中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2025)07-2451-17

论文引用格式:Chen X L, Zhang X G, Du Z L and Wang X. 2025. Distortion semantic aggregation network for salient object detection in 360° omnidirectional images. Journal of Image and Graphics, 30(7):2451-2467(陈晓雷, 张学功, 杜泽龙, 王兴. 2025. 面向360°全景图像显著目标检测的畸变语义聚合网络. 中国图象图形学报, 30(7):2451-2467)[DOI:10.11834/jig.240371]

面向360°全景图像显著目标检测的畸变语义聚合网络

陈晓雷*,张学功,杜泽龙,王兴

兰州理工大学电气工程与信息工程学院,兰州　730050

摘　要:**目的**　为了有效应对360°全景图像的几何畸变和大视野特性带来的挑战,提出一种畸变自适应语义聚合网络(distortion semantic aggregation network,DSANet)。该网络能够提升360°全景图像显著目标检测性能。**方法**　DSANet由3个模块组成:畸变自适应校正模块(distortion aware calibration module,DACM)、多尺度语义注意力聚合模块(multiscale semantic attention aggregation module,MSAAM)以及渐进式细化模块(progressive refinement module,PRM)。DACM模块利用不同扩张率的可变形卷积来学习自适应权重矩阵,校正360°全景图像中的几何畸变;MSAAM模块结合注意力机制和可变形卷积,提取并融合全局语义特征与局部细节特征,生成多尺度语义特征;PRM模块逐层融合多尺度语义特征,进一步提升检测精度。MSAAM模块与PRM模块相配合,解决360°全景图像的大视野问题。**结果**　在两个公开数据集360-SOD和360-SSOD(共计1 605幅图像)上进行的实验表明,DSANet在6种主流评价指标(Max F-measure、Mean F-measure、MAE(mean absolute error)、Max E-measure、Mean E-measure、Structure-measure)上均优于其他方法。**结论**　本文方法在多个客观评价指标上表现突出,同时生成的显著目标图像在边缘轮廓性和空间结构细节信息上更为清晰。

关键词:深度学习;显著目标检测(SOD);360°全景图像;几何畸变;大视野

Distortion semantic aggregation network for salient object detection in 360° omnidirectional images

Chen Xiaolei*, Zhang Xuegong, Du Zelong, Wang Xing

College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou　730050, China

Abstract: **Objective**　By integrating corrected images with spatial and semantic information, salient object detection (SOD) that uses omnidirectional image data significantly outperforms single-modality detection in terms of prediction accuracy. The emergence of deep learning techniques has further accelerated advancements in SOD for omnidirectional images. However, existing models in this area fundamentally disregard the challenges of image distortion and the distinct characteristics of different modalities. They commonly rely on simplistic fusion methods, such as element addition, multiplication, or concatenation, which fail to foster meaningful interactions among the modalities of omnidirectional images. This approach does not effectively utilize complementary information or the potential correlations among them. To rectify this deficiency, developing more robust methods that enhance information interaction across different modalities of omnidirectional images is imperative, leading to superior SOD results. Researchers have successfully designed a distortion semantic

收稿日期:2024-07-15;**修回日期**:2024-12-05;**预印本日期**:2024-12-12

*通信作者:陈晓雷　chenxl1703@lut.edu.cn

基金项目:国家自然科学基金项目(61967012)

Supported by: National Natural Science Foundation of China (61967012)

aggregation network (DSANet). This innovative method is applied for the first time to SOD in omnidirectional images, rigorously analyzes the correlations among various modalities, and leverages these relationships to optimize the fusion and interaction processes. **Method** DSANet consists of three modules: the distortion adaptive correction module (DACM), the multi-scale semantic attention aggregation module (MSAAM), and the progressive refinement module (PRM). First, an aberration problem occurs in transforming omnidirectional images into equal rectangular projection (ERP) images, and the aberration partially affects the accuracy of target detection. Therefore, this study designs the DACM module to adaptively correct the distortion features of the input ERP image to improve detection accuracy. The mechanism of the DACM module learns the adaptive weight matrix by using deformable convolutions with different dilatation rates for the input ERP image to correct geometric distortions in 360° omnidirectional images. Second, the corrected ERP image is fed into the ResNet-50 encoder for feature extraction. Convolutional neural networks can obtain abstract semantic features through multiple convolution and pooling operations, but the diversity of input images results in significant target scale and location uncertainty, and the theoretical receptive field is frequently different from the actual receptive field to a certain extent, resulting in the network not being able to effectively extract global semantic features. Therefore, if single-scale deep features are used directly, then correct salient targets may not be obtained due to the lack of semantic features. Meanwhile, the combination of global semantic features after channel attention and local detailed features after spatial attention can solve the problem. Based on this idea, this study designs MSAAM, which is primarily composed of two sub-modules: global semantic features and local detail features. The mechanism of MSAAM is to input the features extracted by the encoder into the channel attention module to extract the channel weight values through convolutional processing and then multiply them by pixels with the input features to obtain the global semantic features. Subsequently, the global semantic features are inputted into the spatial attention module to obtain the spatial weight values and multiply them by the input feature elements to obtain the detailed features. Finally, the global semantic features are merged with the local detailed features to generate the multi-scale semantic features. The PRM module fuses the multi-scale semantic features layer by layer to further improve detection accuracy. The MSAAM module cooperates with the PRM module to solve the problem of a large field of view in 360° omnidirectional images. Finally, the saliency map generated by the MSAAM module is unclear and internally mutilated due to the contradiction among different feature layers in the fusion process. To obtain more accurate saliency maps, this study designs the PRM in the decoder, which fuses the multi-scale semantic feature maps from the encoder layer by layer from deep to shallow maps, to further refine and enhance the feature maps at each stage. The mechanism of action of PRM up-samples the high-level features and fuses them with the low-level features for elemental multiplication and then multiplies and sums them with each of the high-level and low-level features to obtain the coarse-level features. The coarse-level features are activated to obtain the weight values, inverted, multiplied with the coarse-level features, summed to obtain the fine-level features, and then aggregated from the high level to the low level to obtain the output features. **Result** Experiments on two publicly available datasets, namely, 360-SOD and 360-SSOD (total of 1 605 images), show that DSANet outperforms the other methods in six mainstream evaluation metrics, i. e., max F-measure, mean F-measure, mean absolute error (MAE), max E-measure, mean E-measure, and structure-measure. **Conclusion** The method proposed in this study excels in several objective evaluation metrics while generating salient object images that are clearer in terms of edge contouring and spatial structural detail information.

Key words: deep learning; salient object detection (SOD); 360° omnidirectional image; geometric distortion; large-scale vision

0 引 言

显著目标检测(salient object detection, SOD)专注于定位并识别图像或视频中最能吸引人注意的对象(Itti 等, 1998; Borji 等, 2019)。在增强现实(augmented reality, AR)和虚拟现实(virtual reality, VR)迅猛发展的今天,360°全景图像和视频应用日益流行。戴上头戴式显示器(head-mounted display, HMD),用户可以通过转动头部自由地改变观看视角,享受身临其境的交互式视觉体验。360°全景图像和视频的应用带来了一系列新的挑战,包括360°全景图像和视频的制作(Tang 等, 2019; Li 等, 2018a)、压缩(Fu 等, 2009; Shum 等, 2005)、融合(Xu

等，2021）以及渲染（Zheng等，2022）。在这一背景下，360°全景图像显著目标检测（简称360° SOD）的重要性日益凸显。由于用户可以在HMD中选择性地观察场景，那么识别出最引人注意的物体变得尤为关键。这既是技术挑战，也是为用户打造完美沉浸式体验的关键。因此，开展360° SOD研究，成为实现深度沉浸式体验的一个重要步骤。尽管2D图像显著目标检测（简称2D SOD）已经获得了广泛的研究和深入的探讨（Pan等，2021；Ma等，2021），但由于360°全景图像的以下特殊性，使得传统的2D SOD难以直接适用于360° SOD，亟需研究针对360°全景图像特性的360° SOD方法。

首先，360°全景图像本身是球形图像，目前需要通过等矩形映射投影（equal rectangular projection，ERP）、立方体映射投影（cube-map projection，CMP）等技术将球形图像转换为平面图像再进行后续视觉任务处理，这一过程往往会产生大量几何畸变，而2D SOD针对的是平面图像，因而不需要考虑畸变问题。其次，与2D图像的局部视野不同，360°全景图像能够在360° × 180°的视野下提供更为丰富和全面的目标信息。由于视野范围的差异性，直接将2D SOD应用于360° SOD不仅不能有效地突出前景目标，还可能无法充分抑制背景信息。

针对360°全景图像的几何畸变特性，Ma等人（2020）提出一种新的对象级语义显著性排序模块，用于处理各种视觉扭曲，以准确定位显著视点和对象。Li等人（2020）提出一种畸变自适应模块，可以自适应处理360°全景图像的畸变。Huang等人（2020）提出投影特征自适应和多层次特征自适应模块，对ERP图像和CMP图像的特征进行选择性加权和融合，以减少投影失真。由于这些方法主要侧重于消除360°全景图像投影失真的问题，对图像全局语义信息和局部细节信息的考虑较少，导致面对挑战性的场景时，不能很好地解决显著物体的推理问题，显著目标检测的边缘保持性能较差。针对这一问题，本文提出畸变自适应校正模块（distortion aware calibration module，DACM）。

针对360°全景图像的大视野特性，Ren等人（2021）通过建立像素之间的上下文信息来辅助显著目标检测。Huo等人（2022）通过分组逐点卷积，整合特征图像中的上下文信息，以此提升检测的准确性。Chen等人（2020）通过捕获特征图上的全局上下文信息，并对其赋予不同的权重值，实现高级特征的增强。He等人（2024）提出一个语义和上下文特征聚合网络，通过语义引导模块，将视觉Transformer捕获的ERP图像的全局视觉信息与卷积神经网络（convolutional neural network，CNN）捕获的CMP图像的局部细节信息进行有效融合。叶欣悦 等人（2024）提出通过融合颜色、深度和空间信息，利用RGB_D两种模态数据取得更加准确的预测结果。尽管这些方法在处理大视野问题上取得了进步，但它们大多仅从单一尺度的图像中提取语义信息，而未充分考虑不同尺度之间语义信息的相互作用。针对这一问题，本文提出多尺度语义注意力聚合模块（multiscale semantic attention aggregation module，MSAAM）和渐进式细化模块（progressive refinement module，PRM）。

