

基于多模态多级特征聚合网络的光场显著性目标检测

王安志,任春洪,何淋艳,杨元英,欧卫华

(贵州师范大学 大数据与计算机科学学院, 贵阳 550025)

摘要: 现有基于深度学习的显著性检测算法主要针对二维 RGB 图像设计,未能利用场景图像的三维视觉信息,而当前光场显著性检测方法则多数基于手工设计,特征表示能力不足,导致上述方法在各种挑战性自然场景图像上的检测效果不理想。提出一种基于卷积神经网络的多模态多级特征精炼与融合网络算法,利用光场图像丰富的视觉信息,实现面向四维光场图像的精准显著性检测。为充分挖掘三维视觉信息,设计2个并行的子网络分别处理全聚焦图像和深度图像。在此基础上,构建跨模态特征聚合模块实现对全聚焦图像、焦堆栈序列和深度图3个模态的跨模态多级视觉特征聚合,以更有效地突出场景中的显著性目标对象。在DUTLF-FS和HFUT-Lytro光场基准数据集上进行实验对比,结果表明,该算法在5个权威评估度量指标上均优于MOLF、AFNet、DMRA等主流显著性目标检测算法。

关键词: 深度图;特征融合;光场;聚合网络;显著性目标检测

开放科学(资源服务)标志码(OSID):



中文引用格式: 王安志,任春洪,何淋艳,等.基于多模态多级特征聚合网络的光场显著性目标检测[J].计算机工程,2022,48(7):227-233,240.

英文引用格式: WANG A Z, REN C H, HE L Y, et al. Light field salient object detection based on multi-modal multi-level feature aggregation network[J]. Computer Engineering, 2022, 48(7): 227-233, 240.

Light Field Salient Object Detection Based on Multi-modal Multi-level Feature Aggregation Network

WANG Anzhi, REN Chunhong, HE Linyan, YANG Yuanying, QU Weihua

(School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China)

[Abstract] Most existing deep learning based saliency detection algorithms focus on 2D RGB images. However, they fail to take advantage of 3D visual information of scenes. Most light field saliency detection methods are based on hand-crafted features, whose feature representation capacity is insufficient. These issues lead to poor performance in many challenging scene images. To remedy these problems, this paper proposes a multi-modal multi-level feature aggregation network based on convolutional neural network for light field salient object detection. To fully exploit 3D visual information, two stream sub-network are designed in parallel to handle all-focus images and depth maps separately. Moreover, several feature aggregation modules are developed to aggregate multi-level features to detect the salient objects in scene. Moreover, several cross-modal feature fusion modules are designed to fuse multi-modal features from all-focus images, focal stack, and depth maps, which can highlight a salient object by utilizing 3D visual information. Comprehensive experimental comparisons were performed on the DUTLF-FS and HFUT-Lytro light field benchmark datasets, and the results reveal that the algorithm outperforms the mainstream salient target detection algorithms, such as MOLF, AFNet, and DMRA on five authoritative evaluation metrics.

[Key words] depth map; feature fusion; light field; aggregation network; Salient Object Detection (SOD)

DOI: 10.19678/j.issn.1000-3428.0061811

0 概述

显著性目标检测(Salient Object Detection, SOD)旨在快速检测和分割图像中最吸引人注意的目标对象。在过去十年中,其作为一项重要的计算机视觉任务,在

目标分割、视觉跟踪、行人重识别以及伪装目标检测^[1]等方面进行了成功的应用,引起了人们的广泛关注。传统显著性目标检测方法的RGB图像只包含像素的颜色和强度,但光场图像除了包含像素的颜色和强度信息外,还包含所有入射光的方向。入射光方向指示

基金项目: 国家自然科学基金(62162013, 61962010); 贵州省自然科学基金([2017]1130, [2017]5726-32); 贵州师范大学2019年博士科研启动项目(GZNU[2018]32号); 贵州省大学生创新创业训练计划项目(S202010663031); 贵州师范大学大学生科研训练计划项目(DK2019A059)。

作者简介: 王安志(1986—),男,副教授、博士,主研方向为计算机视觉、深度学习;任春洪、何淋艳、杨元英,本科生;欧卫华,教授、博士。

收稿日期: 2021-06-02 **修回日期:** 2021-08-11 **E-mail:** andyscu@163.com

了场景中物体的三维几何信息。目前,包括 Lytro 和 Raytrix 在内的商用微透镜阵列光场相机越来越流行,光场信息也被用于包括深度估计^[2-3]、超分辨率^[4]等多个视觉任务,为算法提供更丰富的视觉信息,显著提升算法的性能。根据输入图像的类型(如 RGB 图像、RGB-D 图像和光场图像),现有 SOD 算法大致可分为二维、三维和四维的 SOD 算法这三大类。大多数方法都属于第一类,关于后两类方法的研究还较少。作为主流的二维显著性检测算法^[5-7]已经取得了显著的进步,这得益于深度学习技术尤其是卷积神经网络(CNN)的不断进展^[8-10]。但当遇到挑战性真实场景图像时,这些方法通常表现不佳,甚至失效,其原因主要有两点:传统二维 SOD 算法对先验知识的依赖性;RGB 图像缺乏三维信息。由于深度信息有助于理解显著性目标的上下文信息和提高 SOD 的精度,三维 SOD 算法^[11-12]也逐渐引起了学者的关注。然而,深度估计本身是一项极具挑战的课题,低质量的深度图将严重影响三维 RGB-D SOD 算法的性能。

