arXiv:2301.07431v1 [cs.CV] 18 Jan 2023

锐眼：以人类视觉特征相同方式工作的显著目标检测器

朱戈、李金宝，*IEEE* 会员，和郭亚红

摘要—当前方法通过聚合多层级特征，或引入边缘和骨架，来获得更精细的显著图。然而，对于如何在杂乱背景中获取完整的显著目标，现有研究关注较少；在这类场景中，目标通常在颜色和纹理上与背景相似。为处理这一复杂场景，我们提出了一种锐眼网络（SENet），首先将目标从场景中分离出来，然后对其进行精细分割，这符合人类视觉特征，即先观察，再聚焦。不同于以往直接融合边缘或骨架以弥补目标缺陷的方法，所提出的方法旨在利用扩展目标引导网络获得完整预测。具体而言，SENet主要由目标分离（TS）分支和目标分割（OS）分支组成，并通过最小化一种新的分层差异感知（HDA）损失进行训练。在TS分支中，我们构建了一种分形结构，通过扩展真值的监督生成具有扩展边界的显著性特征，从而扩大前景和背景之间的细节差异。在OS分支中，我们首先聚合多层级特征，以自适应地选择互补组件，然后将具有扩展边界的显著性特征输入聚合特征中，引导网络获得完整预测。此外，我们提出 HDA 损失，通过根据每个像素与边界之间的距离分层分配权重，进一步提升显著目标的结构完整性和局部细节。边界区域中外观相似的困难像素将分层获得更多关注，以强调其在完整性预测中的重要性。在五个基准数据集上的综合实验结果表明，所提出的方法在定量和定性方面均优于最先进方法。代码已公开于 https://github.com/lesonly/SENet。

索引词——**计算机视觉、显著目标检测、深度学习、卷积神经网络。**

I. 引言

显著目标检测（SOD）在计算机视觉中通常被视为一个包含两个阶段的过程：1）检测最显著的目标；2）分割该目标的精确区域 [1]。作为一种预处理过程，SOD 广泛应用于下游视觉任务，例如图像理解 [2]、目标跟踪 [3] 和图像检索 [4]。

近年来，随着深度学习技术取得突破，许多基于卷积神经网络（CNN）极大地提升了显著目标检测的性能。然而，在提取图像特征的过程中，CNN 会同时处理形状、颜色、纹理和亮度。在一些背景杂乱的复杂场景中，背景的颜色或纹理往往与前景相似，预测显著图仍会出现不完整预测。如图1所示，不完整预测是指只能预测出显著目标的一部分（第1行），或无法均匀地突出目标（第3行第5列），或无法检测到任何目标（第2行第3列）。

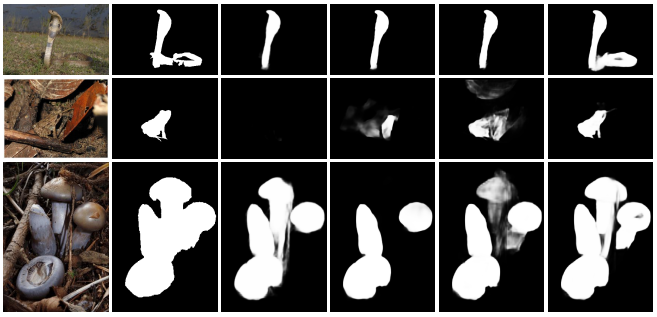


图1：不同方法在背景杂乱的复杂场景中的预测结果示例。第一行中，蛇的身体隐藏在杂草中，其外观与杂草相似。第二行中，青蛙栖息在叶片上，叶片的颜色和纹理与青蛙相似。第三行中，蘑菇散布在灌木丛中，对比度较低。然而，三种近期方法（PurNet [5], DCN [6]和 ICON [7]）在上述场景中得到的预测结果不完整，而我们的方法能够应对这些复杂场景，并生成更加完整的显著图。

为了获得精细的显著图，当前最先进方法主要致力于整合多层级特征，或引入辅助信息（即边缘和骨架）来获取互补特征。为了学习全面的特征，Zhang 等人 [8] 和诸葛等人 [7]提出了聚合网络，以整合多层级特征。然而，由于 CNN模型中反复进行下采样操作，复杂结构显著目标的细节会逐渐减弱，因此难以均匀地突出完整的显著目标。为了获得精确的边界，Zhou 等人 [9] 和魏等人 [10] 将边缘信息引入模型，但边缘信息仅包含整个目标的一小部分，因此难以填充目标内部。

朱戈任职于黑龙江大学电子工程学院和数据科学与技术学院，地址：中国黑龙江省哈尔滨市，邮编 150080（电子邮箱：zhuge@hlju.edu.cn）。

李金宝任职于齐鲁工业大学（山东省科学院）山东人工智能研究院、数学与统计学院，地址：中国山东省济南市，邮编 250353（电子邮箱：li-jinb@sdas.org）。（通讯作者：李金宝。）

郭亚红任职于中国山东省济南市齐鲁工业大学（山东省科学院）计算机科学与技术学院，邮编 250353（电子邮箱：guoyh@qlu.edu.cn）。（通讯作者：郭亚红。）

为了获得更多互补特征，刘等人 [11]和吴等人 [6] 将边缘信息和骨架信息都整合到网络中。然而，特征域的差异可能会带来背景噪声，从而难以辅助网络生成完整的显著图。由于将整张显著图设为前景即可实现 100% 的召回率，因此挑战在于如何在复杂场景中噪声较低的前提下解决预测不完整的问题。简而言之，上述三类方法都是从补充信息的角度出发的。遇到杂乱背景时，很难将目标完整地分离出来。图1展示了不同方法在杂乱场景中的示例结果，包括基于净化机制的 PurNet[5] 、基于引入边界和骨架信息的 DCN [6]，以及基于聚合多层级特征的 ICON[7] 。

此外，现有的大多数模型采用二元交叉熵和交并比作为损失函数，但它们对所有像素一视同仁，且独立处理。由于边界像素与非边界像素的分布不平衡，清晰边界难以预测。为解决这一问题，人们提出了各种损失函数 [12], [13], [14] ，通过为边界附近的像素赋予较高权重来增强边界细节。尽管效果显著，但仅考虑边界周围的像素仍不够全面；位于前景和背景之间边界区域、外观相似的像素容易被错误分类。因此，应关注不同等级的像素，而不是将目标内部的像素一视同仁。

为解决上述问题，我们提出锐眼网络（SENet），采用先分离、后分割的策略，以生成更完整、噪声更少的显著图。与基于补充信息的方法相比，本文提出的方法采用相反的策略，即检测范围由大到小，这与人类视觉的特点一致。首先，为防止背景干扰并定位完整目标，我们设计了目标分离（TS）分支，用于生成带扩展边界的显著性特征（SF w/ EB）。该特征呈分形结构，并在经过膨胀和腐蚀操作编码的扩展真值监督下生成。与边缘或骨架信息相比，本文提出的 SF w/ EB 能够在细节层面扩大显著目标与杂乱背景之间的差异。其次，为了获得精细的显著目标，我们设计了目标分割（OS）分支，该分支由特征聚合（FA）模块和特征交互（FI）模块组成。FA 模块通过渐进式聚合高层语义特征、低层细节特征和全局上下文信息，生成更多互补特征。FI 模块通过逐元素乘法将 SF w/ EB 融入上述聚合特征，以增强显著目标的结构和细节信息。交互特征具有语义一致性和结构信息一致性的特点。此外，为抑制交互特征带来的冗余和噪声，我们使用一系列加权卷积层对其进行细化。最后，FA 模块和 FI 模块将级联三次，生成最终的显著图。在每次级联中，使用原始真值监督 FI 模块的输出，以防止 SF w/ EB 在学习过程中干扰模型。

此外，我们提出一种分层差异感知（HDA）损失，以进一步改善预测的结构完整性和局部细节。不同于以往为边界周围像素赋予较高权重的损失函数，所提出的 HDA 损失根据像素与边界的距离分层分配权重：通过计算膨胀后的真实标注和腐蚀后的真实标注中某个像素与其邻域像素之间差异的总和，为每个像素赋予权重。因此，边界区域中的困难像素将受到更多关注，以区分前景和背景中相似的部分。总而言之，我们的主要贡献如下：

1) 我们提出了一种新颖的显著目标检测器，其工作方式与人类视觉特征相同：先分离目标，再对其进行分割。该框架的检测范围由大到小，分割精度由粗到细，能够整合扩展后的显著性特征与聚合的互补特征，从而生成更完整、噪声更少的显著图。

2) 我们设计了一种层次差异感知损失，根据像素与边界的距离分级赋予像素权重。通过聚焦于难以区分的区域，该损失能够进一步提升结构完整性和局部细节的预测效果。

3) 在五个基准数据集上进行的大量实验结果表明，所提出的模型在定量和定性评估方面均优于15种最先进方法。此外，我们的模型效率更高，处理一幅 320×320图像时的推理速度为39.5帧每秒。

II. 相关工作

过去十年间，人们提出了大量基于图像的显著目标检测方法。传统方法主要依据内在特征生成显著图，例如中心先验 [15], 颜色差异 [16]和对比度。尽管上述方法在大多数简单场景中能够生成良好的显著图，但由于缺乏高层语义信息，它们在某些复杂场景中不够鲁棒。随着深度学习在计算机视觉领域取得突破，许多基于卷积神经网络（CNN）的方法显著提升了显著目标检测的性能。本节简要回顾两类主流的基于CNN的方法，包括基于聚合多层级特征的方法，以及基于引入辅助信息（例如边缘或骨架）的方法。

A. 基于多层级特征聚合的方法

为了识别不同尺度的目标，许多方法通过聚合多层级特征，融合特征金字塔中的语义信息与细节。其中，Hou 等人 [18] 在跳层结构中引入了短连接，以便在每一层生成更高级的表示。Zhang 等人 [19] 提出了一种门控双向消息传递模块，用于自适应地整合多层级特征。Deng 等