为了更好地应对360°全景图像的几何畸变特性和大视野特性，本文在已有研究基础上，提出一种用于360° SOD的畸变语义聚合网络（distortion semantic aggregation network，DSANet）。该网络主要由本文提出的畸变自适应校正模块DACM、多尺度语义注意力聚合模块MSAAM和渐进式细化模块PRM构成。首先ERP图像f'通过DACM模块对畸变区域进行自适应校正获得校正图像f。然后将校正图像f输入到改进后的ResNet-50（residual network 50 layers）骨干网络，提取得到5个层次的特征图f_i（i = 1，2，3，4，5），再将f_i（i = 1，2，3，4，5）输入到多尺度语义注意力聚合模块进行特征提取并通过连接操作获得多尺度语义特征图c_i（i = 1，2，3，4）。接着c_i（i = 1，2，3，4）输入到渐进式细化模块PRM_i（i = 1，2，3）进行特征图逐层细化获得细节图p_i（i =1，2，3），最后将细节图p_i（i =1，2，3）通过连接融合生成清晰准确的显著图S。

本文工作创新点和贡献概括如下：1）提出一种新的360° SOD网络DSANet，该网络有效应对360°全景图像的几何畸变特性和大视野特性，充分考虑图像全局语义信息和局部细节信息，以及不同尺度之间上下文信息的相互作用，实验证明在两个公开的360° SOD数据集上，DSANet综合性能优于13种先进的SOD方法。2）提出畸变自适应校正模块DACM，该模块通过计算输入ERP图像的局部权重矩阵和全局权重矩阵，能够自适应地校正ERP图像

的畸变特征。3)提出多尺度语义注意力聚合模块MSAAM和渐进式细化模块PRM,MSAAM模块通过结合通道注意力后的全局语义特征和空间注意力后的局部细节特征,消除了使用单尺度深层特征带来的语义特征缺失问题,PRM模块对MSAAM模块提取的多尺度语义特征图由深层到浅层渐进式聚合,逐步生成清晰准确的显著图,两个模块联合工作更好地应对了360°全景图像的大视野特性。

1 相关工作

1.1 2D SOD方法

近年来,在2D SOD领域,Transformer框架和卷积神经网络(CNN)框架已成为主流研究方法。

在基于Transformer的2D SOD研究中,代表性成果颇多。例如,Liu等人(2022)在编码阶段利用Swin Transformer网络提取基本特征,并通过注意力机制对齐跨模态特征。Su等人(2024)提出一种Transformer框架,专门用于协同显著性检测和视频显著目标检测,该框架搭载了两个Transformer模块以捕捉不同像素间一组特征的远程依赖关系。Fang等人(2024)设计了一个包含多个Transformer模块的系统,旨在学习多模态和多尺度特征之间的协同信息。Yuan等人(2021)利用T2T-ViT(tokens-to-token vision Transformer)作为骨干网络,并提出一种多任务Transformer解码器,旨在同时检测显著目标及其边界。Zhu等人(2021)基于纹理标签和伪装物体标签之间的互补关系,提出一种交互式引导框架,专门解决具有模糊边界和复杂纹理的SOD问题。Xie等人(2022)提出一种单阶段金字塔嫁接网络,结合了Transformer和CNN的优势,独立从不同分辨率的图像中提取特征,进而通过特征迁移和注意导向损失的监督,有效进行SOD。这些方法不仅展示了Transformer在视觉领域的强大潜力,也促进2D SOD向着更加精准和高效的方向前进。

在CNN驱动的2D SOD领域,多项研究展示了创新性的方法和显著的成果。Cong等人(2022)引入了关系推理网络,该网络结合了基于图卷积的关系推理编码器和多尺度注意力解码器,有效提升了SOD的精准度。Song等人(2023)提出一种级联显著性检测框架,通过细化显著目标的细节图和主体图,生成更为精确的显著图。Li等人(2023a)设计了一种邻域上下文协调网络,通过促进相邻特征之间的互动,增强了上下文信息的捕获能力。Chen等人(2020)通过引入全局上下文感知机制和多尺度信息融合,显著提升了SOD的准确性和鲁棒性。Zhao等人(2015)利用CNN构建了一个融合全局和局部上下文信息的SOD框架。Lee等人(2016)通过结合CNN编码的低级距离图和VGG(Visual Geometry Group)提取的高级特征图,显著提高了显著图的精度。Li和Yu(2015)通过聚合CNN提取的多尺度特征图,生成了高质量的视觉显著图。Hou等人(2019)通过在高低层之间增加短连接,实现了显著目标和边界的精确定位。Lin等人(2022)提出一种轻量级多尺度上下文网络,使用基于注意力的金字塔特征聚合机制和注意力引导网络,进一步提升了检测性能。何伟和潘晨(2022)通过结合通道—空间注意力聚合模块和注意残差细化模块,实现了精准的SOD。Li等人(2023b)提出一种基于语义匹配和边缘对齐的轻量级网络,通过动态语义匹配模块激活高级语义特征中的显著目标位置。Li等人(2024)通过动态互补注意模块实现空间和通道注意权重的实时更新,增强了网络对宽接收域内显著目标的检测能力。

1.2 360° SOD方法

Li等人(2020)开创性地建立了首个公开的360°全景图像SOD数据集,并提出一种新颖的畸变自适应模块,该模块通过切割ERP图像来降低畸变问题。Zhang等人(2022)探索了ERP图像和CMP图像之间的互补信息,设计了一种能够有效融合基于360°多投影特征的通道—空间相互关注模块。Zhang等人(2023)提出一种结合ERP和CMP两种投影方式的双分支混合投影网络,通过像素级乘法捕捉它们之间的非线性关系,并引入压缩扩张模块以增强特征表达。他们还加入了渐进预测模块,以综合不同尺度上的显著性特征。Huang等人(2023)提出一种新的特征提取网络,专门针对ERP和CMP图像,通过特征自适应网络融合技术,成功生成了高质量的显著性图。Huang等人(2023)提出一种轻量级的畸变感知网络,旨在保持网络效率的同时优化性能输出。Cong等人(2024)引入4幅CMP图像,并采用动态加权方法整合CMP图像的特征,有效保持了CMP图像中目标的完整性。Zhao等人(2023)提出一个基于Transformer的畸变自适应模型,专门解决

全景图像畸变问题。He等人(2024)提出一个语义和上下文特征聚合网络，通过语义引导模块，将ViT捕获的ERP图像的全局信息与CNN捕获的CMP图像的局部信息进行有效融合。

2 本文方法

2.1 网络总体结构与工作原理

DSANet网络结构如图1所示。该网络由畸变自适应校正模块(DACM)、多尺度语义注意力聚合模块(MSAAM)和渐进式细化模块(PRM)组成，整体上可分为编码器和解码器两部分。

首先，输入的360°全景图像f'经过预处理后，进入DACM模块。DACM模块利用不同扩张率的可变形卷积(deformable conv，DConv)来学习自适应权重矩阵，校正全景图像中的几何畸变。校正后的特征图f通过DACM输出并传递到编码器部分。

编码器部分的主要功能是提取图像的多尺度语义特征。编码器由多个卷积层组成，不同颜色的卷积块表示不同的卷积操作，用于逐层提取图像特征。每层输出的特征图分别记为$f_i(i=1,2,3,4,5)$。每一层的特征图都会通过MSAAM模块进行处理，MSAAM模块通过结合注意力机制和可变形卷积，提取并融合全局语义和局部细节特征。

在解码器部分，网络通过逐步细化特征图，最终生成显著目标检测结果。编码器输出的多尺度特征$c_i(i=1,2,3,4)$将通过解码器中的PRM模块逐层融合和细化。PRM模块首先处理最高分辨率的特征图c_4，然后在其基础上逐层融合，得到更高层次的特征p_1,p_2,p_3。最终生成的特征图通过解码器进一步上采样，得到高分辨率的显著目标图像。

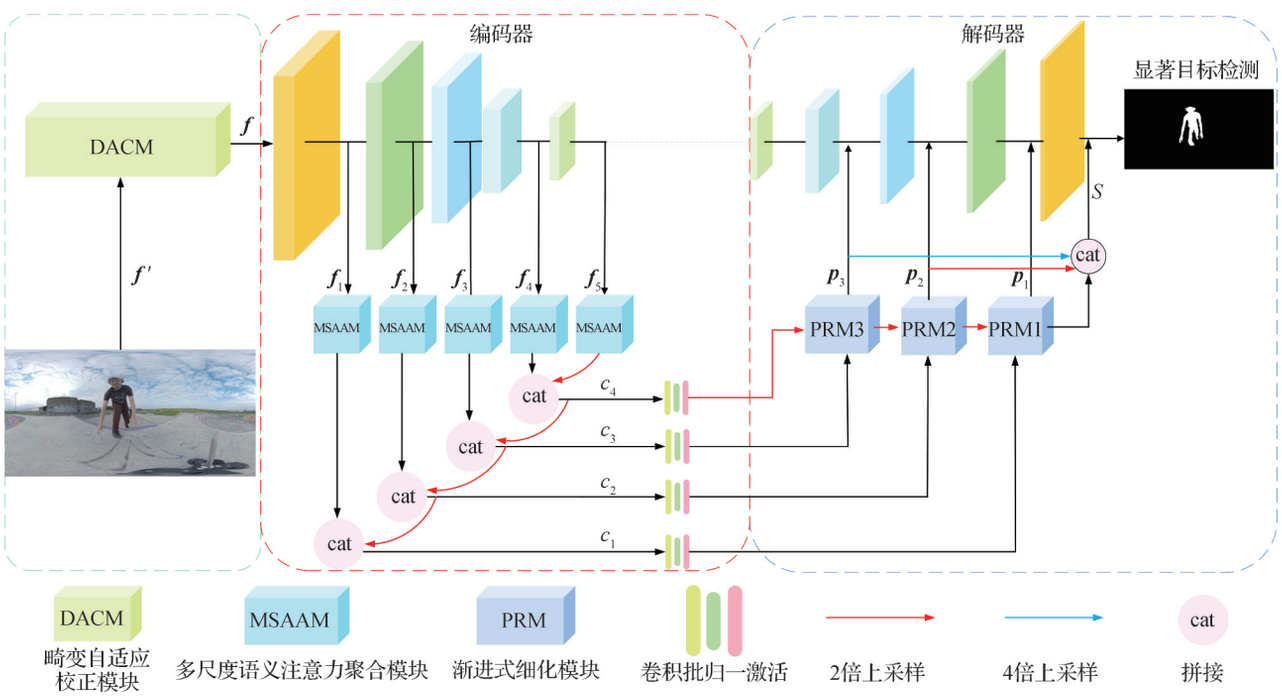

图1 整体网络架构

Fig. 1 Overall network architecture

2.2 畸变自适应校正模块DACM

ERP是最常用的360°全景图像投影方式之一，但它会在图像的不同位置，尤其是极点附近，带来不同程度的畸变，针对ERP图像的投影畸变问题，本文设计了畸变自适应校正模块DACM，模块结构如图2所示。该模块能够对输入ERP图像的畸变特征进行自适应校正。

