·图形图像处理·　　文章编号：1000-3428(2023)05-0255-07　　文献标志码：A　　中图分类号：TP391

基于深度质量感知和分层特征引导的RGB-D显著性检测

宋梦柯，郑元超，陈程立诏

（青岛大学 计算机科学技术学院，山东 青岛 266071）

摘　要：现有基于融合的RGB-D显著性物体检测方法在对跨模态特征进行融合时忽视了RGB和深度图两模态特征的差异性，跨模态特征融合不均衡的问题使得模型不能充分利用跨模态互补特征，而低质量深度图也会对模型性能带来损害。提出一种基于深度质量感知和分层特征引导的RGB-D显著性物体检测算法。算法分为两个阶段：深度质量感知阶段和分层特征引导阶段。在第一阶段，利用深度质量感知从现有的主流RGB-D显著性物体检测训练数据集中挖掘高质量深度图，对训练集进行增强，提升低质量深度图的质量，减少噪声数据对模型性能的损害；在第二阶段，利用特征引导网络对RGB图和深度图进行分层自适应权重动态融合，在有效增加融合效率的同时增强跨模态融合的感知能力。在基准数据集NJUD、NLPR、SSD、STEREO和SIP上的实验结果表明，相比于SSF、CDNet、D3Net、DASNet等方法，该算法能够大幅提升深度图质量，其中在NLPR数据集上F-Measure值为0.934，MAE仅为0.020，综合性能优于其他相关SOTA方法，证明了先挖掘高质量深度图再进行跨模态自适应动态融合算法的有效性。

关键词：深度质量感知；特征引导；跨模态融合；分层融合；RGB-D显著性检测

开放科学（资源服务）标志码（OSID）：

中文引用格式：宋梦柯，郑元超，陈程立诏. 基于深度质量感知和分层特征引导的RGB-D显著性检测[J]. 计算机工程，2023，49(5)：255-261，268.

英文引用格式：SONG M K，ZHENG Y C，CHEN C L Z. RGB-D saliency detection via depth quality perception and hierarchical feature guidance[J]. Computer Engineering，2023，49(5)：255-261，268.

RGB-D Saliency Detection via Depth Quality Perception and Hierarchical Feature Guidance

SONG Mengke，ZHENG Yuanchao，CHEN Chenglizhao

（College of Computer Science and Technology，Qingdao University，Qingdao 266071，Shandong，China）

【Abstract】Existing fusion-based RGB-D saliency object detection methods ignore the differences between RGB and depth map features when fusing cross-modal features.The problems from fusing unbalanced cross-modal features makes the model insufficiently leverage cross-modal complementary features. Moreover，low-quality depth maps can hurt model performance.This paper proposes an RGB-D salient object detection algorithm based on depth quality perception and hierarchical feature guidance. The algorithm is divided into two stages：depth quality perception stage and hierarchical feature guidance stage.In the first stage，depth quality perception is used to mine high-quality depth maps from the existing mainstream RGB-D salient object detection training data sets to enhance the training sets.This process significantly improves the quality of low-quality depth maps and reduces the damage of noise data on model performance.In the second stage，the feature-guidance network is used to perform hierarchical adaptive weight dynamic fusion of the RGB and depth map，which effectively increases the fusion efficiency and enhances the cross-modality fusion perception.The experimental results on five benchmark datasets（NJUD，NLPR，SSD，STEREO，and SIP）show that the proposed algorithm significantly improves the depth map quality compared to methods such as SSF，CDNet，D3Net，and DASNet.Moreover，on the NLPR dataset，the F-Measure value is 0.934，whereas the MAE is only 0.020.The comprehensive performance is better than other related SOTA methods，proving the effectiveness of the proposed algorithm in first mining high-quality depth maps and then performing cross-modal adaptive dynamic fusion.

【Key words】depth quality perception；feature guidance；cross-modal fusion；hierarchical fusion；RGB-D saliency detection