人 [20]采用迭代策略，交替利用低层和高层特征来优化显著图。刘等人 [21]提出了一种注意力网络，用于学习每个像素的全局和局部上下文。Chen 等人[22]提出了上下文感知模块，用于聚合低层外观特征、高层语义特征和全局上下文特征。魏等人 [12] 提出了交叉特征模块，通过乘法消除不同层之间的差异，并以级联方式优化多层级特征。Wang 等人 [23] 提出了一种有效的框架，逐步优化多层级特征，使其具有更强的表征能力。Pang 等人 [24] 设计了聚合模块来整合相邻层级的特征，并利用自交互模块生成多尺度表示。Ma 等人 [25]通过逐层收缩，成对聚合相邻特征，以提取有效的细节和语义。Zhang [8]设计了一种架构搜索框架，以减轻人工专家手动设计多尺度特征融合的工作量。Li 等人 [5]提出了促进与校正注意力，以净化容易出错的区域。诸葛等人 [7]聚合具有不同感受野的特征，以提高特征的多样性。

尽管上述方法取得了很大进展，但受反复下采样操作的限制，具有复杂结构的物体细节信息会有所退化，因此在一些背景杂乱的复杂场景中可能出现预测不完整的问题。

B. 基于引入辅助信息的方法

为补充局部细节，许多方法通过引入边缘和骨架等辅助信息，帮助网络生成显著图。这些方法旨在探索不同信息之间的互补性。刘等人 [26] 利用额外的边缘数据集与显著性数据集进行联合训练，以改善显著目标的局部细节。Zhao 等人 [27]和 Zhao 等人 [28] 直接采用二元交叉熵构建边缘损失，以学习更多边界细节。Feng 等人 [29], 、魏等人 [12] 和 Chen 等人 [13]采用边缘增强损失来解决模糊边缘问题。Su 等人 [30], 、Zhao 等人[31] 和 Zhu 等人 [14] 使用独立分支预测显著边界，从而更好地补充显著目标的边界细节。吴等人 [32] 和周等人 [9]提出双向特征细化机制，在显著性与边缘这两个任务之间传递互补信息。魏等人 [10] 将真实标注分解为主体图和细节图，并采用两个分支分别关注目标的中心区域和边缘。尽管将边缘信息引入框架，但边缘仅包含完整目标的一小部分，因此难以填充目标内部。为同时准确定位目标的边界和内部，近期方法将边缘和骨架信息同时引入框架。刘等人 [11] 提出动态特征集成策略，自动探索显著目标、边缘和骨架检测之间的特征组合。吴等人 [6]提出一种跨多分支解码器，通过迭代利用跨任务聚合来整合显著性、边缘和骨架之间的特征。然而，不同任务之间的差异可能会引入噪声，使网络难以统一突出整个显著目标。

无论是基于聚合多层级特征的方法，还是引入辅助信息的方法，它们关注的都是如何弥补物体的缺陷，因此难以在复杂场景中获得完整预测。与上述通过丰富信息来改进的方法不同，我们提出了一种框架，利用扩展的显著性特征引导网络生成噪声较低的完整显著图。

III. 方法

本节首先概述所提出模型的整体架构。然后详细说明各个分支的工作方式。最后介绍所提出的损失函数。

A. 网络概述

如图 2 所示，所提出的网络采用非对称编码器–解码器架构。编码器基于 ResNet-50 [33] 提取多层级特征。解码器主要由目标分离（TS）分支和目标分割（OS）分支组成。其中，TS分支构建分形结构，以生成扩展边界显著性特征（SF w/ EB）。OS 分支逐步整合 SF w/ EB 和聚合特征，从而获得完整的显著图。为加速优化，所提出的层次差异感知损失和多层级监督策略应用于该网络。

B. 特征编码器

我们使用 ResNet-50 [33] 作为骨干网络，因为它在速度和准确率之间实现了更好的平衡。具体而言，由于低层特征会带来更多计算开销，但提升有限 [34], ，我们仅使用 ResNet-50 的最后四个残差块来获取四级特征，记为 $\{F_i|i=2,3,4,5\}$。对于尺寸为 $H\times W\times 3$的输入图像，第 i 级特征的尺寸为 $\frac{H}{2^i}\times\frac{W}{2^i}\times C_i$，其中 C_i 是第 i级特征的通道数，且 C_i 的值为 64×2^i。

C. 目标分离分支

为了有效地将显著目标与背景分离，我们提出了目标分离（TS）分支，用于生成带扩展边界的显著性特征，从而扩大前景和背景之间的细节差异。如图 2 所示，在我们提出的扩展真值的监督下，TS分支采用自相似分形结构，旨在通过不同路径将高层特征的语义平滑地传递至低层特征。尽管这种结构与我们先前的工作 [14], 在理念上相似，但TS分支在两个方面与之显著不同：1）两个分支的监督信号不同（即两个分支生成的特征不同）。2）所用的子模块

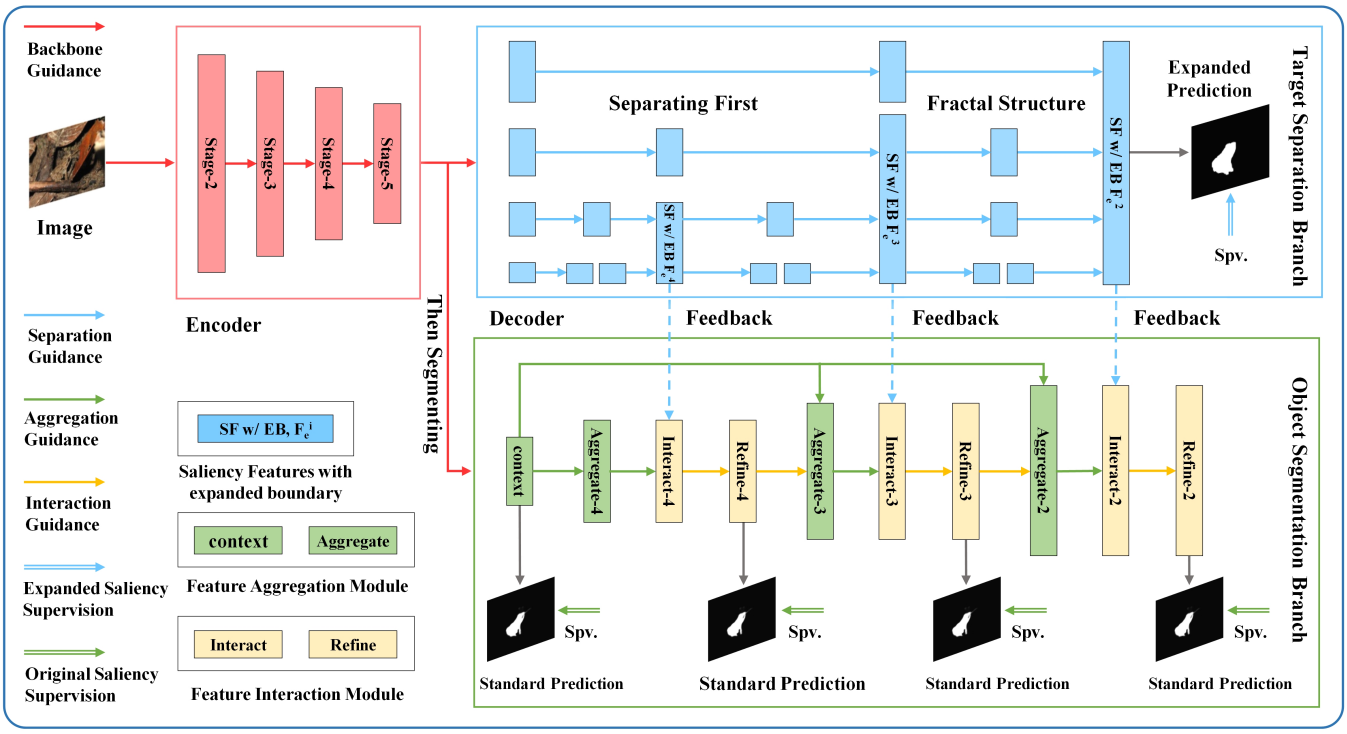


图 2：所提出模型的整体架构。编码器基于 ResNet-50，包括 stage2 至 stage5。解码器由目标分离（TS）分支和目标分割（OS）分支组成。采用多层级监督来优化所提出的网络。

也不相同。由于带扩展边界的显著性特征的范围大于原始显著性特征，TS分支加入了用于扩大视觉感受野的模块，以及用于抑制噪声的注意力模块。扩展真值和带扩展边界的显著性特征的生成细节如下所述。

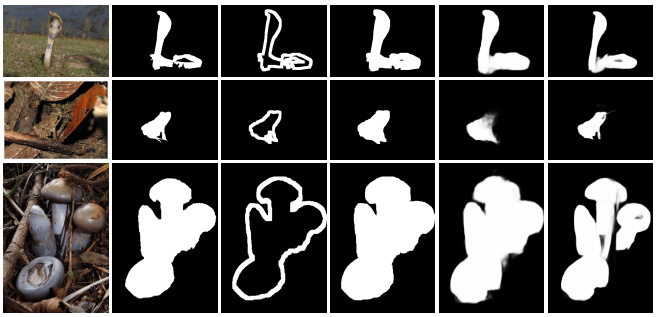
1）扩展真值生成：考虑到边界是区分前景与背景的主要信息，我们构建扩展真值来监督 TS 分支的输出，从而生成带扩展边界的显著性特征。图 3 展示了扩展真值的构建过程，以及网络不同组件的预测结果。可以看出，与原始真值相比，扩展真值包含扩展后的目标边界。在扩展真值的监督下，TS 分支输出带扩展边界的显著性特征。此外，为了体现带扩展边界的显著性特征为网络带来的益处，最后一列展示了整个网络 SENet 的预测结果。

