

文章编号:1671-6833(2022)02-0035-09

基于金字塔特征与边缘优化的显著性对象检测

田 旭¹, 彭 飞¹, 刘 飞¹, 陈庆文², 闫馨宇^{3,4}

(1.国网青海省电力公司经济技术研究院,青海 西宁 810000; 2.中国电建集团西北勘测设计研究院有限公司,陕西 西安 710065; 3.天津大学 智能与计算学部,天津 300350; 4.天津大学 天津机器学习重点实验室,天津 300350)

摘 要:针对图像显著性对象检测领域中多尺度特征提取不充分、对象边缘模糊等问题,提出了一个端到端的基于注意力嵌入的金字塔特征以及渐进边缘优化的显著性对象检测模型。首先,设计了由多个扩张卷积构成的注意力嵌入的密集空洞金字塔模块(AEDAPM),在不减小特征分辨率的前提下,得到丰富且有效的多级多尺度特征;其次,为了解决显著性对象边缘模糊的问题,提出了渐进边缘优化模块(SEOM),在特征恢复分辨率的过程中逐步补充空间细节信息,使模型检测出的显著对象能够拥有清晰的边缘轮廓。在 DUTS-TE、ECSSD、DUT-OMRON、HKU-IS、PASCAL-S 5 个显著性领域公开的数据集上与其他 12 种已有的先进方法在 3 个常用指标下进行了比较,结果表明:所提方法能够得到更加准确、边缘更加清晰的显著性结果。此外,自对比实验也充分证明了提出的注意力嵌入的密集空洞金字塔模块和渐进边缘优化模块的有效性。

关键词:显著性对象检测;多尺度特征提取;全卷积神经网络;边缘特征提取;深度学习

中图分类号: TP391.4 **文献标志码:** A **doi:**10.13705/j.issn.1671-6833.2022.02.003

0 引言

图像显著性检测旨在找出图像中最吸引人视觉注意的区域,因其反映了人视觉机制最本质的机理而成为各种计算机视觉任务中重要的预处理步骤,比如图像识别^[1]、视觉跟踪^[2-3]、对象检测^[4]等。

近年来,解决显著性对象检测的方法已出现很多。早期的显著性对象模型多数是基于非深度学习的传统算法,显著映射的生成都基于浅层的、人工设置的先验特征,这导致算法的鲁棒性及泛化能力不强,并且人工提取特征也非常耗时。随着大数据时代的到来,越来越多的学者开始研究基于深度学习的显著性检测方法。当前,应用最广且最有效的方法是全卷积神经网络(fully convolutional networks, FCNs)^[5],它通过卷积-反卷积结构能够端到端地生成与输入大小相同的显著映射。然而,在全卷积神经网络中,大多是通过依次堆叠多个卷积层和池化层来逐步扩大感受野并提取高级语义特征,这会导致特征映射的分辨率不

断减小,对于显著性对象检测这种像素级的任务是非常不友好的。此外,利用深层卷积提取的高级特征去预测显著映射,会由于缺乏浅层的空间细节信息,使得检测到的显著性对象的边缘十分模糊。

本文针对以上两大问题,提出了一种全新的显著性对象检测方法:基于注意力嵌入的金字塔特征以及渐进边缘优化的显著性对象检测模型。首先,受到 DenseASPP^[6]的启发,设计了一个由多个扩张卷积构成的注意力嵌入的密集空洞金字塔模块(attention embedded dense atrous pyramid module, AEDAPM),通过具有不同扩张率的扩张卷积处理,在不减小特征映射分辨率的前提下,获得具有不同感受野的多级多尺度上下文特征,同时嵌入注意力机制,对这些不同感受野下得到的特征产生不同程度的响应,以此来灵活检测图像中大小、形状不一的显著对象。其次,为了得到清晰的显著性对象的边缘,提出了一个渐进边缘优化模块(steppered edge optimization module, SEOM)。仅仅用高级语义特征去检测显著性对

收稿日期:2021-08-10;修订日期:2021-11-23

基金项目:国家自然科学基金资助项目(62076180)

通信作者:闫馨宇(1996—),女,山西太原人,天津大学博士研究生,主要从事显著性对象检测研究,E-mail:xinyuyan@tju.edu.cn。

象,是无法得到清晰的边界轮廓的,由于浅层特征中保留了更多的空间细节,适合于定位边界,所以借助渐进边缘优化模块,向不同分辨率的显著特征中逐次补充浅层的外观特征,以此得到更加清晰的边缘轮廓。

1 相关工作

1.1 传统的显著性对象检测方法

显著性对象检测的研究已有深远的历史,早期的方法大多靠自底向上的计算模型以及人工设置的较为低级的特征去检测显著性对象,这些特征大多借助于经验,如对比度、中心先验等。这些方法中最经典的要数 Itti 等^[7]提出的模型,其受灵长类动物视觉神经系统及其行为的启发,对多个特征通道和尺度进行分解,再通过滤波得到多个显著特征,最后将其融合得到最终的显著图。Hou 等^[8]认为图像的信息都包含在图像的幅度谱信息中,利用谱残差模型来检测显著性。后来,FT 模型由 Achanta 等^[9]提出,利用颜色以及亮度特征的中心周围算子来得到显著映射。Xie 等^[10]通过利用中低层次线索,在贝叶斯框架内提出了一种自下而上显著性的新模型。更多的总结可以参考 Fan 等^[11]的研究。

