跨模态融合和边界可变形卷积引导的RGB-D显著性目标检测

孟令兵[1],袁梦雅[2],时雪涵[1],张 乐[1],吴锦华[1],程 菲[1,3]
(1. 安徽信息工程学院计算机与软件工程学院,安徽芜湖 241000;2. 安徽信息工程学院电气与电子工程学院,安徽芜湖 241000;
3. 杭州电子科技大学管理学院,浙江杭州 310000)

摘 要: RGB-Depth(RGB-D)显著性目标检测是一项有意义且具有挑战性的任务,基于现有卷积神经网络检测方法在简单场景中获得了良好的检测性能,但不能有效应对背景信息混乱,深度图质量低和目标轮廓复杂的情况. 为应对上述问题,本文提出了一种跨模态融合和边界可变形卷积引导的RGB-D显著性目标检测方法. 首先,本文以Swin-Transformer为特征提取器,分别对RGB模态与深度图模态进行特征提取,并通过跨模态注意力增强特征模块对两种模态特征进行融合以挖掘显著物的共性与互补特征. 接着将提出的相邻多尺度特征增强模块嵌入编码器深层,以获得丰富的全局上下文特征信息,更精准地定位显著物的位置. 然后通过构建一个边界特征提取解码器(U-Net架构)生成显著物的边界线索图,并重复采用跨模态融合特征确保生成显著物边界的完整性. 最后,本文设计了一个边界可变形卷积引导模块,使用边界线索图与可变形卷积引导跨模态融合特征进行解码以得到更加准确的显著图. 通过在6个公开基准数据集上与25种主流方法相比较,本文所提模型在多个指标上均有较明显的提升,从而证明了本文方法的有效性.

关键词: 显著性目标检测;跨模态融合;边界特征;可变形卷积;显著图
基金项目: 安徽省自然科学基金(No.2008085MF201);安徽省教育厅自然科学重点项目(No.2022AH051894,No.2022AH051887);安徽省高校优秀青年人才支持计划(No.gxyq2022147)
中图分类号: TP751 **文献标识码:** A **文章编号:** 0372-2112(2023)11-3155-12
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230042

RGB-D Salient Object Detection Based on Cross-Modal Fusion and Boundary Deformable Convolution Guidance

MENG Ling-bing[1], YUAN Meng-ya[2], SHI Xue-han[1], ZHANG Le[1], WU Jin-hua[1], CHENG Fei[1,3]
(1. *School of Computer and Software Engineering*, *Anhui Institute of Information Technology*, *Wuhu*, *Anhui* 241000, *China*;
2. *School of Electrical and Electronic Engineering*, *Anhui Institute of Information Technology*, *Wuhu*, *Anhui* 241000, *China*;
3. *School of Management*, *Hangzhou Dianzi University*, *Hangzhou*, *Zhejiang* 310000, *China*)

Abstract: RGB-Depth (RGB-D) salient object detection is a meaningful and challenging task. The current method based on convolutional neural networks has achieved good detection performance in simple scenes, but cannot effectively handle scenes with cluttered background information, low-quality depth maps, and complex object contours. In order to solve the above problems, an RGB-D SOD model based on cross-modal fusion and boundary deformable convolution guidance is proposed in this paper. Firstly, the Swin Transformer is used as an extractor to extract features from the RGB modality and depth modality, respectively, which fuse the two modalities by using a cross-modal attention enhancement feature (CMAEF) module, to explore the common and complementary features of salient objects. Then, the proposed adjacent multi-scale feature enhancement (AMFE) module is embedded deep-level into the encoder to obtain rich global contextual feature information, which can locate the position of salient objects more accurately. Next, the boundary cue maps of salient objects are generated by boundary feature extraction decoder (U-Net architecture) constructed and repeated using cross-modal fusion features to ensure the integrity of the generated salient object boundaries. Finally, we designed a boundary deformable convolution guidance (BDCG) module that uses boundary cue maps with deformable convolution to guide the de-

收稿日期:2023-01-10;修回日期:2023-08-28;责任编辑:朱梅玉

coding of cross-modal fusion features to obtain more accurate saliency maps. Comprehensive experiments on six popular benchmark datasets compared with 25 mainstream methods demonstrate that the proposed model shows significant improvement in metrics, which proves the effectiveness of the proposed model.

Key words: salient object detection; cross-modal fusion; boundary features; deformable convolution; saliency map

Foundation Item(s): Natural Science Foundation of Anhui Province (No.2008085MF201); Natural Science Major Project of Anhui Provincial Department of Education (No.2022AH051894, No.2022AH051887); Outstanding Young Talents Support Program Project of Anhui Province (No.gxyq2022147)

1 引言

显著性目标检测(Salient Object Detection,SOD)旨在分割出自然图像中最为引人注目的物体或者区域[1~5],它通过模拟人类的视觉注意机制,仅处理图像中的特定区域,不仅可以提高计算效率,而且可以节省计算和存储成本,因此已经被广泛用于计算机视觉任务中,如:图像裁剪[6]、图像分割[7]和视觉跟踪[8]等. 以RGB数据为研究任务取得了较好的检测效果,但在低光照、背景与显著物颜色或者轮廓相似的场景下难以准确地检测出完整的目标. 深度图(depth)具有良好的几何结构、内部一致性和光照不变性,能够为RGB图像提供辅助信息并提高模型检测的性能. 在RGB图像和深度图融合策略上,大多数模型采用单流和双流两种方式,而后者使用最为广泛,并且大多数双流模型采用多级别特征聚合及跨模态特征融合[9~11]模块策略以提高检测性能,比如,长距离关联度融合模块[9]、一致性-差异聚合模块[10]和分层交替交互模块[11]. 文献[12,13]通过设计自动筛选模块过滤低质量深度图或对深度图质量进行排序以提高模型检测性能. 这些模块都要经过精心设计,并且过于复杂的模块还会引入新的噪声信息. 此外针对多目标和大目标的显著性物体,多种多尺度特征提取模块[14~16],并且将它们嵌入到编码器深层从而获得丰富的上下文信息.

由于跨模态特征融合策略和多尺度特征提取模块等技术的使用,显著性目标检测性能有了较大的提高. 但上述的多尺度特征提取模块未考虑相邻层次特征的交互,当待检测的图像中包含多目标或者大目标时会出现漏检等现象,并且上述技术对于具有复杂边界的物体效果较差. 最近,一些研究人员提出基于边界特征[17~19]学习的方法,例如,文献[17]采用多个显著性检测模块构建边界信息渐进式引导模型. 文献[18]利用边界提取模块提供准确的边缘特征信息,并将其添加到解码阶段的边缘感知位置注意单元. 然而这些模型将边界图直接与特征图级联,未能充分利用边界图约束模型生成更为精准的显著图.

综上所述,现有的检测方法在多种复杂场景下难以精准的分割出显著性物体. 为解决上述的问题,本文提出一种双分支的端到端模型,从而提高检测的性能.

2 本文方法

2.1 整体框架

本文提出的检测模型如图1所示. 首先,我们先采用两个完全相同的Swin Transformer(ST)骨干网络为编码器提取RGB特征与深度图特征,记为 F_i^{R} 和 F_i^{D} ($i\in\{1,2,3,4\}$,i 表示特征图层数,R表示RGB,D表示Depth),输入到模型中的图像大小记为 $C\times W\times H$(C、W 和 H 分别表示特征图的通道数、长度和宽度). 第1层特征图被缩放至 $W/4\times H/4$,后3层的特征图分辨率被缩放为 $W/2^{i+1}\times H/2^{i+1}$($i\in\{2,3,4\}$),其中第 i 层的通道数记为 C_i,$C_i\in\{128,256,512,1\,024\}$. 然后使用跨模态注意力增强特征(Cross-Modal Attention Enhancement Feature module,CMAEF)模块来挖掘RGB和深度图的显著物共性. 接着,我们把得到的深层跨模态特征图输入到相邻多尺度特征增强模块(Adjacent Multi-scale Feature Enhancement module,AMFE)中用于获取丰富的上下文特征信息. 在解码过程中产生两个分支,每个分支均采用编码器-解码器结构. 第一分支是边界线索图分支,通过边界特征提取模块生成显著物的边界线索图. 第二个分支是显著图分支,在此分支中本文提出一个边界可变形卷积引导模块(Boundary Deformable Convolution Guidance module,BDCG),该模块利用生成的边界线索图和可变形卷积引导跨模态融合特征进行逐层解码从而生成清晰、完整和准确的显著图.