具体而言，我们首先分别对原始真值进行膨胀和腐蚀操作，记为 Dilate(.) 和 Erode (.)。然后，计算 Dilate (.) 与 Erode (.) 的差集，得到边界包络，其中包含边界的内侧和外侧。最后，计算边界包络与原始真值的并集，生成扩展真值。边界包络和扩展真值分别定义如公式 1 和公式 2 所示：

$$GT_{be} = \text{Dilate}(GT, k)^i - \text{Erode}(GT, k)^i, \quad (1)$$

$$GT_e = GT_{be} \cup GT, \quad (2)$$

其中，GT 是原始真值，GT_{be} 表示边界包络，GT_e 表示扩展真值；卷积核 k的大小为 3×3，迭代次数 i根据经验设为 5，$-$ 表示减法运算，$\cup$ 表示并集运算。



(a) (b) (c) (d) (e) (f)

图 3：（a）图像。（b）原始真值。（c）边界包络。（d）扩展真值。（e）TS分支的预测。（f）整个网络 SENet 的预测。我们可以观察到，在扩展真值的监督下，TS分支预测的显著目标包含扩展后的边界，这有助于扩大显著目标与背景之间的细节差异，并引导我们的模型生成完整的显著图。

2）带扩展边界生成的显著性特征：所提出的 TS 分支采用整体信息通信策略：如图2所示，通过多条路径将高层特征编码并注入低层特征。具体而言，我们首先使用一个 1×1 卷积层$conv1$处理编码器生成的特征图 $\{F_i|i=2,3,4,5\}$，得到通道数为 256 的压缩特征表示。然后，使用 ASPP模块 [35] 处理高层特征 F_4 和 F_5，以捕获多尺度上下文信息，从而增强高层内部特征之间的语义相关性。接下来，使用 RFB模块[36] 处理低层特征 F_2 和 F_3，以扩大其感受野，从而学习低层特征的全局视野。最后，我们利用分形结构重复上述特征融合过程，直至 F_2 完成融合，以促进信息平滑传递并缓解网络退化。在融合过程中，TS分支将输出三个层级的带扩展边界的显著性特征，并将其嵌入目标分割分支，引导网络生成完整预测。该过程可描述如下：

$$F_e^4 = Ca(\delta(up(\theta(\eta(conv_1(F_5))) + \eta(conv_1(F_4))), \quad (3)$$

$$F_e^3 = Ca(\delta(up(\theta(F_e^4) + conv_1(F_e^4))+\zeta(conv_1(F_3)))), \quad (4)$$

$$\begin{aligned} F_e^2 = Ca(\delta(up(\theta(\delta(\theta(F_e^3) + conv_1(F_e^3)))+ \\ conv_1(\delta(\theta(F_e^3) + conv_1(F_e^3)))+ \\ conv_1(F_e^3)) + \theta^*(\zeta(conv_1(F_2)))), \end{aligned} \quad (5)$$

其中，$\{F_e^i|i=2,3,4\}$ 表示由 TS分支生成的三级带扩展边界的显著性特征，其尺寸分别为$\frac{H}{16}\times\frac{W}{16}\times256$、$\frac{H}{8}\times\frac{W}{8}\times256$ 和 $\frac{H}{4}\times\frac{W}{4}\times256$；$\overline{conv_1}$表示用于通道缩减的 1×1 卷积层；η为 ASPP模块，ζ 表示 RFB模块；θ 表示一系列卷积层，包括一个具有 64 个通道的 1×1 卷积层、一个具有 64 个通道的 3×3 卷积层和一个具有 256 个通道的 1×1卷积层，上述三个卷积层之间设有两个 ReLU函数；$theta^*$ 用于编码 RFB模块处理后的特征，包括一个具有 64 个通道的 3×3 卷积层、一个 ReLU函数和一个具有 256 个通道的 1×1 卷积层；up 表示双线性插值操作，δ 表示 ReLU 激活函数，Ca表示通道注意力模块 [37]。

3) 目标分离分支损失：在所提出的扩展真值监督下，我们使用二元交叉熵（BCE）和交并比（IoU）函数来最小化TS分支的损失 L_c。具体而言，使用一个 3×3卷积层将F_e^2的通道数压缩至 1。然后，通过双线性插值将压缩后的特征图上采样至与真实标注相同的大小，记为F_e^1。最后，使用 Sigmoid 函数将预测值归一化为 [0,1]，以F_e^1。该过程可描述为：

$$F_e^1 = up(conv_3(F_e^2)), \quad (6)$$

$$L_c = \mathrm{Mean}(\mathrm{BCE}(\sigma(F_e^1), GT_e)+\mathrm{IoU}(\sigma(F_e^1), GT_e)), \quad (7)$$

其中，F_e^2和 F_e^1是由TS分支生成的带扩展边界的显著性特征，其大小分别为 $\frac{H}{4}\times\frac{W}{4}\times256$ 和 $H\times W\times1$。up表示双线性插值操作，σ 表示 Sigmoid 函数，Mean(.) 返回数组元素的均值。

D. 目标分割分支

为有选择地整合多层级特征与带边界增强的显著性图，我们提出了目标分割（OS）分支，其中包含特征聚合（FA）模块和特征交互（FI）模块。首先，使用FA模块聚合来自编码器和解码器的多层级特征，以生成互补特征。然后，FI模块采用反馈机制，将TSbranch生成的带边界增强的显著性图馈入上述聚合特征中，以恢复更多结构信息和细节信息。接着，为抑制上述交互特征中包含的冗余，我们使用一系列卷积层[2] 对其进行细化。最后，我们将上述三个步骤逐步级联三次，以学习更具判别性的特征并生成最终显著图。在每次级联中，原始真值都将用于监督FI模块的输出，从而引导网络学习标准的显著性特征。图4展示了FA模块和FI模块的一次级联过程。

1）特征聚合模块：考虑到不同层级的特征包含不同属性，FA模块旨在通过聚合高层特征的语义、低层特征的细节以及全局上下文特征，获得互补特征表示。全局上下文特征用于从图像的全局视野推断不同显著区域之间的相关性。因此，我们可以对具有较大感受野的顶层 F_5 进行编码，以生成全局上下文特征。具体而言，为了使后续模块在较低维度空间中运行，我们首先对顶层 F_5应用一个 1×1卷积层，以获得具有256个通道的压缩特征表示。然后，对上述压缩特征应用ASPP模块，以捕获多尺度上下文信息，记为F_h^5。最后，为恢复原始顶层的信息，我们对 F_h^5应用一个 1×1 卷积层，以获得具有2048个通道的特征，并以残差连接的方式从空间和通道维度为其分配上下文权重，从而抑制因通道增加而产生的冗余和噪声，并学习判别性特征。该过程可描述为：

$$F_h^5 = \eta(\delta(conv_1(F_5))), \quad (8)$$

$$\begin{aligned} \omega_s = \sigma(bn(conv_1(F_h^5))), \\ F_s^5 = \omega_s \odot F_5 + F_5, \end{aligned} \quad (9)$$

$$\begin{aligned} \omega_c=\sigma(conv_1(\delta(conv_1(gap(F_s^5))))), \\ F_{gc} = \omega_c \odot \delta(bn(conv_1(F_s^5))), \end{aligned} \quad (10)$$

其中，F_h^5 是经 ASPP模块处理的多尺度上下文特征，ω_s 和 F_s^5 表示空间权重和

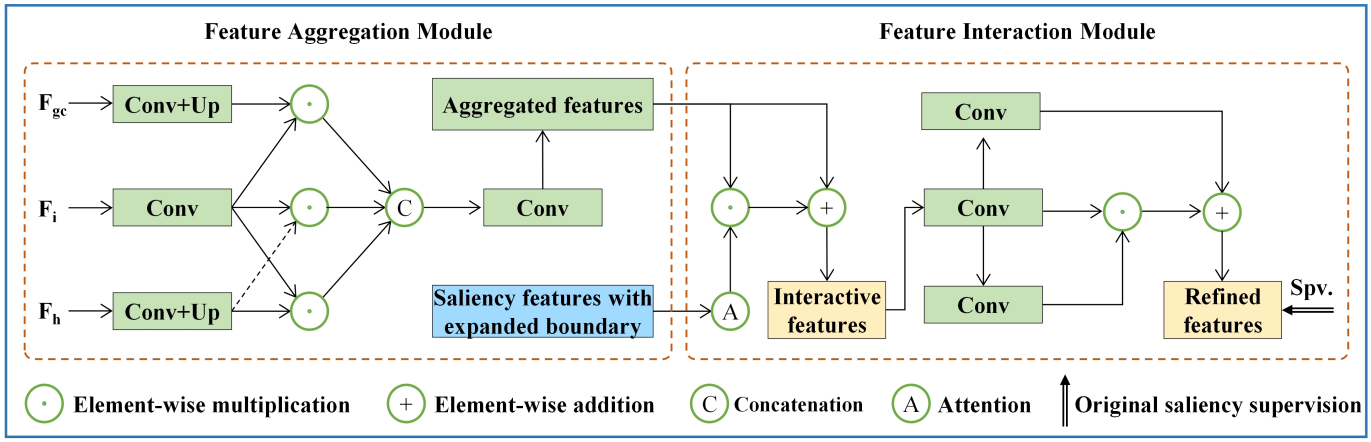


图 4：特征聚合模块和特征交互模块的一次级联过程。

加权特征， ω_c 和 F_{gc} 表示通道权重和最终的全局上下文特征， $conv_1$ 表示 1×1卷积层， η 是 ASPP模块， bn 表示批归一化层， σ 表示 Sigmoid 函数， $\odot$ 表示逐元素乘法， gap 表示全局平均池化， δ 表示 ReLU 激活函数。

基于上述全局上下文特征，我们可以将其与高层特征和低层特征进行聚合。具体而言，如图 4 所示，FA 模块首先接收三个输入，包括前一阶段解码输出的高层特征 $\{F_h^i|i=5,4,3\}$、骨干网络提供的低层特征 $\{F_i|i=4,3,2\}$，以及全局上下文特征 F_{gc}。需要注意的是，高层特征有两种形式：一种是多尺度上下文特征 $\{F_h^i|i=5\}$，另一种是$\{F_h^i|i=4,3\}$，由下文将介绍的特征交互模块生成。换言之，FA 模块输出的聚合特征将输入 FI 模块，与边界扩展后的显著性特征进行交互。FI 模块输出的交互特征将在下一阶段作为高层特征聚合到 FA 模块中。接下来，将上述三部分拼接，并通过一系列卷积和上采样操作，得到聚合特征 $\{F_{fa}^i|i=4,3,2\}$，即 FA 模块的最终输出：

$$\begin{aligned}F_{fa}^i = conv(cat(&conv(F_i)\odot up(conv(F_{gc})),\\ &conv(F_i)\odot up(conv(F_h^{i+1}))),\\ &conv(F_i))\odot conv(F_h^{i+1})))\ \ i=4,3,2,\end{aligned}\tag{11}$$