具体来说，该模块将输入的原始ERP图像f'通过可变形卷积操作学习，得到一个局部权重矩阵f_{local}，该矩阵能够反映每个局部位置的重要性，计算

式为

$$\boldsymbol{f}_{\text{local}} = 2 \times f_{\text{sigmoid}}\left(ReLU\left(BN\left(Conv\left(\sum_{r=3,5,7} DConv_r(\boldsymbol{f}')\right)\right)\right)\right) \tag{1}$$

式中，$\boldsymbol{f}'$ 表示输入的原始ERP图像，$DConv_r()(r=3,5,7)$表示不同扩张率的可变形卷积。$Conv$表示卷积核大小为3×3的卷积，$ReLU$表示ReLU（rectified linear unit）激活函数，sigmoid表示sigmoid激活函数。系数2表示将矩阵归一化到范围[0, 2]，值大于1表示更重要，小于1表示不重要。

同时输入的原始ERP图像通过全局平均池化操作学习得到一个全局权重矩阵$\boldsymbol{f}_{\text{global}}$，该矩阵能够反映全局语义信息的重要性，计算式为

$$\boldsymbol{f}_{\text{global}} = 2 \times f_{\text{sigmoid}}\left(ReLU\left(BN\left(Conv\left(f_{\text{Globalpool}}(\boldsymbol{f}')\right)\right)\right)\right) \tag{2}$$

式中，$f_{\text{Globalpool}}$表示全局平均池化操作，用于提取畸变特征权重矩阵，其余变量含义同式（1）。

局部权重矩阵$\boldsymbol{f}_{\text{local}}$与原始图像$\boldsymbol{f}'$相乘获得局部校正图像，局部校正图像与全局权重矩阵$\boldsymbol{f}_{\text{global}}$相乘再与原始图像$\boldsymbol{f}'$相加获得校正图像$\boldsymbol{f}$。计算式为

$$\boldsymbol{f} = \boldsymbol{f}' \times \boldsymbol{f}_{\text{local}} \times \boldsymbol{f}_{\text{global}} + \boldsymbol{f}' \tag{3}$$

式中，$\boldsymbol{f}_{\text{local}}$表示局部权重矩阵，$\boldsymbol{f}_{\text{global}}$表示全局权重矩阵，$\boldsymbol{f}$表示校正后的ERP图像。

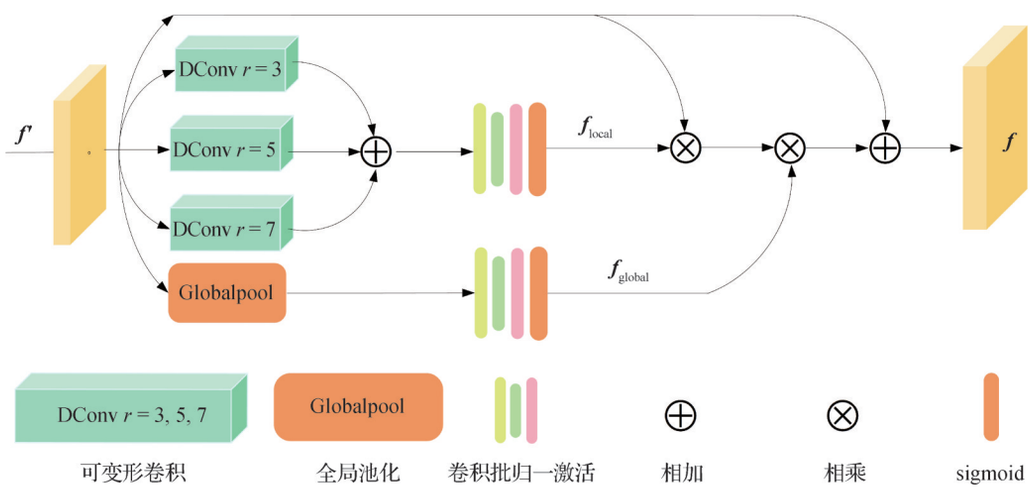

图2　DACM模块结构图

Fig. 2　DACM module structure diagram

2.3　多尺度语义注意力聚合模块MSAAM

卷积神经网络通过多次卷积和池化操作，可以得到抽象的语义特征，但是输入图像的多样性带来显著目标尺度和位置的不确定性，而且理论上的感受野往往与实际感受野有一定的差距，导致网络无法有效提取全局语义特征。因此如果直接使用单尺度的深层特征，可能会因为语义特征缺失导致无法获得正确的显著目标，而结合通道注意力（channel attention mechanism，CAM）后的全局语义特征和空间注意力（spatial attention mechanism，SAM）后的局部细节特征可以解决该问题。基于这一思路，本文设计了一个多尺度语义注意力聚合模块（MSAAM），该模块主要由全局语义特征（global semantic feature，GSF）和局部细节特征（local detail feature，LDF）两个子模块构成。模块结构如图3所示。

2.3.1　GSF子模块

GSF子模块首先将3×3卷积处理后的输入特征图$\boldsymbol{f}_i(i=1,2,3,4,5)$通过全局最大池化（global max pooling，GMP）和全局平均池化（global average pooling，GAP）操作获取图像的通道特征图$\boldsymbol{f}_c$。

然后对通道特征图$\boldsymbol{f}_c$采用多分辨率可变形卷积进行多尺度采样，再将各层次相加后进行sigmoid激活获得通道权重矩阵$\boldsymbol{a}_{t_1}$，最后将3×3卷积处理后的输入特征图$\boldsymbol{f}_i(i=1,2,3,4,5)$与通道权重矩阵$\boldsymbol{a}_{t_1}$相乘获得全局语义特征图$\boldsymbol{f}_{a1}$。上述过程可表达为

$$\boldsymbol{f}_c = GMP\left(Conv_3(\boldsymbol{f}_i)\right) + GAP\left(Conv_3(\boldsymbol{f}_i)\right) \tag{4}$$

$$\boldsymbol{a}_{t_1} = f_{\text{sigmoid}}\left(\sum_{r=3,5,7} DConv_r(\boldsymbol{f}_c)\right) \tag{5}$$

$$\boldsymbol{f}_{a1} = \boldsymbol{a}_{t_1} \times Conv_3(\boldsymbol{f}_i) \tag{6}$$

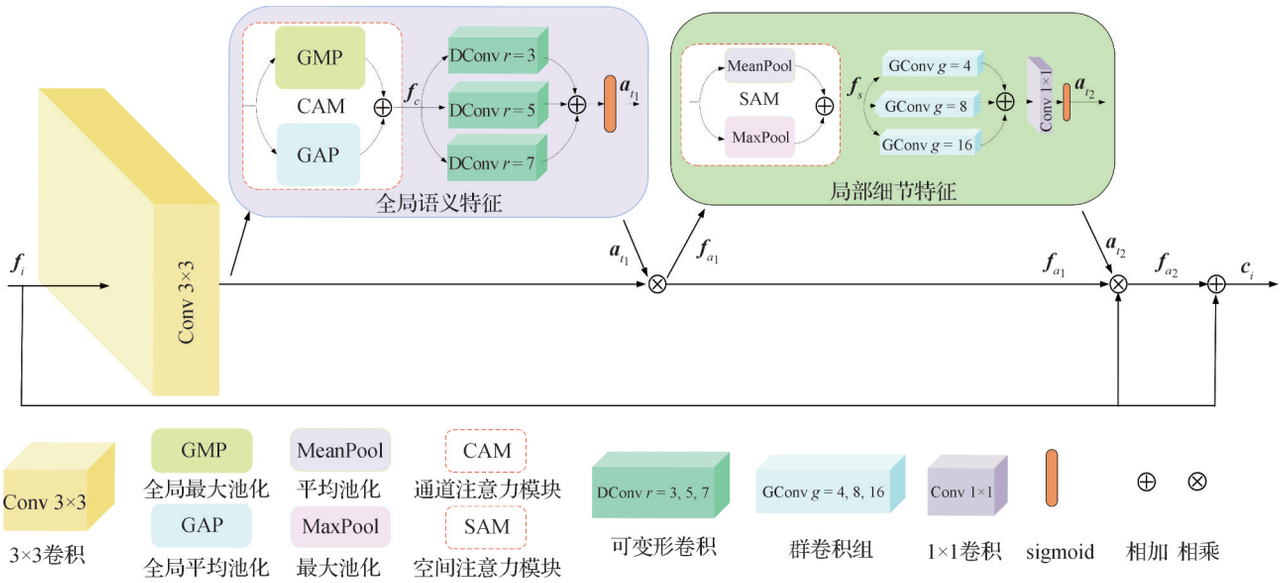

图3　MSAAM模块结构图

Fig. 3　MSAAM module structure diagram

式中，$f_i(i=1,2,3,4,5)$代表编码器提取的5个层次特征图，GAP表示全局平均池化，GMP表示全局最大池化，$DConv_r(r=3,5,7)$表示不同扩张率的可变形卷积，$Conv_3(\ )$表示卷积核大小为3×3的卷积操作，f_{a_1}表示全局语义特征图。

2.3.2　LDF子模块

该模块首先对GSF子模块输出的全局语义特征图f_{a_1}进行平均池化和最大池化获得空间特征图f_s。然后对空间特征图f_s进行卷积核大小为3×3的可变形卷积群组处理后将各层次特征相加，最后通过1×1卷积、sigmoid激活获得细节权重矩阵。上述过程可表达为

$$f_s=f_{\mathrm{MeanPool}}(f_{a1})+f_{\mathrm{MaxPool}}(f_{a_1}) \tag{7}$$

$$a_{t_2}=f_{\mathrm{sigmoid}}\left(Conv_1\left(\sum DConv_g(f_s)\right)\right) \tag{8}$$