2.2 跨模态注意力增强特征模块

本文提出的CMAEF模块如图2所示,它由两部分组成,即跨模态特征融合与注意力机制模块. 首先,分别将RGB特征图和深度特征图输入到2层3×3的卷积层中,然后利用特征交叉融合的策略来探索两种模态间的相关性及互补性,两种模态的特征图进行交叉相乘,从而使不同模态特征图通过交叉融合来互相挖掘和强化显著物共性. 跨模态特征融合如下所示:

$$F_i^{\mathrm{RC}}=\mathrm{Conv}\left(\mathrm{Conv}\left(F_i^{\mathrm{R}}\right)\right) \tag{1}$$

$$F_i^{\mathrm{DC}}=\mathrm{Conv}\left(\mathrm{Conv}\left(F_i^{\mathrm{D}}\right)\right) \tag{2}$$

$$F_i^{\mathrm{RCD}}=F_i^{\mathrm{RC}}\otimes F_i^{\mathrm{D}} \tag{3}$$

$$F_i^{\mathrm{DCR}}=F_i^{\mathrm{DC}}\otimes F_i^{\mathrm{R}} \tag{4}$$

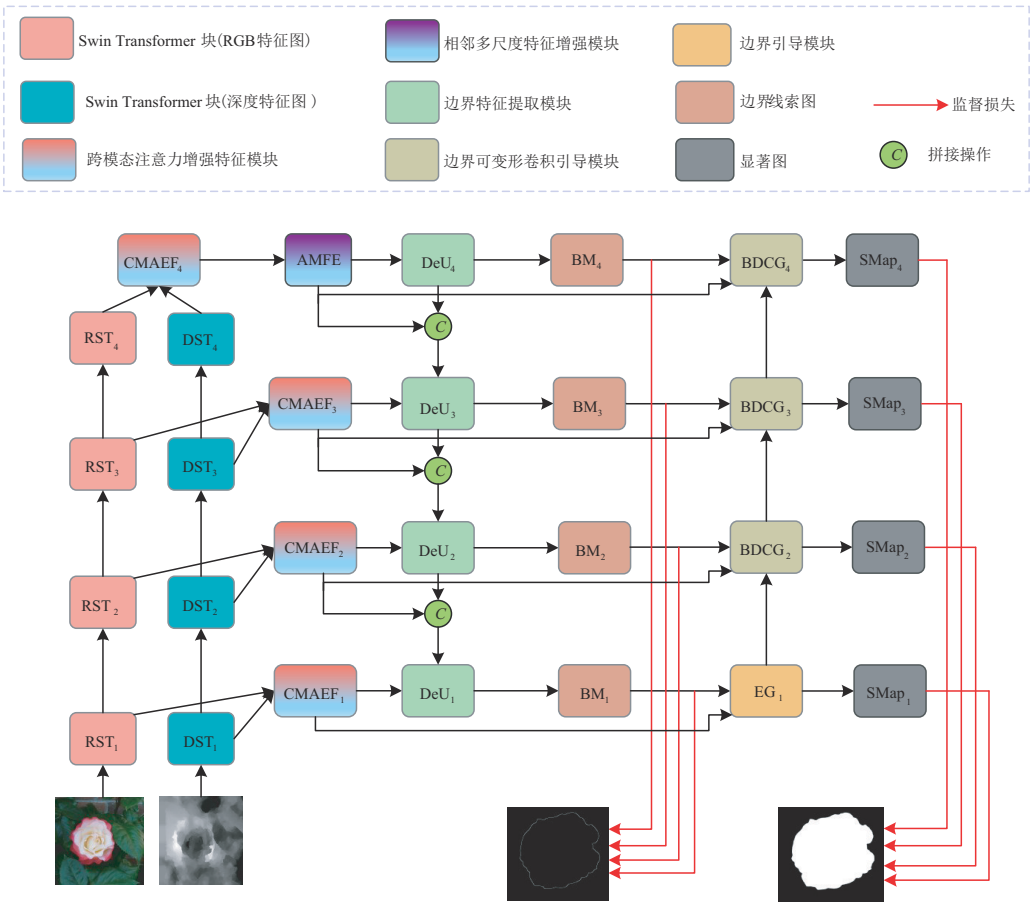


图1　跨模态融合和边界可变形卷积引导的RGB-D显著性目标检测方法

其中,Conv表示卷积运算,上标RC、DC、RCD和DCR表示运算得到不同特征图的标识符,⊗表示对应元素相乘.

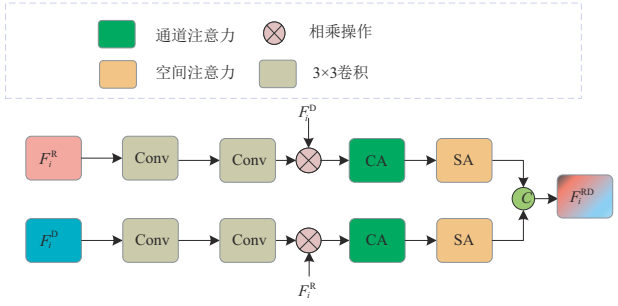


图2　CMAEF模块

正如上文所述,深度图的质量、RGB图像光照和背景色彩等对检测显著物影响较大. 本文利用通道-空间注意力模块[20]来解决上述问题. 通道注意力计算特征图中每个通道的权重,更好地衡量了每个特征通道的重要性,空间注意力则衡量的是不同空间位置的重要性,可以更好地定位出目标的空间区域. 本文通过通道注意力和空间注意力来抑制RGB模态和深度图模态噪声信息和冗余信息,以提高显著性目标特征表达能力,最终通过级联两种模态特征图得到跨模态融合特征(F_i^{RD}). 如式(5)~(7)所示:

$$F_i^{\mathrm{SCRD}} = \mathrm{SA}\left(\mathrm{CA}\left(F_i^{\mathrm{RCD}}\right)\right) \tag{5}$$

$$F_i^{\mathrm{SCDR}} = \mathrm{SA}\left(\mathrm{CA}\left(F_i^{\mathrm{DCR}}\right)\right) \tag{6}$$

$$F_i^{\mathrm{RD}} = C\left(F_i^{\mathrm{SCRD}}, F_i^{\mathrm{SCDR}}\right) \tag{7}$$

其中,SA、CA表示空间注意力和通道注意力,上标SCRD、SCDR和RD表示运算得到不同特征图的标识符.

2.3　相邻多尺度特征增强模块

深层特征图包含丰富的显著性目标位置信息,大卷积核能够获得较大的感受野,有利于捕获大目标的特征,而小卷积核获得较小的感受野,有利于捕获小目标的特征. 很多模型都在网络深层应用多尺度特征提取模块来增强全局上下文特征信息的提取能力,但直接级联多个扩张卷积后的特征图并未考虑各分支之间的语义相关性,不能很好地处理自然场景下显著性目标的数量和大小的多样变化. 因此本文提出一个AMFE模块,利用相邻层次特征交互实现显著性特征互补,从而获取更丰富的全局上下文特征信息,对于捕获具有不同尺寸和不确定数量的显著物至关重要,详细

结构如图3所示.

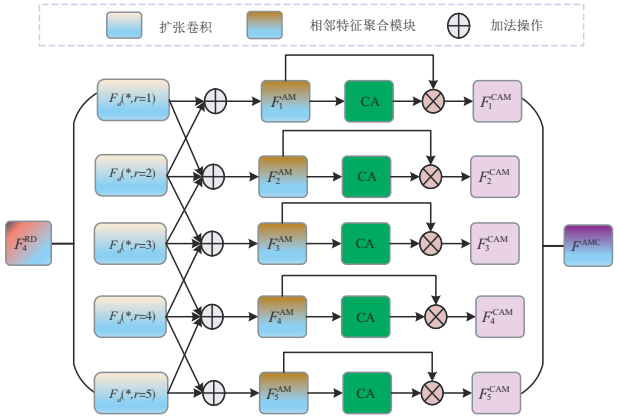


图3 AMFE模块

具体来说,AMFE模块由五个分支组成,每一个分支采用的扩张因子分别为1、2、3、4、5. 输入到该模块的深层特征图为F_4^{RD}. 中间三个分支(F_2^{AM}、F_3^{AM}和F_4^{AM})包含自身分支和与它相邻的两个分支,而F_1^{AM}和F_5^{AM}只包含两个分支:一个自身分支和一个相邻分支. 具体公式表示如下:

$$F_r^{\mathrm{AM}}=\begin{cases}F_d\left(F_4^{\mathrm{RD}},r\right)+F_d\left(F_4^{\mathrm{RD}},r+1\right), r=1\\ \begin{pmatrix}F_d\left(F_4^{\mathrm{RD}},r-1\right)+F_d\left(F_4^{\mathrm{RD}},r\right)\\ +F_d\left(F_4^{\mathrm{RD}},r+1\right)\end{pmatrix}, r\in\{2,3,4\}\\ F_d\left(F_4^{\mathrm{RD}},r-1\right)+F_d\left(F_4^{\mathrm{RD}},r+1\right), r=5\end{cases} \tag{8}$$

其中,F_d表示扩张卷积,r表示扩张因子,上标AM表示相邻特征聚合模块的标识符.

