arXiv:2301.02778v2 [cs.CV] 3 Apr 2023

通过语义匹配与边缘对齐实现光学遥感图像轻量级显著目标检测

Gongyang Li、Zhi Liu，*S enior Member*，*IEEE*，Xinpeng Zhang，*Member*，*IEEE*，以及 Weisi Lin，*Fellow*，*IEEE*

摘要——近年来，依托卷积神经网络（CNNs），人们提出了许多用于光学遥感图像显著性目标检测（ORSI-SOD）的方法。然而，大多数方法忽视了卷积神经网络带来的庞大参数量和计算成本，只有少数方法关注其便携性和移动端适用性。为促进实际应用，本文提出了一种基于语义匹配和边缘对齐的新型轻量级 ORSI-SOD 网络，称为*SeaNet*。具体而言，SeaNet 包含一个用于特征提取的轻量级 MobileNet-V2、一个用于高层特征的动态语义匹配模块（DSMM）、一个用于低层特征的边缘自对齐模块（ESAM），以及一个用于推理的便携式解码器。首先，将高层特征压缩为语义卷积核。随后，通过 DSMM 中的动态卷积操作，利用语义卷积核激活两组高层特征中的显著目标位置。同时，在 ESAM 中，从两组低层特征中提取的跨尺度边缘信息通过 L_2 损失进行自对齐，并用于细节增强。最后，解码器从最高层特征出发，根据两个模块输出中包含的精确位置和精细细节推断显著目标。在两个公开数据集上进行的大量实验表明，轻量级 SeaNet 不仅优于大多数最先进的轻量级方法，而且准确率可与最先进的传统方法相媲美，同时参数量仅为 2.76M，处理 288×288 输入时的浮点运算次数为 1.7G。我们的代码和结果可在 https://github.com/MathLee/SeaNet 获取。

索引词—光学遥感图像、轻量级显著目标检测、语义匹配、边缘对齐。

I. 引言

显著性目标检测（SOD）旨在模拟人类视觉系统，快速定位最吸引注意力的物体/区域 [1]。作为一项重要的预处理步骤，SOD 的成功推动了许多领域的发展，例如图像质量评估 [2], [3], 目标分割 [4], [5], 和目标跟踪 [6]。不同于大多数针对单幅图像 [7],RGB-D/T图像 [8], [9], 和视频 [10] （这些图像和视频均拍摄于自然场景中）提出的显著性目标检测方法，本文聚焦于光学遥感图像中的显著性目标检测

Gongyang Li、Zhi Liu 和 Xinpeng Zhang 就职于特种光纤与光接入网络重点实验室、特种光纤与先进通信联合国际研究实验室、上海先进通信与数据科学研究院、上海大学（中国上海 200444），以及上海大学通信与信息工程学院（中国上海 200444）（电子邮件：ligongyang@shu.edu.cn；liuzhisjtu@163.com；xzhang@shu.edu.cn）。Weisi Lin 就职于南洋理工大学计算机科学与工程学院（新加坡 639798）（电子邮件：wslin@ntu.edu.sg）。通讯作者：*Zhi Liu* 和 *Xinpeng Zhang*。

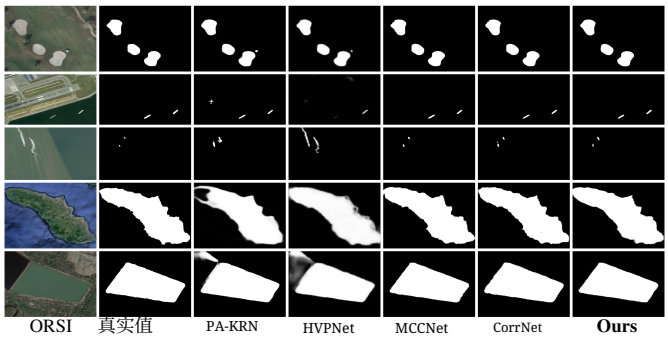


图1。四类方法和我们的方法在光学遥感图像上生成的显著图。PA-KRN [12]是自然场景图像显著性目标检测方法，HVPNet [13] 是轻量级 NSI-SOD 方法，MCCNet [14] 是光学遥感图像显著性目标检测方法，CorrNet [15]是轻量级 ORSI-SOD 方法。请放大查看细节。

图像，简称 ORSI-SOD。顺应自然场景图像显著性目标检测（NSI-SOD）的技术发展趋势 [7]，，我们致力于基于卷积神经网络（CNNs）解决 ORSI-SOD 问题 [11]。

在深度学习时代，人们提出了大量基于 CNN 的 NSI-SOD 方法，检测精度得到了显著提升。在这些方法中，经典的编码器-解码器结构 [16] 是最通用且有效的结构，并且通常结合深度监督等精巧策略 [17],门控机制 [18],边缘辅助 [19], [20], 渐进式架构 [12], 等。尽管 NSI-SOD 方法无法直接克服光学遥感图像的复杂场景问题（如 PA-KRN [12] 在图 1的第三列中所示），但其中包含的策略为基于 CNN 的光学遥感图像显著目标检测方法奠定了基础。针对光学遥感图像显著性目标检测的专用方法考虑了光学遥感图像中显著目标和场景的特性。例如，LVNet [21] 和 EMFINet [22]以多种分辨率的光学遥感图像作为输入，以克服显著目标尺寸不同的问题。MCCNet [14] 综合融合前景、背景、边缘和全局信息，以应对光学遥感图像复杂的背景，从而生成良好的显著图，如图 1的第五列所示。

然而，上述方法可能陷入参数量和计算成本过高的困境，例如图中列出的 PA-KRN 和 MCCNet 的参数量与浮点运算次数。1。为了解决这一问题，轻量级显著性目标检测方法逐渐兴起。与 PA-KRN 相比，轻量级 NSI-SOD 方法的先驱 HVPNet [13] 将

参数量和计算成本减少了数百倍。但 HVPNet 也受限于光学遥感图像，如图1的第四列所示。对于轻量级 ORSI 显著性目标检测方法，如图1的倒数第二列所示，CorrNet [15] 显著减少了参数量并取得了良好的性能，但仍消耗大量计算成本。

基于上述观察，本文提出了一种基于语义匹配和边缘对齐的新型光学遥感图像显著性目标检测方法，称为 *SeaNet*。该方法旨在比 CorrNet 更轻量，同时保持有竞争力的性能。众所周知，特征提取网络提取的特征可分为低层特征和高层特征：前者包含细节和纹理信息，后者包含语义和位置信息。因此，SeaNet 的核心思想是在编码器-解码器结构中采用不同策略处理高层特征和低层特征。