其中， $conv$ 表示一个 3×3 卷积层，后接一个批归一化层和 ReLU 激活函数；$\odot$ 表示逐元素乘法； up 表示双线性插值操作； cat 表示拼接操作。

2）特征交互模块：为生成完整且精细的显著性特征，我们提出了 FI 模块，将带扩展边界的显著性特征（SF w/ EB）输入由 FA 模块生成的聚合特征中。实际上，由于多次下采样操作，图像某些区域的细节会在网络前向传播过程中丢失。即使使用 FA 模块生成互补特征，仍然难以获得杂乱背景中的完整显著目标，尤其是在干扰因素众多且杂乱无序的场景中。考虑到 SF w/ EB 能够扩大显著目标与背景之间的细节差异，我们将其与聚合特征进行交互，帮助网络获得完整预测。此外，为了抑制交互特征带来的冗余和噪声（例如不必要的背景信息），我们对其应用一组加权卷积层。值得注意的是，FI 模块的输出将由原始真值进行监督，因此无需担心 SF w/ EB 会在学习过程中干扰模型。

具体而言，我们首先将坐标注意力 [38] 应用于带扩展边界的显著性特征，以捕获空间方向和位置信息，从而增强感兴趣对象的表示并抑制噪声。随后，通过带残差连接的逐元素乘法，将加权后的带扩展边界的显著性特征融入聚合特征中。上述交互特征具有语义和结构信息一致的特性。接下来，受自细化模块 [22],的启发，我们使用一系列加权卷积层来细化交互特征。最后，将 FA 模块和 FI 模块级联三次，以生成三个层级的特征图 $\{F_h^i|i=4,3,2\}$，然后使用一个 3×3 卷积层、双线性插值和 Sigmoid 函数对多尺度上下文特征 $\{F_h^i|i=5\}$ 和 $\{F_h^i|i=4,3,2\}$ 进行处理，以预测显著图 $\{P_i|i=5,4,3,2\}$，这些显著图是 OS 分支的最终输出。该过程可描述为：

$$F_i^i = coa(conv(F_e^i))\odot F_{fa}^i + F_{fa}^i\ \ i=4,3,2,\tag{12}$$

$$\begin{aligned}F_h^i = \delta(&conv_3(conv(F_i^i))\odot conv(F_i^i)+\\ &conv_3(conv(F_i^i))\ \ i=4,3,2,\end{aligned}\tag{13}$$

$$P_i = \sigma(up(conv_3(F_h^i)))\ \ i=5,4,3,2,\tag{14}$$

其中， F_e^i 表示带扩展边界的显著性特征， F_i^i 表示交互特征，F_h^i 表示细化特征，即上述 FA 模块中所述的高层特征，coa 表示坐标注意力， $conv3$ 表示一个 3×3 卷积层， $conv$ 表示一个后接批归一化层和

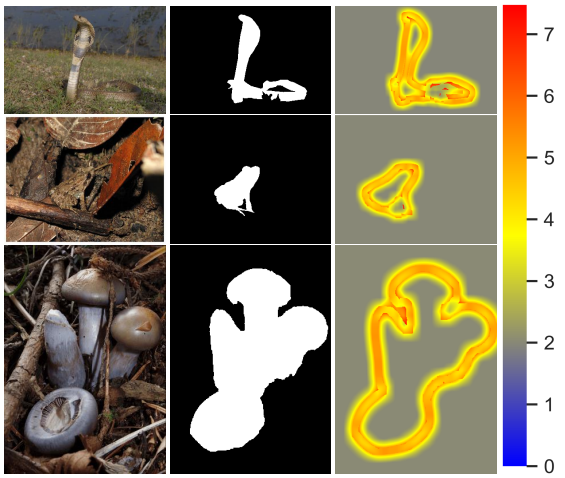


图 5： ω分布的一些视觉示例。我们观察到，显著目标与背景之间边界区域中的像素得到了更多关注，这有助于引导网络识别难以区分的区域，进而提高预测的完整性。

ReLU 激活函数的卷积层，$\odot$ 表示逐元素乘法， δ 表示 ReLU 激活函数， σ 表示 Sigmoid 函数。

E. 层次差异感知损失

为了进一步提高显著目标的完整性，我们提出层次差异感知（HDA）损失，用于监督 OS 分支的输出。在显著目标检测中，二元交叉熵（BCE）和交并比（IoU）是最常用的损失函数。然而，这些函数将每个像素视为彼此独立且同等重要，因此可能会因边界像素与非边界像素分布不平衡而产生模糊边界和不完整预测。受加权像素位置损失 [12], 的启发，我们提出 HDA 损失，引导网络更加聚焦于前景和背景之间的边界区域，从而缓解边界像素与非边界像素的不平衡分布。与以往为边界附近像素赋予较高权重的损失函数不同，所提出的 HDA 损失根据像素与边界之间的距离分级赋予权重：靠近边界的像素赋予较高权重，边界区域中的像素赋予中等权重，远离边界的像素赋予较低权重，如图 5 所示。因此，边界区域中外观相似的像素会得到更多关注，这不仅有助于改善边界细节，也有利于保持全局结构的完整性。

具体而言，所提出的 HDA 损失会为二元交叉熵和交并比损失函数中的每个像素赋予权重 ω_{ij} 。首先，我们对原始真实值（GT）进行膨胀和腐蚀操作，以获得膨胀真实值和腐蚀真实值。然后，分别根据像素与膨胀真实值和腐蚀真实值中其邻域像素之间的差异，计算两种权重，以表示每个像素的重要性。如果某个像素与膨胀/腐蚀真实值中周围像素的差异更大，则会被赋予更高的权重。接下来，我们将上述两种权重相加，以进一步强调靠近边界的像素的重要性。由于它们的边界区域存在交集，两个边界区域交集中的像素会被计算两次，因此会被赋予更高的权重。 ω_{ij} 定义见式 15、式 16 和式 17。

$$\omega_{ij}^{d}=\left|\sum_{x,y\in R_{ij}}[\mathrm{Dilate}(GT_{xy},k)^{n}]-\mathrm{Dilate}(GT_{ij},k)^{n}\right|, \tag{15}$$

$$\omega_{ij}^{e}=\left|\sum_{x,y\in R_{ij}}[\mathrm{Erode}(GT_{xy},k)^{n}]-\mathrm{Erode}(GT_{ij},k)^{n}\right|, \tag{16}$$

$$\omega_{ij}=\omega_{ij}^{d}+\omega_{ij}^{e}, \tag{17}$$

其中， ω_{ij} 是膨胀真实值和腐蚀真实值中像素位置差异的总和，即 ω_{ij}^{d} 和 ω_{ij}^{e}。 R_{ij} 表示以像素 (i, j) 为中心、尺寸为 30×30 的区域。Dilate(.) 和 Erode (.) 分别表示膨胀和腐蚀操作，卷积核 k 的尺寸为 3×3，迭代次数 n 根据经验设为 5。基于所提出的 ω_{ij}，加权二元交叉熵和交并比损失函数可定义为：

$$L_{bce}^{\omega}=\frac{-\sum_{i}^{H}\sum_{j}^{W}\left[\begin{array}{l}(1+\lambda\cdot\omega_{ij})\cdot((GT_{ij}\cdot P_{ij})+\\(1-GT_{ij})\cdot\log(1-P_{ij}))\end{array}\right]}{\sum_{i}^{H}\sum_{j}^{W}[(1+\lambda\cdot\omega_{ij})]}, \tag{18}$$

$$L_{iou}^{\omega}=1-\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\lambda\cdot\omega_{ij})\cdot(GT_{ij}\cdot P_{ij})}{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\lambda\cdot\omega_{ij})\cdot(GT_{ij}+P_{ij}-GT_{ij}\cdot P_{ij})}, \tag{19}$$

其中， λ 设为 5，以偏重最终结果。基于上述定义，所提出的层次差异感知损失 L_{hda} 如式 20 所示，同时考虑了局部结构和全局结构。

$$L_{hda}=\mathrm{Mean}(L_{bce}^{\omega}+L_{iou}^{\omega}), \tag{20}$$

其中，Mean(.) 返回数组元素的平均值。我们采用多层级监督，通过所提出的损失函数优化 OS 分支，该损失函数是预测结果 $\{P_i|i=2,3,4,5\}$和 GT之间四项损失之和，如式 21 所示。

$$L_{f}=\sum_{i=2}^{5}\alpha_{i}\cdot L_{hda}(P_{i},GT), \tag{21}$$

其中， α_i 表示不同损失的权重，从 α_2 到 α_5分别设为 1、0.8、0.6 和 0.4。由于网络的高层特征误差较大，

根据经验，它们应采用较小的权重。整体损失 L通过端到端方式进行最小化，其中包括 TS分支和 OS 分支的损失。

$$L = L_c + \beta \cdot L_f, \tag{22}$$

其中，将 β 设为 1，使上述两种损失的取值范围具有可比性。

IV. 实验

本节首先介绍所用的数据集和评估指标，然后说明实现细节。最后，我们展示大量实验结果，以证明所提方法的有效性。

A. 数据集

为评估模型的性能，我们在五个基准数据集上进行实验。其中，EC- SSD [42]包含 1000 张图像，这些图像语义信息丰富，但物体结构复杂。HKU-IS [43] 由 4,447 个复杂场景组成，其中包含多个彼此分离的物体，且前景和背景的外观相似。PASCAL-S [44]包含 850 张具有挑战性的图像，这些图像是从 PASCAL VOC [45]验证集中精心挑选的。DUT-OMRON [46]包含 5,168 张高质量图像，其背景相对杂乱，内容多样性高。DUTS [47]是目前规模最大的显著性目标检测数据集，包含 10,553 张训练图像（DUTS-TR）和 5,019 张测试图像（DUTS- TE）。训练集和测试集都包含复杂场景。我们使用 DUTS-TR 作为训练数据集，其余数据集作为测试数据集。