这使得F_i^{AM}存在更多的显著性特征,但是同时会造成显著性目标细微特征难以分辨,因此利用通道注意力模块来获得各个分支的权重,从而达到高效特征筛选的目的,然后拼接所有特征图作为最终输出. 其公式如下:

$$F_r^{\mathrm{CAM}}=F_r^{\mathrm{AM}}\otimes\mathrm{CA}\left(F_r^{\mathrm{AM}}\right), r\in\{1,2,3,4,5\} \tag{9}$$

$$F^{\mathrm{AMC}}=C\left(F_1^{\mathrm{CAM}},\cdots,F_r^{\mathrm{CAM}}\right), r=5 \tag{10}$$

其中,上标CAM和AMC表示运算得到不同特征图的标识符.

2.4 边界特征提取模块

利用RGB数据集构建的显著性模型采用边界线索图引导模型可以生成准确的显著图,但是在某些复杂场景下,同样存在边界模糊问题. 受文献[18]的启发,本文通过边界特征提取模块将目标特征信息逐步转化为边界特征信息,从而获得边界线索图,然后利用生成的边界线索图引导模型进行后续解码.

具体来说,第i个边界特征提取模块被定义为DeU_i(其中,第i个模块提取的特征图记为F_i^{BD},$i\in\{1,2,\cdots,4\}$),详细结构如图4所示,对于DeU_i来说,其输入为前一个边界特征提取模块DeU_{i+1}的输出以及第i个跨模态融合特征图F_i^{RD},将二者进行拼接后输入到Conv、BN层和ReLU激活函数、接着通过双线性插值和卷积将特征图调整为通道数为1且大小与输入模型图片相同的特征图,并最终经过Sigmoid函数得到显著物的边界线索图(Boundary Map,BM). 此外,为了准确地生成边界线索图,本文使用监督损失来强化边界特征提取模块对目标边界特征的提取,为了确保显著物边缘结构的完整性,将跨模态融合特征通过级联方式进行复用,再输入到下一个边界特征提取模块中. 所生成的边界线索图如图5所示.

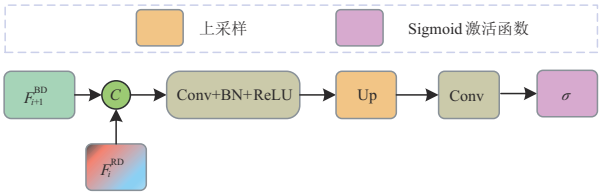


图4 DeU模块

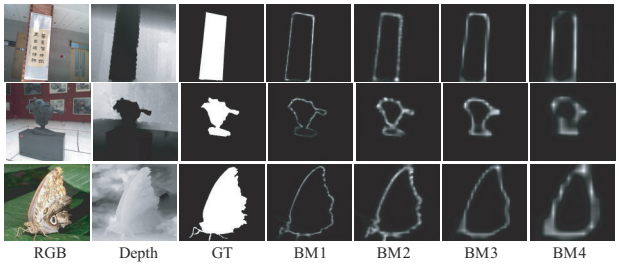


图5 DeU模块生成的BM图

上述过程如下公式所示:

$$F_i^{\mathrm{BD}}=\begin{cases}\mathrm{Conv}\left(F^{\mathrm{AMC}}\right), i=4\\ C\left(C\left(\mathrm{Up}\left(F_{i+1}^{\mathrm{BD}}\right),F^{\mathrm{AMC}}\right),F_i^{\mathrm{RD}}\right), i=3\\ C\left(C\left(\mathrm{Up}\left(F_{i+1}^{\mathrm{BD}}\right),F_i^{\mathrm{RD}}\right),F_i^{\mathrm{RD}}\right), i\in\{1,2\}\end{cases} \tag{11}$$

$$\mathrm{BM}_i=\sigma\left(\mathrm{Conv}\left(\mathrm{Up}_{\times 2^i}\left(f\left(F_i^{\mathrm{BD}}\right)\right)\right)\right), i\in\{1,2,3,4\} \tag{12}$$

其中,Up表示双线性插值操作,下标表示上采样的倍数,f表示Conv、BN层和ReLU激活函数,σ表示Sigmoid激活函数.

本文采用交叉熵损失函数(Binary Cross Entropy,BCE)优化边界线索图,即

$$L_i^{\mathrm{bm}}=-\sum_{j=1}^{W\times H}\left\{\mathrm{GT}_{e+}(j)\log\left(\mathrm{BM}_i(j)\right)+\mathrm{GT}_{e-}(j)\log\left(1-\mathrm{BM}_i(j)\right)\right\} \tag{13}$$

其中,j表示像素值,GT_{e+}和GT_{e-}分别表示显著边缘像素和背景像素,L_i^{bm}表示第i个边界特征提取模块的损失函数.

2.5 边界可变形卷积引导模块

通常情况下,自然图像中许多目标边界轮廓复杂多变,采用常规的卷积操作很难精确地提取边界特征,其原因是常规卷积在构建模型变换时受限于固定的几何结构,这种局限性决定了卷积单元仅能对输入图像进行固定位置采样,导致所提取的特征表示能力较弱.因此,本文提出一个BDCG模块如图6所示,该模块能够有效地提取目标的复杂边界特征,从而提高形变的建模能力,能够自适应地提高显著物的特征表示能力.

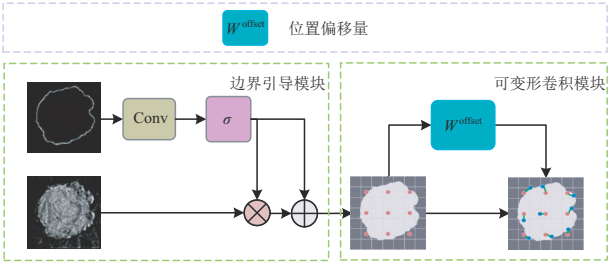


图6　BDCG模块

边界特征信息有助于生成具有精细边界的显著图,本文利用生成的边界线索图引导跨模态融合特征图逐步生成显著图.具体来说,首先将边界线索图输入到3×3的卷积层和Sigmoid激活函数中,从而获得归一化的特征图,该特征图可以视为特征级的注意力映射,将其与跨模态融合特征图相乘再相加,然后送入可变形卷积模块[21]中,最后进行上采样后进入下一层解码模块:

$$F_i^{AE}=\sigma\left(\mathrm{Conv}\left(\mathrm{Ds}_{\times 2^i}\left(\mathrm{BM}_i\right)\right)\right), i\in\{1,2,3,4\} \quad (14)$$

$$F_i^{ARD}=F_i^{AE}\otimes F_i^{RD}+F_i^{RD}, i\in\{1,2,3,4\} \quad (15)$$

$$F_i^{DF}=\mathrm{DFC}\left(F_i^{ARD}\right), i\in\{2,3,4\} \quad (16)$$

其中,Ds表示下采样操作,下标表示下采样的倍数,DFC表示可变形卷积.上标AE、ARD、DF表示运算得到不同特征图的标识符,本文在第2、3和4个显著图解码器块使用可变形卷积,在第1个显著图解码器块未使用可变形卷积.

本文设计的边界可变形卷积引导模块可以将显著性区域特征和边界特征相融合,利用边界线索图屏蔽非显著性区域,并利用可变形卷积提取显著物不规则位置特征以提高复杂场景中模型的性能.

2.6 损失函数

在显著图分支中,使用BCE损失函数和交并比(Intersection-Over-Union,IOU)损失函数优化模型,BCE损失函数可以表示为

$$L_i^{BCE}=-\sum_{x,y}G_{x,y}\log\left(P_{x,y}\right)+\left(1-G_{x,y}\right)\log\left(1-P_{x,y}\right) \quad (17)$$

其中,P和G表示预测的显著图和真值图,$P_{x,y}$和$G_{x,y}$分别表示预测的前景概率值和真值图的前景概率值,L_i^{BCE}表示第i层显著图分支的BCE损失函数.