式中，f_{a_1}为全局语义特征图，MaxPool表示最大池化操作，MeanPool表示平均池化操作，f_s表示空间特征图，$DConv_g(\ )$表示群卷积，$Conv_1$表示卷积核大小为1×1的卷积，a_{t_2}表示细节权重矩阵。

2.3.3　多尺度语义注意力聚合

GSF子模块获得的全局语义特征图f_{a_1}、LDF子模块获得的细节权重矩阵a_{t2}和输入特征图$f_i(i=1,2,3,4,5)$三者相乘获得局部细节特征图f_{a_2}，然后局部细节特征图f_{a_2}与输入特征图$f_i(i=1,2,3,4,5)$相加生成多尺度语义特征图$c_i(i=1,2,3,4)$。其计算过程为

$$f_{a2}=a_{t_2}\times f_{a_1}\times f_i \tag{9}$$

$$c_i=f_{a_2}+f_i \tag{10}$$

2.4　渐进式细化模块PRM

由于不同特征层在融合过程中存在矛盾，MSAAM模块生成的显著图存在不清晰且内部残缺现象。为了得到更准确的显著图，本文在解码器中设计了渐进式细化模块(PRM)，将来自编码器的多尺度语义特征图c_i $(i=1,2,3,4)$由深层图c_4至浅层图c_1逐层融合，以进一步细化和增强每个阶段的特征图。PRM包括PRM3、PRM2和PRM1这3个类似的子模块，结构如图4所示。此处以PRM3模块为例，介绍工作原理。

PRM3模块有两个输入，分别为多尺度语义特征图c_4和c_3。多尺度语义特征图c_4上采样后进行Conv 3×3、BN、ReLU获得特征图c_4'；多尺度语义特征图c_3进行Conv 3×3、BN、ReLU操作获得特征图c_3'。然后特征图c_4'与特征图c_3'相乘获得融合特征图F_{mul}。融合特征图F_{mul}分别与特征图c_3'、c_4'相乘后再相加获得细节特征图F_{sum}。

细节特征图F_{sum}经过Conv 3×3、BN、ReLU后再进行Conv 1×1、sigmoid激活获得权重矩阵图A_t。A_t和$(1-A_t)$分别与细节特征图F_{sum}相乘后再相加，然后进行Conv 3×3、BN、ReLU生成细节图p_3。同理利用PRM2和PRM1获得细节图p_2和p_1。最后p_3进行4倍上采样、p_2进行2倍上采样后和p_1三者通过连接融合，生成清晰准确的显著图S。上述过程可表达为

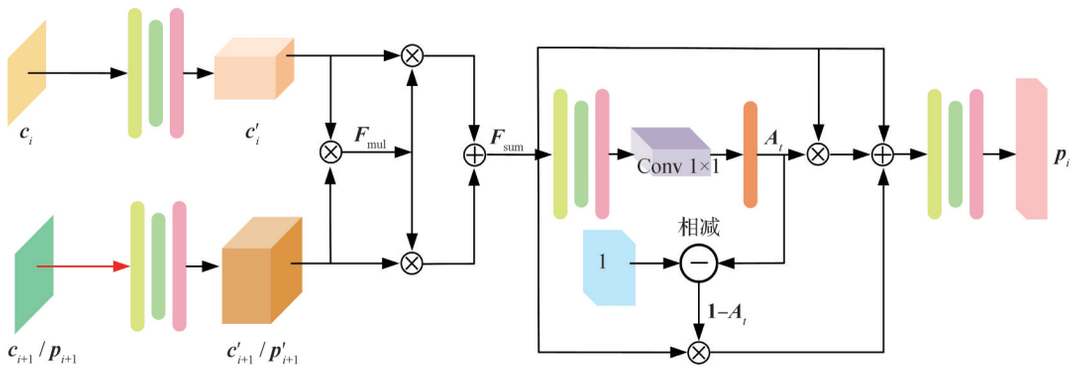

图4　渐进式细化模块结构图

Fig. 4　Progressive refinement module structure diagram

$$c'_i = ReLU\left(BN\left(Conv_3\left(c_i\right)\right)\right) \tag{11}$$

$$c'_{i+1} = ReLU\left(BN\left(Conv_3\left(Up\left(c_{i+1}\right)\right)\right)\right) \tag{12}$$

$$F_{mul} = c'_i \times c'_{i+1} \tag{13}$$

$$F_{sum} = F_{mul} \times c'_i + F_{mul} \times c'_{i+1} \tag{14}$$

$$A_t = f_{sigmoid}\left(Conv_1\left(ReLU\left(BN\left(Conv_3\left(F_{sum}\right)\right)\right)\right)\right) \tag{15}$$

$$p_i = ReLU\left(BN\left(Conv_3\left(A_t \times F_{sum} + \left(1 - A_t\right) \times F_{sum} + F_{sum}\right)\right)\right) \tag{16}$$

式中，c_i表示多尺度语义特征图，c'_i、c'_{i+1}表示特征图，使用PRM2、PRM1时c'_{i+1}变为p'_{i+1}；UP表示上采样操作，$Conv_3$表示3×3卷积操作，F_{mul}表示融合特征图，F_{sum}表示细节特征图，A_t表示权重矩阵图，p_i表示细节图。

2.5　损失函数

本文使用二元交叉熵作为损失函数，引入细节图作为监督（Wei等，2020）。细节图负责捕获显著性物体的细节部分。因此，总损失函数L_{total}定义为

$$loss = loss_{bce} + loss_{iou} + loss_{ssim} \tag{17}$$

$$L_{total} = loss\left(p_{de}, G\right) + loss\left(S, G\right) \tag{18}$$

式中，$loss_{bce}$、$loss_{iou}$、$loss_{ssim}$分别表示BCE（binary cross entropy）损失函数、IOU（intersection over union）损失函数和SSIM（structural similarity index）损失函数，(S, G)表示显著图损失，(p_{de}, G)表示细节图损失。

3　实验结果与分析

3.1　实验设置

3.1.1　数据集

本文在两个公开数据集上对DSANet的性能进行评估，分别是360-SOD（Li等，2020）和360-SSOD（Ma等，2020）。其中，360-SOD包含500幅分辨率为1 024×512像素的ERP全景图像，主要用于全景图像的显著目标检测任务。两个数据集共包含1 605幅同样分辨率的ERP全景图像。

3.1.2　数据预处理

训练前对所有图像进行归一化处理，使其像素值分布在[−1, 1]之间。此外，使用随机水平翻转和旋转等数据增强技术，增加数据的多样性。数据集划分如下：360-SOD数据集中，400幅图像用于训练，100幅图像用于测试；360-SSOD数据集中，850幅图像用于训练，255幅图像用于测试。

3.1.3　评价指标

为了全面评估DSANet的性能，本文采用以下多个主流评价指标：平均绝对误差（mean absolute error, MAE）（Perazzi等，2012）、F-measure（Achanta等，2009）（maxF、meanF）、E-measure（Fan等，2018）（maxE、meanE）和结构测度（structure-measure, Sm）（Fan等，2017）。这些指标分别从不同角度评估了显著目标检测结果的准确性和精确性，能够较为全面地反映模型性能。

3.1.4　训练

DSANet的训练在Pytorch框架下进行，所有实验均在配备了NVIDIA RTX3090 GPU的服务器完成。操作系统为Ubuntu 20.04，Pytorch版本为1.8.1。实验采用随机梯度下降（stochastic gradient descent, SGD）优化器，其中动量参数设为0.9，权重衰减系数设为0.000 1，学习率初始值设为0.000 1，并采用学习衰减策略，根据训练集上的表现调整学习率。实验中，网络训练周期为100个epoch，每个epoch的batchsize设置为4，以平衡训练速度和内存

消耗。

DSANet在训练过程中采用了L2正则化和Dropout技术，以防止模型过拟合。正则化方法可以有效约束模型的复杂度，减少对特定数据样本的过度拟合。同时，在训练中使用了早停策略，即当训练集上的性能在连续10个epoch内没有提升时，停止训练，以防止过拟合。

3.2 对比实验

本文将所提出的模型与目前12种代表性先进方法进行了比较，其中8种方法是传统的2D SOD模型，包括SeaNet（semantic matching and edge alignment）（Li等，2023b）、MSCNet（multi-scale context network）（Lin等，2022）、TSCNet（texture-semantic collaboration network）（Li等，2024）、AGNet_C（attention-guided network）（何伟和潘晨，2022）、ACCoNet（adjacent context coordination network）（Res）（Li等，2023a）、ACCoNet（VGG）（Li等，2023a）、PGNet（pyramid grafting network）（Xie等，2022）和CSDF（cascaded detail modeling and body filling）（Song等，2023）。另外4种方法是360° SOD模型，包括MPFRNet（multi-projection fusion and refinement network）（Cong等，2024）、LDNet（lightweight distortion-aware network）（Huang等，2023）、FANet（features adaptation network）（Huang等，2020）和DATFormer（distortion-aware Transformer）（Zhao等，2023）。为了客观比较，使用公共代码和默认参数来复现作者提供的显著图或结果。

如表1所示，在两个微调360°数据集360-SOD（Li等，2020）和360-SSOD（Ma等，2020）上，DSANet在6个广泛使用的SOD指标方面优于所有2D方法。DSANet模型6个指标中多数超过360° SOD最新的算法。DSANet模型的MAE与MPFR模型差0.000 6，说明DSANet模型在平均绝对误差方面略弱于MPFR。DSANet模型取得了比其他模型更好的结果。

在360-SOD数据集上的主观结果比较如图5所示，DSANet能够检测到分布在ERP图像中的显著目标，对于ERP图像的边缘目标（第2、6行）也有很好的检测效果，并且能够有效地从复杂背景中分离出显著目标而不会引入背景干扰，如第5行中的中间人物目标，其他模型均误将背景信息当做显著目标进行检测，即存在错检问题，而本文模型则没有产生类似的错误。本文模型能够很好地捕捉图像细节，如第6行图像中的左边细节目标。