1.2 基于深度学习的显著性对象检测方法

在最近几年中,深度学习已经在显著性对象检测领域被广泛应用。Wang 等^[12]通过训练 2 个深度神经网络,将超像素的局部估计和全局建议相结合检测显著性。受到全卷积神经网络(fully convolutional networks, FCNs)成功的启发^[5],很多研究开始在全卷积神经网络的基础上展开。Zhang 等^[13]提出一种新颖的全卷积神经网络模型,在解码部分设计了混合上采样方法,在一定程度上解决了反卷积“棋盘效应”的问题。Li 等^[14]提出了基于 FCN 的多任务模型,通过显著性对象检测和图像分割这 2 个相关任务的协作学习,对有效信息进行编码。Hu 等^[15]将深度神经网络和水平集的方法巧妙结合,使显著性对象边界模糊的问题得到了改善。

然而本文认为,在自然图像中,对象的大小不同,形状万千,在这种情况下,多尺度特征的提取尤为重要,并且,在扩大感受野的同时,还要兼顾特征的分辨率以及空间细节特征的保留,如何得到丰富而不冗余的多尺度特征,并保留显著性对象的边缘轮廓特征,从而得到完整、边缘轮廓清晰的显著性对象,仍是一个值得继续研究的问题。

因此,本文提出了基于金字塔特征与边缘优化的显著性对象检测模型,采用注意力嵌入的密集空洞金字塔模块(AEDAPM)并结合渐进边缘优化模块(SEOM),共同检测自然图像中的显著对象。

2 本文方法

2.1 网络架构

本文提出了一个基于注意力嵌入的金字塔特征以及渐进边缘优化的显著性对象检测网络(attention embedded pyramid feature and stepped edge optimized based salient object detection network, ASNet)模型。图 1 展示了它的总体架构。模型是以 VGG-16 为骨干特征提取器的全卷积神经网络结构。首先,将图像输入到 VGG-16 的前 4 个卷积块中,逐层提取图像特征。通常浅层卷积块会提取低级的空间结构特征,深层的卷积结构会提取高级语义特征。接着,将从 VGG-16 中提取到的特征输入到由扩张卷积构成的注意力嵌入的密集空洞金字塔模块(AEDAPM),该模块中,利用一系列带“孔”的卷积去处理特征映射,从而在不减小特征分辨率的前提下,得到丰富的多尺度特征。同时,该模块还嵌入了通道注意力机制,对不同通道的特征进行过滤,从而让多尺度特征更加合理有效。然后,从 AEDAPM 中得到的多尺度特征会传入解码模块中逐步恢复分辨率,在此过程中,网络模型会将从 VGG-16 中得到的空间结构特征相继传入本文设计的渐进边缘优化模块(SEOM),逐步清晰显示对象的边缘轮廓。

2.2 注意力嵌入的密集空洞金字塔模块

在自然图像中,往往包括多个大小不同、形状多变的对象,这就要求深度神经网络可以捕获丰富的多尺度上下文特征。然而,大多数已存在的方法通常直接使用堆叠的卷积层和池化层来解决这一问题,但这样做不仅会丢失空间细节信息,而且无法有效应对复杂多变的情况。对此,Yang 等^[6]提出了用一系列扩张卷积构成 DenseASPP 模块去提取多尺度上下文特征。受此启发,本文设计了一个有注意力嵌入的密集空洞金字塔模块(AEDAPM),在不改变特征分辨率的前提下,多样化扩大感受野,同时嵌入注意力机制,对这些不同感受野下得到的特征产生不同程度的响应,而不是简单地将其均等对待。