DOI：10. 19678/j. issn. 1000-3428. 0064616

基金项目：山东省自然科学基金博士项目（ZR2019BF011）。

作者简介：宋梦柯（1997—），男，硕士研究生，主研方向为计算机视觉、RGB-D显著性检测；郑元超，硕士研究生；陈程立诏，教授、博士。

收稿日期：2022-05-05　　**修回日期：**2022-06-20　　**E-mail：**songsook@163.com

0 概述

RGB-D显著性物体检测(RGB-D SOD)旨在利用RGB图和深度数据在复杂场景中准确地检测和分割最显著的对象,它可以作为目标跟踪[1]、目标检测[2]、内容感知图像编辑[3]、图像检索[4]等任务的预处理步骤。近些年来,RGB-D SOD因深度数据带来的额外空间结构信息而受到越来越多的关注。传统的RGB-D显著性检测方法主要使用人工特征,如对比度先验、颜色先验、纹理先验等来提取低层次特征,然而这种方式对于复杂场景下的显著性区域提取效果并不理想。随着深度卷积神经网络(CNN)的发展,RGB-D显著性检测方法能够有效地利用全卷积网络从RGB图像和深度图中提取多尺度特征。来自深层的高层特征代表图像中的粗尺度和语义信息,而来自浅层的低层特征捕获精细细节以精确定位对象的边界。由于这种互补性,有效地融合多尺度信息成为RGB-D显著性检测成功的关键。

目前主流融合方式分为早期融合、后期融合、中期融合等3类:早期融合[5-6]通过简单地通道叠加直接融合RGB和深度图像,形成四通道输入。但是,这种融合没有考虑两种模态之间的分布差异,可能会导致特征融合不准确;后期融合[7]使用2个并行的网络为RGB和深度数据生成独立的显著图,然后将2个图融合以获得最终的预测图,然而,很难通过这种类型的融合来获取2种模态之间复杂的交互信息;中期融合[8-9]利用2个网络分别编码学习2种模态的中间特征,然后将融合的特征送入后续网络或解码器。

虽然以上方法取得了一定的性能提升,但忽略了对跨模态特征融合的均衡性研究。不同模态之间具有一定的差异性,而且提供的信息也有差异,简单地将两模态特征完全融合会带来信息冗余,甚至会阻碍模型性能提升。因此,提出一种分层特征引导融合方法,将RGB图和深度图进行分层自适应权重动态融合。

此外,通过深度相机获得的深度图并不一定都是高质量的,低质量的深度图会降低模型性能,一些工作针对深度图质量问题也做了相关研究[10-12]。但这些方法都忽略了一个问题:用真实深度图监督RGB估计深度图时,真实深度图质量参差不齐,虽然在后续过程中通过类似误差反传来修正估计结果,但依然不免引入冗余或错误信息,产生不准确的深度图。因此,提出一种新的深度质量感知方法,挖掘高质量的深度图用来对低质量深度图重新估计,以此得到高质量的估计深度图。

为减小RGB-D显著性检测方法中低质量深度图对模型性能的影响并解决跨模态特征融合不均衡的问题,提出一种基于深度质量感知和分层特征引导的RGB-D显著性检测算法。算法分为2个阶段:在第一阶段,利用深度质量感知将低质量的深度图增强为高质量的深度图;在第二阶段,特征引导网络对RGB图和深度图进行分层自适应权重动态融合,达到跨模态均衡融合的目的。

1 相关工作

传统RGB-D SOD方法主要依靠手工特征[13-14],利用大量的显著性先验信息进行图像显著性检测,如对比度先验、图像背景先验、目标先验等。文献[13]提出一种基于深度挖掘利用图像3个不同层的深度线索的多层反向传播显著性检测算法。文献[14]利用中心暗原色先验,基于中心显著性先验和暗原色先验生成中心暗原色映射,将初始显著性映射与中心暗原色映射融合,生成最终显著性映射。然而,这些方法看似简单,但忽略了RGB和深度图之间的差异,因此难以获得可靠的结果。

随着深度学习的出现,许多基于CNN的方法[5-7]主导了这一领域。其中,基于融合的方法[8-9]在RGB-D显著性检测方面贡献较大,取得了引人注目的优异性能。目前主流融合方式主要分为以下3类:

1)早期融合[5-6]:通过简单地通道叠加直接融合RGB和深度图像,形成四通道输入。LIU等[5]提出了单流循环卷积神经网络,将RGB-D四通道输入送入VGG-16网络生成多级特征。REN等[6]设计一个分阶段的RGB-D显著物体检测算法,首先将区域对比度与背景、深度和方向先验相结合以生成显著图,然后提出一种显著性恢复方案,该方案集成了PageRank算法对高置信度区域进行采样并恢复不明确的区域的显著性。但是,早期融合方式没有考虑2种模态之间的分布差异,可能会导致特征融合不准确。

2)后期融合[7]:使用2个并行的网络为RGB和深度数据生成独立的显著图,然后将2个图融合以获得最终的预测图。DING等[7]提出端到端的3个卷积神经网络,使用颜色显著网络、深度显著网络分别获得显著图之后,送入显著融合网络获得最终显著图。然而,很难通过后期融合方式来获取2种模态之间复杂的交互信息。