IOU损失函数定义如下:

$$L_i^{IOU}=1-\frac{\sum_{x=1}^{H}\sum_{y=1}^{W}P_{x,y}\cdot G_{x,y}}{\sum_{x=1}^{H}\sum_{y=1}^{W}\left(P_{x,y}+G_{x,y}-P_{x,y}\cdot G_{x,y}\right)} \quad (18)$$

其中,L_i^{IOU}表示第i层显著图分支的IOU损失函数.

则整个模型的总损失函数为

$$L=\sum_{i=1}^{4}\left(L_i^{BCE}+L_i^{IOU}+L_i^{bm}\right) \quad (19)$$

3 实验结果与分析

3.1 数据集

为充分验证所提模型的有效性,本文在六个公共基准的RGB-D数据集进行了广泛的实验比较,这六个数据集分别是:LFSD[22]、NLPR[23]、NJU2K[24]、SSD[25]、SIP[19]和STERE[26].LFSD数据集是由Lytro相机所采集,其中包括100张图片对,显著性区域由3个人共同分割确定.NLPR包括1 000张分辨率为640×480的图像对,部分图像中存在多个显著物体,深度图通过Microsoft Kinect在不同的光照条件下所获得.NJU2K是最大的RGB-D数据集,包含1 985张图像对.SSD从室内外场景的三部立体电影中选取80张图像构成的测试集,每张图像的分辨率高达960×1 080,但深度图质量较为粗糙.SIP采用华为双摄像头智能手机捕获的929张图像对,图像大小分辨率为992×744,图像包含来自各种视角、姿势、遮挡、低光照和复杂背景的真实场景.STERE是第一个立体图像数据集,共包含1 000对双目图像,这些图像主要从互联网采集.

3.2 评价指标

F-measure(F_β)[27]:F_β是一种综合度量指标,由Precision与Recall加权调和平均而得,即

$$F_\beta=\frac{\left(1+\beta^2\right)\times\mathrm{Precision}\times\mathrm{Recall}}{\beta^2\times\mathrm{Precision}+\mathrm{Recall}} \quad (20)$$

其中,β^2被设置成0.3,F-measure的最大值、平均值和自适应值分别被定义为F_β^{max}、F_β^{mean}和F_β^{adp}.

Mean Absolute Error(M):M是检测的显著图与人工标注的真值图平均绝对误差,其数值越小,表明算法越好,计算公式如下:

$$M=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left|\mathrm{SP}(i,j)-\mathrm{GT}(i,j)\right| \quad (21)$$

其中,SP为检测的显著图,W和H为图像的宽和高,i和j为像素点的横纵坐标.

E-measure(E_ξ)[28]:E_ξ同时结合局部像素显著值和图像级平均显著值来评估预测的显著图与真值图之间

的相似性,表示如下:

$$E_{\xi}=\frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\boldsymbol{V}\left(\mathrm{SP}\left(i,j\right),\mathrm{GT}\left(i,j\right)\right) \tag{22}$$

其中,$\boldsymbol{V}$表示增强对其矩阵,E-measure 的最大值、平均值和自适应值分别被定义为$E_{\xi}^{\max}$、E_{ξ}^{mean}和E_{ξ}^{adp}.

3.3 实验细节

本文模型通过深度学习框架 pytorch 实现,使用机器的 CPU 型号为 Xeon Platinum 8260C,显卡型号为 RTX3090. 同以往方法[9,11]一样,本文以 1 485 张 NJU2K 图像对以及 700 张 NLPR 图像对为训练集,其余 NJU2K 图像对以及 NLPR 图像对做为测试集. 输入模型中图像对大小为 384×384,训练前采用随机裁剪、水平翻转等方法对训练数据进行增强以缓解过度拟合问题. 特征编码骨干初始参数从预训练 ST 中获得,本文设定 ST 嵌入维度为 128、窗口大小为 12. 初始学习率定为 0.000 1,每三轮迭代改变学习率一次,衰减率设置为 0.9,采用 Adam 作为模型的优化器,设置数据批次大小为 6,避免模型出现过拟合的情况,本文在验证集上选择最优模型,我们设置总训练轮次为 200,最终保存的轮次为 60,训练总时间约为 7 小时. 此外,本文方法无需任何预处理和后处理.

3.4 实验比较

本文方法与 25 个 RGB-D 显著性目标检测方法进行了比较,包括 BBS[29]、HDFN[30]、ICNet[31]、CMW[32]、VST[33]、DSAM[34]、DCF[12]、3DNet[13]、HAIN[11]、CDNet[35]、UTA[36]、BTS[37]、DQIF[38]、TTNet[39]、CMDI[40]、SSP[10]、EENet[41]、DIGR[42]、CCAF[43]、LDCM[9]、UIFN[44]、C^{2}DF[45]、CMVM[46]、Swin[47]、SPNet[48]. 所有对比的显著图均由原作者直接提供或者由原作者提供训练好的模型直接生成.

26 种显著性检测方法在 6 个测试集的评价指标如表 1 和 2 所示. 通过 LFSD 数据集对$F_{\beta}^{\max}$、$F_{\beta}^{\mathrm{mean}}$、$F_{\beta}^{\mathrm{adp}}$和$M$进行分析,可以发现本文模型的这 4 项指标优于其他模型,并且达到最优结果,$F_{\beta}^{\max}$、$F_{\beta}^{\mathrm{mean}}$、$F_{\beta}^{\mathrm{adp}}$较次优结果分别提高了 0.005、0.01、0.001,M较次优结果降低 0.003. 在 NLPR、NJU2K、STERE 三个数据集上,本文的 7 项指标均达到最优效果,并且三个数据集中大多图像具有复杂的背景信息,表明本文模型较其他模型更能适应于复杂环境下的检测. 在 SIP 数据集上,本文各项指标也都取得较好的检测结果,虽然本文的指标低于 Swin 和