B. 评估指标

为评估我们方法的性能，我们采用五种广泛使用的指标进行全面比较。

精确率-召回率（PR）曲线根据二值显著图和真实标注计算得出:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{23}$$

其中，TP、TN、FP、FN分别表示真正例、真负例、假正例和假负例。我们使用一组从 0 到 255 的阈值对预测进行二值化。对于每个阈值，都可以得到一组精确率/召回率，并据此绘制 PR 曲线。

平均绝对误差（MAE）用于衡量预测与真实标注之间的平均 L1 距离:

$$\text{MAE} = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} |P(i,j) - GT(i,j)|, \tag{24}$$

其中， P 是预测显著图， GT 是真实标注。

F值 通过计算加权调和均值，同时考虑精确率和召回率:

$$F_\beta = \frac{(1+\beta)^2 \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}, \tag{25}$$

其中， β 设为 0.3，以强调精确率而非召回率。我们绘制完整的 F值曲线，并分别报告最高 F值分数（F_β）。

结构相似度（S_α） [48]用于评估预测显著图与真实标注之间的结构相似度。 S_α分别考虑目标感知（S_o）和区域感知（S_r）结构相似度:

$$S_m = \alpha \cdot S_o + (1-\alpha) \cdot S_r, \tag{26}$$

其中， α 设为 0.5，以平衡 S_o 和 S_r。

C. 实现细节

我们使用 Pytorch 1.3 实现所提出的模型。骨干网络参数使用在 ImageNet [49],上预训练的 ResNet-50 初始化，其他参数则随机初始化。训练时，我们将每张图像调整为 320×320 ，并进行随机水平翻转和隐藏块 [50], ，随后随机裁剪出大小为 288×288的图像块。骨干网络的最大学习率设为 5e-3，其他部分设为 5e-2。我们采用预热和线性衰减策略调整学习率。整个网络通过随机梯度下降进行端到端训练。动量和权重衰减分别设为 0.9 和 5e-4。批量大小设为 64，最大训练轮次设为 36。测试时，仅将每张图像调整为320×320，然后输入网络进行预测。我们不使用多尺度训练或任何后处理步骤（例如 CRF [51]）。

D. 与最先进方法的比较

我们将所提出的方法与 15 种基于 ResNet骨干网络的最先进显著目标检测方法进行比较 [33], 包括 CPD-R [34], BASNet [39], PoolNet [26],SIBI [30], CAGNet[40], ITSD [9], MINet [24],GateNet [41], GCPA [22], F3Net[12], PurNet [5], GFINet [14], MSFNet [8],DCN [6] 和 ICON [7]。为确保比较公平，我们使用相同的代码评估作者提供的最佳显著图。

1）定量比较：表I展示了在五个基准数据集上，采用三种评估指标进行定量比较的结果。可以看出，在大多数数据集上，我们的方法在三项指标上始终优于其他竞争方法。与其他方法相比，尽管它们在某个数据集的某项指标上表现出优势（例如，MSFNet [8] 在平均绝对误差方面），我们的方法具有更好的泛化能力：在五个数据集的三项指标上，大多数情况下排名第一，少数情况下排名第二。由于 F_β 对显著图中的各种缺陷都很敏感， S_m 用于评估预测显著图的结构完整性，而平均绝对误差能够反映背景噪声的强度，这三项指标可以全面反映一种方法能否在背景噪声较低的前提下实现相对完整的预测。与基于多层特征聚合方法的ICON [7] 、引入边缘信息的GFINet [14] 以及同时引入边缘和骨架信息的DCN [6]相比，我们的方法在以下方面的性能平均分别提升了2.1%、1.9%和0.3%

表 I：在五个数据集上与 14 种最先进方法的性能比较。采用最大 F值（F_β，越大越好）、S值（S_m，越大越好）和平均绝对误差（越小越好）来衡量模型性能。最佳的三项结果分别以红色和绿色标出。

方法	HKU-IS（编号 4447）			DUTS-TE（编号 5019）			DUT-OMRON（编号 5168）			PASCAL-S（编号 850）			ECSSD（编号 1000）		
	F_β	S_m	MAE	F_β	S_m	MAE	F_β	S_m	MAE	F_β	S_m	MAE	F_β	S_m	MAE
CPD-R [34]	0.925	0.906	0.034	0.865	0.869	0.043	0.797	0.825	0.057	0.864	0.842	0.072	0.939	0.918	0.037
BASNet [39]	0.928	0.909	0.032	0.859	0.866	0.048	0.805	0.836	0.057	0.857	0.832	0.076	0.943	0.916	0.037
PoolNet [26]	0.933	0.917	0.032	0.880	0.883	0.040	0.808	0.836	0.056	0.869	0.845	0.074	0.944	0.921	0.039
SIBI [30]	0.931	0.913	0.032	0.872	0.879	0.040	0.803	0.832	0.059	0.870	0.848	0.070	0.946	0.924	0.035
CAG-R [40]	0.926	0.904	0.030	0.866	0.864	0.040	0.791	0.815	0.054	0.866	0.836	0.066	0.937	0.908	0.037
GateNet [41]	0.934	0.915	0.033	0.887	0.885	0.040	0.818	0.838	0.055	0.945	0.920	0.040	0.875	0.852	0.068
ITSD [9]	0.934	0.917	0.031	0.883	0.885	0.041	0.821	0.840	0.061	0.876	0.856	0.064	0.947	0.925	0.035
MINet [24]	0.935	0.919	0.029	0.884	0.884	0.037	0.810	0.833	0.055	0.873	0.851	0.064	0.948	0.925	0.034
GCPA [22]	0.938	0.920	0.031	0.888	0.891	0.038	0.812	0.839	0.056	0.876	0.861	0.061	0.949	0.927	0.035
F3Net [12]	0.937	0.917	0.028	0.891	0.888	0.036	0.813	0.839	0.053	0.878	0.855	0.062	0.945	0.924	0.033
PurNet [5]	0.935	0.915	0.031	0.878	0.881	0.039	0.814	0.841	0.051	0.873	0.843	0.069	0.945	0.925	0.035
MSFNet [8]	0.927	0.907	0.027	0.877	0.875	0.034	0.799	0.820	0.046	0.870	0.845	0.061	0.941	0.914	0.033
GFINet [14]	0.939	0.921	0.028	0.890	0.889	0.038	0.823	0.843	0.054	0.877	0.856	0.064	0.948	0.929	0.032
DCN [6]	0.939	0.922	0.027	0.894	0.891	0.035	0.823	0.845	0.051	0.878	0.857	0.062	0.952	0.928	0.032
ICON [7]	0.939	0.920	0.029	0.892	0.889	0.037	0.825	0.844	0.057	0.881	0.857	0.063	0.950	0.929	0.032
SENet（我们的方法）	0.941	0.922	0.027	0.899	0.896	0.034	0.827	0.847	0.051	0.882	0.859	0.062	0.950	0.929	0.032

在五个数据集上的 F_β、 S_m 和平均绝对误差方面。尤其是在规模较大的数据集（DUT-OMRON [46] 和 DUTS-TE [47]）上，这些数据集包含杂乱背景以及具有复杂结构的多个显著目标；与上述三种方法相比，我们的方法优势更加明显，模型性能在 F_β、 S_m 和平均绝对误差方面平均分别提升了3.9%、3.2%和0.7%。上述三项指标的持续提升表明，我们的方法能够在背景噪声较低的前提下有效提高预测的完整性。

此外，我们绘制了上述方法的精确率-召回率曲线和F度量曲线，以进行整体比较，如图6所示。从这些曲线可以看出，在大多数情况下，不同阈值下我们的方法均优于其他竞争方法，这与表I中报告的指标一致。

此外，我们在 DUT- OMRON [46]数据集上，将本方法的推理速度和边界精度与最先进的模型进行了比较。采用平均绝对误差$_b$ [14]评估边界质量，该指标表示距边界 10 个像素范围内区域（包括边界内外区域）的平均绝对误差。所有模型均采用 ResNet-50 作为编码器，其解码器由多个分支组成，包括用于预测显著目标的分支，以及用于预测边界或骨架的分支。为确保比较公平，我们使用相同的代码和环境测试这些模型。如表 II 所示，输入图像为 320×320 时，我们的 SENet 在单块 2080Ti GPU 和 Intel Xeon CPU E5-2678 v3 @ 2.50GHz 上的运行速度为 39.5 帧每秒。与同时引入边缘信息和骨架信息的 DCN [6]方法相比，我们的 SENet 在帧每秒、平均绝对误差$_b$ 和 F_β方面均表现更优。与引入边缘信息的 GFINet [14] 方法相比，我们的模型在平均绝对误差$_b$方面表现出色。虽然本方法的速度略慢于 GFINet [14]，但其预测的完整性更好。在解码过程中，我们的模型逐步将 TSbranch 生成的扩展显著性特征馈入 OSbranch，这也符合人类视觉系统的工作方式。与其他同类方法相比，我们的模型不仅边界精度和整体精度更高，而且效率也很高。

表II：DUT-OMRON数据集上不同显著目标检测方法的推理速度和准确率比较 [46] 。所有方法均采用Res Net-50作为编码器，其解码器由多个用于预测的树枝组成。排名前三的结果以红色和绿色标出。

方法	编码器	输入尺寸	FPS	MAE$_b$	F_β
EGNet$_{19}$ [28]	ResNet-50	380×320	12.4	0.022	0.815
SIBI$_{19}$ [30]	ResNet-50	400×300	15.1	0.023	0.803
SCRN$_{19}$ [32]	ResNet-50	352×352	37.1	0.024	0.812
ITSD$_{20}$ [9]	ResNet-50	288×288	38.4	0.024	0.821
DCN$_{21}$ [6]	ResNet-50	352×352	31.0	0.020	0.823
GFINet$_{21}$ [14]	ResNet-50	320×320	46.5	0.019	0.823
SENet（我们的方法）	ResNet-50	320×320	39.5	0.019	0.827