3)中期融合[8-9]:利用2个网络分别编码学习2种模态的中间特征,然后将融合的特征送入后续网络或解码器。ZHAO等[9]将对比度先验引入基于CNN的架构以增强深度信息,然后使用流体金字塔集成模块将增强的深度与RGB特征集成。CHEN等[8]提出一种互补感知融合模块,在每个模块中引入跨模态残差函数和互补感知监督,通过级联该模块从深到浅的密集添加逐级监督,逐步选择和组合跨层级的补充信息。

虽然这些基于不同阶段融合的方法带来了性能的提升,但对跨模态特征融合均衡性的研究却鲜有涉及。不同模态之间具有一定的差异性,而且提供的信息也有差异,简单地将两模态特征完全融合会带来信息冗余,甚至会阻碍模型性能提升。因此,提出一种分层特征引导融合方法,实现对RGB图和深度图进行分层自适应权重动态融合。

近年来,研究者针对深度图质量问题也做了相关研究[10-12]。ZHAO等[10]提出一种深度感知显著性目标检测框架,用RGB图估计出深度图并用真实深度图做监督,使深度信息用作误差加权映射来改进分割过程。ZHANG等[11]提出一种深度质量启发的特征操作用于过滤深度图特征,主要借助低层RGB和深度特征对齐以及深度流的整体注意来显式地控制和增强跨模态融合。JIN等[12]提出利用RGB先估计出深度图,再将估计出的深度图和真实深度图进行动态融合来对深度图进行增强。然而,这些方法在用RGB估计深度图时,不同质量的深度图会引入冗余或错误信息,产生不准确的结果。因此,提出一种新的深度质量感知方法,挖掘高质量的深度图用来对低质量深度图重新估计,以此得到高质量的估计深度图。

2　本文方法

2.1　整体结构

本文方法分为2个阶段:深度质量感知阶段和分层特征引导阶段。深度质量感知阶段目的是增强现有RGB-D SOD数据集低质量深度图的质量,包含训练过程和测试过程。在训练过程中,深度过滤器用于过滤出高质量深度图,此部分高质量深度图用于监督高质量深度图生成器生成高质量估计深度图。在测试过程中,高质量深度图生成器用于将低质量的深度图增强为高质量的深度图。分层特征引导阶段由双流网络、特征引导模块和解码器组成。双流网络分别提取RGB和深度图特征,并分别生成对应的显著图,用于计算损失,指导误差反传。特征引导模块利用基于图表示的方式获取跨模态信息之间的关系,并得到分层特征自适应融合权重,该权重用于分层指导跨模态特征融合。解码器采用级联解码的方式由高到低逐层解码,得到最终显著图。下文第2.2节详述了深度质量感知阶段,第2.3节详述了分层特征引导过程。

2.2　深度质量感知

深度信息在显著性物体检测方法中具有重要作用,原因在于深度信息可以作为颜色信息的补充,为模型提供边缘和轮廓信息。高质量的深度图能够提供更准确的边缘和轮廓信息,有助于模型学习与训练。反之,低质量的深度图边缘和轮廓模糊,不仅不能帮助颜色信息定位显著物体,甚至会带来冗余信息,影响模型性能。因此,提出一种增强现有RGB-D SOD数据集低质量深度图质量的方法,称为深度质量感知。图1对比了传统训练方法和深度质量感知训练方法的整体结构流程。具体来说,深度质量感知核心思想就是挖掘高质量的深度图用于对低质量深度图重新估计,以此得到高质量的估计深度图。该方法主要包含深度过滤器和高质量深度图生成器2个部分,前者用于挖掘高质量的深度图,后者用于对低质量深度图重新估计,得到高质量的估计深度图,实现低质量深度图的增强。

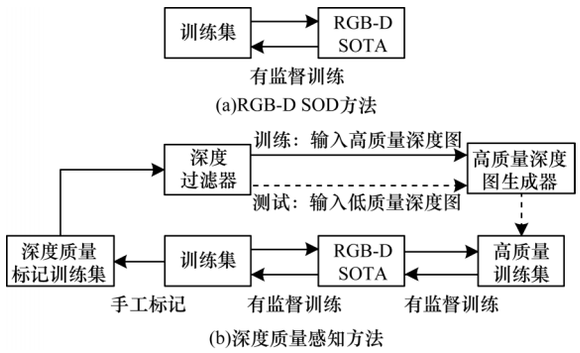

图1　RGB-D SOD方法和深度质量感知方法的整体结构流程

Fig.1　Overall structure flow of RGB-D SOD method and deep quality perception method