表1 数据集 LFSD、NLPR 和 SIP 在 8 种评价指标下的定量比较

方法	LFSD							NLPR							SIP						
	$F_{\beta}^{\max}\uparrow$	$F_{\beta}^{\mathrm{mean}}\uparrow$	$F_{\beta}^{\mathrm{adp}}\uparrow$	$M\downarrow$	$E_{\xi}^{\max}\uparrow$	$E_{\xi}^{\mathrm{mean}}\uparrow$	$E_{\xi}^{\mathrm{adp}}\uparrow$	$F_{\beta}^{\max}\uparrow$	$F_{\beta}^{\mathrm{mean}}\uparrow$	$F_{\beta}^{\mathrm{adp}}\uparrow$	$M\downarrow$	$E_{\xi}^{\max}\uparrow$	$E_{\xi}^{\mathrm{mean}}\uparrow$	$E_{\xi}^{\mathrm{adp}}\uparrow$	$F_{\beta}^{\max}\uparrow$	$F_{\beta}^{\mathrm{mean}}\uparrow$	$F_{\beta}^{\mathrm{adp}}\uparrow$	$M\downarrow$	$E_{\xi}^{\max}\uparrow$	$E_{\xi}^{\mathrm{mean}}\uparrow$	$E_{\xi}^{\mathrm{adp}}\uparrow$
C^{2}DF	0.867	0.859	0.862	0.066	0.902	0.897	0.883	0.917	0.903	0.894	0.022	0.961	0.955	0.957	0.877	0.864	0.864	0.053	0.916	0.912	0.915
CCAF	0.832	0.824	0.831	0.088	0.876	0.865	0.871	0.909	0.891	0.877	0.027	0.957	0.949	0.950	0.880	0.864	0.863	0.054	0.917	0.909	0.914
CDNet	0.761	0.750	0.779	0.109	0.824	0.809	0.833	0.921	0.898	0.878	0.024	0.964	0.953	0.952	0.878	0.861	0.860	0.056	0.914	0.905	0.911
CMDI	0.874	0.863	0.870	0.063	0.914	0.905	0.904	0.916	0.898	0.883	0.024	0.960	0.951	0.952	0.884	0.868	0.866	0.054	0.915	0.908	0.910
DCF	0.841	0.835	0.841	0.075	0.883	0.877	0.883	0.912	0.897	0.887	0.022	0.963	0.957	0.956	0.884	0.874	0.874	0.052	0.922	0.915	0.920
DQIF	0.865	0.852	0.863	0.069	0.902	0.895	0.891	0.912	0.893	0.880	0.024	0.961	0.952	0.952	0.889	0.873	0.874	0.049	0.926	0.918	0.922
EENet	0.841	0.825	0.820	0.080	0.888	0.879	0.880	0.908	0.886	0.874	0.025	0.961	0.950	0.951	0.880	0.862	0.861	0.053	0.916	0.908	0.914
HDFN	0.862	0.833	0.822	0.077	0.896	0.882	0.879	0.917	0.892	0.889	0.023	0.963	0.953	0.957	0.894	0.873	0.872	0.048	0.930	0.921	0.923
HAIN	0.853	0.842	0.852	0.080	0.886	0.878	0.877	0.915	0.901	0.897	0.024	0.960	0.955	0.957	0.892	0.876	0.875	0.053	0.922	0.916	0.919
ICNet	0.870	0.852	0.861	0.071	0.903	0.892	0.891	0.908	0.885	0.870	0.028	0.952	0.941	0.944	0.857	0.835	0.836	0.069	0.903	0.891	0.899
LDCM	0.874	0.847	0.832	0.069	0.909	0.891	0.887	0.905	0.872	0.849	0.029	0.954	0.936	0.937	0.872	0.842	0.837	0.062	0.911	0.894	0.901
SSP	0.729	0.719	0.739	0.126	0.798	0.779	0.802	0.899	0.882	0.879	0.027	0.949	0.941	0.946	0.874	0.862	0.868	0.058	0.910	0.900	0.910
UIFN	0.879	0.869	0.875	0.058	0.918	0.912	0.913	0.911	0.896	0.885	0.023	0.959	0.953	0.952	0.830	0.821	0.821	0.080	0.873	0.861	0.871
CMVM	0.877	0.861	0.839	0.064	0.912	0.905	0.893	0.921	0.903	0.896	0.021	0.966	0.959	0.960	0.902	0.886	0.880	0.043	0.935	0.928	0.928
BBS	0.858	0.843	0.857	0.072	0.901	0.884	0.889	0.918	0.896	0.882	0.023	0.961	0.951	0.952	0.883	0.869	0.872	0.055	0.922	0.906	0.916
CMW	0.882	0.862	0.870	0.067	0.912	0.900	0.891	0.903	0.878	0.859	0.029	0.951	0.940	0.940	0.874	0.852	0.851	0.062	0.913	0.900	0.906
3DNet	0.854	0.843	0.854	0.074	0.891	0.885	0.876	0.919	0.903	0.892	0.022	0.965	0.957	0.958	0.889	0.874	0.874	0.048	0.924	0.919	0.920
BTS	0.873	0.859	0.869	0.070	0.906	0.890	0.893	0.923	0.904	0.892	0.023	0.965	0.955	0.956	0.901	0.886	0.885	0.044	0.933	0.924	0.926
SPNet	0.755	0.740	0.759	0.120	0.833	0.800	0.816	0.919	0.907	0.904	0.021	0.962	0.957	0.958	0.904	0.893	0.893	0.043	0.933	0.929	0.930
DSAM	0.889	0.878	0.881	0.055	0.924	0.918	0.923	0.906	0.896	0.890	0.024	0.952	0.948	0.949	0.875	0.866	0.864	0.057	0.912	0.907	0.908
DIGR	0.865	0.851	0.863	0.067	0.902	0.893	0.888	0.928	0.905	0.890	0.023	0.965	0.955	0.956	0.897	0.880	0.879	0.052	0.925	0.913	0.918
UTA	0.835	0.827	0.832	0.089	0.874	0.871	0.853	0.926	0.917	0.917	0.020	0.965	0.961	0.962	0.884	0.871	0.872	0.048	0.926	0.923	0.921
VST	0.892	0.871	0.863	0.054	0.928	0.917	0.913	0.920	0.897	0.882	0.024	0.962	0.950	0.954	0.915	0.894	0.889	0.040	0.944	0.933	0.937
TTNet	0.870	0.864	0.868	0.066	0.905	0.901	0.889	0.924	0.910	0.909	0.020	0.966	0.961	0.960	0.899	0.891	0.892	0.043	0.930	0.926	0.924
Swin	0.889	0.874	0.879	0.059	0.921	0.912	0.899	0.936	0.920	0.908	0.018	0.974	0.966	0.967	0.927	0.913	0.912	0.035	0.950	0.943	0.943
Ours	0.894	0.884	0.889	0.051	0.926	0.921	0.905	0.938	0.927	0.925	0.016	0.974	0.970	0.973	0.914	0.905	0.905	0.041	0.937	0.932	0.931

注:前三个结果以红色、绿色和蓝色显示.

表2　数据集 NJU2K、STERE和SSD在8种评价指标下的定量比较

方法	NJU2K							STERE							SSD						
	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$F_\beta^{adp}\uparrow$	$M\downarrow$	$E_\xi^{max}\uparrow$	$E_\xi^{mean}\uparrow$	$E_\xi^{adp}\uparrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$F_\beta^{adp}\uparrow$	$M\downarrow$	$E_\xi^{max}\uparrow$	$E_\xi^{mean}\uparrow$	$E_\xi^{adp}\uparrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$F_\beta^{adp}\uparrow$	$M\downarrow$	$E_\xi^{max}\uparrow$	$E_\xi^{mean}\uparrow$	$E_\xi^{adp}\uparrow$
C^2DF	0.909	0.897	0.897	0.039	0.942	0.936	0.919	0.897	0.881	0.880	0.038	0.943	0.936	0.925	0.860	0.846	0.845	0.048	0.917	0.910	0.910
CCAF	0.910	0.895	0.895	0.038	0.944	0.936	0.919	0.887	0.869	0.868	0.045	0.934	0.926	0.920	0.749	0.738	0.754	0.075	0.861	0.835	0.869
CDNet	0.879	0.863	0.864	0.048	0.927	0.916	0.907	0.904	0.884	0.882	0.039	0.946	0.936	0.927	0.813	0.794	0.789	0.060	0.900	0.884	0.901
CMDI	0.921	0.908	0.907	0.035	0.951	0.944	0.925	0.776	0.763	0.775	0.077	0.872	0.860	0.875	0.846	0.831	0.827	0.056	0.899	0.889	0.899
DCF	0.915	0.902	0.901	0.036	0.950	0.944	0.924	0.901	0.886	0.883	0.039	0.945	0.939	0.929	0.851	0.835	0.826	0.050	0.909	0.902	0.897
DQIF	0.907	0.893	0.890	0.043	0.946	0.936	0.918	0.904	0.878	0.875	0.040	0.948	0.936	0.918	0.781	0.764	0.762	0.072	0.877	0.860	0.876
EENet	0.912	0.892	0.857	0.038	0.950	0.940	0.916	0.895	0.875	0.847	0.041	0.940	0.930	0.916	0.847	0.827	0.816	0.052	0.914	0.900	0.899
HDFN	0.910	0.889	0.852	0.039	0.944	0.934	0.910	0.900	0.867	0.843	0.042	0.943	0.929	0.915	0.870	0.846	0.831	0.046	0.925	0.911	0.906
HAIN	0.915	0.903	0.900	0.038	0.944	0.938	0.922	0.906	0.887	0.885	0.040	0.944	0.936	0.925	0.838	0.829	0.836	0.052	0.903	0.894	0.901
ICNet	0.890	0.867	0.863	0.052	0.924	0.912	0.905	0.898	0.870	0.865	0.045	0.942	0.927	0.915	0.841	0.816	0.799	0.064	0.902	0.888	0.879
LDCM	0.910	0.875	0.821	0.046	0.947	0.925	0.892	0.906	0.869	0.837	0.043	0.946	0.926	0.912	0.867	0.835	0.816	0.054	0.921	0.899	0.895
SSP	0.902	0.887	0.853	0.043	0.938	0.930	0.908	0.883	0.867	0.857	0.047	0.930	0.921	0.919	0.763	0.751	0.756	0.079	0.866	0.841	0.864
UIFN	0.910	0.898	0.896	0.039	0.942	0.936	0.921	0.883	0.869	0.867	0.047	0.933	0.925	0.915	0.854	0.837	0.829	0.048	0.918	0.909	0.902
CMVM	0.924	0.910	0.872	0.032	0.924	0.910	0.872	0.912	0.893	0.869	0.034	0.953	0.945	0.931	0.859	0.846	0.835	0.042	0.926	0.916	0.911
BBS	0.920	0.903	0.902	0.035	0.949	0.938	0.924	0.903	0.883	0.885	0.041	0.942	0.929	0.925	0.859	0.844	0.849	0.044	0.919	0.905	0.912
CMW	0.902	0.882	0.880	0.046	0.936	0.923	0.911	0.901	0.873	0.869	0.043	0.944	0.929	0.917	0.871	0.837	0.820	0.051	0.930	0.906	0.900
3DNet	0.914	0.901	0.901	0.037	0.947	0.940	0.918	0.906	0.888	0.886	0.037	0.947	0.939	0.927	0.851	0.830	0.829	0.048	0.904	0.892	0.905
BTS	0.927	0.911	0.908	0.035	0.957	0.947	0.926	0.911	0.890	0.889	0.038	0.949	0.938	0.931	0.758	0.741	0.741	0.077	0.867	0.830	0.867
SPNet	0.928	0.917	0.917	0.029	0.957	0.952	0.932	0.906	0.890	0.888	0.037	0.949	0.943	0.930	0.863	0.853	0.855	0.044	0.920	0.915	0.911
DSAM	0.907	0.898	0.897	0.040	0.938	0.934	0.922	0.900	0.892	0.892	0.039	0.942	0.937	0.927	0.863	0.851	0.847	0.048	0.913	0.909	0.903
DSAM	0.907	0.898	0.897	0.040	0.938	0.934	0.922	0.900	0.892	0.892	0.039	0.942	0.937	0.927	0.863	0.851	0.847	0.048	0.913	0.909	0.903
DIGR	0.936	0.919	0.917	0.028	0.963	0.953	0.928	0.914	0.892	0.889	0.037	0.954	0.941	0.927	0.846	0.830	0.823	0.052	0.898	0.889	0.889
UTA	0.914	0.904	0.908	0.038	0.952	0.948	0.919	0.912	0.903	0.905	0.033	0.949	0.946	0.928	0.842	0.834	0.838	0.049	0.898	0.895	0.892
VST	0.919	0.898	0.856	0.035	0.951	0.939	0.913	0.907	0.879	0.843	0.038	0.951	0.937	0.917	0.876	0.849	0.818	0.045	0.935	0.920	0.907
TTNet	0.926	0.918	0.919	0.030	0.955	0.951	0.924	0.911	0.893	0.893	0.033	0.953	0.948	0.927	0.873	0.865	0.865	0.041	0.934	0.929	0.926
Swin	0.938	0.924	0.921	0.027	0.963	0.956	0.933	0.918	0.896	0.893	0.033	0.956	0.947	0.929	0.878	0.866	0.863	0.040	0.925	0.917	0.912
Ours	0.941	0.932	0.932	0.025	0.964	0.960	0.938	0.922	0.907	0.908	0.030	0.956	0.951	0.932	0.882	0.869	0.869	0.037	0.928	0.924	0.923