具体而言，我们提出动态语义匹配模块，用于对高层特征进行语义匹配：首先压缩语义信息，再将其与高层特征匹配，以感知显著目标的位置。我们还提出了边缘自对齐模块，用于对低层特征进行边缘对齐：对从低层特征中提取的跨尺度边缘信息进行对齐，以纠正边缘误差，并利用这些信息增强特征。为提高效率，我们采用 MobileNet-V2 [23]作为骨干网络，并采用深度可分离卷积 [23], [24]作为基本卷积组件，以控制参数量和计算成本。因此，SeaNet 仅含 2.76M 个参数量，运算量为 1.7G 浮点运算次数，并能生成准确的显著图，如图1最右列所示。具体而言，与最先进的轻量级方法相比，SeaNet 在检测精度方面具有竞争力；与最先进的传统方法相比，SeaNet 在计算复杂度方面具有竞争力。

我们的主要贡献有以下三点：

- 我们采用不同策略探索 MobileNet-V2 的高层和低层特征，并提出一种基于语义匹配和边缘对齐的新型轻量级光学遥感图像显著性目标检测网络，称为 *SeaNet*，其参数量仅为 2.76M，每张 288×288 图像的浮点运算次数为 1.7G。
- 我们为高层语义特征提出了一个动态语义匹配模块。DSMM 通过语义卷积核执行动态卷积来感知显著目标的位置，这不仅提高了特征交互的灵活性，还有效减少了参数量。此外，该模块执行通道相关性运算，以激活逐通道交互。
- 我们为低层细节特征提出了一个边缘自对齐模块。ESAM 专注于提取边缘信息以增强细节，并通过 L_2 损失对齐跨尺度边缘信息，从而纠正边缘误差。与 DSMM 类似，该模块也引入了通道相关性。

本文其余部分安排如下。在第II节中，我们回顾面向自然场景图像和光学遥感图像的基于 CNN 的显著性目标检测方法。在第III节中，我们详细介绍 SeaNet。在第IV节中，我们展示全面的实验结果。在第V节中，我们给出结论。

II. 相关工作

A. 基于 *CNN* 的显著性目标检测

自然场景图像中的显著目标检测 [7] 已取得显著成功，尤其是基于 CNN 的方法。在 [17],Hou 等人创新性地将深度监督引入自然场景图像显著性目标检测，显著增强了显著目标多尺度特征的表征能力。该方法有效提高了检测精度，并对后续基于 CNN 的方法产生了深远影响。Hu 等人 [25] 和 Deng等人 [26] 专注于循环机制。前者首先拼接多层级特征，再将其与来自不同层级的特征相结合；后者则交替使用低层特征和高层特征。此外，一些研究人员关注边缘/边界信息。Feng 等人。 [27] 提出边界增强损失以提高目标边界的完整性，并将其与深度监督相结合。Qin 等人。 [28]提出了一种混合损失，结合二元交叉熵、结构相似性和交并比损失，以分割具有精细结构和清晰边界的显著目标。Zhao等人。 [19] 通过边缘监督对显式边缘进行建模，以保留显著目标的边界。Zhou 等人。 [20] 提出了显著性到轮廓及轮廓到显著性策略，以实现快速显著性检测。Lee 等人 [29]提出了注意力引导追踪模块，通过显式边缘突出显著目标。

除了上述经典策略外，Pang 等人. [30]增强了自然场景图像显著性目标检测中多尺度特征之间的交互。Chen 等人. [31] 考虑了低层、高层和全局信息，以提高显著性图的完整性。Xu 等人. [12] 在知识回顾网络中采用全局定位与局部分割策略，以避免显著信息被稀释。Li 等人. [32],同时捕获了特征的多感受野信息和跨层级特征的互补信息。Qiu等人. [33] 扩展了空洞空间金字塔池化，并将通道与空间注意力嵌入其中，以探索空间和通道中的信息依赖关系。

尽管上述基于 CNN 的方法在自然场景图像显著性目标检测中表现出色，但它们无法有效处理光学遥感图像的独特属性，且往往生成不理想的显著图。此外，这些方法通常只关注准确率，而忽视计算复杂度。尽管如此，它们的策略仍为我们的方法提供了启发，例如经典的深度监督、边缘辅助和差异化特征处理。

B. 光学遥感图像中的基于 *CNN* 的显著性目标检测

光学遥感图像中的显著目标检测是显著性目标检测领域的一颗新星，近年来提出了许多基于 CNN 的方法。作为先驱，Li 等。 [21] 提出了首个基于 CNN 的光学遥感图像显著性目标检测方法，名为 LVNet。该方法通过将双流金字塔模块与带有嵌套连接的编码器-解码器模块相结合，感知不同尺寸的显著目标。此外，Li 等。 [34]

探究了光学遥感图像显著性目标检测中跨层级特征之间的交互。在 [35], Huang 等首先在语义引导解码器中粗略定位显著目标，然后以自顶向下方式细化粗略显著图。Cong 等。 [36] 为高层特征设计了关系推理编码器，并在多尺度注意力解码器中推断显著目标。Li 等。 [37] 采用两组相邻特征辅助当前特征，捕获上下文信息，以应对具有挑战性的光学遥感图像场景。

与自然场景图像显著性目标检测类似，一些研究者将边缘/边界信息引入光学遥感图像显著性目标检测。Zhang 等 [38]构建了一种多任务结构，可同时预测边缘图和显著性图。Tu 等 [39]和 Zhou 等 [22]，随后 [19]，基于低层和高层特征提取边界信息，并在两个解码器中保留显著目标的边界。Li 等 [14]将边缘与前景、背景和全局信息相结合，并充分考虑这些信息之间的互补性，以适应光学遥感图像。

上述针对光学遥感图像显著性目标检测的专用方法取得了令人满意的性能。然而，这些方法通常参数量大、计算成本高，不适用于航天设备，也难以在实际应用中推广。为此，我们提出了基于 MobileNet-V2 [23] 和两个轻量且有效模块的 SeaNet；该网络适用于移动设备，同时能够取得有竞争力的性能。