近年来,随着 Lytro 和 Raytrix 等光场相机的普及,光场信息相对容易获得。光场图像包含丰富的视觉信息,由全聚焦图像、焦堆栈序列和深度图组成,其中焦堆栈包含了丰富的三维几何和外观信息。因其独特的视觉特性,四维光场 SOD^[13-15]具有良好的发展前景。然而,现有光场 SOD 算法大多基于手工设计的特征来实现显著性目标的检测,包括对比度、纹理特征、物体完整性、聚焦性先验等。由于高维光场数据比二维 RGB 图像更难处理,导致面向光场图像的 SOD 方法研究仍较少,基于 CNN 的光场 SOD 算法仅有 MoLF^[16]和 DLLF^[13]等极少数的模型。

面向四维光场图像的基于 CNN 的 SOD 框架具有重要的研究意义和实用价值。为了将 CNN 框架应用于光场 SOD,本文提出一种用于光场 SOD 的端到端的多模态多级特征聚合检测算法。该算法包括两个并行的 CNN 子网络,分别从焦堆栈序列、全聚焦图像和深度图中提取多级多尺度的视觉特征,并利用不同模态间视觉特征的互补性,构建多模态特征聚合模块,以实现更精准的显著性目标检测。

1 相关工作

本文主要介绍二维 RGB^[17]、三维 RGB-D、四维光场 SOD 这三类方法。上述方法又可分为传统的方法和基于深度学习的方法。前者主要基于手工设计的特征,不能满足挑战性场景下的显著性检测需求,本文主要讨论基于深度学习的 SOD 算法。

1.1 面向二维 RGB 图像的 SOD

随着 CNN 的发展,大量基于 CNN 的 SOD 算法被提出。这些算法主要结合后处理步骤^[18]、上下文特征^[19-20]、注意机制^[9,20-23]和循环细化模型^[10,24]。文献[25]提出一种端到端深度对比度网络,其能产生像素级显著图,然后通过全连接的条件随机场后处理步骤进一步改善显著图。文献[5]构建一个 HED 架构,在该架构的跳层结构中引入了短连接。文献[26]提出一种通用的聚合多级卷积网络,该框架将多级特征组合起来预测显著图。随后,DENG 等提出一种带有残差精炼模块的递归残差网络,残差精炼模块用于学习中间预测结果的互补性显著性信

息。文献[12]创建一个双分支的 SOD 网络,同时预测显著性目标的轮廓和显著图。文献[20]提出一种像素级的上下文注意网络来学习上下文特征,以生成包括全局和局部特征的显著图。

综上所述,CNN 可以自动提取多级视觉特征,并直接学习图像到显著图的映射,但基于 CNN 的 SOD 方法面对挑战性的复杂场景仍表现不佳,而将现有基于 CNN 的二维 SOD 模型^[27]直接应用于光场图像也并不可取。因此,有必要开展基于 CNN 的光场 SOD 研究。

1.2 面向 RGB-D 图像的三维 SOD

QU 等^[28]设计一种基于 CNN 的 RGB-D SOD 算法自动学习交互机制,并利用手工设计的特征训练基于 CNN 的 SOD 模型。CHEN 等^[29-31]利用多级特征互补性和跨模态特征互补性,设计了一个多路的多尺度融合网络来预测显著图。CHEN 等^[31]还提出一种三流的注意力融合网络来提取 RGB-D 特征,并引入通道注意机制自适应地选择互补的视觉特征。ZHU 等^[32]提出一个独立的编码器网络来处理深度线索,并利用基于 RGB 的先验模型指导模型的学习过程。WANG 等^[33]提出一种双流 CNN 自适应融合框架,将 RGB 模态和深度模态产生的显著图进行后融合。PIAO 等^[34]提出深度诱导的多尺度递归注意力网络,该网络包括一个深度精炼模块,用于提取并融合互补的 RGB 和深度特征、深度诱导的多尺度加权模块,以及一个递归的注意力模块,以粗到细的方式生成更准确的显著性预测结果。

1.3 面向光场的四维 SOD

目前,只有较少的 SOD 算法设计光场图像,大部分方法都基于手工设计的特征。尽管如此,这些方法在一些复杂场景上表现出了较好的效果。LI 等^[35]的工作显示了利用光场图像进行显著性检测的实用性,首先估计聚焦性和物体完整性线索,然后将其与对比度线索结合以检测显著性目标对象,他们还建立了光场显著性数据集 LFSO。随后,LI 等^[36]提出一种加权稀疏编码框架来处理不同类型的输入(RGB 图像、RGB-D 图像和光场图像)。ZHANG 等^[37]引入位置先验、背景先验和深度线索,扩展了基于颜色对比度的 SOD 方法来实现四维光场 SOD。随后,ZHANG 等^[38]集成了多种源自光场的视觉特征和先验,提出一种集成的计算方案来检测显著性目标,并构建一个基准数据集 HFUT-Lytro。WANG 等^[13]提出一种双流的融合框架,以焦堆栈和全聚焦图像为输入,并使用对抗样本来帮助训练深度网络。ZHANG 等^[16]提出一种面向记忆的光场 SOD 网络,利用 Mo-SFM 模块的特征融合机制和 Mo-FIM 模块的特征集成机制,能够准确预测显著性目标。另外还引入了一个新的光场数据集 DUTLF-FS。李等^[39]提出一种基于聚焦度和传播机制的光场图像显著性检测方法,使用高斯滤波器对焦堆栈图像的聚焦度信息进行衡量,确定前景图像和背景图像。图1给出了相关的实例,与二维和三维的 SOD 算法相比,得益于光场图像丰富的视觉信息,四维光场显著性检测方法在挑战性场景上具有更好的检测性能。然而,现有四维光场 SOD 算法仍然没有充分考虑所有的光场输入信息以及光场视觉特征之间的互补性,导致多模态融合不充分,检测性能仍不够理想。