首先,对现有的RGB-D SOD数据集进行标注,将数据集分为高质量深度图和低质量深度图2个部分,称为深度质量标记训练集。具体地,分别将RGB和深度图输入预训练的RGB显著性检测(如CPD[15])模型中,得到RGB显著图和深度显著图。理论上,如果RGB显著图和深度显著图质量相近,则认为该深度图是高质量的;反之,如果RGB显著图和深度显著图质量相差过大,则认为该深度图是低质量的。显著图质量可通过计算和真值图之间的S-measure指标获得。

得到深度质量标记训练集之后,使用深度过滤器过滤出具有低质量深度图的数据。该深度过滤器本质上是一个二元分类器,由特征编码器和分类头组成,输出0(低质量)和1(高质量)。该分类器的输入为所有的深度图(包括高质量和低质量),特征编码器采用预训练的RGB SOD模型(如CPD[15]),以此得到显著图。分类头采用S-measure,将得到的显著图和对应的真值图作比较,如果S-measure值大于一定的阈值(γ),认为该深度图是高质量的,分类头输出1,反之,输出0。计算公式如下:

$$\mathrm{DF}(\boldsymbol{d})=\mathrm{Sm}\left(\mathrm{FE}(\boldsymbol{d})\right) \tag{1}$$

其中:DF(·)表示深度过滤器,输出0或1;FE(·)表示特征编码器;Sm(·)表示S-measure;$\boldsymbol{d}$表示深度图。

通过深度过滤器得到高质量深度图之后,利用高质量深度图训练一个高质量深度图生成器来增强低质量的深度图,该高质量深度图生成器可采用现有单目深度估计方法(如Monodepth2)。与单目深度估计不同的是,高质量深度图生成的输入是高质量深度图对应的RGB图,而不是任意的RGB图;用于监督的深度图是高质量的深度图,而不是任意质量的深度图。具体细节可以表示为:

$$\mathrm{HDE}\left(\boldsymbol{H}_{\mathrm{r}}\right)=\mathrm{MDE}\left(\boldsymbol{H}_{\mathrm{r}}\right) \tag{2}$$

其中:HDE(·)表示高质量深度图生成器;MDE(·)表示单目深度估计;$\boldsymbol{H}_{\mathrm{r}}$表示高质量深度图对应的RGB图。在测试阶段,HDE(·)的输入是低质量深

度图对应的RGB图。经过以上操作，可以对低质量的深度图进行重新估计，得到高质量深度图，以此增强低质量深度图的质量，如图2所示，高质量深度图生成器生成的深度图质量明显高于原始低质量深度图的质量。

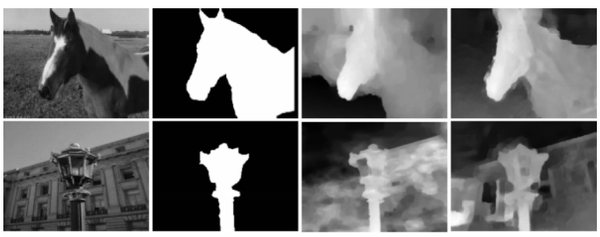

图2 深度质量感知生成的深度图

Fig.2 Depth maps generated by depth quality perception

需要指出的是，深度过滤器中更好的特征编码器以及性能更好的单目深度估计方法对低质量深度图的质量有更好的提升，但这不是本节所讨论的重点，可作为未来工作进一步研究。并且，深度质量感知模块可作为即插即用模块部署在现有RGB-D SOTA模型中提升SOTA模型性能。

2.3 分层特征引导

为了解决不同模态之间的差异性，以及简单地将两模态特征完全融合带来的信息冗余问题，提出一种分层特征引导方法，用于对RGB图和深度图进行分层自适应权重动态融合，如图3所示。其中，双流网络均采用ResNet50[16]结构（去除分类部分，保留特征提取结构），解码器部分采用级联解码方式由高到低逐层解码，得到最终显著图。受注意力机制启发，特征引导模块从RGB和深度图两模态融合特征中学习权重，并用该权重指导跨模态融合。

