模型的训练集,使用剩余的 100 对图像作为修正模型的测试集,训练后的模型用 PoolNet+、EGNet+ 和 CPDNet+ 表示。

3.2.1 定性实验评估

所提算法的定性测试结果如图 4 所示,在多种场景中对各种算法进行了比较。第 1 行和第 2 行图片为光照与对比度较好的环境中对行人的检测结果,在此场景中,所提算法可以很好地检测显著性目

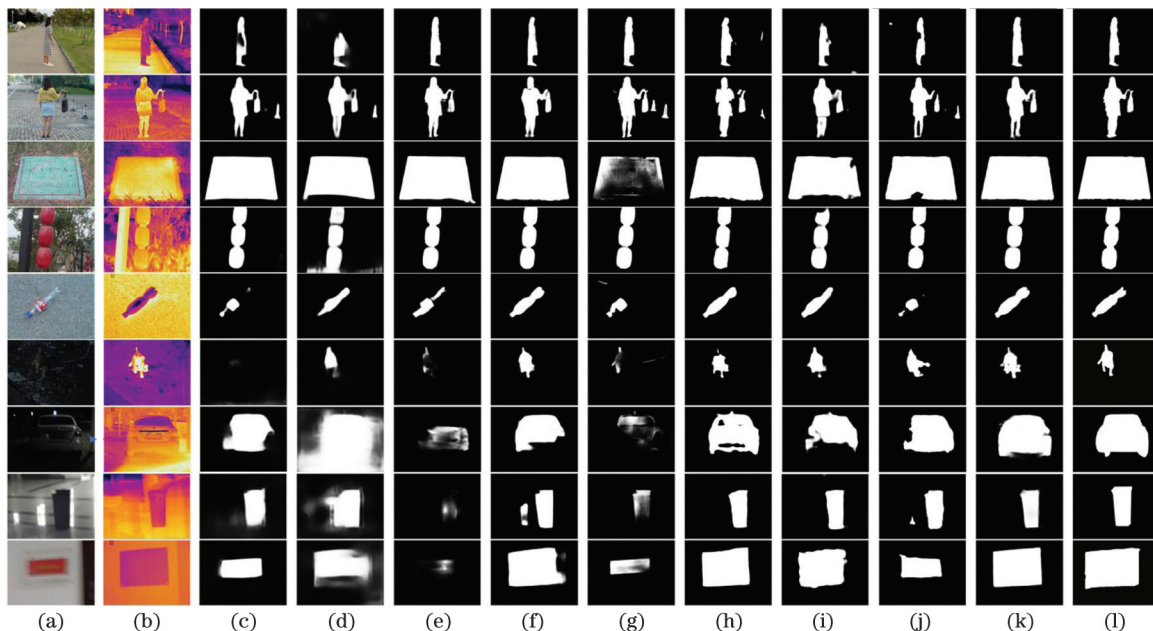


图 4 不同算法的显著图比较。(a) RGB 图;(b) T 图;(c) EGNet;(d) EGNet+;(e) CPDNet;(f) CPDNet+;(g) PoolNet;(h) PoolNet+;(i) DMRA;(j) A2dele;(k)所提算法;(l) GT

Fig.4 Saliency map comparison between different algorithms. (a) RGB images; (b) T images; (c) EGNet; (d) EGNet+; (e) CPDNet; (f) CPDNet+; (g) PoolNet; (h) PoolNet+; (i) DMRA; (j) A2dele; (k) proposed algorithm; (l) GT