表1 两个数据集不同方法客观指标对比

Table 1 Comparison of objective indicators of different methods on two datasets

类别	方法	360-SSOD数据集						360-SOD数据集					
		MAE ↓	maxF ↑	meanF ↑	maxE ↑	meanE ↑	Sm ↑	MAE ↓	maxF ↑	meanF ↑	maxE ↑	meanE ↑	Sm ↑
2D SOD	TSC2023	0.027 7	0.706 1	0.693 8	0.873 8	0.864 8	0.792 0	0.030 5	0.624 8	0.617 2	0.844 0	0.829 0	0.745 3
	Sea2023	0.036 3	0.573 0	0.556 1	0.828 3	0.811 7	0.706 3	0.056 3	0.356 7	0.306 1	0.708 0	0.541 1	0.596 3
	AG2022	0.025 3	0.754 3	0.709 2	0.877 6	0.856 9	0.800 1	0.029 7	0.598 8	0.585 3	0.827 2	0.808 7	0.756 0
	PG2023	0.023 2	0.679 5	0.667 0	0.839 4	0.802 8	0.775 2	0.033 8	0.527 6	0.502 7	0.743 2	0.693 7	0.686 2
	MSC2022	0.090 8	0.671 0	0.543 6	0.844 6	0.729 3	0.646 8	0.031 9	0.622 3	0.608 2	0.836 1	0.810 8	0.764 7
	CSDF2022	0.027 7	0.660 1	0.651 6	0.859 6	0.808 9	0.757 8	0.032 6	0.549 0	0.521 1	0.811 7	0.739 9	0.694 2
	ACC_r2022	0.032 6	0.507 1	0.482 6	0.810 6	0.694 8	0.673 8	0.031 2	0.604 3	0.590 4	0.829 7	0.793 8	0.735 8
	ACC_v2022	0.025 6	0.675 4	0.670 1	0.869 0	0.847 0	0.787 3	0.036 5	0.575 2	0.555 5	0.831 2	0.792 9	0.734 1
360° SOD	DAT2023	0.020 4	0.763 8	0.745 9	0.890 4	0.877 5	0.835 2	0.028 7	0.641 9	0.609 9	0.850 9	0.812 8	0.760 1
	MPFR2023	**0.019 0**	0.765 3	0.755 6	0.885 4	0.875 0	0.841 6	–	–	–	–	–	–
	LD2023	0.028 9	0.656 1	0.641 4	0.865 5	0.843 7	0.768 0	0.034 1	0.586 8	0.569 5	0.841 5	0.802 1	0.725 2
	FA2022	0.026 4	0.675 0	0.651 8	0.868 3	0.818 0	0.759 3	0.042 1	0.593 9	0.521 5	0.842 6	0.732 4	0.718 6
	本文	0.019 6	**0.781 4**	**0.771 9**	**0.908 0**	**0.900 2**	**0.844 0**	**0.027 5**	**0.642 6**	**0.628 4**	**0.851 4**	**0.817 8**	**0.770 0**

注：加粗字体表示各列最优结果，“–”表示因没有源代码而无法测试 。“↑”表示值越高越好，“↓”表示值越低越好。

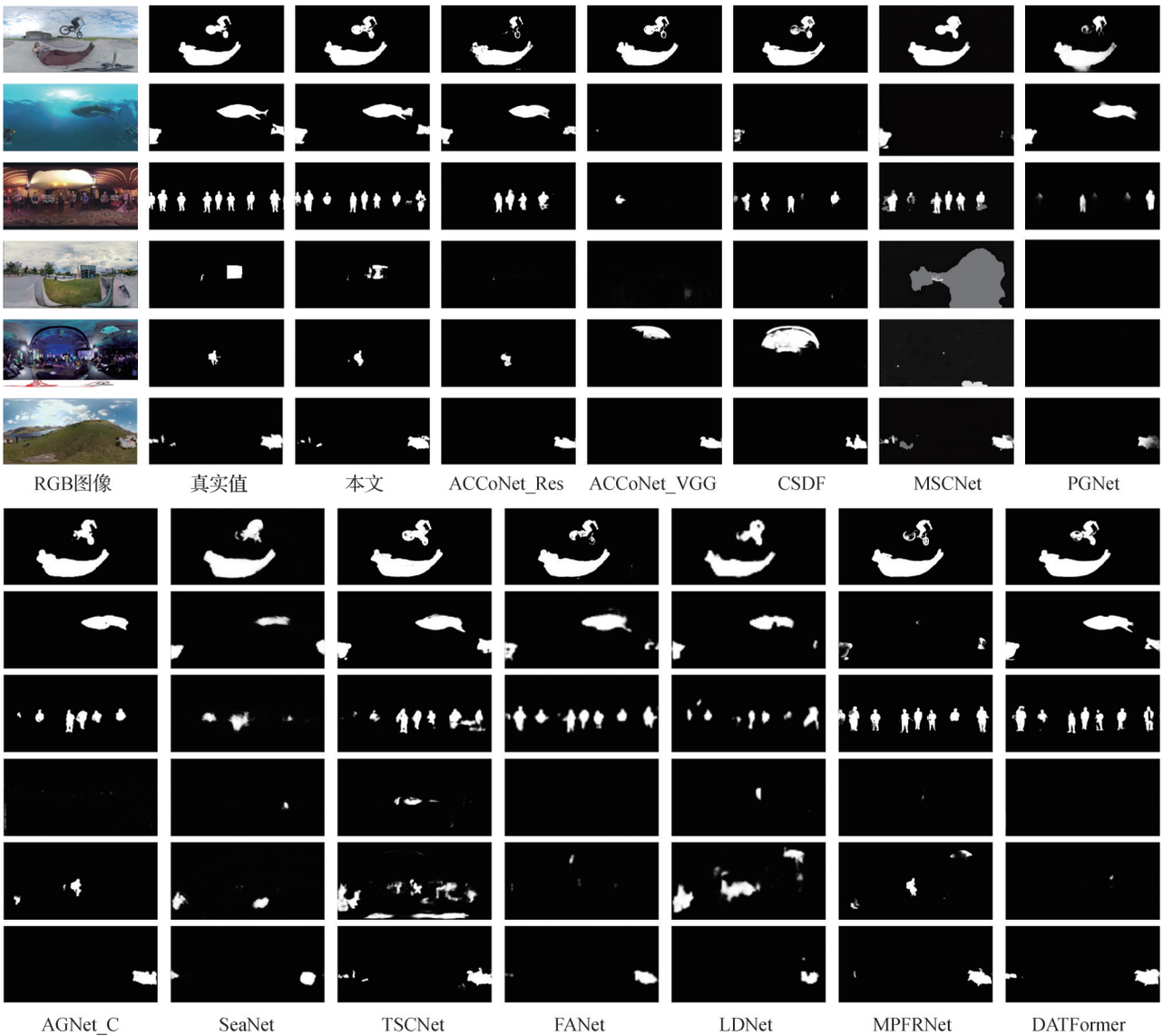

图5　360-SOD数据集上不同模型的主观可视化对比结果

Fig. 5　Subjective visualization comparison of different models on the 360-SOD dataset

在360-SSOD数据集上的主观结果比较如图6所示，本文方法能够较好地屏蔽背景信息的干扰（如图6第1，2，4行），而其他方法或多或少地受到了背景信息的干扰，存在多检测或漏检测的现象。另外，本文模型能够很好地检测出细节显著目标（第6行）中的细绳。最后，根据主客观性能分析，DSANet 360-SOD和360-SSOD显著目标检测能力表现最佳。

3.3　DACM模块和MSAAM模块与常规卷积注意力模块的对比

在实际应用中，模型的计算复杂度与性能之间的平衡至关重要，因此，本文进行了DACM模块和MSAAM模块与常规卷积注意力模块SE、CA、CBAM的复杂度和检测精度对比实验，实验结果如表2所示。从表2数据中，可以观察到不同注意力模块（SE、CA、CBAM、DACM、MSAAM）的每秒浮点运算次数（floating point operations per second，FLOPs）和参数量（Params）基本保持一致，FLOPs均在57.28 B～57.35 B范围内，参数量则在24.54 M～24.55 M之间。综合来看，本文提出的DACM模块和MSAAM模块在FLOPs和参数量略高于或等同于其他传统模块的前提下，检测精度表现更佳，显著优于其他传统模块。

3.4　消融实验

为了更好地测试本文所提各个模块的有效性，本文针对所提模型中关键的3个模块进行详细的消融实验，评估其对模型性能的贡献，实验结果如表3所示。

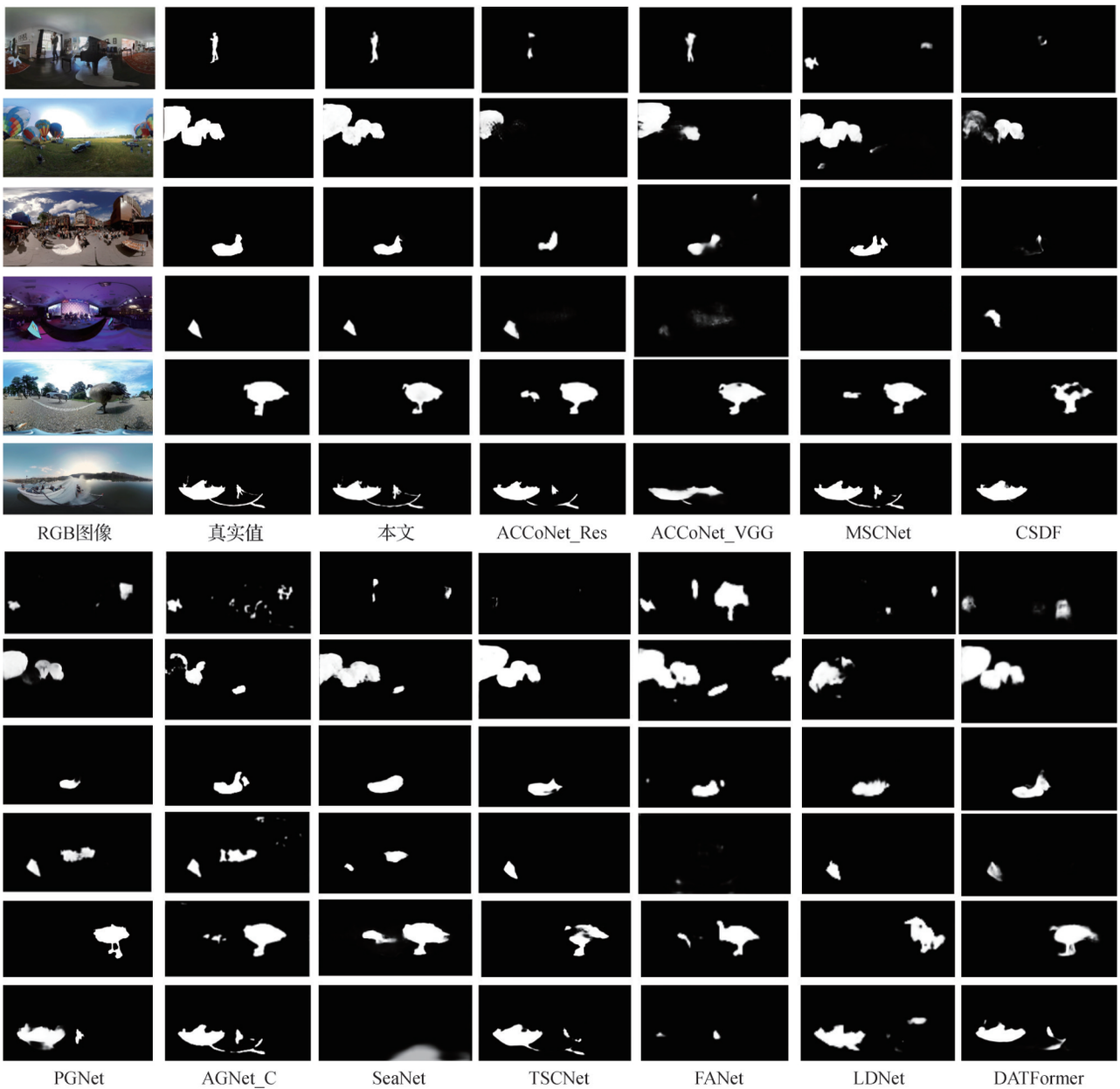

图6　360-SSOD数据集上不同模型的主观可视化对比结果

Fig. 6　Subjective visualization comparison of different models on the 360-SSOD dataset