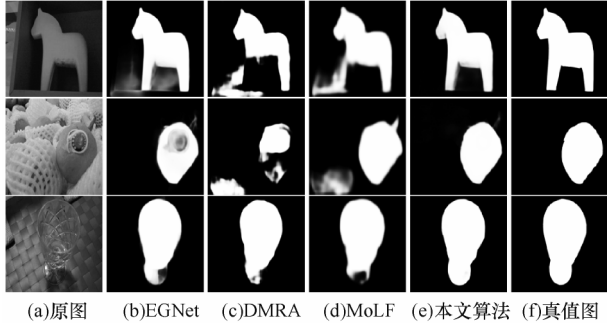


图1 不同SOD算法的实例结果

Fig.1 Sample results of different SOD algorithms

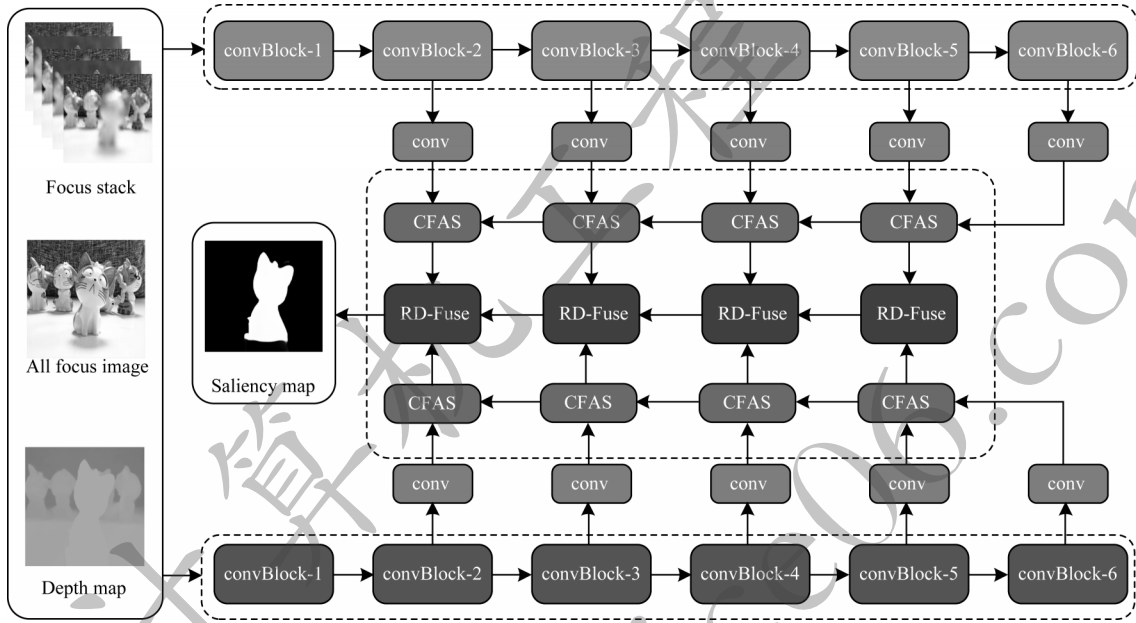


图2 本文算法的整体流程

Fig.2 Overall procedure of the proposed algorithm

2.1 多级特征聚合模块

上分支子网络以全聚焦图像和焦堆栈为输入,下分支以深度图为输入,将全聚焦图像的RGB三通道和焦堆栈的每个切片通道进行连接输入上分支网络。如文献[29,31]将深度图编码成三通道的HHA表示并送入下分支网络。如图2所示,采用ResNet-50基网络提取多级特征。与原始模型不同,借鉴EGNet^[42]的做法,在每条边路径上插入3个卷积层,以获得更具判别性和鲁棒性的上下文特征,其中在每个卷积层后添加一个ReLU层以保证网络的非线性能力。为了简洁起见,将这些卷积层和ReLU层组合表示为图2中的一个conv模块。高级特征包含抽象的语义信息,有助于定位显著目标和去除噪声,而低层特征可以提供边缘、纹理等更详细的空间结构细节,但包含部分背景噪声,这两级视觉特征是互补的^[40-41]。此外,全局上下文信息有助于从全局角度检测更完整、更准确的显著对象。因此,引入一个多级特征聚合模块,将这三级特征有效聚合以得到更具判别性的聚合特征。多个多级特征聚合模块从