标的区域,同时排除非显著性物体对模型的干扰;第 3 行和第 4 行图片为环境条件较好时的显著性目标检测结果,可以发现,所有算法基本上能够较为准确地完成透明物体的显著性检测;第 6 行和第 7 行图片为光照较差场景中的显著性检测结果,其余大部分算法只能检测出部分的显著性区域,所提算法可以在较为准确地检测出显著性目标区域的同时检测到显著目标边界;第 8 行和第 9 行图片为模糊物体的显著性检测结果。从 5~9 行可以看到:基于 RGB、T 图像的算法难以在复杂场景中检测显著性目标,修正的 RGB-T 算法可以更好地检测显著性目标区域;修正的 RGB-T 算法与基于 RGB-D 的算法检测结果比较粗糙,相比之下,所提算法更好地突出了显著性目标的边缘细节信息,更加接近真实值。以上定性实验的结果表明,所提算法具有有效性,特别在透明物体、光照条件较差、模糊物体等具有挑战性的环境中,相比其他算法具有更强的鲁棒性。

3.2.2 定量实验评估

各种算法在 VT821 与 VT1000 数据集上的 F_{max} 、 F_{ave} 和 MAE 的测试结果如表 5 所示。可以发现,深度学习算法的精度整体要低于基于 RGB-T 图像的传统算法,所提算法在两个数据集上的整体性能要优于其他对比算法。CPDNet+、PoolNet+ 与 EGNet+ 算法相较于原来算法的提升不是很明显。可能因为 VT821 与 VT1000 数据集中的大部分图片都是在光照环境与背景环境较好的条件下拍摄的,在这种环境下,T 图像的显著性检测表现可能不如

RGB 图像;并且原模型是基于 RGB 图像的显著性检测模型,而本文对 RGB 图像通道与 T 图像通道进行拼接,没有进行进一步优化,影响了融合网络的精度。EGNet+ 模型相比原模型精度有所降低,可能因为 EGNet 网络在细化模块中需要大量的图像纹理信息,而 T 图像中纹理信息较少,因此 EGNet+ 网络模型的精度在整个数据集中反而降低了。各种算法在 VT821 与 VT1000 数据集上的 $P-R$ 曲线如图 5 所示。 $P-R$ 曲线是利用不同阈值下的 $P-R$ 值来确定的,在利用不同阈值确定 $P-R$ 曲线时,相比其他算法,所提算法的 $P-R$ 曲线处于比较好的位置并且跨度区间较窄,说明了所提算法的有效性。

表 5 不同算法的定量比较

Algorithm	VT821			VT1000		
	F_{max}	F_{ave}	MAE	F_{max}	F_{ave}	MAE
EGNet	0.7717	0.7459	0.0512	0.8796	0.8637	0.0400
EGNet+	0.6994	0.6475	0.116	0.7840	0.7453	0.0938
CPDNet	0.7971	0.7882	0.0425	0.8869	0.8621	0.0326
CPDNet+	0.7915	0.7796	0.0471	0.8738	0.8570	0.0333
PoolNet	0.8192	0.8046	0.0485	0.8795	0.8651	0.0411
PoolNet+	0.8177	0.8086	0.0411	0.8759	0.8656	0.0313
DMRA	0.8411	0.8270	0.0417	0.8613	0.8444	0.0381
A2dele	0.7513	0.7482	0.0621	0.8609	0.8575	0.0401
CGL	0.780	0.744	0.0849		0.727	
CRA	0.747	0.739	0.1083		0.693	
Proposed algorithm	0.8433	0.8291	0.0375	0.9095	0.8943	0.0268