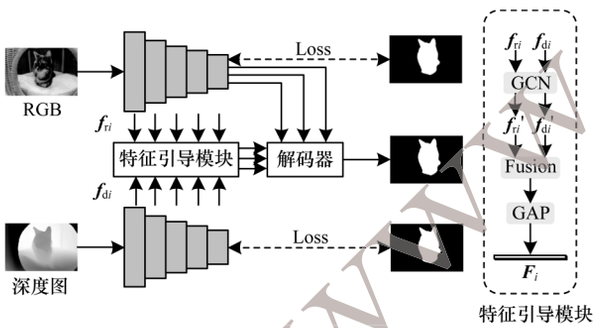

图3 分层特征引导网络原理示意图

Fig.3 Schematic diagram of the principle of hierarchical feature-guided network

具体来说，RGB流网络输出5层RGB特征，分别为$\boldsymbol{f}_{ri}(i=1,2,\cdots,5)$，同样，深度流网络输出特征为$\boldsymbol{f}_{di}(i=1,2,\cdots,5)$，以分层融合的方式将$\boldsymbol{f}_{ri}$和$\boldsymbol{f}_{di}$输入特征引导模块中得到分层特征引导权重$\boldsymbol{F}_i(i=1,2,\cdots,5)$。特别地，采用基于图知识表示（例如GCN）的方式对RGB和深度图特征进行融合，原因在于图知识表示可以在全局范围内有效地获取节点（此处为特征图的每个像素）之间的多重关系，能够更好地融合传递跨模态信息。具体过程分为2个部分：图构建和消息传递。在图构建过程中，分别将RGB和深度图的每一个像素点作为图的节点。为了降低参数量，对RGB和深度图特征进行重采样操作，重采样后每层特征尺寸分别为(28,28)，即每层图节点个数为28×28=784，GCN的输入特征形状为“(节点数,特征数)”。在此基础上，分别将RGB和深度图的图特征叠加在一起构成一个融合图特征，作为消息传递过程的输入。在消息传递过程中，针对融合图特征采用2个1维卷积操作，即在2个方向上分别做1维卷积，对两模态特征在图结构上进行传递和融合。具体过程可表示为：

$$\mathrm{GCN}(\boldsymbol{f}_{ri},\boldsymbol{f}_{di})=\mathrm{ConvY}(\mathrm{ConvX}(\mathrm{CAT}(\boldsymbol{f}_{ri},\boldsymbol{f}_{di}))) \tag{3}$$

其中：CAT(·)表示在特征宽度或者高度上的叠加操作；ConvX(·)表示在宽度方向上的一维卷积；ConvY(·)表示在高度方向上的一维卷积。经过该处理过程，可以得到基于图表示融合之后得图融合特征$\boldsymbol{f}'_{ri}(i=1,2,\cdots,5)$和$\boldsymbol{f}'_{di}(i=1,2,\cdots,5)$，然后通过二维卷积、矩阵相加和矩阵相乘等操作，将图融合特征进一步融合，得到空间特征，经过全局平均池化GAP(·)之后得到特征引导权重$\boldsymbol{F}_i(i=1,2,\cdots,5)$。具体细节可以表示为：

$$F_{\varphi}(\boldsymbol{f}_{ri},\boldsymbol{f}_{di})=\left(\mathrm{Conv}\left(\boldsymbol{f}_{ri}\right)\otimes\mathrm{Conv}\left(\boldsymbol{f}_{di}\right)\right)\otimes\left(\mathrm{Conv}\left(\boldsymbol{f}_{ri}\right)\oplus\mathrm{Conv}\left(\boldsymbol{f}_{di}\right)\right) \tag{4}$$

$$\boldsymbol{F}_i=\mathrm{GAP}(F_{\varphi}(\boldsymbol{f}_{ri},\boldsymbol{f}_{di})) \tag{5}$$

其中：Conv(·)表示二维卷积；$F_{\varphi}(\cdot)$表示特征融合；GAP(·)表示全局平均池化。特征引导权重$\boldsymbol{F}_i(i=1,2,\cdots,5)$是一个$1\times n$（$n$表示维数）的向量，用来决定后续融合过程中应该包含多少RGB特征和深度图特征。融合方式如下：

$$\boldsymbol{F}_{ui}=(\boldsymbol{F}_i\otimes\boldsymbol{f}_{ri})\oplus(\boldsymbol{F}_i\otimes\boldsymbol{f}_{di}) \tag{6}$$

其中：$\boldsymbol{F}_{ui}(i=1,2,\cdots,5)$表示特征引导融合特征；$\otimes$表示按元素乘法；$\oplus$表示矩阵加法。

在解码阶段，采用级联解码的方式，即由高到低依次融合解码。具体地，分为两阶段融合，如图3所示，首先将RGB流的最后三层的高层语义特征$\boldsymbol{F}_{u3}$、$\boldsymbol{F}_{u4}$、$\boldsymbol{F}_{u5}$相加，然后将相加后的特征再分别与低层细节特征相乘，最后经过二维卷积和上采样得到最终显著图。需要注意的是，在RGB流和深度流最后也生成了对应的显著图（具体操作在图3中省略）。因此，该分层特征引导方法生成3种显著图，分别用真值图对其进行监督。其中，对RGB流和深度流生成的显著图进行监督用于指导双流网络的误差反传过程，使得网络更好地拟合训练集的数据分布。