2）定性比较：为了进一步说明我们模型的优势，我们展示了 11 种复杂场景下的视觉比较，如图 7 所示。从上到下依次为视野外、大型物体、小型物体、多物体、细长物体、低对比度、形状复杂度、杂乱背景、遮挡、反射和前景干扰。这些场景具有一些共同特征，包括杂乱背景、低对比度，以及前景和背景之间颜色或纹理相似。可以观察到，所提出的方法在检测和分割显著目标方面表现更好，不仅能获得完整的显著图

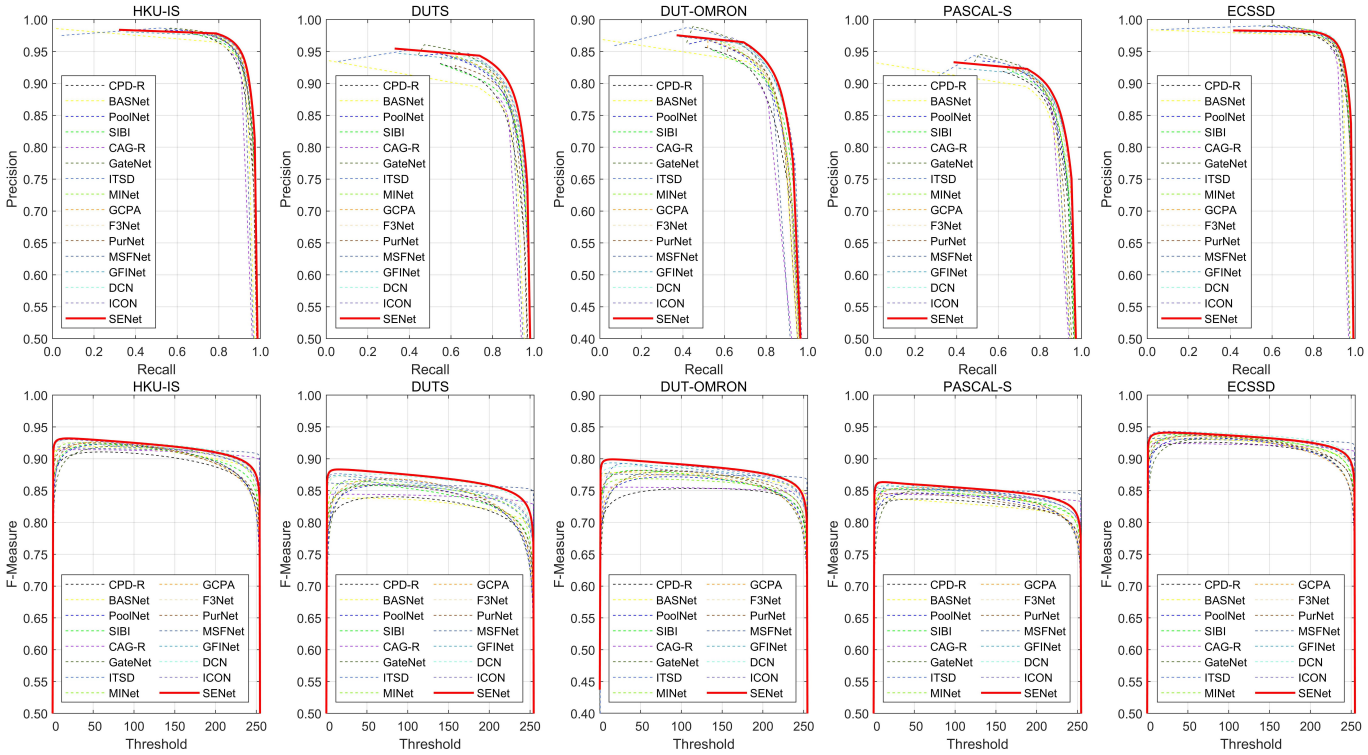


图 6：与 15 种最先进方法在五个数据集上的性能比较。第一行展示精确率-召回率曲线的比较，第二行展示不同阈值下 F 值曲线的比较。

，还能有效抑制背景噪声，例如图像边界处的汽车（第 1 行）、道路上的公共汽车（第 2 行）、云中的鸟类（第 3 行）、展厅里的支架（第 5 行）、叶片上的青蛙（第 6 行）、树枝上的花朵（第 7 行）、灌木丛中的蘑菇（第 8 行）以及草坪周围的建筑物（第 9 行）。这说明了所提出的先分离后分割策略的强大之处，该策略能够利用扩展后的显著性特征引导网络获得完整且噪声较少的显著图。

E. 消融研究

为探究所提出模型中不同组件的有效性，我们在基准数据集上进行了一系列实验，以比较模型在不同设置下的性能变化。

1）关键组件的有效性：为了探究所提出的不同组件的有效性，我们构建了四种模型进行比较。第一种设置为基线模型，我们移除了目标分离（TS）分支，并将所提出的分层差异感知（HDA）损失替换为二元交叉熵（BCE）与交并比（IoU）的组合，记为“基线”。为验证由 TS 分支生成的扩展边界显著性特征（SF w/ EB）和 HDA 损失的有效性，我们分别在第二种和第三种模型中加入 TS 分支（记为“基线 + T”）和 HDA（记为“基线 + H”）。最后一种设置是我们的模型，包含了所有提出的模块，记为“基线 + T+ H”。这四种模型的参数保持一致。

上述模型的比较结果列于表 III。具体而言，在两个大型数据集上的 F_β、 S_m 和平均绝对误差（MAE）方面，“基线 +H”的平均性能额外提升了 3.2%，而“基线+ T”的平均性能额外提升了 4.6%。此外，我们观察到“基线 + T” 的性能提升优于 “ 基线 + H”，这表明扩展边界显著性特征对于我们的模型获得完整预测十分重要。如表 III 所示，我们的整体模型（即“基线 + T +H”）取得了更好的性能，在 F_β、 S_m 和平均绝对误差（MAE）方面平均额外提升了 7%，验证了所提出的扩展边界显著性特征（SF w/ EB）和 HDA 损失的有效性。

此外，按照 [52], 的建议，我们计算显著图中高亮像素的平均分数（ASHP）。给定一个预测显著图 P，ASHP 可表示为:

$$\text{ASHP} = \frac{1}{N_P}\sum_{i=1}^{H}\sum_{j=1}^{W}|P(i,j)|, \tag{27}$$

其中， N_P表示满足条件 $P(i,j)>0$的像素数量。然而，由于缺少原始真值作为衡量标准，ASHP 只能部分描述模型的显著性程度。我们的实验结果表明，当预测显著图包含更多噪声或输出尺寸较大时，ASHP 的值也会更高。因此，在模型的消融研究中，我们仅将 ASHP / 平均绝对误差（A/M）用作辅助指标，用于在一定程度上评估模型的噪声和显著性程度。如表 III 所示，与“基线”相比，“基线 + T + H”在两个大型数据集上的 A/M 平均额外提升了 20.4%。

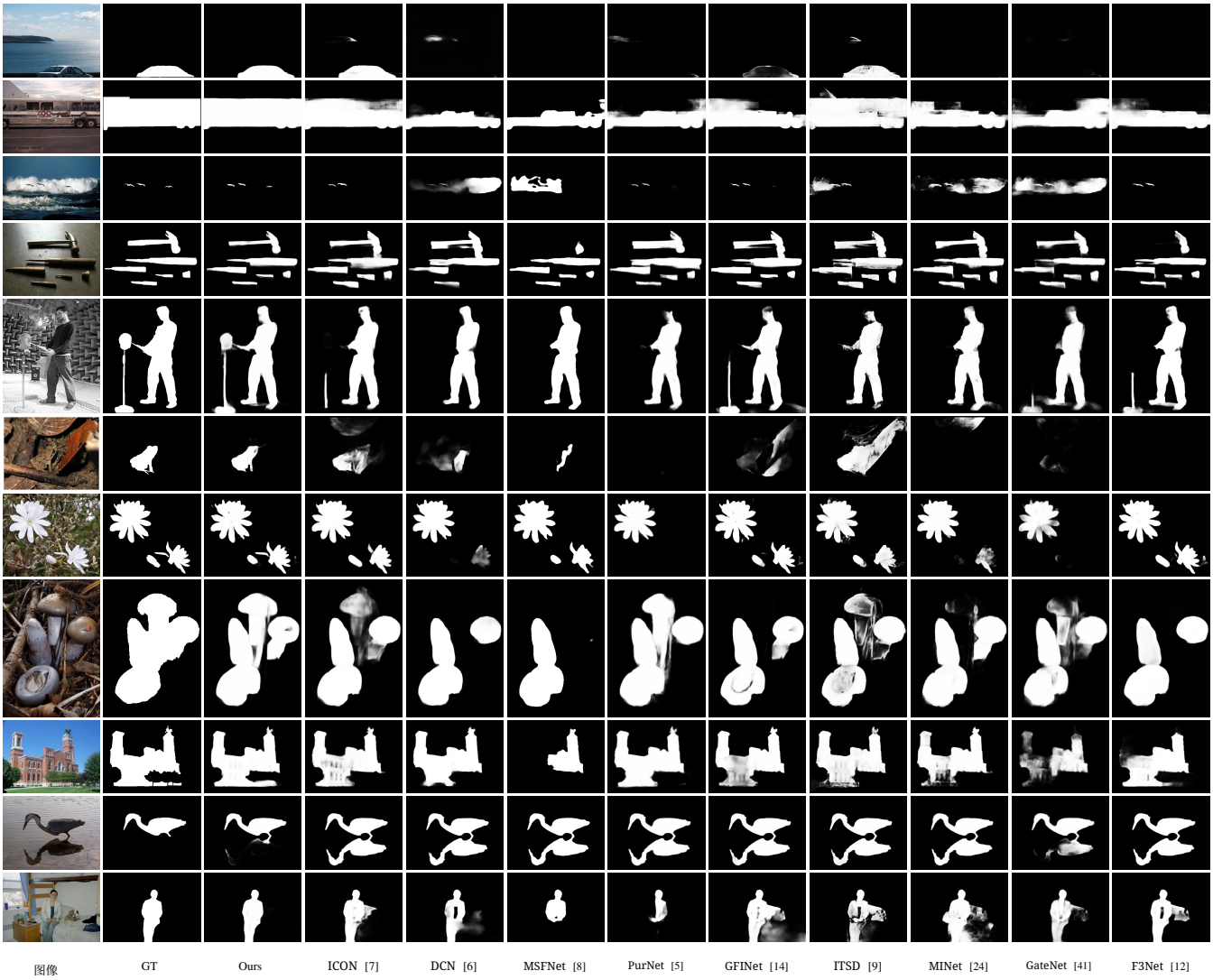


图 7：不同复杂场景下与最先进方法的定性比较。从上到下依次为视野外、大型物体、小型物体、多物体、细长物体、低对比度、形状复杂度、杂乱背景、遮挡、反射和前景干扰。

这表明，在背景噪声较低的前提下，我们的模型能够提高预测显著图的显著性程度。

表 III：在两个背景相对杂乱、内容多样性较高的大规模基准数据集上，对不同模块进行消融研究。Baseline、T 和 H 分别表示基线网络、由 TSbranch 生成的带 EB 的 SF，以及 HDA 损失。最佳结果以**粗体**标出。