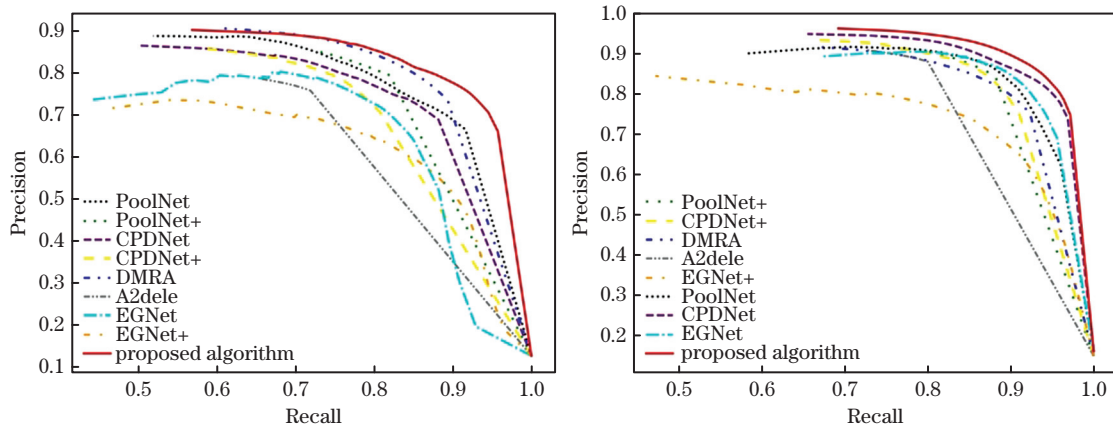


图 5 不同算法在不同数据集下的 $P-R$ 曲线。(a)VT821;(b)VT1000

Fig.5 $P-R$ curves of different algorithms at different datasets. (a) VT821; (b) VT1000

使用 VT1000 数据集中的低光照数据子集对所提算法进行了测试。为了更好地展示多源图像融合的优势,在 CPDNet、EGNet 和 PoolNet 的基础上增加了 T 图像进行测试的对比实验,分别用

CPDNet++、EGNet++ 和 PoolNet++ 表示。VT1000 数据集的低光照条件数据子集包括 56 对在低光照条件下的 RGB、T 图像,含有低温物体、图像失焦等多种情况。在该数据集中,使用了 5-折交

叉验证(5-fold cross-validation),即随机将数据集分为 5 份,每份包含 11 张图片,轮流从中取出 4 份作为训练数据,剩下一份加上多出的一张图片用于测试,最后对实验结果进行平均作为最终结果。为了减小数据量较小导致的误差,使用随机增强策略对图像进行镜像翻转和 180 度旋转,将训练数据量扩大了 3 倍。除了减小学习率使模型更易收敛之外,其他的模型配置均与 CPDNet、EGNet 和 PoolNet 对应论文中的配置一致。修正的 RGB-T 网络在光照条件较差的环境中的检测精度要明显好于原网络模型,如表 6 所示,表现出更强的鲁棒性,证明了融合算法的有效性。

表 6 VT1000 低光照数据子集中的测试结果

Table 6 Test results in VT1000 low light data subset

Algorithm	F_{max}	F_{ave}	MAE
EGNet	0.6602	0.5825	0.1100
EGNet+	0.8393	0.8136	0.0616
EGNet++	0.8225	0.8036	0.0538
CPDNet	0.7662	0.7092	0.1060
CPDNet+	0.8071	0.7827	0.0721
CPDNet++	0.8164	0.8038	0.0546
PoolNet	0.5803	0.5578	0.1140
PoolNet+	0.7922	0.7683	0.0622
PoolNet++	0.7532	0.7135	0.0940
Proposed algorithm	0.8743	0.8651	0.0386

3.3 消融实验

为了进一步说明所提算法中各个模块的作用,本部分进行了针对各个模块的消融实验。消融实验的定性分析结果如表 7 所示,使用各个模块的首字母代表某个模块,I 代表图像光照分级模块,A 代表图像光照融合模块,H 代表图像细节信息提取模

块,E 代表图像细节信息融合模块,另外“-”表示不使用某模块,“+”代表使用某个模块。Baseline 算法表示使用了 RGB、T 图像逐像素等比例融合的方法进行显著性检测,为本次消融实验的基准。+I 表示使用 β' 值作为融合图像中 RGB 图像所占比例,从而引导 RGB-T 图像的融合。+IA 表示在使用 IMJ 模块计算 RGB、T 光照评分的基础上使用所提光照融合算法进行 RGB-T 图像融合。所提算法在前者的基础上使用 RGB-T 细节图像进行融合,获得最终的 RGB-T 融合图像。-E 表示不使用 EIFT 模块融合 RGB、T 细节图像,只使用 RGB 细节图像与 RGB-T 光照图像进行融合生成最终的 RGB-T 融合图像。通过消融实验,证明了设计的 IMJ、ALF、HFIE 和 EIF 模块的有效性。