C. 轻量级显著目标检测

轻量级显著目标检测是一项新兴任务，最初在自然场景图像显著性目标检测中得到探索。Gao 等 [40] 提出了一种灵活的自适应卷积层，具有强大的多尺度表征能力，并构建了一个极轻量级网络（即 CSNet），用于自然场景图像显著性目标检测。Liu 等 [13] 提出了一种基于密集连接的分层视觉感知（HVP）模块，并基于HVP模块和残差注意力构建了轻量级HVPNet，以有效学习多尺度上下文。同时，Liu 等 [41] 提出了一种立体注意多尺度（SAM）模块，并构建了轻量级SAMNet，用于多层级、多尺度学习。总之，以上三项工作侧重于学习有效的多层级和多尺度信息，并构建强大的特征提取骨干网络，以实现轻量级自然场景图像显著性目标检测。然而，它们并未进一步处理和增强不同层级提取的特征，而是直接基于这些特征推断显著目标。此外，这些方法处理的是自然场景图像，而非光学遥感图像，缺乏针对性。因此，本文着重于充分利用现有特征提取骨干网络不同层级的特征，并针对遥感图像特征开发专用、轻量且有效的模块，从而实现更有效的专用光学遥感图像显著性目标检测方案。

针对光学遥感图像显著性目标检测，Li 等 [15]提出了首个轻量级方法，即 CorrNet。他们精简了原版 VGG-16 [42]，以实现高效的特征提取，并采用由粗到细的策略，结合密集的轻量级细化模块，检测光学遥感图像中的显著目标。CorrNet 的参数量大幅减少至仅 4.09M，但其计算成本仍然很高，浮点运算次数达到 21.1G。针对 RGB-D 显著性目标检测，Wu 等 [43] 提出了首个轻量级方法，即*i.e.*，MobileSal，该方法基于 MobileNet-V2 [23]。

受 [43], SeaNet 的启发，我们采用 MobileNet-V2 作为骨干网络，以解决现有轻量级 ORSI-SOD 方法（即 CorrNet）计算成本过高的问题。此外，为了减少参数量，我们提出了两个轻量且有效的模块，即 DSMM 和 ESAM。DSMM 将动态卷积扩展为动态深度卷积，而 ESAM 则通过自对齐的方式校正跨尺度边缘特征。

III. 提出的方法

本节详细介绍所提出的轻量级 SeaNet。在第 III-A 节中，我们介绍所提出 SeaNet 的网络概述。在第 III-B 节和第 III-C 节中，我们分别详细介绍两个轻量级模块，即。：DSMM 和 ESAM。在第 III-D 节中，我们介绍解码器和损失函数。

A. 网络概述

如图 2 所示，所提出的轻量级 SeaNet 基于显著性目标检测中常用的编码器-解码器结构 [8], [9], [47], [48]。SeaNet 包含一个编码器、一个语义知识压缩（SKC）单元、一个动态语义匹配模块（DSMM）、一个边缘自对齐模块（ESAM）和一个轻量级解码器。它首先通过语义匹配实现显著目标的位置激活，然后通过边缘对齐增强细节。

我们的 SeaNet 输入尺寸为 3×288×288。对于 SeaNet 的编码器，我们采用轻量级的 MobileNet-V2 [23]，即保留前十七个倒置残差瓶颈，并截去最后三层，*i.e.*，即两个卷积层和一个平均池化层。我们根据第一个、第三个、第六个、第十三个和最后一个瓶颈，将 MobileNet-V2 划分为五个块，分别记为 E^t ($t = 1,2,3,4,5$)。输出的五级特征记为 $\boldsymbol{f}_{\mathrm{e}}^t \in \mathbb{R}^{c_t \times h_t \times w_t}$，其中 h_t 和 w_t 为 $\frac{288}{2^t}$，以及 $c_t \in \{16, 24, 32, 96, 320\}$。针对 ORSI-SOD，我们采用不同策略来处理生成的高层和低层特征。对于高层特征，我们首先在语义知识压缩（SKC）单元中，将最高层特征 $\boldsymbol{f}_{\mathrm{e}}^5$ 压缩为两个语义卷积核，即*i.e.*，$\boldsymbol{k}_3$和 $\boldsymbol{k}_4$，然后将其用作动态逐通道卷积 [24], [44] 的卷积核，分别与两组高层特征*i.e.*， $\boldsymbol{f}_{\mathrm{e}}^3$和 $\boldsymbol{f}^4{}_{\mathrm{e}}$进行卷积，此过程在 DSMM 中完成。除上述空间语义匹配外，我们还通过通道相关性 [45], [46],增强通道交互，生成 $\boldsymbol{f}_{\mathrm{dsmm}}$。同时，对于低层特征，我们通过池化减法操作 [27], 获取边缘信息，并以自对齐的方式纠正边缘误差。随后，我们采用校正后的边缘特征，在 ESAM 的基于边缘的增强单元（EEU）中进行特征增强，并计算通道相关性，生成$\boldsymbol{f}_{\mathrm{esam}}$。SeaNet 的解码器由三个轻量级块组成，分别记为 $\mathrm{D}^{1\text{-}2}$、$\mathrm{D}^{3\text{-}4}$和 D^5。利用 $\boldsymbol{f}_{\mathrm{e}}^5$、$\boldsymbol{f}_{\mathrm{dsmm}}$和 $\boldsymbol{f}_{\mathrm{esam}}$，我们逐步突出显著目标，以更好地恢复分辨率。我们还引入了