2.4　损失函数

本文所采用的损失函数为二元交叉熵损失,具体公式如下:

$$L_{\text{bce}}=-\sum_i g_i\times\ln p_i-\sum_i(1-g_i)\times\ln(1-p_i)\qquad(7)$$

其中:g_i代表真值图的第i个像素值;p_i代表预测图第i个像素值;×表示矩阵点乘。

3　实验

3.1　数据集

为了评估本文方法的性能,在5个基准数据集上进行实验,分别为SIP[17]、SSD[18]、NJUD[19]、NLPR[20]和STEREO[21]。对NJUD的1 500个样本和NLPR的700个样本进行训练,将这2个数据集的其余图像和其他3个数据集中的图像用于测试。SIP[17]数据集由1 000张高分辨率图像组成,涵盖了不同视角、姿势、遮挡、光照和背景的不同现实场景。SSD[18]数据集由80张立体电影图像组成。NJUD[19]数据集包括各种分辨率的2 003个立体图像对。在这些图像对中,1 400对作为训练集,100对作为验证集,其余作为测试集。NLPR[20]数据集由11种室内和室外场景的1 000张图像组成。其中650个作为训练集,50个作为验证集,剩下的300个作为测试集。STEREO[21]数据集有797张立体图像。这些图像主要来源于互联网和3D电影,并利用光学方法生成深度图。

3.2　评价指标

评价指标采用F-measure和平均绝对误差(Mean Absolute Error,MAE),其中F-measure越高越好,MAE越低越好。

3.3　实验设置

采用在ImageNet数据集上预先训练的ResNet-50作为骨干网络,本文所有实验均使用Python以PyTorch(版本号1.7.1,Cuda版本号10.1)实现,并且所有的训练过程都在NVIDIA GeForce RTX 2080工作站(Win 10操作系统,32 GB内存,8 GB显存,Intel® Xeon® W-2133 CPU @3.60 GHz)上进行。在训练阶段,初始学习率设置为0.001,每训练50轮学习率衰减1/10,batchsize大小为3,输入图像下采样为352×352像素,采用Adam作为优化器。

3.4　实验结果与分析

3.4.1　定量分析与可视化分析

为验证所提方法的有效性,在3种评价指标上与9种基于主流RGB-D SOD方法进行对比,包括CDNet[12]、D3Net[17]、DMRA[22]、A2dele[23]、SSF[24]、TriNet[25]、CoNet[26]、DFM[27]、DASNet[28]等,所有比较方法的显著性图都由作者提供或通过运行其发布的代码获得。不同方法的定量分析结果如表1所示。可以看出,本文方法基本优于所有比较的方法,除NLPR外,在其余4个数据集上有明显的效果提升。相比于基于深度感知的方法CDNet、DFM和DASNet,本文基于深度质量感知和分层特征引导的方法性能更好,在NJUD、SIP、STEREO数据集上F-measure指标平均提升1.4%,MAE平均降低0.3%左右。取得较好性能的主要原因在于本文方法有效地挖掘出了高质量深度图并对低质量深度图进行了显著增强,同时分层自适应权重动态融合解决了跨模态特征差异性的问题。

表1　不同方法定量对比结果

Table 1　Quantitative comparison results of different methods

方法	NJUD		NLPR		SSD		SIP		STEREO	
	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE
DMRA	0.873	0.043	0.865	0.031	0.844	0.058	0.811	0.085	0.869	0.050
BBSNet	0.902	0.035	0.896	0.023	0.843	0.044	0.868	0.055	0.883	0.041
CoNet	0.882	0.045	0.865	0.031	0.819	0.060	0.845	0.063	0.887	0.037
TriNet	0.917	0.034	0.909	0.026	0.876	0.049	0.892	0.043	0.891	0.045
SSF	0.896	0.043	0.896	0.027	0.762	0.054	0.880	0.053	0.840	0.065
D3Net	0.879	0.046	0.872	0.030	0.815	0.050	0.838	0.063	0.891	0.046
CDNet	0.919	0.036	0.920	0.025	0.871	0.045	0.885	0.047	0.898	0.040
DFM	0.913	0.039	0.912	0.024	0.875	0.047	0.890	0.049	0.904	0.040
DASNet	0.911	0.042	0.929	0.021	0.881	0.042	0.892	0.047	0.915	0.037
本文方法	0.925	0.031	0.934	0.020	0.883	0.040	0.897	0.041	0.919	0.034