表 7 消融分析的定量比较结果

Table 7 Quantitative comparison results of ablation analysis

Algorithm	VT821			VT1000		
	F_{max}	F_{ave}	MAE	F_{max}	F_{ave}	MAE
Baseline	0.8256	0.8003	0.0445	0.8396	0.8315	0.0361
+I	0.8326	0.8125	0.0429	0.8459	0.8361	0.0342
+IA	0.8493	0.8139	0.0417	0.8956	0.8827	0.0330
Proposed algorithm	0.8433	0.8291	0.0375	0.9095	0.8943	0.0268
-E	0.8366	0.8193	0.0411	0.9051	0.8819	0.0271

图 6 为视觉对比效果,第 1、2 行表示显著性物体与背景相似的场景,第 3 行中图像较为模糊,第 4 行中图片背景干扰信息较多。在具有挑战性的场景中,对比算法大多只能检测显著性物体的一部分,相比之下所提算法更加接近真实值。另外,如图 6 (e)、(f)两列所示,所提 HFIE 与 EIF 模块在挑战性的场景中可以很好地突出显著性目标。

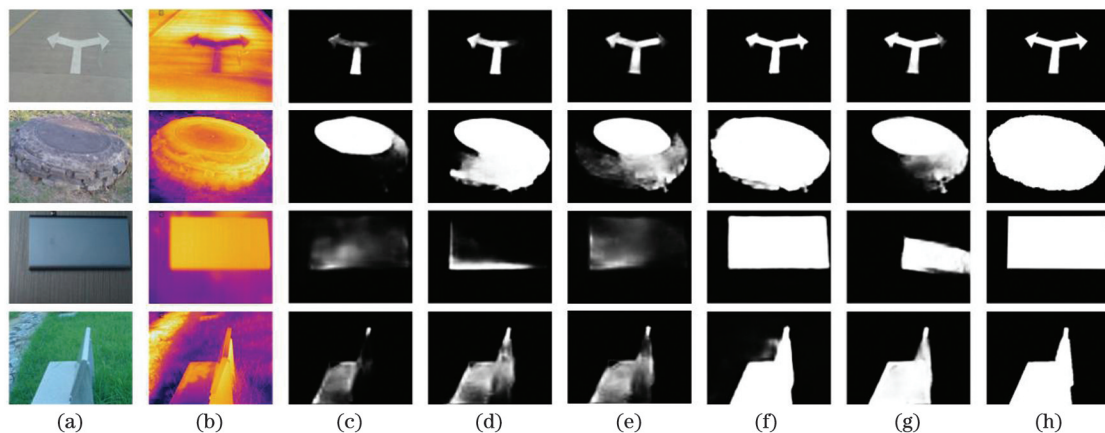


图 6 消融分析显著图的定性比较。(a) RGB 图;(b) T 图;(c) 基准图;(d) +I;(e) +IA;(f) 所提算法;(g) -E;(h) GT
Fig. 6 Qualitative comparison of salient maps for ablation analysis. (a) RGB images; (b) T images; (c) Baseline; (d) +I; (e) +IA; (f) proposed algorithm; (g) -E; (h) GT

4 结 论

在亮度较低、背景颜色相近、图片遮挡严重、雾霾天气等条件下,使用 RGB 图像难以检测显著性目标区域。T 图像由于具有物体红外热辐射,仅仅使用 T 图像难以检测显著性目标边界;一般的 RGB-T 图像融合算法难以针对显著性检测任务同时提高显著性区域与边界检测精度。针对这些问题,提出了一种 RGB-T 光照图像融合算法(IMJ、ALF),利用 T 图像的红外热辐射信息提高了检测显著性目标区域的能力;提出了一种 RGB-T 细节图像融合算法(HFIE、EIF),提高了检测显著性目标边界的能力。