模型	DUT-OMRON			DUTS-TE		
	F_β	S_m	平均绝对误差 A/M	F_β	S_m	平均绝对误差 A/M
基线	0.818	0.836	0.061 12.049	0.888	0.889	0.040 19.025
基线 + H	0.823	0.842	0.057 12.912	0.893	0.892	0.036 21.306
基线 + T	0.826	0.844	0.055 13.636	0.895	0.894	0.035 22.000
基线 +T +H	**0.827**	**0.847**	**0.051 14.569**	**0.899**	**0.896**	**0.034 22.853**

此外，由于“基线 + H”的改进是与“Baseline + T”相比结果较差，因此我们使用不同的损失函数训练其他模型 GCPA [22]，以进一步验证所提出的层次差异感知损失的效果。对比结果列于表 IV。如表 IV 所示，与其他边缘增强损失相比，所提出的层次差异感知损失取得了更好的性能。具体而言，就F_β、 S_m 和平均绝对误差而言，相比原始二元交叉熵和交并比损失函数，性能平均提升了 7.1%，这进一步证明了所提出的 HDA 损失的有效性。

此外，为了进一步验证由扩展真值监督生成的带扩展边界的显著性特征，我们在 TS 分支中用原始真值替换扩展真值，记为“Baseline + O”。这意味着“Baseline + O”将输出标准显著性特征。如表 V 所示，就 F_β、 S_m 和平均绝对误差而言，“Baseline + T”的性能平均比“Baseline + O”提升了 2.4%，这进一步证明了

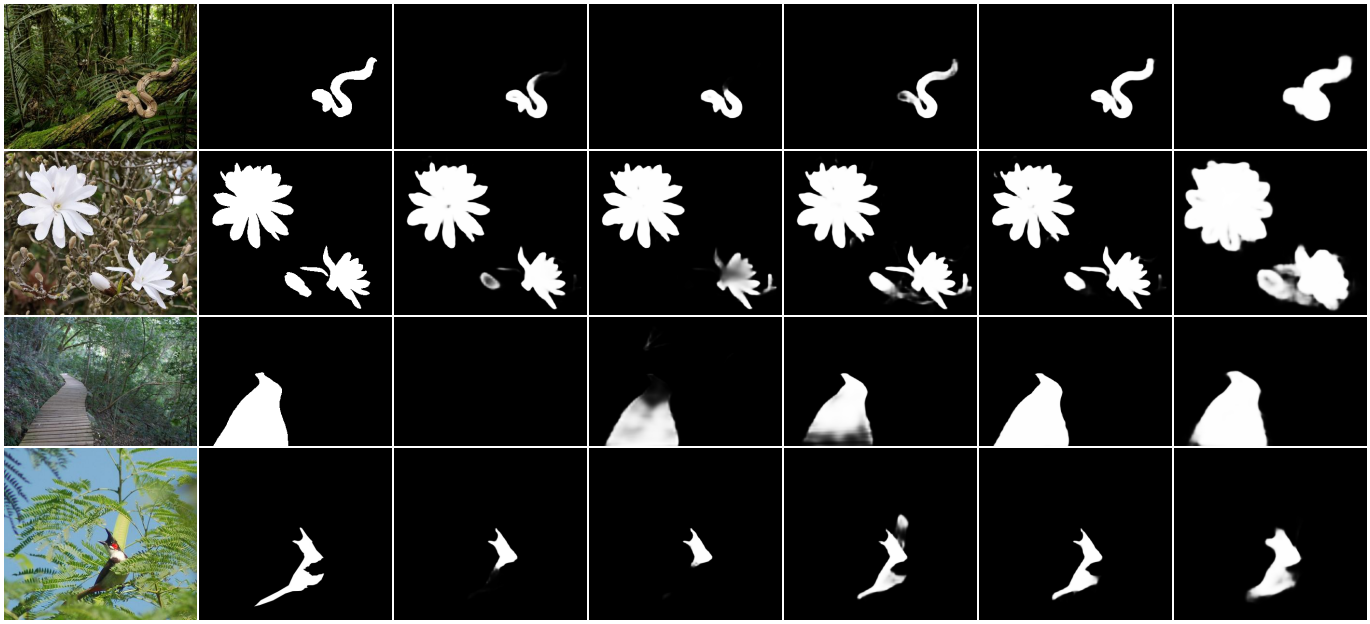


图8：展示所提出组件优势的视觉比较。基线+H：采用所提出的层次差异感知损失的基线预测结果。基线+T：采用TS分支的基线预测结果。基线+T+H：包含所有所提出组件的模型预测结果。C：TS分支的预测结果。

表IV：采用不同损失函数在其他模型GCPA [22]上训练得到的比较结果。最佳结果以**粗体**标出。

模型	DUT-OMRON			DUTS-TE		
	F_β	S_m	MAE	F_β	S_m	MAE
二元交叉熵 + 交并比	0.815	0.834	0.057	0.876	0.880	0.041
PPA [12]	0.816	0.835	0.055	0.886	0.885	0.039
BRA [14]	0.819	**0.841**	0.057	0.887	0.886	0.040
HDA（我们的方法）	**0.821**	**0.841**	**0.052**	**0.888**	**0.888**	**0.037**

提出的带扩展边界的显著性
特征。

表 V：不同监督信号生成的显著性特征对性能的影响。“Baseline +O”表示使用原始真值监督 TS 分支，“Baseline + T”表示使用所提出的扩展真值监督 TS 分支。最佳结果以**粗体**标出。

模型	DUT-OMRON			DUTS-TE		
	F_β	S_m	MAE	F_β	S_m	MAE
基线 +O	0.819	0.840	0.056	0.884	0.886	0.039
基线 + T	**0.826**	**0.844**	**0.055**	**0.895**	**0.894**	**0.035**

2）不同组件的视觉比较：图8展示了不同组件生成的预测结果的视觉比较。与“基线”相比，引入 TS 分支生成的带扩展边界的显著性特征后，“基线 + T”能够有效提升预测结果的完整性。加入 HDA 损失后，“基线 + T + H”不仅能够检测出完整的显著目标，还能更好地抑制噪声。此外，我们展示了 TS 分支预测的扩展显著目标（第7列）。尽管 TS 分支的预测结果包含一些噪声，但目标结构相对完整，这进一步验证了所提出模型的有效性。

V. 结论

本文分析了基于补充信息的方法难以从复杂场景中检测出完整显著目标的问题，尤其是在杂乱背景下。为解决这些问题，我们提出了一种名为 SENet 的新型网络，用于生成噪声较低的完整显著图。该网络首先采用分离扩展的方式处理显著目标，然后对精细目标进行分割。此外，为了识别前景和背景之间难以区分的区域，我们提出了一种分层差异感知损失，根据像素与边界的距离将其划分为不同等级并赋予相应权重，从而进一步提升预测的结构完整性和局部细节。与 15 种最先进方法相比，其中包括基于聚合多层特征以及引入边缘或骨架信息的方法，在五个基准数据集上的实验结果表明，所提模型的性能优于这些方法。此外，我们的模型效率更高，在单块 NVIDIA RTX 2080Ti GPU 上的推理速度达到 39.5 帧每秒。