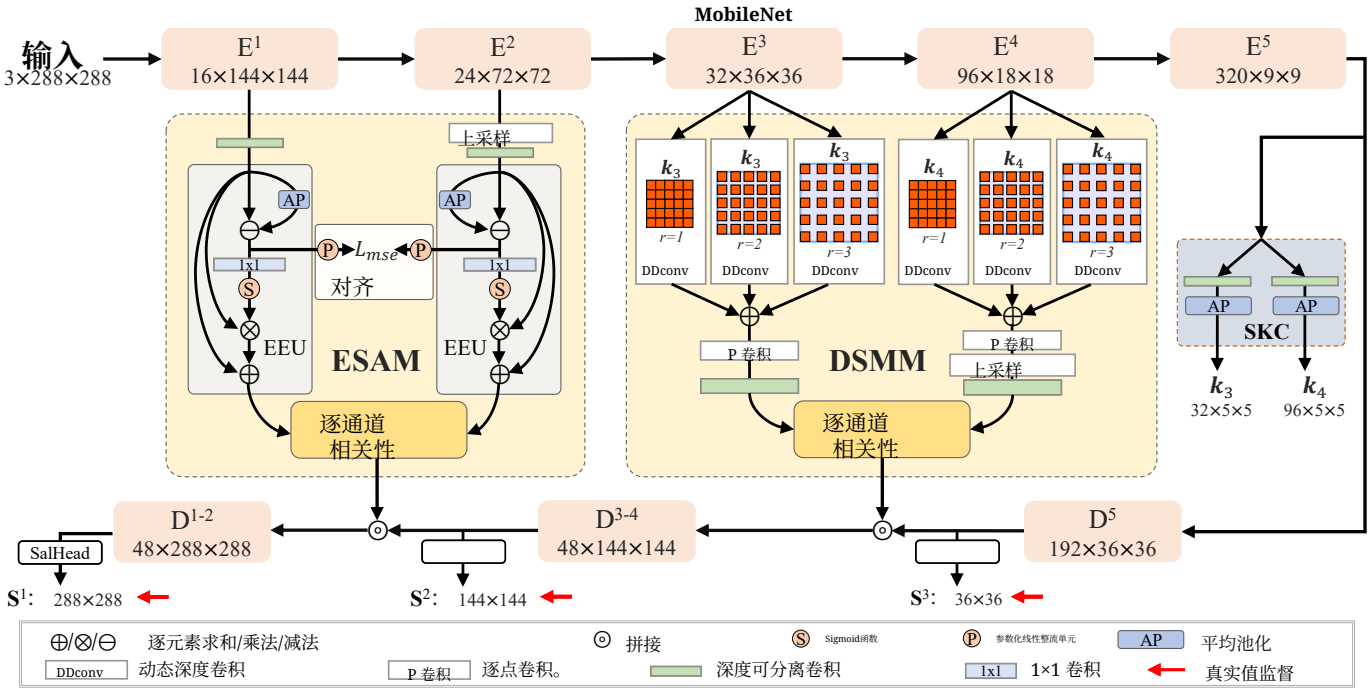


图2. 所提出的轻量级 SeaNet 流程图，该网络采用编码器-解码器架构。首先，MobileNet-V2 [23] 从输入中提取基本特征嵌入，得到五级特征。接着，在语义知识压缩（SKC）单元中获取两个语义卷积核。利用语义卷积核在动态语义匹配模块（DSMM）中与高层特征进行卷积，以激活显著目标的位置。随后，我们在 DSMM 中执行通道相关性 [45], [46] 。同时，在边缘自对齐模块（ESAM）中，我们采用基于边缘的增强单元（EEU），从低层特征中提取边缘特征以进行细节增强，并通过 L_{mse} 损失（即*e.*， L_2 损失）对边缘特征进行对齐。通道相关性也被引入 ESAM。最后，解码器基于最高层特征以及 DSMM 和 ESAM 的输出推断显著目标，并生成输出显著性图$\mathbf{S}^1$。

在每个解码器块之后添加显著性推理头（SalHead），以进行深度监督 [17] 并生成最终显著性图。

B. 动态语义匹配模块

高层特征包含丰富的语义信息，有利于显著目标定位。在此，我们提出 DSMM，以有限的参数量和计算成本，利用高层特征有效激活显著区域。受动态卷积启发 [44],，我们使用现有特征而非参数初始化来生成卷积核，从而减少参数量。此外，我们将动态卷积与深度可分离卷积（DSconv） [24], 相结合，提出动态深度卷积（DDconv），以同时降低计算成本。DDconv 在 DSMM 的空间语义匹配中发挥着重要作用。然而，仅考虑空间交互并不足够，因此我们进一步在 DSMM 中引入通道交互，通过通道相关性 [45], [46]增强通道依赖关系。

我们在图2的中间部分展示了 DSMM 的详细结构。我们的 DSMM 可分为两部分，*i.e.*，即空间语义匹配和逐通道相关性。下文将基于这两个部分介绍 DSMM。由于 SKC 单元为 DSMM 生成语义卷积核，我们首先详细介绍该单元。

1) SKC 单元。如图2右侧所示，SKC 单元直接压缩 $\boldsymbol{f}_{\mathrm{e}}^5$，但效果显著。由于 SKC 单元会为作为 DSMM 输入的 $\boldsymbol{f}_{\mathrm{e}}^3$ 和 $\boldsymbol{f}^4{}_{\mathrm{e}}$ 生成语义卷积核，我们首先通过两个并行的深度可分离卷积层压缩 $\boldsymbol{f}_{\mathrm{e}}^5$ 的通道数，使其与 $\boldsymbol{f}_{\mathrm{e}}^3$ 和 $\boldsymbol{f}_{\mathrm{e}}^4$ 的通道数相匹配，然后通过两个并行的自适应平均池化层将分辨率压缩至合适大小，从而生成两个语义卷积核 $\boldsymbol{k}_3 \in \mathbb{R}^{32\times5\times5}$ 和 $\boldsymbol{k}_4 \in \mathbb{R}^{96\times5\times5}$。SKC 单元的形式化定义如下：

$$\boldsymbol{k}_t = \mathrm{AP}(\mathrm{DSconv}(\boldsymbol{f}_{\mathrm{e}}^5)), \quad t = 3, 4, \tag{1}$$

其中，DSconv(·) 表示 3×3 DSconv层，AP(·) 表示自适应平均池化层。

2) 空间语义匹配。我们采用 $\boldsymbol{k}_3$ 和 $\boldsymbol{k}_4$ 包含全局信息的特征作为动态卷积层的卷积核 [44]以减少参数量。然而，传统动态卷积层的计算成本与常规卷积层相同。受 DSconv 启发 [24], ，我们将动态卷积改进为 DDconv，使其以逐通道方式执行动态卷积，即，也就是说，将动态卷积的组数设为输入特征的通道数。此外，我们将空洞机制 [49] 引入 DDconv，并利用多个空洞感受野充分感知显著目标，从而实现精确定位。这有助于应对 ORSI 中存在多个目标以及目标尺寸不一的场景。

如图2中的 DSMM 所示，我们分别采用三个空洞率为 $\{1, 2, 3\}$ on$\boldsymbol{f}_{\mathrm{e}}^3$ 和 $\boldsymbol{f}^4{}_{\mathrm{e}}$ 的空洞动态深度卷积层。事实上，DDconv 是一种无参数的语义匹配过程，它打破了对