注:前三个结果以红色、绿色和蓝色显示.

VST模型,但是也超过了大部分的模型.在SSD数据集中的F_β^{max}、F_β^{mean}和F_β^{adp}三项指标中,本文模型取得了最好的结果,且与Swin相比较分别高出了0.004、0.003、0.004.将全部数据集与评取得了最好的结果,且与次优结果相比较分别高出价指标进行综合分析可以发现:本文方法在六个数据集测评的7项指标均取得了较好的结果,表明本文方法具有较好的检测性能,同时上述对比结果也表明本文方法具有较强的泛化能力.

图7展示了本方法与其他12种方法的可视化显著图.从1~6行图像可以看出显著性目标边界轮廓较复杂,对于第2行的目标,许多模型不但预测的显著物边界不够精确,而且还遗漏了枝条的信息,相比较之下,由于本文将边界线索图引入模型中,从而能够使生成的显著图具有清晰完整的边界轮廓.由第10、11、12行可看出显著物和背景颜色极其相似,在第12行中,许多方法生成的显著图均会丢失显著物的脚,尽管VST方法预测出显著物的脚,但是其边界轮廓也比较模糊,而本文所预测的显著图具有完整和清晰的边界.另外,受益于相邻多尺度特征增强模块的效果,本文方法在对大目标(第7和8行)、小目标(第9行)和多目标(最后1行)的预测上也具有良好的检测结果.综合来看,本文的可视化显著图要显著优于其他模型,并且比其他模型更接近真值图.

3.5　消融实验

3.5.1　各个模块的有效性

通过消融实验来了解不同模块对模型所产生的效果,主要考虑如下四种模型:(1)基线BL模型(U-Net编码器-解码器结构,定义为A模型);(2)BL+CMAEF(定义为B模型);(3)BL+CMAEF+AMFE(定义为C模型);(4)BL+CMAEF+AMFE+BDCG(定义为D模型).

实验结果如表3所示,从6个数据集上我们可以观察到:从模型(A)~(D)各个指标均有不同程度的上升或者持平.另外,图8还展示了4种模型所产生的显著图,从图8(e)~(h)中可以直观地看到显著图效果逐渐接近真值图.上述定量与定性实验表明将每个模块加入模型中都能提高检测的性能.

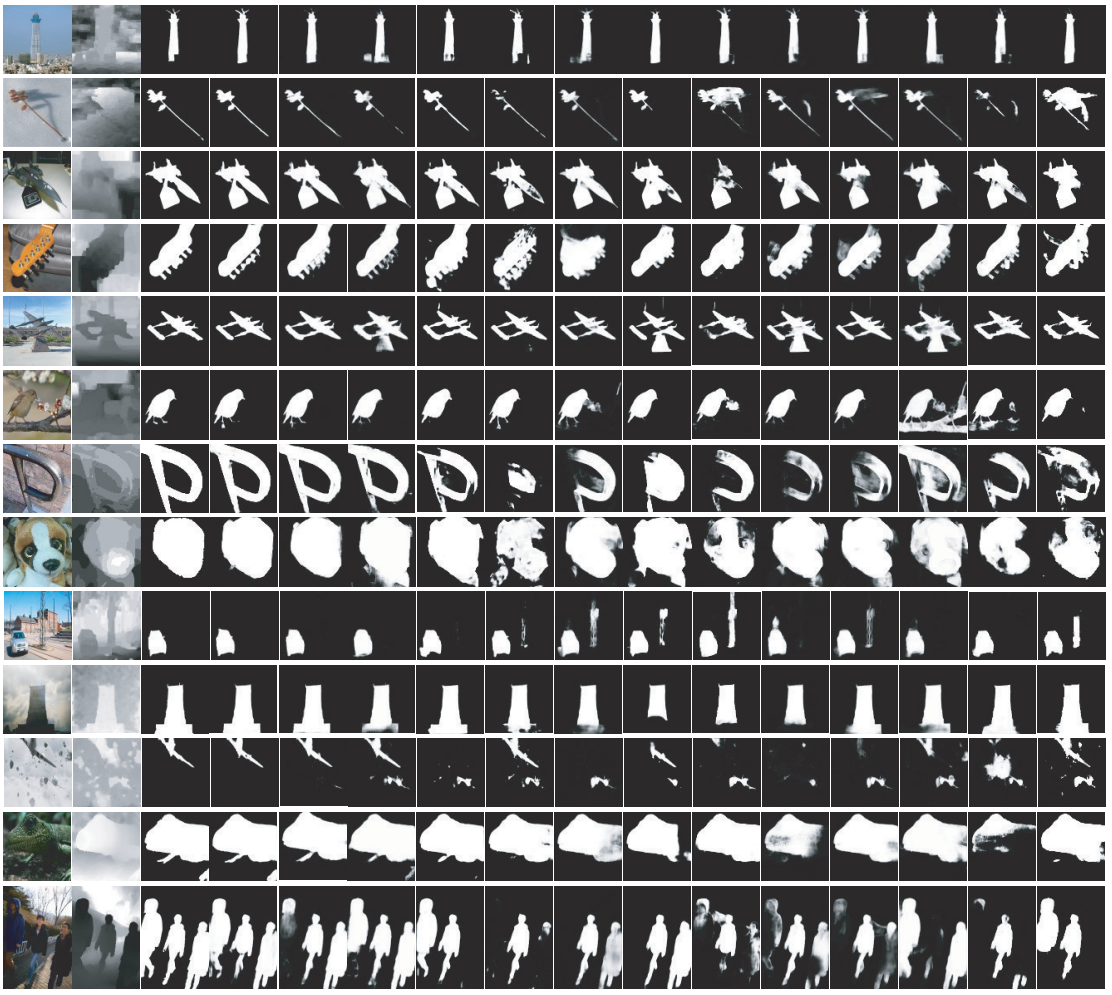
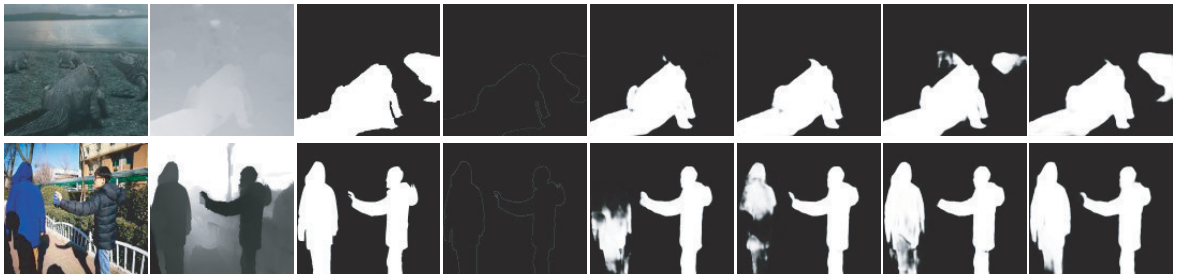