致谢

本研究获国家自然科学基金（NSFC）项目（批准号：62172243）和国家重点研发计划项目（批准号：

2020年YFB1710200），以及中国黑龙江省自然科学基金项目（批准号：LH2022F044）资助。

参考文献

[1] A. Borji、M. Cheng、Q. Hou、H. Jiang 和 J. Li，《显著目标检测：综述》，《计算机视觉与媒体》，第5卷，150，2019年2，117期，第–页。[2] F. Zhang、B. Du 和 L. Zhang，《显著性引导的无监督特征学习用于场景分类》，《IEEE地球科学与遥感汇刊》，第53卷，第4期，第2175–2184页，2015年。[3] S. Hong、T. You、S. Kwak 和 B. Han，《通过卷积神经网络学习判别式显著图进行在线跟踪》，载于国际机器学习会议，《机器学习研究期刊》研讨会与会议论文集，第37卷。JMLR.org，2015年，第597–606页。[4] Y. Gao、M. Shi、D. Tao 和 C. Xu，《用于快速图像检索的数据库显著性》，《IEEE多媒体汇刊》，第17卷，第3期，第359–369页，2015年。[5] J. Li、J. Su、C. Xia、M. Ma 和 Y. Tian，《采用净化机制和结构相似性损失的显著目标检测》，《IEEE图像处理汇刊》，第30卷，第6855–6868页，2021年。[6] Z. Wu、L. Su 和 Q. Huang，《用于显著目标检测的分解与补全网络》，《IEEE图像处理汇刊》，第30卷，第6226–6239页，2021年。[7] M. Zhuge、D.-P. Fan、N. Liu、D. Zhang、D. Xu 和 L. Shao，《通过完整性学习进行显著目标检测》，《IEEE模式分析与机器智能汇刊》，2022年。[8] M. Zhang、T. Liu、Y. Piao、S. Yao 和 H. Lu，《Auto-msfnet：用于显著目标检测的多尺度融合网络搜索》，载于 MM ’21：ACM。ACM，2021年，第667–676页。[9] H. Zhou、X. Xie、J. Lai、Z. Chen 和 L. Yang，《用于精准快速显著性检测的交互式双流解码器》，载于 CVPR。IEEE，2020年，第9138–9147页。[10] J. Wei、S. Wang、Z. Wu、C. Su、Q. Huang 和 Q. Tian，《用于显著目标检测的标签解耦框架》，载于 CVPR。IEEE，2020年，第13 022–13 031页。[11] J. Liu、Q. Hou 和 M. Cheng，《动态特征整合：同时检测显著目标、边缘和骨架》，《IEEE图像处理汇刊》，第29卷，第8652–8667页，2020年。[12] J. Wei、S. Wang 和 Q. Huang，《F^3net：显著目标检测的融合、反馈与聚焦》，载于AAAI。AAAI 出版社，2020 年，第 12 321–12 328 页。[13] Z. Chen 、H. Zhou、J. Lai、L. Yang 和 X. Xie，《轮廓感知损失：用于显著目标分割的边界感知学习》，《IEEE图像处理汇刊》，第30卷，第431–443页，2021年。[14] G. Zhu、J. Li 和 Y. Guo，《补充与抑制：边界和非边界都有助于显著目标检测》，《IEEE神经网络与学习系统汇刊》，2021年。[15] J. Wang、H. Jiang、Z. Yuan、M. Cheng、X. Hu 和 N. Zheng，《显著目标检测：一种判别式区域特征整合方法》，《国际计算机视觉杂志》，第123卷，第2期，第251–268页，2017年。[16] R. Achanta、S. S. Hemami、F. J. Estrada 和 S. S¨usstrunk，《频率调谐的显著区域检测》，载于2009年IEEE计算机学会计算机视觉与模式识别会议（CVPR 2009），2009年6月20日至25日，美国佛罗里达州迈阿密。IEEE计算机学会，2009年，第1597–1604页。[17] M. Cheng、N. J. Mitra、X. Huang、P. H. S. Torr 和 S. Hu，《基于全局对比度的显著区域检测》，《IEEE模式分析与机器智能汇刊》，第37卷，第3期，第569–582页，2015年。[18] Q. Hou、M. Cheng、X. Hu、A. Borji、Z. Tu 和 P. H. S. Torr，《通过短连接实现深度监督显著目标检测》，《IEEE模式分析与机器智能汇刊》，第41卷，第4期，第815–828页，2019年。[19] L. Zhang、J. Dai、H. Lu、Y. He 和 G. Wang，《用于显著目标检测的双向消息传递模型》，载于 CVPR。IEEE计算机学会，2018年，第1741–1750页。[20] Z. Deng、X. Hu、L. Zhu、X. Xu、J. Qin、G. Han 和 P. Heng，《R^3net：用于显著性检测的循环残差细化网络》，载于国际人工智能联合会议。ijcai.org，2018年，第684–690页。[21] N. Liu 、J. Han 和 M. Yang，《Picanet：用于精准显著性检测的逐像素上下文注意力学习》，《IEEE图像处理汇刊》，第29卷，第6438–6451页，2020年。[22] Z. Chen、Q. Xu、R. Cong 和 Q. Huang，《用于显著目标检测的全局上下文感知渐进聚合网络》，载于 AAAI。AAAI出版社，2020年，第10 599–10 606页。[23] B. Wang、Q. Chen、M. Zhou、Z. Zhang、X. Jin 和 K. Gai，《用于显著目标检测的渐进式特征精炼网络》，载于 AAAI。AAAI出版社，2020年，第12 128–12 135页。

[24] Y. Pang、X. Zhao、L. Zhang 和 H. Lu，《多尺度交互网络用于显著目标检测》，载于 CVPR。IEEE，2020年，第 9410–9419 页。[25] M. Ma、C. Xia 和 J. Li，“用于显著目标检测的金字塔特征收缩”，载于 AAAI。AAAI出版社，2021年，第 2311–2318 页。[26] J. Liu、Q. Hou、M. Cheng、J. Feng 和 J. Jiang，“一种用于实时显著目标检测的简单池化设计”，载于 CVPR。计算机视觉基金会 / IEEE，2019年，第 3917–3926 页。[27] T. Zhao 和 X. Wu，“ 用于显著性检测的金字塔特征注意力网络”，载于 CVPR。计算机视觉基金会 / IEEE，2019年，第 3085–3094 页。[28] J. Zhao、J. Liu、D. Fan、Y. Cao、J. Yang 和 M. Cheng，“Egnet：用于显著目标检测的边缘引导网络”，载于 ICCV。IEEE，2019年，第 8778–8787 页。[29] M. Feng、H. Lu 和 E. Ding，“用于边界感知显著目标检测的注意力反馈网络”，载于 CVPR。计算机视觉基金会 / IEEE，2019年，第 1623–1632 页。[30] J. Su、J. Li、Y. Zhang、C. Xia 和 Y. Tian，“选择性还是不变性：边界感知显著目标检测”，载于 ICCV。IEEE，2019年，第 3798–3807 页。[31] Z. Zhao、C. Xia、C. Xie 和 J. Li，“用于快速准确显著目标检测的互补三边解码器”，载于 MM ’21：ACM。ACM，2021年，第 4967–4975 页。[32] Z. Wu、L. Su 和 Q. Huang，“用于边缘感知显著目标检测的堆叠交叉细化网络”，载于 ICCV。IEEE，2019年，第 7263–7272 页。[33] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，载于 CVPR。IEEE 计算机学会，2016年，第 770–778 页。[34] Z. Wu、L. Su 和 Q. Huang，“用于快速准确显著目标检测的级联部分解码器”，载于 CVPR。计算机视觉基金会 / IEEE，2019年，第 3907–3916 页。[35] L. Chen、G. Papandreou、F. Schroff 和 H. Adam，“重新思考用于语义图像分割的空洞卷积”，CoRR，卷 abs/1706.05587，2017年。[36] S. Liu、D. Huang 和 Y. Wang，“用于准确快速目标检测的感受野模块网络”，载于 ECCV，丛书《计算机科学讲义》，卷 11215。施普林格，2018 年，第 404–419 页。[37] J. Hu、L. Shen 和 G. Sun，“压缩与激励网络”，载于 CVPR。计算机视觉基金会 / IEEE计算机学会，2018年，第 7132–7141 页。[38] Q. Hou、D. Zhou 和 J. Feng，“用于高效移动网络设计的坐标注意力”，载于 CVPR。计算机视觉基金会 / IEEE，2021年，第 13 713–13 722 页。[39] X. Qin、Z. V. Zhang、C. Huang、C. Gao、M. Dehghan 和 M. J¨agersand，“Basnet：边界感知显著目标检测”，载于 CVPR。计算机视觉基金会 / IEEE，2019年，第 7479–7489 页。[40] S. Mohammadi、M. Noori、A. Bahri、S. G. Majelan 和 M. Havaei，“ Cagnet：用于显著目标检测的内容感知引导”，《模式识别》，卷 103，第 107303 页，2020年。[41] X. Zhao、Y. Pang、L. Zhang、H. Lu 和 L. Zhang，“抑制与平衡：一种用于显著目标检测的简单门控网络”，载于 ECCV (2)，丛书《计算机科学讲义》，卷 12347。施普林格，2020年，第 35–51 页。[42] Q. Yan、L. Xu、J. Shi 和 J. Jia，“层次显著性检测”，载于 CVPR。IEEE计算机学会，2013年，第 1155–1162 页。[43] G. Li 和 Y. Yu，“基于多尺度深度 CNN 特征的视觉显著性检测”，《IEEE图像处理汇刊》，卷 25，第 11 期，第 5012–5024 页，2016年。[44] Y. Li、X. Hou、C. Koch、J. M. Rehg 和 A. L. Yuille，“显著目标分割的奥秘”，载于 CVPR。IEEE计算机学会，2014年，第 280–287 页。[45] M. Everingham、L. V. Gool、C. K. I. Williams、J. M. Winn 和 A. Zisserman，“PASCAL视觉目标类别（VOC）挑战赛”，《国际计算机视觉杂志》，卷 88，第 2 期，第 303–338 页，2010年。[46] C. Yang、L. Zhang、H. Lu、X. Ruan 和 M. Yang，“基于图的流形排序的显著性检测”，载于 CVPR。IEEE计算机学会，2013年，第 3166–3173 页。[47] L. Wang、H. Lu、Y. Wang、M. Feng、D. Wang、B. Yin 和 X. Ruan，“通过图像级监督学习检测显著目标”，载于 CVPR。IEEE计算机学会，2017年，第 3796–3805 页。[48] D. Fan、M. Cheng、Y. Liu、T. Li 和 A. Borji，“结构度量：评估前景图的新方法”，载于 ICCV。IEEE计算机学会，2017年，第 4558–4567 页。[49] J. Deng、W. Dong、R. Socher、L. Li、K. Li 和 F. Li，“ImageNet：大规模层次图像数据库”，载于 CVPR。IEEE计算机学会，2009年，第 248–255 页。

[50] K. K. Singh、H. Yu、A. Sarmasi、G. Pradeep 和 Y. J. Lee，“藏猫猫：一种用于弱监督定位及其他任务的数据增强技术”，*CoRR*，卷 abs/1811.02545，2018年。[51] P. Kr¨ahenb¨uhl 和 V. Koltun，“使用高斯边缘势对全连接条件随机场进行高效推理”，载于 神经信息处理系统大会，2011年，第 109–117 页。[52] R. Wu、M. Feng、W. Guan、D. Wang、H. Lu 和 E. Ding，“一种通过交织式多重监督进行显著目标检测的互学习方法”，载于 *CVPR*。计算机视觉基金会 / IEEE，2019年，第 8150–8159 页。