具体来说,AEDAPM 的网络结构如图 2 所示。首先,将来自于骨干特征提取网络的各个卷积块中池化层输出的特征定义为 $\{E_1, E_2, E_3,$

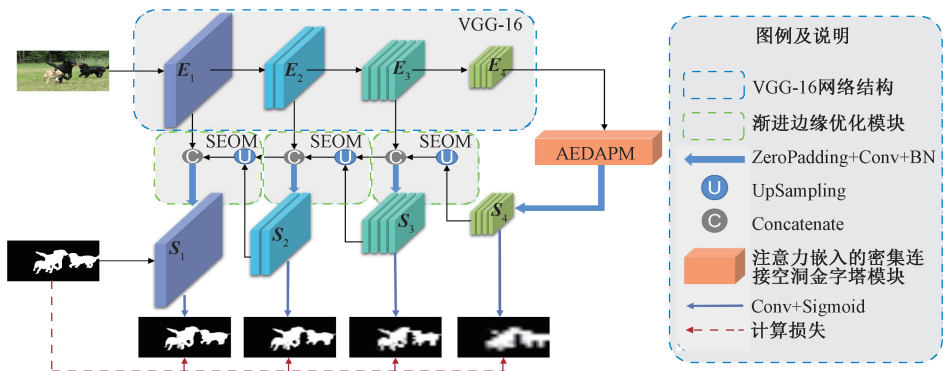


图 1 ASNet 总体架构

Figure 1 ASNet overall architecture

E_4 }, 将 E_4 作为该模块的输入, 为了使最终提取的高级特征能够适应对象大小、形状的多变性, 应用了多个具有不同扩张率 ($r=3, 6, 12, 18, 24$) 的空洞卷积块来处理输入的特征, 每个卷积块都包含一个输出通道为 64、扩张率为 r 且卷积核大小为 3×3 的空洞卷积层, 用于防止过拟合的 BN 层以及为了提高模型的表达能力的 ReLU 激活层。空洞卷积是指在一般的卷积核中加入不同大小的“孔”, 从而让该卷积的感受野增大, 并且不会减小特征的分辨率。然而, 空洞卷积也存在天然的缺陷, 即“棋盘效应”, 对此, 通过密集连接的方式将较浅的扩张卷积层的输出与其输入相连接, 并继续输送到更深的扩张卷积层中, 如式(1)所示:

$$y_l = \text{Conv}_1([y_{l-1}, y_{l-2}, \dots, y_0])。 \quad (1)$$

式中: y_l 为扩张率为 1 的空洞卷积块的输入; Conv_1 代表包含输出通道为 256 且卷积核大小为 1×1 的卷积层、BN 层、以及 ReLU 激活层的卷积块; $[y_{l-1}, y_{l-2}, \dots, y_0]$ 表示将之前各层的输出串联起来形成特征, 其目的是为了不同感受野下得到的特征更好地融合, 进而输入到下一个空洞卷积块中。

密集的结构让不同扩张率的扩张卷积之间可以相互依赖, 构成密集特征金字塔, 缓解“棋盘效应”, 同时也可以获得更大的感受野。如式(2)所示, 注意力嵌入的密集空洞金字塔模块最大的感受野为 128。

$R_{\max} = R_{3,3} + R_{3,6} + R_{3,12} + R_{3,18} + R_{3,24} = 128。$ (2)

式中: $R_{K,d}$ 表示卷积核大小为 K , 扩张率为 d 的扩张卷积的感受野。

在原始的 DenseASPP 中, 当特征通过最后一个空洞卷积块后, 会将每一个空洞卷积块以及该模块所输入的初始特征相串联, 作为 DenseASPP 模块最终输出的多尺度的上下文特征。然而, 由于不同通道的特征映射了不同的语义, 如果将它们不加以区别地集成为多尺度特征, 会导致信息冗余。所以本文与 DenseASPP 有所区别的是嵌入了通道注意力机制, 对不同通道的特征进行过滤, 向对显著对象有较高响应的通道分配较高的权值, 同时向表示非显著区域特征的通道分配较低的权值, 削弱非显著的背景信息, 产生更加合理的多尺度特征。

将原始 DenseASPP 的输出 $f \in \mathbf{R}_{W \times H \times C}$ 展开为 $f =$

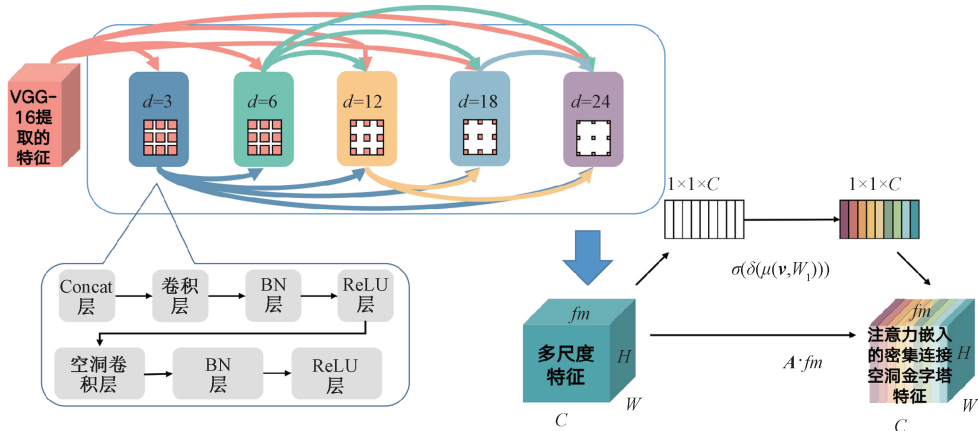


图 2 注意力嵌入的密集空洞金字塔模块

Figure 2 Attention embedded dense atrous pyramid module

$[f_1, f_2, \dots, f_c]$, 其中 $f_i \in \mathbf{R}_{W \times H}$ 是 f 的第 i 个切片, C 是 f 的总通道数。首先, 对每个 f_i 进行平均池化从而获得通道级特征向量 $\mathbf{v} \in \mathbf{R}_c$, 之后, 利用全连接层去捕获通道间的相关性, 接着, 通过 sigmoid 激活函数将特征映射到 $[0, 1]$, 从而获得每个通道的注意力权值 \mathbf{A} 。具体过程为

$$\mathbf{A} = F(\mathbf{v}, \mathbf{W}) = \sigma(\delta(\mu(\mathbf{v}, \mathbf{W}_1)))。 \quad (3)$$

式中: \mathbf{A} 为每个通道的注意力权值; \mathbf{v} 表示通道级特征向量; \mathbf{W} 为该通道注意力机制中的参数; σ 为 sigmoid 激活函数; δ 为 ReLU 激活函数; μ 为全连接层; \mathbf{W}_1 为全连接层中的参数。