表2　DACM模块和MSAAM模块与常规卷积注意力模块对比

Table 2　Comparison of DACM and MSAAM modules with conventional convolutional attention modules

方法	FLOPs/B	Params/M	MAE ↓	maxF ↑	meanF ↑
SE	**57.28**	**24.54**	0.023 0	0.753 4	0.732 2
CA	57.29	**24.54**	0.022 2	0.769 2	0.751 7
CBAM	57.31	24.55	0.024 8	0.749 8	0.734 8
DACM	57.29	**24.54**	0.021 0	0.760 8	0.756 4
MSAAM	57.35	24.55	**0.020 2**	**0.773 1**	**0.758 6**

注：加粗字体表示各列最优结果，“↑”表示值越高越好，“↓”表示值越低越好。

表3　相关模块以及多分支结构的消融实验结果

Table 3　Ablation results of related modules as well as multi-branched structures

结果	Baseline	DACM	MSAAM	PRM	MAE ↓	maxF ↑	meanF ↑	maxE ↑	meanE ↑	Sm ↑
Ⅰ	√	–	–	–	0.023 0	0.712 1	0.704 5	0.858 7	0.840 0	0.795 3
Ⅱ	√	√	–	–	0.021 8	0.746 5	0.739 4	0.884 6	0.858 3	0.813 2
Ⅲ	√	–	√	–	0.022 0	0.740 0	0.731 6	0.872 0	0.853 3	0.811 5
Ⅳ	√	–	–	√	0.020 8	0.753 5	0.743 6	0.883 0	0.853 6	0.824 6
Ⅴ	√	√	√	–	0.020 7	0.766 7	0.756 4	0.898 0	0.888 7	0.825 7
Ⅵ	√	√	–	√	0.021 0	0.760 8	0.756 4	0.887 8	0.874 7	0.828 8
Ⅶ	√	–	√	√	0.020 2	0.773 1	0.758 6	0.900 3	0.883 9	0.833 7
本文	√	√	√	√	**0.019 6**	**0.781 4**	**0.771 9**	**0.908 0**	**0.900 1**	**0.844 0**

注：加粗字体表示各列最优结果。“√”表示使用对应模块，“–”表示未使用。“↑”表示值越高越好，“↓”表示值越低越好。

1）DACM模块的有效性。引入DACM模块显著提高了模型在各个评价指标上的表现（如表3实验结果II和本文方法），尤其是在meanF和Sm上，证明了DACM模块在改善全景图像几何畸变方面的有效性。

2）MSAAM模块的有效性。加入MSAAM模块后，多个指标（如maxF和meanE）均有显著提升（如表3实验结果III、VI和本文方法），表明MSAAM模块在处理大视野全景图像中的有效性，特别是在结合全局语义信息和局部细节特征方面。

3）PRM模块的有效性。使用本文方法时，MAE达到了0.019 6，是所有实验结果中最小的。这表明PRM模块在减少误差、提高显著性目标检测精度方面的有效性。

由图7主观实验结果可以看到，不同的消融策

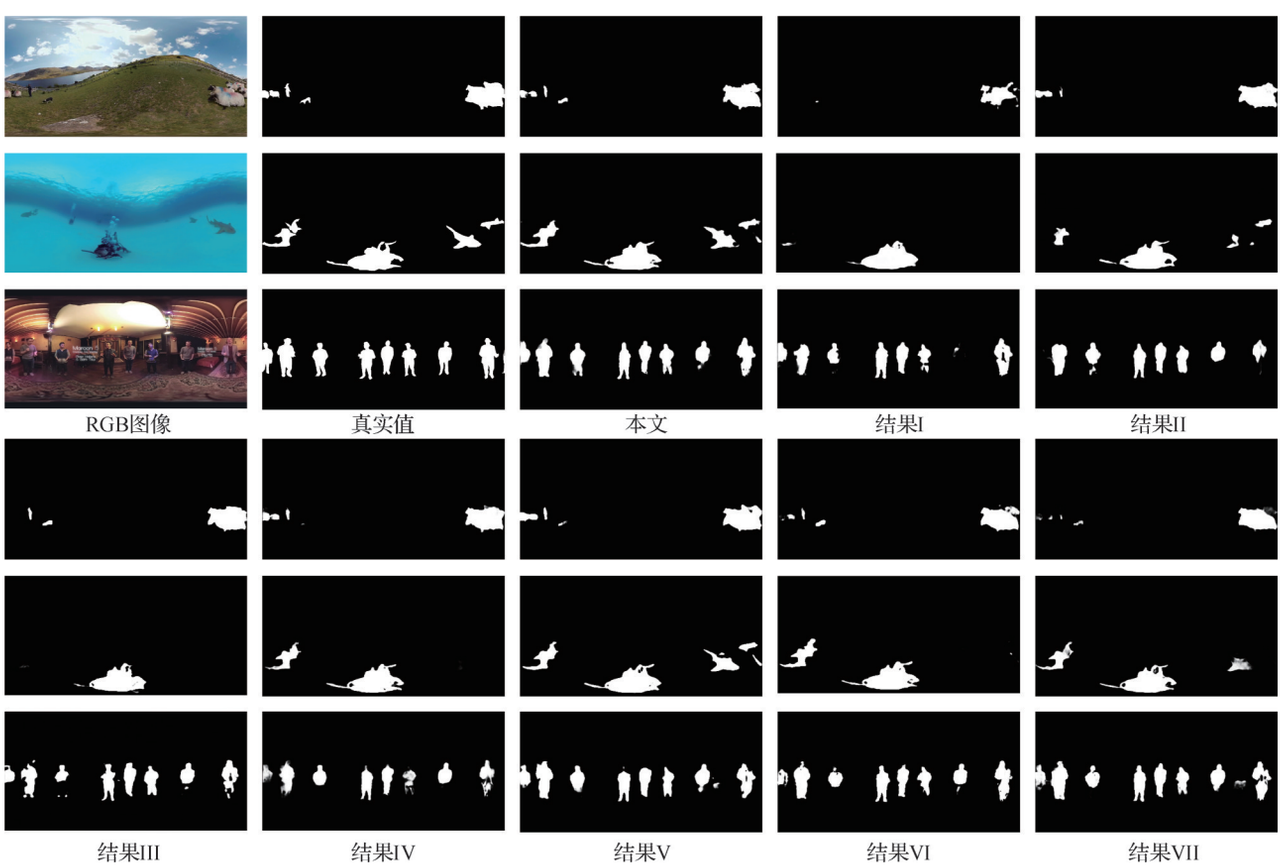

图7　相关模块以及结构的消融实验可视化对比

Fig. 7　Visual comparison of ablation experiments of related modules and structures

略下网络均不能得到令人满意的视觉效果，因此所有评估的消融实验对本文方法起到了正向的贡献。例如，第1幅图在逐步加入3个模块后，检测结果相较于baseline越来越精细，尤其当加入DACM、PRM模块后模型能够更好地检测出图像中的边缘小目标，当3个模块协调作用时模型能够很好地检测出显著目标。因此，本文的完备模型能够准确定位显著目标，且能检测出显著目标的细节，检测结果更贴近人工标注的真实值(groundtruth，GT)。

综上，图表实验结果都证明了本文提出的3个模块能够明显提高各项性能指标得分，增强网络显著目标检测性能，对获得高质量的360°全景图像显著目标检测效果是有效的。

3.5 模型在复杂场景下的鲁棒性

表4对比了不同方法在360-SOD和360-SSOD数据集上的交叉验证结果。从实验结果可见，本文方法在多个关键指标上均取得了最佳表现，特别是在maxF和meaF两个衡量显著性目标检测准确性的指标上显著优于其他方法。这表明DSANet不仅在单一数据集上表现优异，在跨数据集的泛化能力上也表现突出。交叉验证的结果进一步验证了DSANet方法的鲁棒性和广泛适用性。

表4 360-SOD和360-SSOD数据集上交叉验证不同方法客观指标对比

Table 4 Comparison of objective metrics for different methods of cross-validation on 360-SOD and 360-SSOD datasets