2 本文方法

本文构建一个双流编解码器网络架构,并提出端到端的多模态多级特性聚合检测算法,算法总体架构如图2所示。双流子网络采用相互独立且相同的网络结构,分别进行多级特征融合。这些来自不同模态的多级聚合特征被进一步融合生成预测显著图。和其他基于CNN的SOD模型^[40-41]类似,本文提出网络也使用ResNet-50作为编码器,用于提取多级特征。和PCA^[29]和TANet^[31]相同,添加一个 15×15 的卷积层作为第6个卷积块,提取全局上下文特征。

上到下串联,自顶向下逐步聚合和精炼多级特征,其主要过程如式(1)~式(5)所示:

$$f_l^{i+1'} = \text{conv}(f_l^{i+1}) \quad (1)$$

$$f_h^{i+1} = \text{Up}(\text{conv}(f^i)) \quad (2)$$

$$f^{i+1} = \text{ReLU}(f_l^{i+1'} \odot f_h^{i+1}) \quad (3)$$

$$f_2^{i+1} = \text{ReLU}(\text{Up}(f^i) \odot \text{conv3}(f_l^{i+1'})) \quad (4)$$

$$f_l^{i+1'} = \text{conv3}(\text{Concat}(f_l^{i+1}, f_2^{i+1})) \quad (5)$$

其中: $\text{conv}(\cdot)$ 为普通卷积; $\text{Up}(\cdot)$ 为上采样运算; $\text{ReLU}(\cdot)$ 为ReLU激活函数; conv3 为 3×3 标准卷积层; \odot 为逐元素相乘运算; $\text{Concat}(\cdot)$ 为拼接。多级特征聚合模块接收通过卷积块的特征 f_l^{i+1} 和从邻近上层多级特征聚合模块的输出特征 f_h^{i+1} ; f_l^i 为卷积块 i 提取的特征。最上层多级特征聚合模块以通过卷积块的输出特征 f_l^5 和卷积块6的输出特征为输入。通过这种自上而下的监督方式,多级特征逐渐被聚合和精炼。不同于之前(例如:PoolNet, BASNet)的拼接或直接相加的聚合方式,本文采用更有效的直接相乘运算方式来增强检测响应并抑制背景噪声,如图3所示。

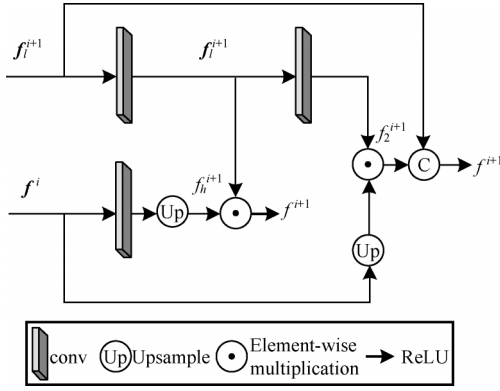


图3 多级特征聚合模块的网络结构

Fig.3 Network structure of multi-level feature aggregation module

2.2 跨模态特征融合模块

为了充分利用不同模态间特征的互补性信息,与文献[29,31]类似,本文设计一种跨模态特征融合模块,该模块可以有效地捕获跨模态特征和多级特征之间的互补性。如图2所示,将多个多模态特征融合模块放置于双流子网络成对的多级特征聚合模块之间,自顶向下进行级联,进一步对不同模态间的多级特征进行精炼和融合,并生成最终的显著预测图。为了充分获取多模式互补信息,跨模态残差连接和互补性监督信息同时被用来进一步提升不同模态的特征互补性。第 i 级由 1×1 卷积层挑选出的深度特征 F_D^i ,与另一个支流网络中残差连接得到的特征 F_R^i 进行逐元素相加。增强的特征 $F_R^{i'}$ 作为两个 1×1 卷积层的输入,减少训练过程中计算量。过程描述如式(6)所示:

$$F_R^{i'} = F_R^i + RC(\text{conv}(F_D^i)) \quad (6)$$

类似地,跨模态残差连接和互补性监督也被用来增强深度特征 F_D^i ,计算过程如式(7)所示:

$$F_D^{i'} = F_D^i + RC(\text{conv}(F_R^i)) \quad (7)$$

增强的特征 $F_R^{i'}$ 、 $F_D^{i'}$ 与相邻的上层融合特征进行连接,并经过一个 1×1 卷积,生成更具判别性的显著性预测图,如式(8)所示:

$$\text{Pred}_{RD}^{i+1} = \text{conv}(\text{Concat}(F_R^{i'}, F_D^{i'})) \quad (8)$$

2.3 Adabelief优化器

优化器的选择影响神经网络训练的检测精度和可靠性。现有方法通常采用SGD、Adam等优化器。实验结果表明,SGD具有较好的泛化能力,但收敛速度慢、训练不稳定,需要仔细调参;Adam无需仔细调参即可达到较好的效果,但很难得到最佳效果,且泛化性不够理想。本文引入了更强壮的AdaBelief^[44]优化器,类似于Adam,无需细心调参便能得到更好的效果,其定义如下:

$$\theta_t = \theta(t-1) - \alpha \frac{m_t}{\sqrt{s_t}} \quad (9)$$

其中: α 为学习率; m_t 为指数移动。

3 实验结果与分析

3.1 实验设置

为了评估所提算法的性能,在现有的光场显著性检测基准数据集DUTLF-FS^[13,16]和HFUT-Lytro^[38]上进行实验对比。性能评估度量除了采用权威的准确度-召回率曲线(PRC)、F-measure(F_β)、加权F-measure(WF_β)^[44]和平均绝对误差(MAE)^[45]外,还采用了结构相似性指标(Sm)^[46]和增强匹配指标(Em)^[47]共计6个指标全面评估所提出算法的性能。本文提出的算法基于深度学习框架PyTorch实现,在Nvidia GTX 3090 GPU上进行训练。