在不同场景下对比本文方法与其他方法的显著性可视化结果,如图4所示。可以看出,本文方法在多种场景下均取得了较好的效果。在第1行中,RGB中显著性物体和背景对比度不高,但本文方法仍能较好地检测出边缘部分,验证了基于深度质量感知的有效性。在第3、4行中,depth质量较低,但本文方法生成的显著性图依然能精确地检测出完整的物体,说明本文提出的分层特征引导能自适应选择性融合RGB和depth特征,避免了低质量depth对检测结果的不良影响。

图4 不同方法的显著性可视化结果

Fig.4 Saliency visualization results of different methods

3.4.2 基于深度质量感知的RGB-D SOTA模型重训练

为了验证所提出的深度质量感知模块作为即插即用模块的泛化性，本文将该模块部署在现有针对深度图质量进行研究的RGB-D SOTA模型中（例如CDNet、DFM和DASNet），对SOTA模型进行重新训练，实验结果见表2，其中，CDNet+、DFM+、DASNet+表示部署深度质量感知模块进行重训练。可以看出，基于深度质量感知模块进行重训练的SOTA模型在5个基准数据集上的性能表现均明显优于原始SOTA模型的性能，如在NJUD数据集上，3种方法的F-measure指标分别提升3%、2%、6%，验证了深度质量感知模块作为即插即用模块的泛化性。原因在于深度质量感知模块可以过滤出低质量的深度图，并通过高质量深度图生成器将低质量深度图增强为高质量深度图，从而减少了低质量深度图对模型带来的负面效果。

表2 不同深度图增强方法对比结果

Table 2 Comparison results of different depth enhancement methods

方法	NJUD		NLPR		SSD		SIP		STEREO	
	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE
CDNet	0.919	0.036	0.920	0.025	0.871	0.045	0.885	0.047	0.898	0.040
CDNet+	0.922	0.034	0.921	0.023	0.875	0.042	0.887	0.044	0.901	0.038
DFM	0.913	0.039	0.912	0.024	0.875	0.047	0.890	0.049	0.904	0.040
DFM+	0.915	0.034	0.920	0.023	0.878	0.045	0.894	0.046	0.909	0.035
DASNet	0.911	0.042	0.929	0.021	0.881	0.042	0.892	0.047	0.915	0.037
DASNet+	0.917	0.040	0.932	0.021	0.882	0.041	0.894	0.043	0.918	0.035
本文方法	0.925	0.031	0.934	0.020	0.883	0.040	0.897	0.041	0.919	0.034

3.5 消融分析

3.5.1 深度质量感知的有效性

为了验证所提出的深度质量感知方法的有效性，将本文方法与3种深度图增强方法（空间注意力、通道注意力和混合注意力）进行比较，实验结果如表3所示。可以看出，通过使用本文深度质量感知方法的性能在5个数据集上均优于另外3种深度图增强方法。其中，在NJUD数据集上F-measure指标分别增加0.8%、1.0%、0.3%，MAE指标分别降低0.8%、1.1%、0.7%。实验结果表明，仅通过单一或者混合注意力机制对深度图增强的效果有限，不如通过挖掘高质量深度图并用高质量深度图对低质量深度图进行增强的深度质量感知方法。

表3 深度质量感知消融分析结果

Table 3 Ablation analysis results of depth quality perception

方法	NJUD		NLPR		SSD		SIP		STEREO	
	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE
空间注意力	0.915	0.042	0.907	0.035	0.866	0.048	0.885	0.047	0.898	0.042
通道注意力	0.913	0.045	0.914	0.031	0.870	0.050	0.880	0.049	0.893	0.043
混合注意力	0.920	0.037	0.925	0.022	0.880	0.043	0.894	0.045	0.913	0.038
本文方法	0.925	0.031	0.934	0.020	0.883	0.040	0.897	0.041	0.919	0.034

3.5.2　特征引导模块的有效性

为了验证特征引导模块的有效性,将本文方法与2种直接融合方式(加法、乘法)进行比较,实验结果如表4所示。可以看出,采用非特征引导方式将特征直接相加的性能不如将特征直接相乘,其中,在SSD数据集上F-measure指标降低了0.5%。相比直接相加,乘法操作对特征的特异性损失较小。而采用特征引导的方式效果最优,原因在于与非特征引导简单融合相比,特征引导融合考虑了跨模态特征的差异性问题,对不同重要性的特征进行了加权,对不重要的特征分配较小的权重,使得针对更重要特征融合得更加充分。