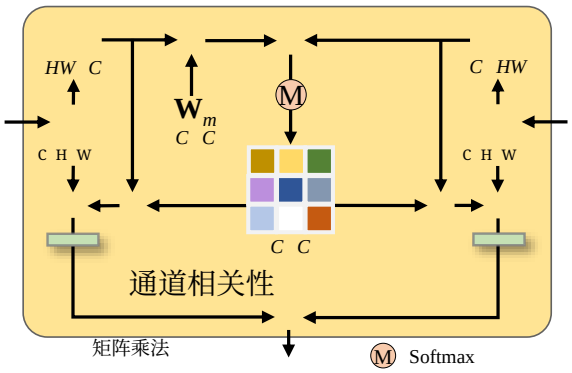


图3. 通道相关性的示意图。

训练参数的限制，提高了模块乃至整个网络的灵活性。不同于依次执行逐通道卷积和逐点卷积的深度可分离卷积，我们通过逐元素求和整合三个空洞DD卷积的输出特征，然后执行逐点卷积以融合多感受野特征。这不仅进一步减少了参数量，也有助于不同感受野生成的特征之间进行交互。空间语义匹配的公式如下：

$$\boldsymbol{f}_{\mathrm{sm}}^{t}=\mathrm{Pconv}(\sum_{i=1}^{3}\mathrm{DDconv}(\boldsymbol{f}_{\mathrm{e}}^{t};\boldsymbol{k}_{t},r_{i})),\quad t=3,4, \tag{2}$$

其中，$\boldsymbol{f}_{\mathrm{sm}}^{t}\in\mathbb{R}^{c_t\times h_t\times w_t}$表示语义匹配的输出特征，DDconv(·;$\boldsymbol{k}_t$，$r_i$)表示具有动态卷积核 $\boldsymbol{k}_t$ 和空洞率 $r_i\in\{1,2,3\}$的 DDconv 层，$\sum$ 表示多个特征的逐元素求和，Pconv(·)表示逐点卷积层。

3）通道相关性。除了空间交互之外，我们还在 DSMM 中执行由空间协同注意力扩展而来的通道相关性 [45], [46]。我们首先通过一个 DSconv 层和一次上采样操作，在通道和分辨率上对齐$\boldsymbol{f}_{\mathrm{sm}}^{3}$ 和 $\boldsymbol{f}_{\mathrm{sm}}^{4}$，以获得通道相关性的输入特征，*i.e.*，即 $\{\hat{}\boldsymbol{f}^{3}{}_{\mathrm{sm}},\hat{}\boldsymbol{f}_{\mathrm{sm}}^{4}\}\in\mathbb{R}^{c_4\times h_3\times w_3}$。随后，我们将通道相关性记作 CCorr(·)，并定义如下：

$$\boldsymbol{f}_{\mathrm{dsmm}}=\mathrm{CCorr}(\hat{\boldsymbol{f}}_{\mathrm{sm}}^{3},\hat{\boldsymbol{f}}_{\mathrm{sm}}^{4}), \tag{3}$$

其中，$\boldsymbol{f}_{\mathrm{dsmm}}\in\mathbb{R}^{(2\times c_4)\times h_3\times w_3}$ 表示 DSMM 的输出特征。

我们在图 3中展示了通道相关性的结构，为简洁起见，我们将输入特征简化为 $\{\boldsymbol{f}_1,\boldsymbol{f}_2\}\in\mathbb{R}^{C\times H\times W}$。首先，我们调整 $\boldsymbol{f}_1$ 和 $\boldsymbol{f}_2$的尺寸，并得到展平后的 $\hat{}\boldsymbol{f}_1\in\mathbb{R}^{HW\times C}$和 $\hat{}\boldsymbol{f}_2\in\mathbb{R}^{C\times HW}$。然后，我们通过矩阵乘法将 $\hat{}\boldsymbol{f}_1$ 与可训练矩阵 $\mathbf{W}_m\in\mathbb{R}^{C\times C}$ 相乘，以自适应地学习特征变换。$\mathbf{A}\in\mathbb{R}^{C\times C}$是 $\hat{}\boldsymbol{f}_1$和 $\hat{}\boldsymbol{f}_2$的通道亲和矩阵，可通过矩阵乘法计算如下：

$$\mathbf{A}=\hat{\boldsymbol{f}}_2\circledast(\hat{\boldsymbol{f}}_1\circledast\mathbf{W}_m), \tag{4}$$

其中，$\circledast$ 表示矩阵乘法。通过这种方式，我们对沿通道方向的特征依赖关系进行建模。

然后，我们分别采用按行和按列的 Softmax 函数对亲和矩阵进行归一化，并将建立的通道依赖关系传递至 $\boldsymbol{f}_1$ 和 $\boldsymbol{f}_2$。此外，我们引入短连接和一个 3×3 DSconv层，以融合原始的空间增强特征（*i.e.*，$\boldsymbol{f}_1$ 和 $\boldsymbol{f}_2$）与上述通道增强特征。我们将上述过程形式化如下：

$$\boldsymbol{f}_{\mathrm{sc}}^{1}=\mathrm{DSconv}\big((\mathrm{M}_r(\mathbf{A})\circledast\hat{\boldsymbol{f}}_1^{\top})\oplus\boldsymbol{f}_1\big), \tag{5}$$

$$\boldsymbol{f}_{\mathrm{sc}}^{2}=\mathrm{DSconv}\big(((\mathrm{M}_c(\mathbf{A}))^{\top}\circledast\hat{\boldsymbol{f}}_2)\oplus\boldsymbol{f}_2\big), \tag{6}$$

其中，$\{\boldsymbol{f}_{\mathrm{sc}}^{1},\boldsymbol{f}_{\mathrm{sc}}^{2}\}\in\mathbb{R}^{C\times H\times W}$表示空间增强和通道增强后的特征，$\mathrm{M}_r(\cdot)$ 和 $\mathrm{M}_c(\cdot)$ 分别表示按行和按列的 Softmax 函数，$\top$ 表示矩阵转置操作，$\oplus$ 表示逐元素求和。值得注意的是，为简洁起见，我们省略了特征尺寸变换。最后，我们将 $\boldsymbol{f}^{1}{}_{\mathrm{sc}}$ 和 $\boldsymbol{f}_{\mathrm{sc}}^{2}$拼接起来，得到通道相关性的输出特征 $\boldsymbol{f}_{\mathrm{ccorr}}\in\mathbb{R}^{2C\times H\times W}$，即. DSMM 的输出特征 $\boldsymbol{f}_{\mathrm{dsmm}}$。

总之，我们以轻量化方式实现了 DSMM，参数量和计算成本均有限。此外，我们在 DSMM 中充分考虑了空间交互和通道交互，从而全面、准确地定位显著目标，有助于应对光学遥感图像中的复杂场景。