最后, 使用注意力权值 \mathbf{A} 对原始多尺度特征 f_m 进行加权, 从而获得注意力嵌入的密集连接空洞金字塔特征 \tilde{f}_m :

$$\tilde{f}_m = \mathbf{A} \cdot f_m。 \quad (4)$$

图 3 展示了该模块前后的特征可视化结果。在该模块之前, 网络提取到的特征比较杂乱, 不能区分明显的显著对象, 在该模块之后, 由于密集空洞卷积对感受野的多样化扩大, 以及注意力机制对特征的有效选择, 使网络提取的特征对显著区域具有了高响应性, 从而准确地获得了多个大小不同、位置不一的显著性对象。

2.3 渐进边缘优化模块

显著性对象的边缘模糊一直是该领域的热点问题。本文认为, 深层特征中往往包含更多的语义特征, 而浅层特征中包含更多的空间结构特征, 仅仅用高级语义特征去检测显著性对象无法得到清晰的边界轮廓。由于浅层的特征中保留了更多的空间细节, 适合于定位边界, 所以借助渐进边缘优化模块, 向最终进行显著性检测的特征中依次补充从骨干特征提取网络中提取的各级浅层特

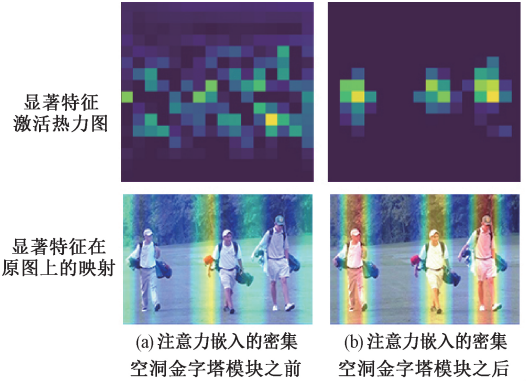


图 3 注意力嵌入的密集空洞金字塔模块前后的特征可视化

Figure 3 Visualization of feature maps around attention embedded dense atrous pyramid module

征, 以此来得到更加清晰的边缘轮廓。

渐进边缘优化模块的结构如图 4 所示。将来自于注意力嵌入的密集连接空洞金字塔模块特征 \tilde{f}_m 作为该模块的输入, 并定义该模块中的卷积块 H_i 为 $\text{ZeroPadding}(1, 1) \rightarrow \text{Conv}(3 \times 3, N_i, \text{'valid'}) \rightarrow \text{BN}$, N_i 为与当前特征分辨率相同的底层特征 C_i 的通道数。首先, 将 \tilde{f}_m 通过卷积块得到显著特征 S_4 ; 然后, 通过式 (5) 依次得到显著特征 $\{S_3, S_2, S_1\}$ 。

$S_{i-1} = H_{i-1}(\text{Concat}(\text{UpSampling}(S_i, C_i)))$ 。 (5) 式中: H_{i-1} 为第 $(i-1)$ 个卷积块; Concat 为连接层; UpSampling 为上采样层; S_i 为第 i 个显著特征; C_i 为第 i 个来自于骨干特征提取网络的特征。即每一次将特征上采样后, 都将其与来自底层网络的相同分辨率的特征相串联, 以补充相应的空间信息, 并经过卷积层进行充分融合。

为了使每一次边缘优化都能够从底层网络中获取对显著性检测有用的特征, 而不是引入噪声,

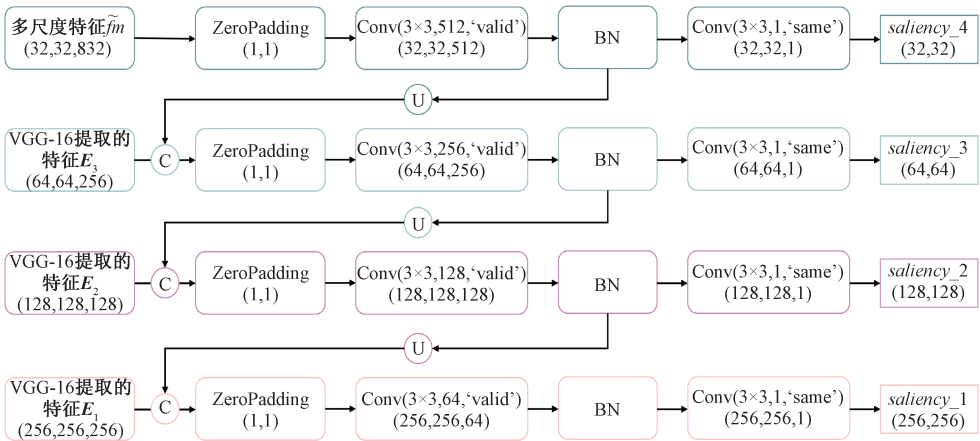


图 4 渐进边缘优化模块

Figure 4 Stepped edge optimization module

因此在每一个 S_i 之后,都设置了一个输出通道为 1、卷积核大小为 3×3 的卷积层,并将得到的显著映射 $\{saliency_4, saliency_3, saliency_2, saliency_1\}$ 作为输出,将原始的标注结果分别调整到与各个显著映射大小相同的分辨率 (256×256 、 128×128 、 64×64 、 32×32),并计算其损失,监督每一个边缘优化块从底层网络中提取出有效的空间细节信息,从而使得显著映射中显著性对象的边缘逐步清晰。