3.2 消融实验

本文在权威的光场显著性检测数据集DUTLF-FS和HFUT-Lytro上进行了消融实验,并采用 F_β 、MAE、Em和Sm这4个广泛使用的评价指标对算法中的多级特征聚合模块(CFAS)和跨模态特征融合模块(RD-Fuse)进行有效性验证。表1所示为在DUTLF-FS和HFUT-Lytro两个数据集上的测试结果。其中, \checkmark 为添加模块,在HFUT-Lytro中,基网络(第1行)在MAE评价指标为0.122;第2行是在基模型基础上添加CFAS模块的结果,其MAE指标降低到0.095;第3行是在基模型基础上使用RD-Fuse得到的结果,其MAE指标降低到0.106;第4行为同时组合CFAS模块和RD-Fuse模块的结果,其MAE评价指标进一步降低到0.083。上述结果表明,CFAS模块和RD-Fuse模块对检测性能均有提升,且是相容的。

表1 CFAS和RD-Fuse模块的测试结果

Table 1 Test results of CFAS and RD-Fuse modules

数据集	Base	CFAS	RD-Fuse	F_β	MAE	Em	Sm
DUTLF-FS	\checkmark			0.795	0.081	0.846	0.756
	\checkmark	\checkmark		0.857	0.058	0.896	0.823
	\checkmark		\checkmark	0.841	0.051	0.887	0.804
	\checkmark	\checkmark	\checkmark	0.912	0.038	0.946	0.925
HFUT-Lytro	\checkmark			0.578	0.122	0.706	0.738
	\checkmark	\checkmark		0.621	0.095	0.753	0.796
	\checkmark		\checkmark	0.604	0.106	0.741	0.782
	\checkmark	\checkmark	\checkmark	0.697	0.083	0.798	0.812

3.3 定量分析

为全面评估所提出算法的性能,与目前主流显著性目标检测方法MOLF^[16]、AFNet^[33]、DMRA^[34]、LFS^[35]、WSC^[36]、DILF^[37]、F3Net^[41]、CPD^[48]这8种先进的二维、三维和四维SOD算法进行了定量的实验对比。为保证实验对比的公平性,所有对比算法的显著图或由作者直接提供,或由作者官方源代码生

成。如图4所示,本文算法精度基本都超过其他主流算法。观察图5可以看出,本文算法的灰色矩形条高度最低,即MAE误差最小;而黑色矩形条值最高,即 WF_β 指标更具有优势。从图6的PRF值可知,本文算法也明显优于其他算法,原因是本文采用的多级特征聚合模块在确保较高的召回率下能获得更好精度和 F_β 值结果。