类别	方法	360-SOD数据集						360-SSOD数据集					
		MAE↓	maxF↑	meanF↑	maxE↑	meanE↑	Sm↑	MAE↓	maxF↑	meanF↑	maxE↑	meanE↑	Sm↑
2D SOD	TSC2023	0.056 2	0.459 2	0.453 2	0.743 0	**0.738 2**	0.655 7	0.033 5	0.613 1	0.607 9	0.830 2	0.790 6	0.729 3
	Sea2023	0.059 6	0.424 9	0.414 8	0.730 8	0.712 6	0.629 2	0.037 6	0.536 6	0.519 5	0.811 1	0.753 0	0.669 2
	AG2022	0.049 8	0.426 5	0.420 9	0.713 6	0.683 0	0.656 6	0.034 6	0.563 6	0.526 0	0.753 5	0.689 2	0.687 7
	PG2023	0.051 1	0.413 2	0.394 7	0.684 4	0.650 9	0.631 8	0.031 5	0.642 0	0.621 7	0.817 5	**0.804 9**	0.749 1
	MSC2022	0.127 8	0.402 5	0.336 3	0.701 7	0.618 3	0.559 3	0.032 5	0.669 6	0.653 4	0.833 0	0.813 1	0.751 1
	CSDF2022	0.049 7	0.404 3	0.385 1	0.739 6	0.653 1	0.614 4	0.034 5	0.559 5	0.533 9	0.767 0	0.682 1	0.689 9
	ACC_r2022	0.051 4	0.430 5	0.413 2	0.737 6	0.672 6	0.639 0	0.031 6	0.558 1	0.537 7	0.793 8	0.705 9	0.701 5
	ACC_v2022	0.050 7	0.432 7	0.408 5	**0.747 7**	0.670 5	0.637 8	0.037 2	0.583 9	0.565 7	0.829 7	0.767 3	0.732 1
360° SOD	DAT2023	0.050 5	0.438 0	0.432 0	0.706 4	0.697 7	0.651 2	0.028 2	0.693 0	0.656 5	**0.864 2**	0.801 7	0.759 6
	MPFR2023	–	–	–	–	–	–	–	–	–	–	–	–
	LD2023	–	–	–	–	–	–	–	–	–	–	–	–
	FA2022	0.050 4	0.444 9	0.424 7	0.726 1	0.690 9	0.645 4	0.044 1	0.585 8	0.558 8	0.810 2	0.775 8	0.710 6
	本文	**0.049 4**	**0.469 8**	**0.457 6**	0.729 9	0.707 9	**0.672 5**	**0.027 5**	**0.700 2**	**0.670 4**	0.853 5	0.795 2	**0.767 7**

注：加粗字体表示各列最优结果，“–”表示因没有源代码而无法测试。“↑”表示值越高越好，“↓”表示值越低越好。

360-SOD训练360-SSOD测试的可视化对比如图8所示。从图中可以看出，DSANet在处理复杂场景(如水下、室内、户外等)时，能够更准确地分割出显著目标，尤其是在细节处理和边缘保持方面表现尤为突出。例如，在第1行和第4行的水下和室内场景中，DSANet能够较好地分割出目标物体的完整轮廓，而其他方法则出现了不同程度的目标分割不完整或背景误分割的情况。相比之下，PGNet和AGNet_C在这些复杂场景下表现较差，容易产生过度分割或遗漏目标的现象。360-SSOD训练360-SOD测试的可视化对比也表明，相比其他方法，DSANet能够更加准确地分割出显著目标，处理动态背景、光照变化及复杂纹理的能力更强，限于篇幅不再赘述。

交叉验证的主观可视化对比进一步说明本文方法在复杂场景下具有更强的鲁棒性和适应性。

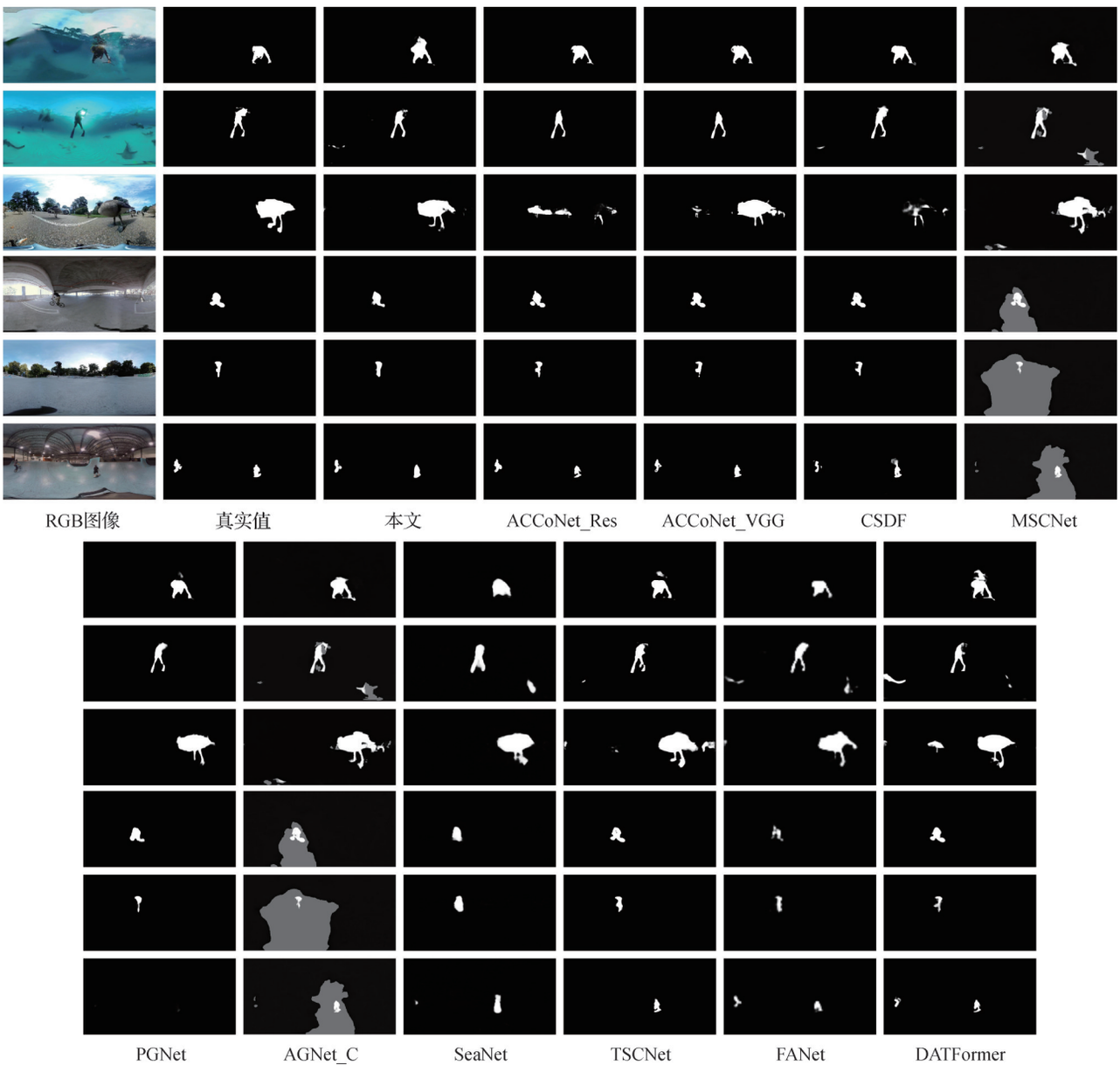

图8　360-SOD训练360-SSOD测试可视化对比

Fig. 8　360-SOD training 360-SSOD test visualization comparison

4　结　论

针对360° SOD任务，本文首先回顾了SOD任务中图像信息融合方面的现有工作，并总结了基于两类骨干网络对性能提升的重要性。在探讨了现有方法尚不能很好地处理360°全景图像的几何畸变和大视野问题之后，本文分析了现有网络，提出一种新的360° SOD模型，该模型结合ResNet-50主干网络与设计的畸变自适应校正模块（DACM）、多尺度语义注意力聚合模块（MSAAM）和渐进式细化模块（PRM）等一起构成了多尺度类U型结构框架。其中，畸变自适应校正模块解决几何畸变问题；多尺度语义注意力聚合模块能够根据全景图像的特点及具体输入图像，提取全局语义与局部细节特征；渐进式细化模块（PRM）通过逐步补充深层特征来弥补浅层特征的不足，并利用此模块过滤深层特征中的冗余信息和背景噪声，后两个模块配合解决大视野问题。在2个广泛使用的公开数据集上的定量和定性的实验结果表明，以6种主流测度作为评估方法，本文提出方法相比较参与对比的方法取得了更优秀的性能。同时，通过消融实验进一步分析了本文方法的

网络结构中关键部分对网络整体性能的贡献。定量和定性的消融实验结果表明这些关键部分对网络性能提升均起到正向的作用。

未来计划在以下3个方向继续开展深入研究，以进一步提高360° SOD的检测性能。1)增强模型的上下文感知能力。引入更加先进的上下文注意力机制，使模型能够更好地理解遮挡场景中的上下文信息，从而提高检测准确性。2)多视角或多模态数据的融合。通过引入多视角或多模态数据(如深度信息、红外图像等)，使得模型能够从更多维度获取信息，减少由于遮挡引起的检测误差。3)遮挡情况下的自适应学习。探索基于自监督或半监督学习的方法，使模型在遮挡情况下能够自适应调整，以提升在复杂场景中的检测效果。

参考文献(References)

Achanta R, Hemami S, Estrada F and Susstrunk S. 2009. Frequency-tuned salient region detection//Proceedings of 2009 CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, USA: IEEE: 1597-1604 [DOI: 10.1109/CVPR.2009.5206596]

Borji A, Cheng M M, Hou Q B, Jiang H Z and Li J. 2019. Salient object detection: a survey. Computational Visual Media, 5(2): 117-150 [DOI: 10.1007/s41095-019-0149-9]

Chen Z Y, Xu Q Q, Cong R M and Huang Q M. 2020. Global context-aware progressive aggregation network for salient object detection//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 10599-10606 [DOI: 10.1609/AAAI.v34i07.6633]

Cong R M, Huang K, Lei J J, Zhao Y, Huang Q M and Kwong S. 2024. Multi-projection fusion and refinement network for salient object detection in 360°omnidirectional image. IEEE Transactions on Neural Networks and Learning Systems, 35(7): 9495-9507 [DOI: 10.1109/TNNLS.2022.3233883]

Cong R M, Zhang Y M, Fang L Y, Li J, Zhao Y and Kwong S. 2022. RRNet: relational reasoning network with parallel multiscale attention for salient object detection in optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60: #5613311 [DOI: 10.1109/TGRS.2021.3123984]

Fan D P, Cheng M M, Liu Y, Li T and Borji A. 2017. Structure-measure: a new way to evaluate foreground maps//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4548-4557 [DOI: 10.1109/ICCV.2017.487]

Fan D P, Gong C, Cao Y, Ren B, Cheng M M and Borji A. 2018. Enhanced-alignment measure for binary foreground map evaluation//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: IJCAI: 698-704 [DOI: 10.24963/ijcai.2018/97]

Fang X, Jiang M F, Zhu J C, Shao X L and Wang H P. 2024. Group-TransNet: group transformer network for RGB-D salient object detection. Neurocomputing, 594: #127865 [DOI: 10.1016/j.neucom.2024.127865]