C. 边缘自对齐模块

低层特征包含丰富的纹理和目标细节信息，有助于描绘显著目标的精细结构。我们提出 ESAM，以高效、有效地挖掘边缘信息并增强细节，从而保留光学遥感图像中显著目标的复杂形状。不同于一些基于边缘的光学遥感图像显著目标检测方法 [14][22], [38], [39]，我们的 ESAM 轻量且无需边缘监督即可提取边缘信息，使用起来更加便捷。与 DSMM 类似，ESAM 也充分考虑了特征的空间交互和通道交互。如图 2左侧所示，ESAM 由两个基于边缘的增强单元（EEU）和一个通道相关性模块组成。下面依次介绍它们。

1) 带自对齐的基于边缘的增强单元。ESAM 的输入特征为 $\boldsymbol{f}_{\mathrm{e}}^{1}$和 $\boldsymbol{f}_{\mathrm{e}}^{2}$。我们通过 DSconv层和上采样操作对它们进行对齐，得到$\{\hat{}\boldsymbol{f}_{\mathrm{e}}^{1}、\hat{}\boldsymbol{f}_{\mathrm{e}}^{2}\}\in\mathbb{R}^{c_2\times h_1\times w_1}$，它们是 EEU 的输入特征。我们采用池化减法操作 [27]提取两组边缘特征 $\{\hat{}\boldsymbol{f}_{\mathrm{edge}}^{1}、\hat{}\boldsymbol{f}_{\mathrm{edge}}^{2}\}\in\mathbb{R}^{c_2\times h_1\times w_1}$，分别来自$\hat{}\boldsymbol{f}^{1}{}_{\mathrm{e}}$和$\hat{}\ \boldsymbol{f}_{\mathrm{e}}^{2}$，如下所示：

$$\boldsymbol{f}_{\mathrm{edge}}^{t}=\hat{\boldsymbol{f}}_{\mathrm{e}}^{t}\ominus\mathrm{AP}(\hat{\boldsymbol{f}}_{\mathrm{e}}^{t}),\quad t=1,2, \tag{7}$$

其中 $\ominus$表示逐元素减法。$\boldsymbol{f}^{t}{}_{\text{边缘}}$随后用于描绘$\hat{}\boldsymbol{f}^{t}{}_{\mathrm{e}}$中的边缘区域。然而，由于没有边缘监督，所获得的边缘信息不可避免地容易受到误差影响。针对这一问题，我们提出了一种基于均方误差损失（即，L_2 或 L_{mse} 损失）的新型自对齐机制，并将其应用于边缘特征，如图2中的 ESAM 所示。通过这种方式，我们可以在$\hat{}\ \boldsymbol{f}_{\mathrm{edge}}^{1}$ 与$\hat{}\ \boldsymbol{f}_{\mathrm{edge}}^{2}$ 之间进行自适应边缘误差校正，并在训练阶段获得准确且一致的边缘信息。

表 I 三个解码器块的详细结构和参数。（3×3，320，192）表示深度可分离卷积的卷积核大小为 3×3，输入通道数为 320，输出通道数为 192。

方面	D^5	$D^{3\text{-}4}$	$D^{1\text{-}2}$
输入尺寸	$[320\times9\times9]$	$[384\times36\times36]$	$[96\times144\times144]$
深度可分离卷积	$(3\times3,320,320)$	$(3\times3,384,192)$	$(3\times3,96,48)$
深度可分离卷积	$(3\times3,320,320)$	$(3\times3,\ 192,\ 192)$	$(3\times3,\ 48,\ 48)$
上采样	$4\times$	$4\times$	$2\times$
深度可分离卷积	$(3\times3,320,192)$	$(3\times3,192,48)$	$(3\times3,\ 48,\ 48)$
输出尺寸	$[192\times36\times36]$	$[48\times144\times144]$	$[48\times288\times288]$

基于校正后的边缘信息，我们对$\hat{}$ $\boldsymbol{f}_{\mathrm{e}}^{t}$ 进行边缘增强，并获得边缘增强单元（EEU）的输出特征，记为 $\boldsymbol{f}_{\mathrm{eeu}}^{t}\in\mathbb{R}^{c_2\times h_1\times w_1}$，如下所示：

$$\boldsymbol{f}_{\mathrm{eeu}}^{t}=\mathrm{conv}_{1\times1}^{s}(\boldsymbol{f}_{\mathrm{edge}}^{t})\otimes\hat{\boldsymbol{f}}_{\mathrm{e}}^{t}\oplus\hat{\boldsymbol{f}}_{\mathrm{e}}^{t},\quad t=1,2, \tag{8}$$

其中，$\mathrm{conv}_{1\times1}^{s}(\cdot)$是带有Sigmoid激活函数的 1×1 卷积层，$\otimes$ 表示逐元素乘法。

2）逐通道相关性。EEU侧重于在空间层面增强特征，我们通过以下通道相关性来增强其输出 $\boldsymbol{f}_{\mathrm{eeu}}^{1}$ 和 $\boldsymbol{f}_{\mathrm{eeu}}^{2}$ 之间的通道交互：

$$\boldsymbol{f}_{\mathrm{esam}}=\mathrm{CCorr}(\boldsymbol{f}_{\mathrm{eeu}}^{1},\boldsymbol{f}_{\mathrm{eeu}}^{2}), \tag{9}$$

其中，$\boldsymbol{f}_{\mathrm{esam}}\in\mathbb{R}^{(2\times c_2)\times h_1\times w_1}$ 是ESAM的输出特征。

通过这种方式，ESAM节省了大量参数和计算成本，同时为准确突出ORSI中显著目标的复杂几何结构和拓扑结构提供了有力支持。

D. 解码器与损失函数

1）解码器。基于上述两个模块提供的显著目标位置和细节信息，我们设计了一种轻量级解码器来生成显著图。如图2底部所示，我们的轻量级解码器由三个模块组成，即，$D^{1\text{-}2}$、$D^{3\text{-}4}$和D^5。每个解码器块依次包含两个DSconv层、一个上采样操作和另一个DSconv层。其详细参数见表I。具体而言，我们在这三个解码器块之后设置SalHeads，以生成三张分辨率不同的显著图，即，$\mathbf{S}^3\in[0,1]^{1\times36\times36}$、$\mathbf{S}^2\in[0,1]^{1\times144\times144}$和$\mathbf{S}^1\in[0,1]^{1\times288\times288}$，其中前两张用于深度监督，最后一张是SeaNet的最终输出。SalHead由一个丢弃层 [51] 和一个 1×1卷积层组成。