表4　不同特征融合方式对比结果

Table 4　Comparison results of different feature fusion modes

融合方式	NJUD		NLPR		SSD		SIP		STEREO	
	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE
矩阵加法	0.901	0.046	0.901	0.033	0.828	0.053	0.875	0.057	0.870	0.049
矩阵乘法	0.903	0.045	0.902	0.031	0.832	0.052	0.878	0.055	0.875	0.043
本文方法	0.925	0.031	0.934	0.020	0.883	0.040	0.897	0.041	0.919	0.034

4　结束语

本文提出一种深度图质量增强方法——深度质量感知,不同于现有的基于注意力机制的深度图质量增强方法仅针对单一图进行增强,该方法借助高质量深度图对低质量深度图进行增强,作用更加明显。同时,设计一个分层特征引导网络,对RGB图和深度图进行分层自适应权重动态融合,解决跨模态特征差异性的问题。在5个基准数据集上的实验结果表明,本文方法性能优于NJUD、SIP等深度图增强方法。后续将对GCN做进一步研究,使用更有效的GCN对跨模态特征进行融合,同时在不降低精度的前提下减少模型的参数量。

参考文献

[1] LEE H, KIM D. Salient region-based online object tracking[C]// Proceedings of IEEE Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2018: 1170-1177.

[2] XU K, BA J L, KIROS R, et al. Show, attend and tell: neural image caption generation with visual attention[C]//Proceedings of the 32nd International Conference on Machine Learning. New York, USA: ACM Press, 2015: 2048-2057.

[3] 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.

[4] HE J F, FENG J Y, LIU X L, et al. Mobile product search with bag of hash bits and boundary reranking[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 3005-3012.

[5] LIU Z, SHI S, DUAN Q, et al. Salient object detection for RGB-D image by single stream recurrent convolution neural network[J]. Neurocomputing, 2019, 363: 46-57.

[6] REN J Q, GONG X J, LU Y, et al. Exploiting global priors for RGB-D saliency detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 25-32.

[7] DING Y, LIU Z, HUANG M, et al. Depth-aware saliency detection using convolutional neural networks[J]. Journal of Visual Communication and Image Representation, 2019, 61: 1-9.

[8] CHEN H, LI Y F. Progressively complementarity-aware fusion network for RGB-D salient object detection[C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 3051-3060.

[9] ZHAO J X, CAO Y, FAN D P, et al. Contrast prior and fluid pyramid integration for RGBD salient object detection[C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 3922-3931.

[10] ZHAO J W, ZHAO Y F, LI J, et al. Is depth really necessary for salient object detection? [C]//Proceedings of the 28th ACM International Conference on Multimedia. New York, USA: ACM Press, 2020: 1745-1754.

[11] 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, USA: ACM Press, 2021: 731-740.

[12] 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: A Publication of the IEEE Signal Processing Society, 2021, 30: 3376-3390.

[13] ZHU C B, LI G. A multilayer backpropagation saliency detection algorithm and its applications[J]. Multimedia Tools and Applications, 2018, 77(19): 25181-25197.

[14] ZHU C B, LI G, WANG W M, et al. An innovative salient object detection using center-dark channel prior [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2018: 1509-1515.

[15] WU Z, SU L, HUANG Q M. Cascaded partial decoder for fast and accurate salient object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 3902-3911.

[16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.

[17] 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.

(下转第268页)

（上接第261页）

[18] LI G, ZHU C B. A three-pathway psychobiological framework of salient object detection using stereoscopic technology[C]// Proceedings of IEEE International Conference on Computer Vision Workshops. Washington D. C., USA: IEEE Press, 2018: 3008-3014.

[19] JU R, GE L, GENG W J, et al. Depth saliency based on anisotropic center-surround difference[C]//Proceedings of IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2015: 1115-1119.

[20] PENG H, BING L, XIONG W, et al. RGBD salient object detection: a benchmark and algorithms[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 92-109.

[21] NIU Y Z, GENG Y J, LI X Q, et al. Leveraging stereopsis for saliency analysis[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 454-461.

[22] PIAO Y R, JI W, LI J J, et al. Depth-induced multi-scale recurrent attention network for saliency detection[C]// Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2020: 7253-7262.

[23] PIAO Y R, RONG Z K, ZHANG M, et al. A2dele: adaptive and attentive depth distiller for efficient RGB-D salient object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 9057-9066.

[24] ZHANG M, REN W S, PIAO Y R, et al. Select, supplement and focus for RGB-D saliency detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 3469-3478.

[25] 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, USA: ACM Press, 2021: 4481-4490.

[26] JI W, LI J J, ZHANG M, et al. Accurate RGB-D salient object detection via collaborative learning[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 52-69.

[27] 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, USA: ACM Press, 2021: 731-740.

[28] 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.

编辑 金胡考