Fu C W, Wan L, Wong T T and Leung C S. 2009. The rhombic dodecahedron map: an efficient scheme for encoding panoramic video. IEEE Transactions on Multimedia, 11(4): 634-644 [DOI: 10.1109/TMM.2009.2017626]

He W and Pan C. 2022. The salient object detection based on attention-guided network. Journal of Image and Graphics, 27(4): 1176-1190 (何伟, 潘晨. 2022. 注意力引导网络的显著性目标检测. 中国图象图形学报, 27(4): 1176-1190) [DOI: 10.11834/jig.200658]

He Z T, Shao F, Chen G, Chai X L and Ho Y S. 2024. SCFANet: semantics and context feature aggregation network for 360° salient object detection. IEEE Transactions on Multimedia, 26: 2276-2288 [DOI: 10.1109/TMM.2023.3293994]

Hou Q B, Cheng M M, Hu X W, Borji A, Tu Z W and Torr P H S. 2019. Deeply supervised salient object detection with short connections. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(4): 815-828 [DOI: 10.1109/TPAMI.2018.2815688]

Huang M K, Li G Y, Liu Z and Zhu L C. 2023. Lightweight distortion-aware network for salient object detection in omnidirectional images. IEEE Transactions on Circuits and Systems for Video Technology, 33(10): 6191-6197 [DOI: 10.1109/TCSVT.2023.3253685]

Huang M K, Liu Z, Li G Y, Zhou X F and Meur O L. 2020. FANet: features adaptation network for 360° omnidirectional salient object detection. IEEE Signal Processing Letters, 27: 1819-1823 [DOI: 10.1109/LSP.2020.3028192]

Huo F S, Zhu X G, Zhang L, Liu Q F and Shu Y. 2022. Efficient context-guided stacked refinement network for RGB-T salient object detection. IEEE Transactions on Circuits and Systems for Video Technology, 32(5): 3111-3124 [DOI: 10.1109/TCSVT.2021.3102268]

Itti L, Koch C and Niebur E. 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11): 1254-1259 [DOI: 10.1109/34.730558]

Lee G Y, Tai Y W and Kim J. 2016. Deep saliency with encoded low level distance map and high level features//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 660-668 [DOI: 10.1109/CVPR.2016.78]

Li G B and Yu Y Z. 2015. Visual saliency based on multiscale deep feature//Proceedings of 2015 IEEE Conference on Computer Vision

and Pattern Recognition. Boston, USA: IEEE: 5455-5463 [DOI: 10.1109/CVPR.2015.7299184]

Li G Y, Bai Z and Liu Z. 2024. Texture-semantic collaboration network for ORSI salient object detection. IEEE Transactions on Circuits and Systems II: Express Briefs, 71(4): 2464-2468 [DOI: 10.1109/TCSII.2023.3333436]

Li G Y, Liu Z, Zeng D, Lin W S and Ling H B. 2023a. Adjacent context coordination network for salient object detection in optical remote sensing images. IEEE Transactions on Cybernetics, 53(1): 526-538 [DOI: 10.1109/TCYB.2022.3162945]

Li G Y, Liu Z, Zhang X P and Lin W S. 2023b. Lightweight salient object detection in optical remote-sensing images via semantic matching and edge alignment. IEEE Transactions on Geoscience and Remote Sensing, 61: #5601111 [DOI: 10.1109/TGRS.2023.3235717]

Li J, Su J M, Xia C Q and Tian Y H. 2020. Distortion-adaptive salient object detection in 360° omnidirectional images. IEEE Journal of Selected Topics in Signal Processing, 14(1): 38-48 [DOI: 10.1109/JSTSP.2019.2957982]

Li J, Wang Z M, Lai S M, Zhai Y P and Zhang M J. 2018a. Parallax-tolerant image stitching based on robust elastic warping. IEEE Transactions on Multimedia, 20(7): 1672-1687 [DOI: 10.1109/TMM.2017.2777461]

Lin Y H, Sun H, Liu N Z, Bian Y T, Cen J and Zhou H Y. 2022. A lightweight multi-scale context network for salient object detection in optical remote sensing images//Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal, Canada: IEEE: 238-244 [DOI: 10.1109/ICPR56361.2022.9956350]

Liu Z Y, Tan Y C, He Q and Xiao Y. 2022. SwinNet: swin transformer drives edge-aware RGB-D and RGB-T salient object detection. IEEE Transactions on Circuits and Systems for Video Technology, 32(7): 4486-4497 [DOI: 10.1109/TCSVT.2021.3127149]

Ma G X, Li S, Chen C L Z, Hao A M and Qin H. 2020. Stage-wise salient object detection in 360° omnidirectional image via object-level semantical saliency ranking. IEEE Transactions on Visualization and Computer Graphics, 26(12): 3535-3545 [DOI: 10.1109/TVCG.2020.3023636]

Ma G X, Li S, Chen C L Z, Hao A M and Qin H. 2021. Rethinking image salient object detection: object-level semantic saliency reranking first, pixelwise saliency refinement later. IEEE Transactions on Image Processing, 30: 4238-4252 [DOI: 10.1109/TIP.2021.3068649]

Pan C, Liu J F, Yan W Q, Cao F L, He W and Zhou Y X. 2021. Salient object detection based on visual perceptual saturation and two-stream hybrid networks. IEEE Transactions on Image Processing, 30: 4773-4787 [DOI: 10.1109/TIP.2021.3074796]

Perazzi F, Krähenbühl P, Pritch Y and Hornung A. 2012. Saliency filters: contrast based filtering for salient region detection//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 733-740 [DOI: 10.1109/CVPR.2012.6247743]

Ren Q H, Lu S J, Zhang J X and Hu R J. 2021. Salient object detection by fusing local and global contexts. IEEE Transactions on Multimedia, 23: 1442-1453 [DOI: 10.1109/TMM.2020.2997178]

Song Y, Tang H, Sebe N and Wang W. 2023. Disentangle saliency detection into cascaded detail modeling and body filling. ACM Transactions on Multimedia Computing, Communications and Applications, 19(1): #7 [DOI: 10.1145/3513134]

Su Y K, Deng J L, Sun R Z, Lin G S, Su H J and Wu Q Y. 2024. A unified transformer framework for group-based segmentation: co-segmentation, co-saliency detection and video salient object detection. IEEE Transactions on Multimedia, 26: 313-325 [DOI: 10.1109/TMM.2023.3264883]

Shum H Y, Ng K T and Chan S C. 2005 A virtual reality system using the concentric mosaic: construction, rendering, and data compression. IEEE Transactions on Multimedia, 7(1): 85-95 [DOI: 10.1109/TMM.2004.840591]

Tang M H, Wen J T, Zhang Y, Gu J W, Junker P, Guo B C, Jhao G, Zhu Z Y and Han Y X. 2019. A universal optical flow based real-time low-latency omnidirectional stereo video system. IEEE Transactions on Multimedia, 21(4): 957-972 [DOI: 10.1109/TMM.2018.2867266]

Wei J, Wang S H, Wu Z, Chi S, Huang Q M and Tian Q. 2020. Label decoupling framework for salient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13022-13031 [DOI: 10.1109/CVPR42600.2020.01304]

Xie C X, Xia C Q, Ma M C, Zhao Z R, Chen X W and Li J. 2022. Pyramid grafting network for one-stage high resolution saliency detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 11707-11716 [DOI: 10.1109/CVPR52688.2022.01142]

Xu H, Zhang H and Ma J Y. 2021. Classification saliency-based rule for visible and infrared image fusion. IEEE Transactions on Computational Imaging, 7: 824-836 [DOI: 10.1109/TCI.2021.3100986]

Ye X Y, Zhu L, Wang W W and Fu Y. 2024. RGB_D salient object detection algorithm based on complementary information interaction. Journal of Image and Graphics, 29(5): 1252-1264 (叶欣悦, 朱磊, 王文武, 付云. 2024. 互补特征交互融合的RGB_D实时显著目标检测. 中国图象图形学报, 29(5): 1252-1264) [DOI: 10.11834/jig.230583]

Yuan L, Chen Y P, Wang T, Yu W H, Shi Y J, Jiang Z H, Tay F E H, Feng J S and Yan S C. 2021 Tokens-to-token ViT: training vision transformers from scratch on ImageNet//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 558-567 [DOI: 10.1109/ICCV48922.2021.00060]

Zhang J, Zhang Q D, Shen X L and Wang X. 2023. Salient object detection on 360° omnidirectional image with Bi-branch hybrid projection

network//Proceedings of the 25th IEEE International Workshop on Multimedia Signal Processing (MMSP). Poitiers, France: IEEE: 1-5 [DOI: 10.1109/MMSP59012.2023.10337695]

Zhang Y, Hamidouche W and Deforges O. 2022. Channel-spatial mutual attention network for 360° salient object detection//Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal, Canada: IEEE: 3436-3442 [DOI: 10.1109/ICPR56361.2022.9956354]

Zhao R, Ouyang W L, Li H S and Wang X G. 2015. Saliency detection by multi-context deep learning//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1265-1274 [DOI: 10.1109/CVPR.2015.7298731]

Zhao Y J, Zhao L C, Yu Q, Sheng L, Zhang J and Xu D. 2023. Distortion-aware transformer in 360° salient object detection//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 499-508 [DOI: 10.1145/3581783.3612025]

Zheng B L, Chen Q, Yuan S X, Zhou X F, Zhang H, Zhang J Y, Yan C G and Slabaugh G. 2022. Constrained predictive filters for single image bokeh rendering. IEEE Transactions on Computational Imaging, 8: 346-357 [DOI: 10.1109/TCI.2022.3171417]

Zhu J C, Zhang X Y, Zhang S and Liu J N. 2021. Inferring camouflaged objects by texture-aware interactive guidance network//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Austin, USA: AAAI Press: 3599-3607 [DOI: 10.1609/AAAI.v35i4.16475]

作者简介

陈晓雷，男，副教授，硕士生导师，主要研究方向为人工智能与计算机视觉。E-mail: chenxl703@lut.edu.cn

张学功，男，硕士研究生，主要研究方向为全景图像显著目标检测。E-mail: 864613727@qq.com

杜泽龙，男，硕士研究生，主要研究方向为全景视频显著性检测。E-mail: 1132911812@qq.com

王兴，男，硕士研究生，主要研究方向为全景图像轻量级显著目标检测。E-mail: 19312938573@163.com