2）损失函数。我们采用二元交叉熵（BCE）损失和交并比（IoU）损失对 SeaNet 进行联合训练。因此，我们的总损失由两部分组成，即显著性损失和边缘对齐损失。此外，我们引入损失权重，以区别对待这两种损失，从而获得更好的训练效果。总损失函数 L_{total}可表示如下：

$$L_{\mathrm{total}}=\sum_{i=1}^{3}(L_{bce}^{i}+L_{iou}^{i})+\lambda\cdot L_{mse}(\mathrm{P}(\boldsymbol{f}_{\mathrm{edge}}^{1}),\mathrm{P}(\boldsymbol{f}_{\mathrm{edge}}^{2})), \tag{10}$$

其中，L_{bce}^{i} 和 L_{iou}^{i} 分别为二元交叉熵损失和交并比损失，它们根据真实值（GT）监督$\mathbf{S}^i$；λ 为损失权重，设为 0.5；L_{mse}（·）为均方误差损失；P（·）为参数化修正线性单元 [52]。

IV. 实验

A. 实验设置

1）数据集。我们在 ORSSD [21]和 EORSSD [38] 数据集上开展实验。ORSSD 数据集包含 800 幅光学遥感图像及其对应的像素级标注。该数据集分为两部分，即：训练集（600 幅图像）和测试集（200 幅图像）。EORSSD 数据集在 ORSSD 数据集的基础上增加了 1200 幅光学遥感图像，总计 2000 幅。该数据集也分为两部分，即：训练集包含 1400 幅图像，测试集包含 600 幅图像。

2) 评估指标。我们从定量评价指标和计算复杂度指标两个方面，对我们的方法及所有对比方法进行评估。

定量评价指标包括 S度量（S_α，α =0.5）、[53], F度量（F_β，β^2 = 0.3）、[54], E度量（E_ξ）[55],和平均绝对误差（$\mathcal{M}$）。前三项指标越高越好，最后一项则相反。我们报告 F度量和 E度量的最大值、均值和自适应值。

计算复杂度指标包括以每秒帧数（fps）衡量的推理速度、以百万（M）为单位衡量的参数量（#Param），以及以十亿（G）为单位衡量的浮点运算次数（FLOPs）。第一项指标越高越好，后两项则相反。推理速度在批量大小为 1 且不计 I/O 时间的情况下进行报告。

3) 训练方案。我们基于 PyTorch [56]，在配备 NVIDIA Titan X GPU（12GB 显存）的计算机上开展实验。训练细节和参数如下：输入尺寸为 288×288，数据增强策略包括翻转和旋转，网络优化器为 Adam [57]，，批量大小为 8，基础学习率为$1e^{-4}$，学习率每 30 个训练轮次衰减至 1/10，训练轮次为 50。此外，我们使用预训练参数初始化 MobileNet-V2，并使用“Kaiming”方法 [52]初始化两个模块和解码器中新添加的层。值得注意的是，对于每个数据集，我们均使用其自身的训练集进行训练，并使用其自身的测试集进行测试，如 [15], [22], [38]所示。

B. 性能分析

我们将 SeaNet 与 17 种最先进的传统显著性目标检测方法进行了比较，其中包括 11 种传统 NSI-SOD 方法（即，DSS [17], RADF [25], R3Net [26], Pool- Net [50], EGNet [19], GCPA [31], MINet [30], ITSD [20],GateNet [18], SUCA [32], 和 PA-KRN [12]），以及 6 种传统 ORSI-SOD 方法（即，LVNet [21], DAFNet [38],

表II Q与当前最先进的---方法 自然场景图像显著性目标检测 方法、光学遥感图像显著性目标检测 方法、以及轻量级方法在 EORSSD 和 ORSSD 数据集上的定量与计算复杂度比较。↑ 表示数值越高越好，而 ↓ 则相反。 T前三名结果分别以**红色**、**蓝色** 和**绿色**标出。