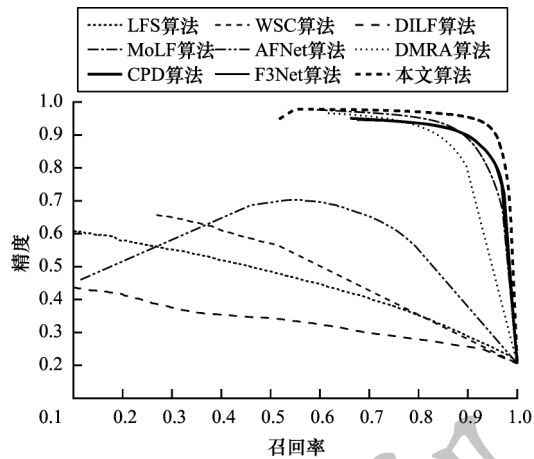


图4 在DUTFS-FS数据集上的PR曲线
Fig.4 PR curves on DUTFS-FS dataset

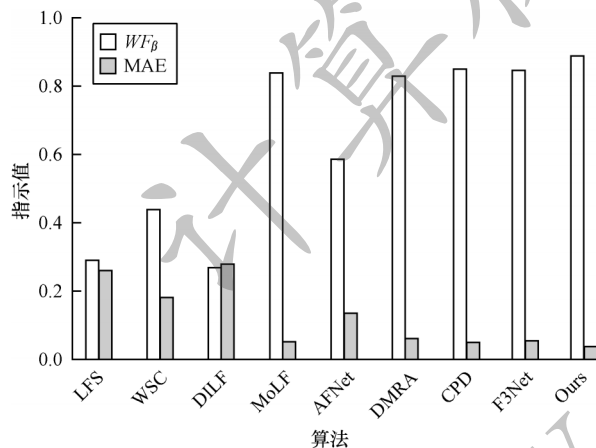


图5 在DUTFS-FS数据集上的 WF_β 和MAE值
Fig.5 WF_β and MAE values on DUTFS-FS dataset

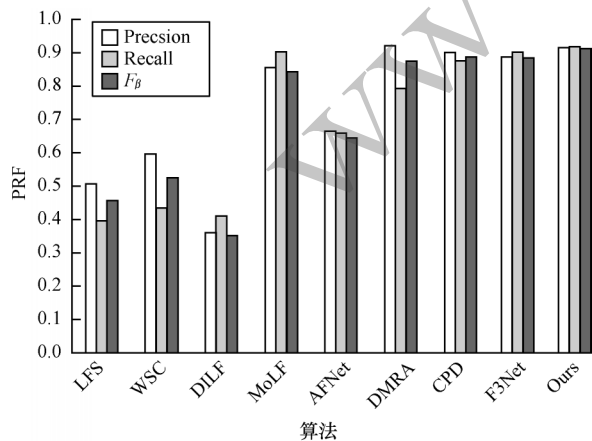


图6 在DUTFS-FS数据集上的精度、召回率和 F_β 值
Fig.6 Precision, recall, and F_β values on DUTFS-FS dataset

图7~图9分别给出了在另一个数据集HFUT-Lytro上的PR曲线、 WF_β 和MAE值以及PRF值,结果表明本文算法的性能更优。

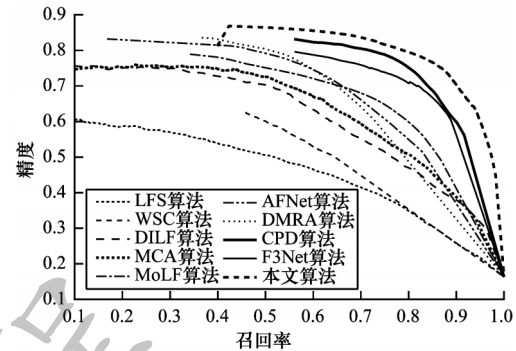


图7 在HFUT-Lytro数据集上的PR曲线
Fig.7 PR curves on HFUT-Lytro dataset

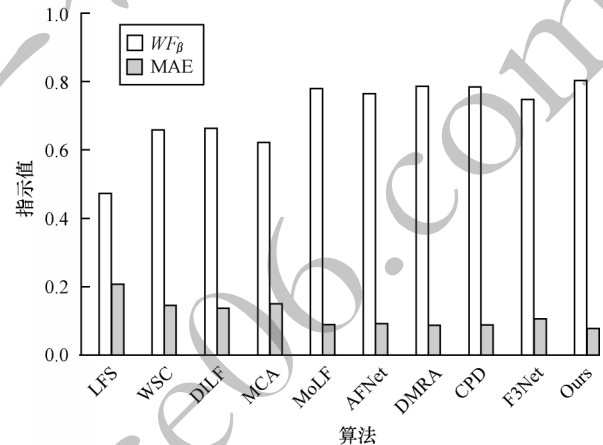


图8 在HFUT-Lytro数据集上的 WF_β 和MAE值
Fig.8 WF_β and MAE values on HFUT-Lytro dataset

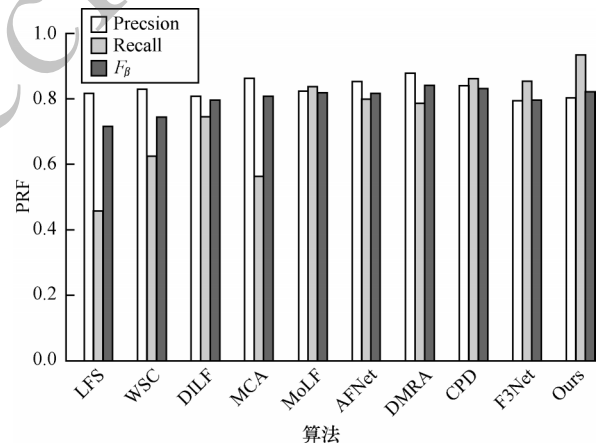


图9 在HFUT-Lytro数据集上的精度、召回率、 F_β 值
Fig.9 Precision, recall rate and F_β value on HFUT-Lytro

3.4 定性分析

图10所示为不同SOD算法视觉对比,给出5个具有代表性的样本实例。第1行为前/背景颜色相似的情况,F3Net检测完全失效;在第2行~第4行为背景杂乱的情形,F3Net检测出现部分噪声未完全抑制,其虽然也采用多级特征聚合方式,但由于CNN