图 5 展示了渐进边缘优化模块中每一次优化后所得到的显著映射的边缘轮廓,可以看出,得益于各级底层网络特征的补充,显著性对象的边缘轮廓也逐渐清晰,一步步接近真值图的边缘轮廓。该模块对网络性能指标的提升将在 3.3 节中详细介绍。

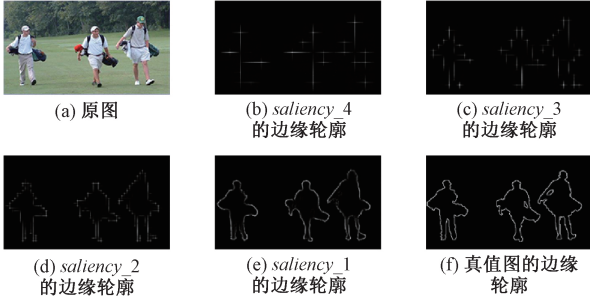


图 5 渐进边缘优化模块中每一次优化所得到的边缘轮廓可视化

Figure 5 Visualization of edges obtained with every optimization in stepped edge optimization module

2.4 实现细节

2.4.1 损失函数

受 Wang 等^[16]的启发,考虑多种评价指标,利用融合的损失函数进行显著性预测,如式(6)所示:

$$Loss = L_C + L_M. \quad (6)$$

式中: L_C 表示交叉熵损失函数; L_M 表示平均绝对误差损失函数。

L_C 的定义如下:

$$L_C = - \sum_{x,y} l_{x,y} \lg P_{x,y} + (1 - l_{x,y}) \lg(1 - P_{x,y}). \quad (7)$$

L_M 的定义如下:

$$L_M = \sum_{x,y} |l_{x,y} - P_{x,y}|. \quad (8)$$

式中: $l_{x,y} \in \{0, 1\}$ 表示像素 (x, y) 的标签; $P_{x,y}$ 表示像素 (x, y) 的预测值。

2.4.2 训练数据集

使用 DUTS-train 数据集^[17]来训练网络,它包含 10 553 张图像。

为了使模型拥有足够强的泛化能力,在训练时使用了数据增强技术。具体来说,使用了随机角度变换、水平或垂直方向上平移、随机错切变换角度、图像水平翻转,从而使图像增强至 65 k。

2.4.3 其他细节

本文模型使用 Keras 框架来实现。主干特征提取网络的参数由在 ImageNet 上预训练的 VGG-16 进行初始化,对于其他卷积层,采用截断法初始化它们的参数。训练时,所有的图像和真值图的分辨率都被调整为 256×256 。模型的初始学习率被设置为 10^{-4} ,使用 Adam 算法在 NVIDIA GTX 1080 Ti GPU 显卡上进行训练,直至模型参数收敛。

3 实验部分

3.1 数据集和评价指标

在显著性领域中常用的 5 个公开数据集上对本文的网络模型进行性能评价,5 个数据集为 DUTS-TE^[17]、HKU-IS^[18]、DUT-OMRON^[19]、PASCAL-S^[20]和 ECSSD^[21]。

与其他最新的显著性对象检测方法相同,本文选用了 3 种流行的评价指标对性能进行评估:精确率-召回率曲线、 F 值以及平均绝对误差 MAE 。

3.2 性能比较

将本文的方法与其他 12 种经典、主流的显著性对象检测方法在 5 个公开数据集上进行了比较,对象检测方法分别是:AMU^[22]、DCL^[23]、DLS^[15]、DS^[14]、ELD^[24]、KSR^[25]、LEGS^[12]、MCDL^[26]、MDF^[18]、UCF^[13]、CapSal^[27]、CAN^[28]。

3.2.1 定量评价

表 1 和图 6 展示了本文方法与其他方法在 5 个数据集上定量比较的结果。表 1 显示了不同方法在不同数据集上所测得的 F 值、平均绝对误差 MAE 、模型参数量和检测时间,图 6 则展示了对应的精确率-召回率曲线。总体而言,本文方法使用所有的评价指标在 5 个数据集中都取得了很好的结果。

3.2.2 定性评价

图 7 显示了本文模型与其他显著性对象检测模型所生成的显著映射的对比结果。可以看出,本文方法相比于其他方法来说,无论是单一物体的简单场景(图 7 第 4、5 行)、存在多个大小不同的对象的场景(图 7 第 3、7 行)、存在多个断开连接的物体的情况(图 7 第 1、2、6 行)、对象与背景

的色彩对比度差异较低的情况(图 7 第 8 行),不仅可以准确地检测出其中所有的显著对象,并且可以获得更加清晰的边缘轮廓。这得益于本文提出的 AEDAPM 和 SEOM 网络结构,使网络具有更