主要针对显著性检测任务设计了多源图像光照及边缘融合的显著性检测网络,对比现有的 RGB-T 图像融合算法,提高了显著检测精度。在 VT821、VT1000 数据集上与其他的显著性检测算法进行比较,所提算法在 F_{\max} 、 F_{ave} 及 MAE 指标中均取得了最优结果。

参 考 文 献

- [1] 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.
- [2] Jin Y K, Yu F Q. Saliency detection based on background continuity priori knowledge[J]. Laser & Optoelectronics Progress, 2019, 56(12): 121006.
金一康, 于凤芹. 基于背景连续性先验知识的显著性检测[J]. 激光与光电子学进展, 2019, 56(12): 121006.
- [3] Wu Z, Su L, Huang Q M. Cascaded partial decoder for fast and accurate salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3902-3911.
- [4] Liu J J, Hou Q B, Cheng M M, et al. A simple pooling-based design for real-time salient object detection [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3912-3921.
- [5] Zhao J X, Liu J J, Fan D P, et al. EGNNet: edge guidance network for salient object detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8778-8787.
- [6] Zhang S L, Xie L B. Salient detection based on cascaded convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(7): 071501.
张松龙, 谢林柏. 基于级联全卷积神经网络的显著性检测 [J]. 激光与光电子学进展, 2019, 56(7): 071501.
- [7] Shi F F, Zhang S L, Peng L. Salient object detection based on deep residual networks and edge supervised learning [J]. Laser & Optoelectronics Progress, 2019, 56(15): 151502.
时斐斐, 张松龙, 彭力. 基于深度残差网络与边缘监督学习的显著性检测 [J]. 激光与光电子学进展, 2019, 56(15): 151502.
- [8] Wang G Z, Li C L, Ma Y P, et al. RGB-T saliency detection benchmark: dataset, baselines, analysis and a novel approach[M]//Wang Y T, Jiang Z G, Peng Y X. Image and graphics technologies and applications. Communications in computer and information science. Singapore: Springer, 2018, 875: 359-369.
- [9] Tu Z Z, Xia T, Li C L, et al. RGB-T image saliency detection via collaborative graph learning [J]. IEEE Transactions on Multimedia, 2020, 22(1): 160-173.
- [10] Ben H A, He Y, Krim H, et al. A multiscale approach to pixel-level image fusion [J]. Integrated Computer-Aided Engineering, 2005, 12(2): 135-146.
- [11] Yang S Y, Wang M, Jiao L C, et al. Image fusion based on a new contourlet packet [J]. Information Fusion, 2010, 11(2): 78-84.
- [12] Bavisetti D P, Dhuli R. Two-scale image fusion of visible and infrared images using saliency detection [J]. Infrared Physics & Technology, 2016, 76: 52-64.
- [13] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.
- [14] Wang J, Jiang X Y, Ji B G. An algorithm to fuse gray-scale infrared and visible light image based on perceptual color space [J]. Journal of Optoelectronics · Laser, 2008, 19(9): 1261-1264.
王加, 蒋晓瑜, 纪伯公. 基于感知颜色空间的灰度可见光与红外图像融合算法 [J]. 光电子·激光, 2008, 19(9): 1261-1264.
- [15] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4700-4708.
- [16] Wang L J, Lu H C, Wang Y F, et al. Learning to detect salient objects with image-level supervision [C]//

- 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3796-3805.
- [17] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [18] Qin X B, Zhang Z C, Huang C Y, et al. BASNet: boundary-aware salient object detection [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7471-7481.
- [19] Tang J, Fan D Z, Wang X X, et al. RGBT salient object detection: benchmark and a novel cooperative ranking approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(12): 4421-4433.
- [20] Piao Y R, Ji W, Li J J, et al. Depth-induced multi-scale recurrent attention network for saliency detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 7253-7262.
- [21] Piao Y R, Rong Z K, Zhang M, et al. A2dele: adaptive and attentive depth distiller for efficient RGB-D salient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 9057-9066.