网络提取特征有限,仅依靠CNN的SOD方法并不能达到特别好的效果。总体来看,本文算法可以在前/背景相似、背景杂乱等挑战性场景图像上能更有效

地抑制背景,精确地检测出完整的显著性目标对象,这主要得益于CNN强大的特征表示能力以及光场丰富的视觉特征。

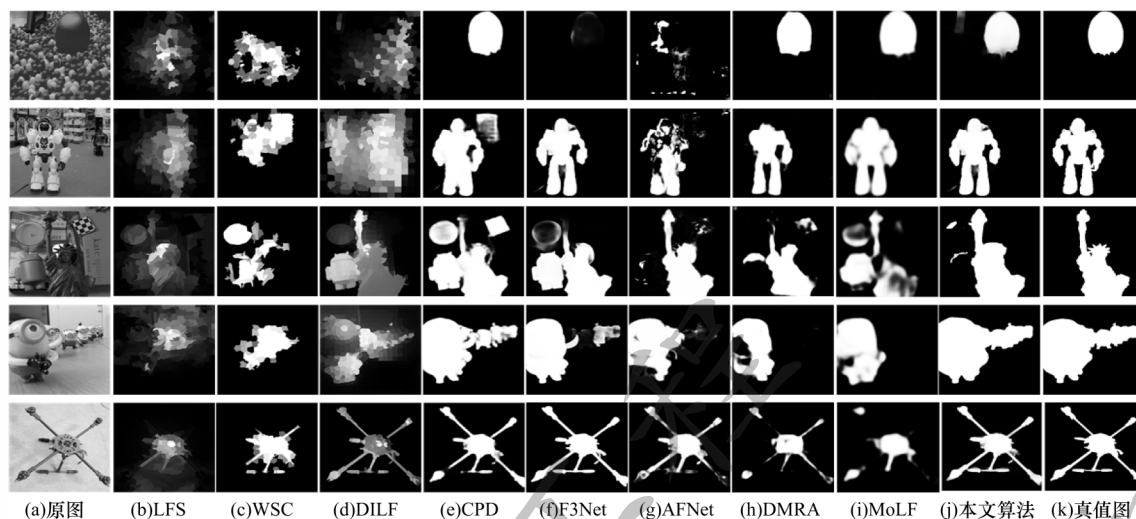


图10 不同SOD算法的视觉对比

Fig.10 Visual comparison of different SOD algorithms

4 结束语

本文提出一种多模态多级特征聚合网络算法来检测显著性目标对象。利用光场图像单模态内的多级多尺度特征和不同模态间互补的多模态多级特征,检测各类挑战性场景中的显著性目标,并在DUTLF-FS、HFUT-Lytro光场基准数据集上与8种目前先进的二维、三维和四维SOD算法进行综合的性能对比。实验结果表明,该算法在各个权威性能评价指标上均取得了更好的结果。由于光场图像丰富的视觉信息有助于解决复杂自然场景下的显著性目标检测问题,因此下一步将挖掘光场中更丰富的视觉信息与特征,以更精准地检测出完整的显著性目标对象。

参考文献

- [1] FAN D P, JI G P, SUN G L, et al. Camouflaged object detection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2020: 2774-2784.
- [2] ZHOU W H, LIANG L K, ZHANG H, et al. Scale and orientation aware EPI-patch learning for light field depth estimation [C]//Proceedings of the 24th International Conference on Pattern Recognition. Beijing, China; [s. n.], 2018: 2362-2367.
- [3] SONG G, LEE K M. Depth estimation network for dual defocused images with different depth-of-field [C]//Proceedings of the 25th IEEE International Conference on Image Processing. Washington D. C. , USA; IEEE Press, 2018: 1563-1567.
- [4] YEUNG H W F, HOU J H, CHEN X M, et al. Light field spatial super-resolution using deep efficient spatial-angular separable convolution [J]. IEEE Transactions on Image Processing, 2019, 28(5): 2319-2330.
- [5] HOU Q B, CHENG M M, HU X W, et al. Deeply supervised salient object detection with short connections [C]//Proceedings of IEEE Conference on Pattern Analysis and Machine Intelligence. Washington D. C. , USA; IEEE Press, 2017: 815-828.
- [6] ZHANG X N, WANG T T, QI J Q, et al. Progressive attention guided recurrent network for salient object detection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018: 714-722.
- [7] DENG Z J, HU X W, ZHU L, et al. R³Net: recurrent residual refinement network for saliency detection [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden; [s. n.], 2018: 684-690.
- [8] 花卉. 多视觉特征结合有约束简化群优化的显著性目标检测[J]. 计算机工程, 2015, 41(11): 257-262.
- [9] 李东民, 李静, 梁大川, 等. 基于多尺度先验深度特征的多目标显著性检测方法[J]. 自动化学报, 2019, 45(11): 2058-2070.
- [10] LI D M, LI J, LIANG D C, et al. Multiple salient objects detection using multi-scale prior and deep features [J]. Acta Automatica Sinica, 2019, 45(11): 2058-2070. (in Chinese)
- [11] 张晴, 李云, 李文举, 等. 融合深度特征和多核增强学习的显著目标检测[J]. 中国图象图形学报, 2019, 24(7): 1096-1105.
- [12] ZHANG Q, LI Y, LI W J, et al. Salient object detection via deep features and multiple kernel boosting learning [J]. Journal of Image and Graphics, 2019, 24(7): 1096-1105. (in Chinese)
- [13] WANG A Z, WANG M H. RGB-D salient object detection via minimum barrier distance transform and saliency fusion [J]. IEEE Signal Processing Letters, 2017, 24(5): 663-667.
- [14] HAN J W, CHEN H, LIU N, et al. CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion [J]. IEEE Transactions on Cybernetics, 2018, 48(11): 3171-3183.