好的特征提取能力和边缘检测能力,从而让显著映射具有更完整的显著区域和更清晰的边缘轮廓。

表 1 不同显著性检测方法在 5 个公开数据集上的 F 值和 MAE 、参数量以及检测时间的对比结果

Table 1 The maximum F , MAE , parameter quantity and detection speed of different saliency detection methods on five released saliency detection datasets												
方法	参数量/ MB	检测时 间/s	DUT-OMRON		DUTS-TE		ECSSD		HKU-IS		PASCAL-S	
			F	MAE	F	MAE	F	MAE	F	MAE	F	MAE
AMU	132.6	0.07	0.715	0.098	0.750	0.085	0.905	0.059	0.887	0.051	0.805	0.100
DCL	71.41	0.39	0.699	0.080	0.742	0.088	0.882	0.068	0.885	0.048	0.787	0.107
DLS	—	0.08	0.644	0.090	—	—	0.826	0.086	0.807	0.069	0.712	0.130
DS	134.26	0.13	0.708	0.120	0.747	0.090	0.868	0.122	0.848	0.078	0.718	0.175
ELD	43.15	0.55	0.678	0.091	0.697	0.092	0.849	0.078	0.868	0.063	0.782	0.111
KSR	113.74	49.64	0.639	0.131	0.660	0.123	0.801	0.133	0.760	0.120	0.733	0.154
LEGS	18.40	1.54	0.631	0.133	0.612	0.137	0.805	0.118	0.736	0.119	0.714	0.154
MCDL	116.82	2.38	0.671	0.752	0.634	0.105	0.816	0.101	0.787	0.092	0.706	0.143
MDF	56.94	7.83	0.643	0.092	0.657	0.114	0.797	0.105	0.839	0.129	0.704	0.142
UCF	117.9	0.04	0.698	0.120	0.742	0.112	0.890	0.069	0.874	0.062	0.787	0.115
CapSal	91.09	0.37	0.535	0.101	0.770	0.063	0.813	0.077	0.840	0.058	0.827	0.073
CAN	—	—	0.594	0.138	—	—	0.765	0.143	—	—	0.725	0.181
本文方法	59.99	0.15	0.721	0.078	0.796	0.062	0.951	0.030	0.890	0.047	0.810	0.087