图7 本文方法与其他12种先进方法的可视化显著图

表3 每个模块的有效性

模型	BL	CMAEF	AMFE	BDCG	LFSD			NLPR			SIP			NJU2K			STERE			SSD		
					$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$
A	√				0.886	0.914	0.055	0.930	0.969	0.020	0.907	0.928	0.045	0.933	0.954	0.028	0.914	0.950	0.033	0.864	0.915	0.044
B	√	√			0.889	0.918	0.054	0.933	0.972	0.018	0.910	0.930	0.043	0.936	0.959	0.026	0.917	0.953	0.032	0.872	0.919	0.042
C	√	√	√		0.890	0.920	0.053	0.934	0.971	0.017	0.911	0.932	0.043	0.938	0.961	0.026	0.918	0.952	0.031	0.876	0.922	0.041
D	√	√	√	√	**0.894**	**0.926**	**0.051**	**0.938**	**0.974**	**0.016**	**0.914**	**0.937**	**0.041**	**0.941**	**0.964**	**0.025**	**0.922**	**0.956**	**0.030**	**0.882**	**0.928**	**0.037**



(*a*) RGB (*b*) Depth (*c*) GT (*d*) BM (*e*) BL (*f*) BL+CMAEF (*g*) BL+CMAEF +AMFE (*h*) BL+CMAEF +AMFE+BGCG

图8 可视化不同模块效果的显著图

3.5.2 CMAEF模块组件的有效性

本文对CMAEF模块组件的有效性进行了研究,我们探究了三种情况,分别移除特征交互模块(w/o FI)、通道注意力(w/o CA)和空间注意力(w/o SA). 实验结果如表4所示,可以看到特征交互模块被移除后,$F_\beta^{\max}$和$E_\xi^{\max}$在LFSD、NLPR和SIP这3个数据集上的性能较CMAEF有所下降. 另外,发现当CMAEF模块去除CA与SA时,各数据集上指标均有所下降,该结果表明各组件对于检测均存在一定的影响,因此完整地利用这三个部分能够取得最佳效果.

表4　CMAEF模块组件的有效性

模块	LFSD			NLPR			SIP			SSD		
	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$
w/o FI	0.892	0.921	0.052	0.937	0.974	0.016	0.910	0.932	0.043	0.884	0.930	0.036
w/o CA	0.892	0.920	0.053	0.937	0.974	0.017	0.911	0.933	0.042	0.880	0.926	0.038
w/o SA	0.891	0.919	0.053	0.937	0.973	0.017	0.911	0.932	0.043	0.881	0.926	0.038
CMAEF	**0.894**	**0.926**	**0.051**	**0.938**	**0.974**	**0.016**	**0.914**	**0.937**	**0.041**	**0.882**	**0.928**	**0.037**

3.5.3 AMFE模块组件的有效性

为验证相邻多尺度特征增强模块的效果,本文对比了文献[49]、文献[14]和文献[15]提出的各种多尺度特征提取模块,分别对应表5中的PPM、PAFEM和MSR模块. 从表中数据可以看出:本文提出的AMFE模块在LFSD、SIP和SSD数据集上的评价指标均高于其他文献提出的模块,表明本文提出的AFME模块能捕获更丰富的全局上下文特征信息,从而更加有效地定位显著性目标位置. 另外,本文对相邻特征聚合模块(F_i^{AM})与CA模块的重要性进行了研究,由表5可知,在移除任意一个模块后,各数据集指标均有所下降,这表明相邻特征聚合模块和通道注意对于捕获多尺度特征至关重要.

表5　AMFE模块组件的有效性

模块	LFSD			NLPR			SIP			SSD		
	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$
PPM	0.890	0.922	0.055	0.936	0.973	0.017	0.911	0.936	0.041	0.879	0.925	0.038
PAFEM	0.892	0.924	0.051	0.938	0.973	0.017	0.912	0.935	0.042	0.880	0.924	0.038
MSR	0.888	0.919	0.055	0.936	0.971	0.017	0.911	0.935	0.042	0.879	0.924	0.038
AMFE	**0.894**	**0.926**	**0.051**	**0.938**	**0.974**	**0.016**	**0.914**	**0.937**	**0.041**	**0.882**	**0.928**	**0.037**
w/o F_i^{AM}	0.892	0.923	0.053	0.935	0.971	0.018	0.911	0.934	0.042	0.879	0.924	0.038
w/o CA	0.893	0.925	0.053	0.937	0.973	0.017	0.912	0.936	0.042	0.882	0.927	0.037

3.5.4 BDCG模块组件的有效性

我们分别移除边界引导模块(w/o BG)和可变形卷积模块(w/o DFG)以验证其作用. 实验结果如表6所示:当移除BG模块和DFG模块后,检测指标都明显降低或持平,其中边界引导模块降低更为显著,表明其对模型的贡献更大. 因此可以得出结论:同时使用这两个模块可以能达到最好的效果.

表6　BDCG模块组件的有效性

模块	LFSD			NLPR			SIP			SSD		
	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$	$F_\beta^{\max}\uparrow$	$E_\xi^{\max}\uparrow$	$M\downarrow$
w/o BG	0.890	0.921	0.052	0.935	0.970	0.017	0.911	0.935	0.041	0.879	0.923	0.038
w/o DFG	0.892	0.924	0.051	0.938	0.973	0.016	0.912	0.935	0.041	0.880	0.924	0.037
BDCG	**0.894**	**0.926**	**0.051**	**0.938**	**0.974**	**0.016**	**0.914**	**0.937**	**0.041**	**0.882**	**0.928**	**0.037**

3.5.5 本文方法的不足

尽管本文方法具有良好的检测效果,但是在某些特定场景中还存在局限性,图9显示了本文方法检测失败的显著图,从图中可见本文方法在目标遮挡(第1行),低光照(第2行)和图像光照变化不均(第3行)情况下取得了较差的检测效果. 如第1行所示,本文方法误将树木作为待检测对象. 另外,如第2行所示在亮度极低的场景下,本文方法仅能检测到突出目标而无法

有效应对左右两黑服装角色. 以上研究表明,本文方法仍有较大提升空间. 我们认为可通过低光照图像增强技术和数据增强技术来减轻图像光照与遮挡所导致的问题,从而提高模型检测的性能.

图9 本文方法预测失败的显著图

4 总结

本文提出了一种两分支的RGB-D显著性目标检测方法,分别为边界线索图分支和显著图分支. 此外,本文还提出了3个模块用于提高检测的性能. 实验结果表明本文所提模型在多个评价指标均具有良好的表现,并且模型预测出的显著图与其他模型相比更接近真值图. 接下来的工作中,我们将研究如何利用不同层次的特征进行融合以此提高在多目标和低光照场景下的检测性能,从而更好的推动该任务的广泛应用.

参考文献

[1] CHEN H, LI Y F. Three-stream attention-aware network for RGB-D salient object detection[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2825-2835.

[2] WANG J, SONG K C, BAO Y Q, et al. CGFNet: Cross-guided fusion network for RGB-T salient object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5): 2949-2961.

[3] 梁大川, 李静, 刘赛, 等. 基于图和稀疏主成分分析的多目标显著性检测[J]. 计算机研究与发展, 2018, 55(5): 1078-1089.
LIANG D C, LI J, LIU S, et al. Multiple object saliency detection based on graph and sparse principal component analysis[J]. Journal of Computer Research and Development, 2018, 55(5): 1078-1089. (in Chinese)

[4] 张荣国, 贾玉闪, 胡静, 等. 超像素内容感知先验的多尺度贝叶斯显著性检测方法[J]. 电子学报, 2020, 48(8): 1509-1515.
ZHANG R G, JIA Y S, HU J, et al. Superpixel content-aware priors based multi-scale Bayesian saliency detection[J]. Acta Electronica Sinica, 2020, 48(8): 1509-1515. (in Chinese)

[5] LI J X, PAN Z F, LIU Q S, et al. Stacked U-shape network with channel-wise attention for salient object detection[J]. IEEE Transactions on Multimedia, 2021, 23: 1397-1409.