- [13] WANG T T, PIAO Y R, LU H C, et al. Deep learning for light field saliency detection[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2019: 8837-8847.
- [14] WANG A Z, WANG M H, LI X Y, et al. A two-stage Bayesian integration framework for salient object detection on light field[J]. Neural Processing Letters, 2017, 46(3): 1083-1094.
- [15] WANG A Z, WANG M H, PAN G, et al. Salient object detection with high-level prior based on Bayesian fusion[J]. IET Computer Vision, 2017, 11(3): 199-206.
- [16] ZHANG LI M I, WEI J, et al. Memory-oriented decoder for light field salient object detection [C]//Proceedings of International Conference on Neural Information Processing Systems. Washington D. C. , USA; IEEE Press, 2019: 896-906.
- [17] FAN D P, CHENG M M, LIU J J, et al. Salient objects in clutter: bringing salient object detection to the foreground[C]//Proceedings of European Conference on Computer Vision. Munich, Germany: Springer, 2018: 186-202.
- [18] LI G B, YU Y Z. Visual saliency detection based on multiscale deep CNN features[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5012-5024.
- [19] ZHAO R, OUYANG W L, LI H S, et al. Saliency detection by multi-context deep learning[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2015: 1265-1274.
- [20] LIU N, HAN J W, YANG M H. PiCANet: learning pixel-wise contextual attention for saliency detection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018: 3089-3098.
- [21] 张青博,王斌,崔宁宇,等. 基于注意力机制的规范化矩阵分解推荐算法[J]. 软件学报, 2020, 31(3): 778-793.
ZHANG Q B, WANG B, CUI N N, et al. Attention-based regularized matrix factorization for recommendation[J]. Journal of Software, 2020, 31(3): 778-793. (in Chinese)
- [22] 周雨佳,窦志成,葛松玮,等. 基于递归神经网络与注意力机制的动态个性化搜索算法[J]. 计算机学报, 2020, 43(5): 812-826.
ZHOU Y J, DOU Z C, GE S W, et al. Dynamic personalized search based on RNN with attention mechanism [J]. Chinese Journal of Computers, 2020, 43(5): 812-826. (in Chinese)
- [23] 冯兴杰,张乐,曾云泽. 基于多注意力CNN的问题相似度计算模型[J]. 计算机工程, 2019, 45(9): 284-290.
FENG X J, ZHANG L, ZENG Y Z. Question similarity calculation model based on multi-attention CNN [J]. Computer Engineering, 2019, 45(9): 284-290. (in Chinese)
- [24] WANG L Z, WANG L J, LU H C, et al. Salient object detection with recurrent fully convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1734-1746.
- [25] LI G B, YU Y Z. Deep contrast learning for salient object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016: 478-487.
- [26] ZHANG P P, WANG D, LU H C, et al. Amulet: aggregating multi-level convolutional features for salient object detection[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2017: 202-211.
- [27] WANG W G, LAI Q X, FU H Z, et al. Salient object detection in the deep learning era: an in-depth survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3239-3259.
- [28] QU L Q, HE S F, ZHANG J W, et al. RGBD salient object detection via deep fusion[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2274-2285.
- [29] 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.
- [30] CHEN H, LI Y F, SU D. Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection[J]. Pattern Recognition, 2019, 86: 376-385.
- [31] 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.
- [32] ZHU C B, CAI X, HUANG K, et al. PDNet: prior-model guided depth-enhanced network for salient object detection [C]//Proceedings of International Conference on Multimedia and Expo. Shanghai, China: [s. n.], 2019: 199-204.
- [33] WANG N N, GONG X J. Adaptive fusion for RGB-D salient object detection[J]. IEEE Access, 2019, 7: 55277-55284.
- [34] 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, 2019: 7253-7262.
- [35] LIN Y, YE J W, JI Y, et al. Saliency detection on light field [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2014: 2806-2813.
- [36] LI N Y, SUN B L, YU J Y. A weighted sparse coding framework for saliency detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2015: 5216-5223.
- [37] ZHANG J, WANG M, GAO J, et al. Saliency detection with a deeper investigation of light field [C]//Proceedings of the 24th International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina: [s. n.], 2015: 2212-2218.
- [38] ZHANG J, WANG M, LIN L, et al. Saliency detection on light field: a multi-cue approach[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 13(3): 32.
- [39] 李爽,邓慧萍,朱磊,等. 联合聚焦度和传播机制的光场图像显著性检测[J]. 中国图象图形学报, 2020, 25(12): 2578-2586.
LI S, DENG H P, ZHU L, et al. Saliency detection on a light field via the focusness and propagation mechanism[J]. Journal of Image and Graphics, 2020, 25(12): 2578-2586. (in Chinese)
- [40] CHEN Z Y, XU Q Q, CONG R M, et al. Global context-aware progressive aggregation network for salient object detection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10599-10606.
- [41] WEI J, WANG S H, HUANG Q M. F³Net: fusion, feedback and focus for salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12321-12328.

(上接第233页)

- [42] ZHAO J X, LIU J J, FAN D P, et al. EGNNet: edge guidance network for salient object detection[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019: 8778-8787.
- [43] ZUANG J T, TANG T, DING Y F, et al. Optimizer: adapting stepsizes by the belief in observed gradients[C]//Proceedings of Conference on Neural Information Processing Systems. Washington D. C. , USA: IEEE Press, 2020: 136-145.
- [44] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2014: 248-255.
- [45] PERAZZI F, KRÄHENBÜHL P, PRITCH Y, et al. Saliency filters: contrast based filtering for salient region detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2012: 733-740.
- [46] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: a new way to evaluate foreground maps[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2017: 4558-4567.
- [47] FAN D P, GONG C, CAO Y, et al. Enhanced-alignment measure for binary foreground map evaluation[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden; [s. n.], 2018: 698-704.
- [48] 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, 2019: 3907-3916.

编辑 索书志