注:红色(加粗)字体表示最优结果,蓝色(浅色)字体表示次优结果。

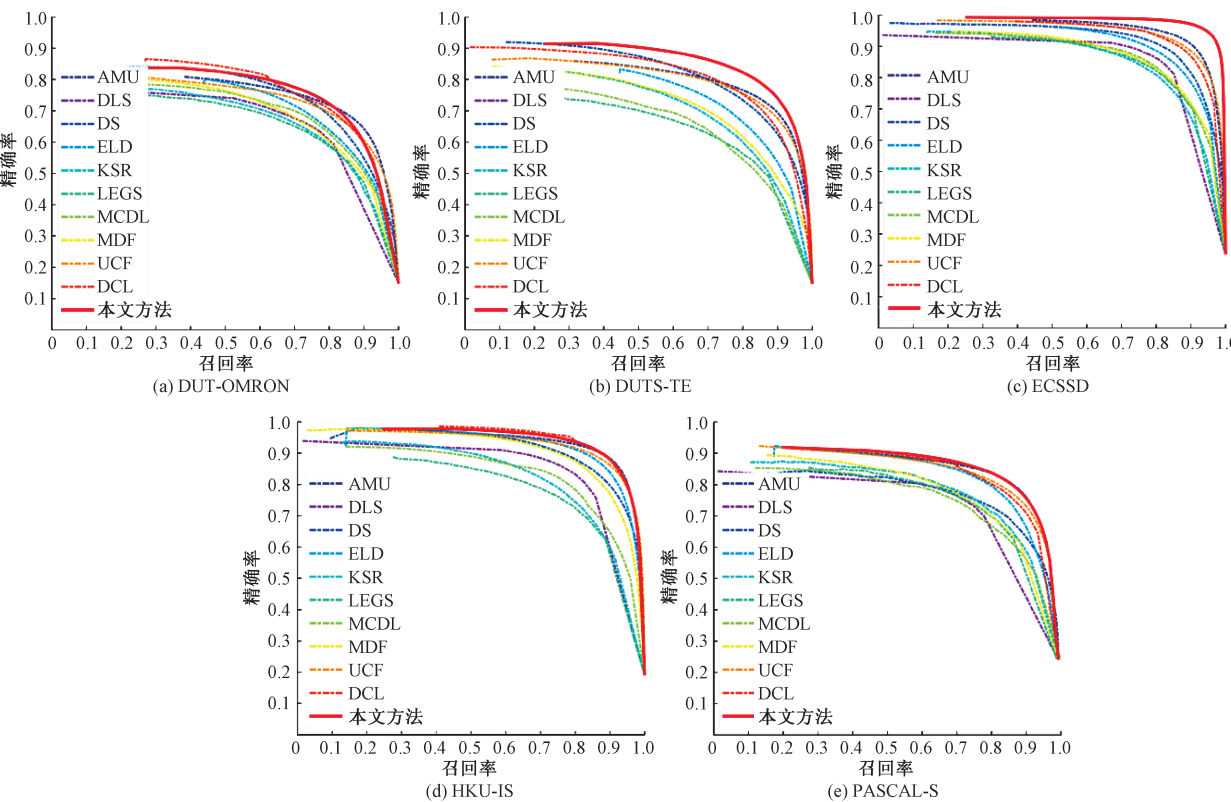


图 6 5 个数据集上的精确率-召回率曲线

Figure 6 Precision-recall curves on five datasets

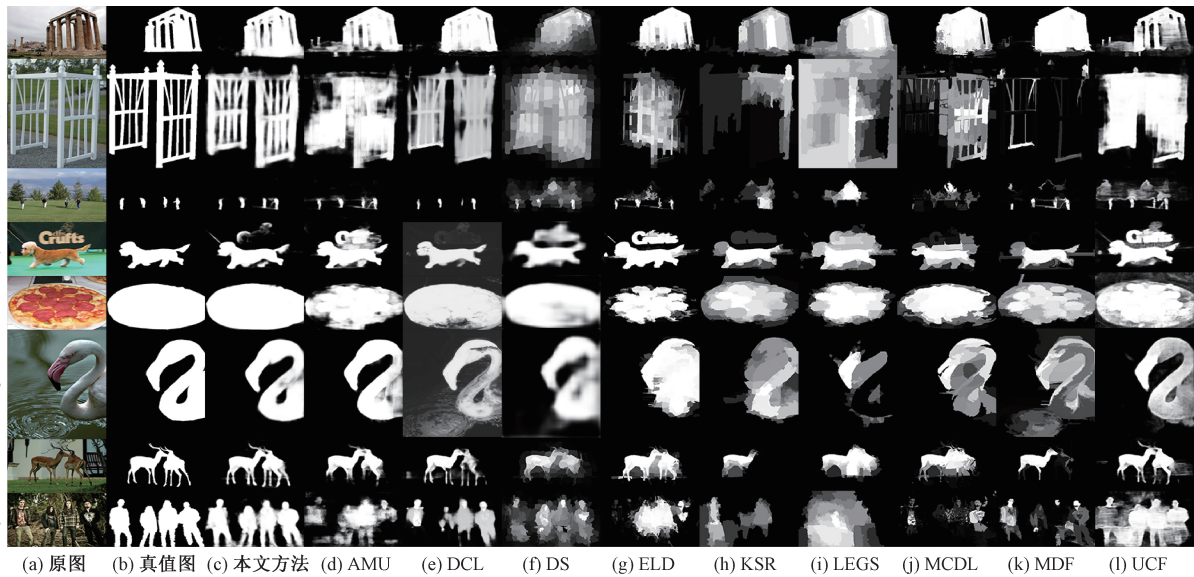


图 7 不同方法的实验结果对比

Figure 7 Qualitative comparisons of different SOD methods

3.3 消融实验

本文提出的网络模型主要由 2 个模块组成：注意力嵌入的密集空洞金字塔模块 (AEDAPM) 和渐进边缘优化模块 (SEOM)。为了检测 2 个模块的有效性,进行了以下的消融实验。如表 2 所示,当模型包含所有模块时,能够取得最好的性能表现,当分别去掉了 AEDAPM 和 SEOM 时,精度都会有一定程度的下降,这说明所有组件对于该方法获得最佳的结果而言都是必要的。

表 2 AEDAPM 和 SEOM 的消融实验

Table 2 Ablation study for AEDAPM and SEOM

序 号	全卷积神 经网络	AEDAPM		SEOM	F	MAE
		空洞卷积	注意力机制			
1	√				0.658	0.143
2	√	√	√		0.729	0.109
3	√	√		√	0.751	0.078
4	√		√	√	0.738	0.098
5	√			√	0.712	0.126
6	√	√	√	√	0.796	0.062

注:√表示该实验的网络包含这一模块,加粗字体为最优值。

首先使用基本的全卷积网络作为基本模型。其次,将 AEDAPM 单独添加到网络中(对比第 1 行和第 2 行),F 值以及 MAE 分别优化了 10.79% 和 23.78%,这说明了本文提出的 AEDAPM 能够获得更加丰富的多尺度特征,从而提高模型的表征能力。此外,将 SEOM 单独添加到基本的全卷积网络中(对比第 1 行和第 5 行),F 值以及 MAE 分别优化了 8.2% 和 11.89%,这表明 SEOM 能够让网络补充浅层的空间细节信息,同时避免冗杂

背景信息的干扰。

此外,对 AEDAPM 中空洞卷积和注意力机制的影响作出了进一步的对比和分析。由表 2 的第 3、4 行所示,空洞卷积比注意力机制对网络的影响更大,这说明空洞卷积对于网络感受野的丰富要比注意力机制对特征产生不同程度的响应更加重要。

4 结论

本文提出了一个新颖的基于金字塔特征与边缘优化的显著性对象检测模型,采用由扩张卷积构成的注意力嵌入的密集空洞金字塔模块 (AEDAPM) 获取不同感受野下不同尺度的上下文信息,以此检测不同大小、形状的显著性对象,同时设计了渐进边缘优化模块 (SEOM),用于补充更加丰富而不冗余的空间细节特征,使显著性对象的边缘轮廓更加清晰。本文提出的方法在多个公开的显著性对象检测数据集上与其他主流方法相比,都能得到更加准确、对象边缘更加清晰的结果。

参考文献:

[1] 杨文柱,刘晴,王思乐,等.基于深度卷积神经网络的羽绒图像识别[J].郑州大学学报(工学版),2018,39(2):11-17.
[2] 张震,李浩方,李孟洲,等.改进 YOLOv3 算法与人体信息数据融合的视频监控检测方法[J].郑州大学学报(工学版),2021,42(1):28-34.
[3] BORJI A,FRINTROP S,SIHITE D N,et al.Adaptive

- object tracking by learning background context [C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2012: 23–30.
- [4] 魏宏彬, 张端金, 杜广明, 等. 基于改进型 YOLO v3 的蔬菜识别算法 [J]. 郑州大学学报(工学版), 2020, 41(2): 7–12, 31.
- [5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 3431–3440.
- [6] YANG M K, YU K, ZHANG C, et al. DenseASPP for semantic segmentation in street scenes [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3684–3692.
- [7] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE transactions on pattern analysis and machine intelligence, 1998, 20(11): 1254–1259.
- [8] HOU X D, ZHANG L Q. Saliency detection; a spectral residual approach [C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2007: 1–8.
- [9] 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.
- [10] XIE Y, LU H, YANG M H. Bayesian saliency via low and mid level cues [J]. IEEE transactions on image processing, 2013, 22(5): 1689–1698.
- [11] FAN D P, CHENG M M, LIU J J, et al. Salient objects in clutter: bringing salient object detection to the foreground [C]// European Conference on Computer Vision. Springer: Cham, 2018: 438–445.
- [12] WANG L J, LU H C, RUAN X, et al. Deep networks for saliency detection via local estimation and global search [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 3183–3192.
- [13] ZHANG P P, WANG D, LU H C, et al. Learning uncertain convolutional features for accurate saliency detection [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 212–221.
- [14] LI X, ZHAO L M, WEI L N, et al. DeepSaliency: multi-task deep neural network model for salient object detection [J]. IEEE transactions on image processing, 2016, 25(8): 3919–3930.
- [15] HU P, SHUAI B, LIU J, et al. Deep level sets for salient object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 540–549.
- [16] WANG W G, SHEN J B, DONG X P, et al. Salient object detection driven by fixation prediction [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1711–1720.
- [17] WANG L J, LU H C, WANG Y F, et al. Learning to detect salient objects with image-level supervision [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 3796–3805.
- [18] LI G B, YU Y Z. Visual saliency based on multiscale deep features [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 5455–5463.
- [19] YANG C, ZHANG L H, LU H C, et al. Saliency detection via graph-based manifold ranking [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3166–3173.
- [20] LI Y, HOU X D, KOCH C, et al. The secrets of salient object segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 280–287.
- [21] YAN Q, XU L, SHI J P, et al. Hierarchical saliency detection [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 1155–1162.
- [22] ZHANG P P, WANG D, LU H C, et al. Amulet: aggregating multi-level convolutional features for salient object detection [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 202–211.
- [23] LI G B, YU Y Z. Deep contrast learning for salient object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 478–487.
- [24] LEE G, TAI Y W, KIM J. Deep saliency with encoded low level distance map and high level features [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 660–668.
- [25] WANG T T, ZHANG L H, LU H C, et al. Kernelized subspace ranking for saliency detection [M]. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 450–466.
- [26] ZHAO R, OUYANG W L, LI H S, et al. Saliency detec-

tion by multi-context deep learning[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 1265-1274.

[27] ZHANG L, ZHANG J M, LIN Z, et al. CapSal: leveraging captioning to boost semantics for salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE, 2019:6017-6026.

[28] 刘尚旺,赵欣莹,杨磊.基于全局和局部信息融合的显著性检测[J]. 河南师范大学学报(自然科学版), 2020, 48(3): 26-33.

Salient Object Detection Based on Pyramid Features and Edge Optimization

TIAN Xu¹, PENG Fei¹, LIU Fei¹, CHEN Qingwen², YAN Xinyu^{3,4}

(1.State Grid Qinghai Electric Power Company Economic and Technical Research Institute, Xining 810000, China; 2.Northwest Engineering Corporation Limited, Xi'an 710065, China; 3.College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; 4.Tianjin Key Lab of Machine Learning, Tianjin University, Tianjin 300350, China)

Abstract: To solve the problems of insufficient multiscale feature extraction and object edge blur in image-based salient object detection, an end-to-end salient object detection model was proposed based on attention embedding pyramid feature and stepped edge optimization. Firstly, the attention embedded dense atrous Pyramid Module (AEDAPM) composed of multiple dilated convolutions was designed to obtain rich and effective multi-level multi-scale features without reducing the feature resolution; Secondly, in order to solve the problem of blurring the edges of salient objects, a stepped edge optimization module (SEOM) is proposed, which gradually supplements spatial detail information during the process of feature restoration resolution, so that the salient objects detected by the model could have clear edge contours. The method in this paper was compared with 12 state-of-the-art saliency methods under 3 common indicators on 5 public datasets, such as DUTS-TE, ECSSD, DUT-OMRON, HKU-IS, and PASCAL-S. The experimental results show that the method proposed in this paper can obtain more accurate and clearer saliency results. In addition, the ablation study also fully proved the effectiveness of the AEDAPM and the SEOM proposed in this study.

Keywords: salient object detection; multi-scale feature extraction; fully convolutional networks; edge feature extraction; deep learning