[6] WANG W G, SHEN J B, LING H B. A deep network solution for attention and aesthetics aware photo cropping[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1531-1544.

[7] CHEN Z X, ZHOU H J, LAI J H, et al. Contour-aware loss: Boundary-aware learning for salient object segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 431-443.

[8] LEE M S, SHIN W, HAN S W. TRACER: Extreme attention guided salient object tracing network (student abstract) [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(11): 12993-12994.

[9] WANG F Y, PAN J S, XU S K, et al. Learning discriminative cross-modality features for RGB-D saliency detection [J]. IEEE Transactions on Image Processing, 2022, 31: 1285-1297.

[10] ZHAO X Q, PANG Y W, ZHANG L H, et al. Self-supervised pretraining for RGB-D salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 3463-3471.

[11] LI G Y, LIU Z, CHEN M Y, et al. Hierarchical alternate interaction network for RGB-D salient object detection [J]. IEEE Transactions on Image Processing, 2021, 30: 3528-3542.

[12] JI W, LI J J, YU S, et al. Calibrated RGB-D salient object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 9466-9476.

[13] FAN D P, LIN Z, ZHANG Z, et al. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(5): 2075-2089.

[14] CHEN S H, FU Y. Progressively guided alternate refinement network for RGB-D salient object detection[C]// Computer Vision — ECCV 2020. Cham: Springer International Publishing, 2020: 520-538.

[15] ZHAO X Q, ZHANG L H, PANG Y W, et al. A single stream setwork for robust and real-time RGB-D salient object detection[C]//Computer Vision — ECCV 2020.

Cham: Springer International Publishing, 2020: 646-662.

[16] ZHAO X Q, PANG Y W, ZHANG L H, et al. Suppress and balance: A simple gated network for salient object detection[C]//Computer Vision — ECCV 2020. Cham: Springer International Publishing, 2020: 35-51.

[17] YAO Z J, WANG L P. Boundary information progressive guidance network for salient object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 4236-4249.

[18] ZHOU X F, SHEN K Y, WENG L, et al. Edge-guided recurrent positioning network for salient object detection in optical remote sensing images[J]. IEEE Transactions on Cybernetics, 2023, 53(1): 539-552.

[19] ZHOU X F, SHEN K Y, LIU Z, et al. Edge-aware multiscale feature integration networkfor salient object detection in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.

[20] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]//Computer Vision — ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

[21] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 764-773.

[22] LI N Y, YE J W, JI Y, et al. Saliency detection on light field[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 2806-2813.

[23] PENG H W, LI B, XIONG W H, et al. RGBD salient object detection: A benchmark and algorithms[C]//Computer Vision — ECCV 2014. Cham: Springer International Publishing, 2014: 92-109.

[24] JU R, GE L, GENG W J, et al. Depth saliency based on anisotropic center-surround difference[C]//2014 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2015: 1115-1119.

[25] LI G, ZHU C B. A three-pathway psychobiological framework of salient object detection using stereoscopic technology[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Piscataway: IEEE, 2018: 3008-3014.

[26] NIU Y Z, GENG Y J, LI X Q, et al. Leveraging stereopsis for saliency analysis[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 454-461.

[27] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1597-1604.

[28] FAN D P, GONG C, CAO Y, et al. Enhanced-alignment measure for binary foreground map evaluation[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Freiburg: Morgan Kaufmann, 2018: 698-704.

[29] ZHAI Y J, FAN D P, YANG J F, et al. Bifurcated backbone strategy for RGB-D salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 8727-8742.

[30] PANG Y W, ZHANG L H, ZHAO X Q, et al. Hierarchical dynamic filtering network for RGB-D salient object detection[C]//Computer Vision — ECCV 2020. Cham: Springer International Publishing, 2020: 235-252.

[31] LI G Y, LIU Z, LING H B. ICNet: Information conversion network for RGB-D based salient object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 4873-4884.

[32] LI G Y, LIU Z, YE L W, et al. Cross-modal weighting network for RGB-D salient object detection[C]//Computer Vision — ECCV 2020. Cham: Springer International Publishing, 2020: 665-681.

[33] LIU N, ZHANG N, WAN K Y, et al. Visual saliency transformer[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 4702-4712.

[34] SUN P, ZHANG W H, WANG H Y, et al. Deep RGB-D saliency detection with depth-sensitive attention and automatic multi-modal fusion[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 1407-1417.

[35] JIN W D, XU J, HAN Q, et al. CDNet: Complementary depth network for RGB-D salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 3376-3390.

[36] ZHAO Y F, ZHAO J W, LI J, et al. RGB-D salient object detection with ubiquitous target awareness[J]. IEEE Transactions on Image Processing, 2021, 30: 7717-7731.

[37] ZHANG W B, JIANG Y, FU K R, et al. BTS-net: Bi-directional transfer-and-selection network for RGB-D salient object detection[C]//2021 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2021: 1-6.

[38] ZHANG W B, JI G P, WANG Z, et al. Depth quality-inspired feature manipulation for efficient RGB-D salient object detection[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 731-740.

[39] LIU Z Y, WANG Y, TU Z Z, et al. TriTransNet: RGB-D

salient object detection with a triplet transformer embedding network[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 4481-4490.
[40] ZHANG C, CONG R M, LIN Q W, et al. Cross-modality discrepant interaction network for RGB-D salient object detection[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 2094-2102.
[41] WU Y H, LIU Y, XU J, et al. MobileSal: Extremely efficient RGB-D salient object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 10261-10269.
[42] CHENG X L, ZHENG X, PEI J L, et al. Depth-induced gap-reducing network for RGB-D salient object detection: An interaction, guidance and refinement approach[J]. IEEE Transactions on Multimedia, 2023, 25: 4253-4266.
[43] ZHOU W J, ZHU Y, LEI J S, et al.CCAFNet: Crossflow and cross-scale adaptive fusion network for detecting salient objects in RGB-D images[J]. IEEE Transactions on Multimedia, 2022, 24: 2192-2204.
[44] GAO W, LIAO G B, MA S W, et al. Unified information fusion network for multi-modal RGB-D and RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(4): 2091-2106.
[45] ZHANG M, YAO S Y, HU B Q, et al. C^2DFNet: Criss-cross dynamic filter network for RGB-D salient object detection[J]. IEEE Transactions on Multimedia, 2023, 25: 5142-5154.
[46] PANG Y W, ZHAO X Q, ZHANG L H, et al. CAVER: Cross-modal view-mixed Transformer for bi-modal salient object detection[J]. IEEE Transactions on Image Processing, 2023, 32: 892-904.
[47] LIU Z Y, TAN Y C, HE Q, et al. SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 4486-4497.
[48] ZHOU T, FU H Z, CHEN G, et al. Specificity-preserving RGB-D saliency detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 4661-4671.
[49] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 6230-6239.

作者简介

孟令兵 男，1994 年 11 月出生，安徽霍邱人. 现为安徽信息工程学院计算机与软件工程学院助教. 主要研究方向为计算机视觉(显著性目标检测、医学图像分割等).
E-mail: lbmeng@iflytek.com

袁梦雅 女，2003 年 4 月出生，安徽合肥人. 现为安徽信息工程学院本科生. 主要研究方向为计算机视觉、传感器网络.
E-mail: 1464616739@qq.com

时雪涵 女，1986 年 11 月出生，安徽阜阳人. 现为安徽信息工程学院计算机与软件工程学院高级工程师. 主要研究方向为计算机视觉、信息安全、软件测试等.
E-mail: xhshi3@iflytek.com

张 乐 女，1997 年 10 月出生，安徽池州人. 现为安徽信息工程学院计算机与软件工程学院助教，主要研究方向为智能感知与物体识别.
E-mail: 1ezhang7@iflytek.com

吴锦华 男，1991 年 12 月出生，安徽省枞阳人. 现为安徽信息工程学院计算机与软件工程学院讲师. 主要研究方向为模式识别.
E-mail: jhwu3@iflytek.com

程 菲(通讯作者) 女，1968 年 7 月出生，安徽黄山人. 现为安徽信息工程学院大数据与人工智能学院副教授. 主要研究方向为智能控制.
E-mail: feicheng6@iflytek.com