方法类型		输入 size	速度 (帧/秒)↑	#参数量 (M)↓	浮点运算次数 (G)↓	EORSSD [38]								ORSSD [21]							
						$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$F_\beta^{\text{mean}}\uparrow$	$F_\beta^{\text{adp}}\uparrow$	$E_\xi^{\max}\uparrow$	$E_\xi^{\text{mean}}\uparrow$	$E_\xi^{\text{adp}}\uparrow$	$\mathcal{M}\downarrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$F_\beta^{\text{mean}}\uparrow$	$F_\beta^{\text{adp}}\uparrow$	$E_\xi^{\max}\uparrow$	$E_\xi^{\text{mean}}\uparrow$	$E_\xi^{\text{adp}}\uparrow$	$\mathcal{M}\downarrow$
DSS_{17} [17]	CN	400×300	8	62.23	114.6	.7868	.6849	.5801	.4597	.9186	.7631	.6933	.0186	.8262	.7467	.6962	.6206	.8860	.8362	.8085	.0363
$RADF_{18}$ [25]	CN	400×400	7	62.54	214.2	.8179	.7446	.6582	.4933	.9140	.8567	.7162	.0168	.8259	.7619	.6856	.5730	.9130	.8298	.7678	.0382
$R3Net_{18}$ [26]	CN	300×300	2	56.16	47.5	.8184	.7498	.6302	.4165	.9483	.8294	.6462	.0171	.8141	.7456	.7383	.7379	.8913	.8681	.8887	.0399
$PoolNet_{19}$ [50]	CN	400×300	25	53.63	123.4	.8207	.7545	.6406	.4611	.9292	.8193	.6836	.0210	.8403	.7706	.6999	.6166	.9343	.8650	.8124	.0358
$EGNet_{19}$ [19]	CN	~380×320	9	108.07	291.9	.8601	.7880	.6967	.5379	.9570	.8775	.7566	.0110	.8721	.8332	.7500	.6452	.9731	.9013	.8226	.0216
$GCPA_{20}$ [31]	CN	320×320	23	67.06	54.3	.8869	.8347	.7905	.6723	.9524	.9167	.8647	.0102	.9026	.8687	.8433	.7861	.9509	.9341	.9205	.0168
$MINet_{20}$ [30]	CN	320×320	12	47.56	146.3	.9040	.8344	.8174	.7705	.9442	.9346	.9243	.0093	.9040	.8761	.8574	.8251	.9545	.9454	.9423	.0144
$ITSD_{20}$ [20]	CN	288×288	16	17.08	54.5	.9050	.8523	.8221	.7421	.9556	.9407	.9103	.0106	.9050	.8735	.8502	.8068	.9601	.9482	.9335	.0165
$GateNet_{20}$ [18]	CN	384×384	25	100.02	108.3	.9114	.8566	.8228	.7109	.9610	.9385	.8909	.0095	.9186	.8871	.8679	.8229	.9664	.9538	.9428	.0137
$SUCA_{21}$ [32]	CN	256×256	24	117.71	56.4	.8988	.8229	.7949	.7260	.9520	.9277	.9082	.0097	.8989	.8484	.8237	.7748	.9584	.9400	.9194	.0145
$PA\text{-}KRN_{21}$ [12]	CN	600×600	16	141.06	617.7	.9192	.8639	.8358	.7993	.9616	.9536	.9416	.0104	.9239	.8890	.8727	.8548	.9680	.9620	.9579	.0139
$LVNet_{19}$ [21]	CR	128×128	1.4	-	-	.8630	.7794	.7328	.6284	.9254	.8801	.8445	.0146	.8815	.8263	.7995	.7506	.9456	.9259	.9195	.0207
$DAFNet_{21}$ [38]	CR	128×128	26	29.35	68.5	.9166	.8614	.7845	.6427	**.9861**	.9291	.8446	**.0060**	.9191	.8928	.8511	.7876	**.9771**	.9539	.9360	.0113
$SARNet_{21}$ [35]	CR	336×336	47	25.91	129.7	.9240	.8719	**.8541**	**.8304**	.9620	.9555	.9536	.0099	.9134	.8850	.8619	.8512	.9557	.9477	.9464	.0187
$MJRBM_{22}$ [39]	CR	352×352	32	43.54	95.7	.9197	.8656	.8239	.7066	.9646	.9350	.8897	.0099	.9204	.8842	.8566	.8022	.9623	.9415	.9328	.0163
$EMFINet_{22}$ [22]	CR	256×256	25	107.26	480.9	**.9290**	**.8720**	.8486	.7984	**.9711**	.9604	.9501	.0084	**.9366**	**.9002**	**.8856**	.8617	.9737	.9671	.9663	.0109
$MCCNet_{22}$ [14]	CR	256×256	95	67.65	112.8	**.9327**	**.8904**	**.8604**	**.8137**	**.9755**	**.9685**	**.9538**	**.0066**	**.9437**	**.9155**	**.9054**	**.8957**	**.9800**	**.9758**	**.9735**	**.0087**
$CSNet_{20}$ [40]	LN	224×224	38	0.14	0.7	.8364	.8341	.7656	.6319	.9535	.8929	.8339	.0169	.8910	.8790	.8285	.7615	.9628	.9171	.9068	.0186
$SAMNet_{21}$ [41]	LN	336×336	44	1.33	0.5	.8622	.7813	.7214	.6114	.9421	.8700	.8284	.0132	.8761	.8137	.7531	.6843	.9478	.8818	.8656	.0217
$HVPNet_{21}$ [13]	LN	336×336	26	1.23	1.1	.8734	.8036	.7377	.6202	.9482	.8721	.8270	.0110	.8610	.7938	.7396	.6726	.9320	.8717	.8471	.0225
$CorrNet_{22}$ [15]	LR	256×256	100	4.09	21.1	**.9289**	**.8778**	**.8620**	**.8311**	.9696	**.9646**	**.9593**	.0083	**.9380**	**.9129**	**.9002**	**.8875**	**.9790**	**.9746**	**.9721**	**.0098**
SeaNet（我们的方法）	LR	288×288	96	2.76	1.7	.9208	.8649	.8519	**.8304**	.9710	**.9651**	**.9602**	**.0073**	.9260	.8942	.8772	**.8625**	.9767	**.9722**	**.9670**	**.0105**

CN：中国；CR：基于 CNN 的 ORSI-SOD 方法；LN：轻量级 NSI-SOD 方法；LR：轻量级 ORSI-SOD 方法 od.

SARNet [35], MJRBM [39],EMFINet [22],和 MCC-Net [14]）。此外，我们还将轻量级 SeaNet 与 4 种最先进的轻量级显著性目标检测方法进行了比较，其中包括 3 种轻量级 NSI-SOD 方法（即，CSNet [40],SAM- Net [41]，和 HVPNet [13]），以及轻量级 ORSI-SOD 方法 CorrNet [15]。上述 NSI-SOD 方法均采用默认参数设置，在与 SeaNet 相同的 ORSI-SOD 数据集上重新训练，以生成显著图。其他方法的显著图则取自其作者或公开源代码。

*1）与传统显著性目标检测方法的比较。*表II 上半部分展示了 SeaNet 与传统显著性目标检测方法在自然场景图像和光学遥感图像上的定量评估结果及计算复杂度评估结果。与传统 NSI-SOD 方法相比，SeaNet 在准确性和计算复杂度方面均取得了最佳表现。例如，与性能最佳的 NSI-SOD 方案 PA-KRN 相比， [12], SeaNet 在 $\mathcal{M}$方面具有显著优势，例如，在 EORSSD 数据集上为 0.0073 对 0.0104，在 ORSSD 数据集上为 0.0139 对 0.0105；此外，SeaNet 还具有 6×更快的推理速度，达到 51.1× 更少的参数量，以及 363.3× 更少的浮点运算次数。这体现了专用方法的优势；即使是我们的轻量级专用方法，也能胜过传统 NSI-SOD 方案。

与传统的光学遥感图像显著性目标检测方法相比，我们的 SeaNet 准确率相当，但计算复杂度显著更低。例如，与 EMFINet [22], 相比，SeaNet 的准确率相近，即.如，$E_\xi^{\max}$：在 EORSSD 数据集上为 0.9710 对 0.9711，在 ORSSD 数据集上为 0.9767 对 0.9737；而 SeaNet 的推理速度快 3.6× 倍，参数量和浮点运算次数分别少 24.5× 和 282.8× 。与表现最佳的 MCCNet [14], 相比，尽管 SeaNet 的准确率不占优势，但其较低的计算复杂度仍然十分突出。

*2）与轻量级显著性目标检测方法的比较。*在表II底部，我们报告了 SeaNet 与四种用于自然场景图像和光学遥感图像的轻量级显著性目标检测方法的比较结果。与轻量级 NSI-SOD 方法相比，SeaNet 在参数量和浮点运算次数方面并无优势，但在推理速度方面具有明显优势。这意味着我们的轻量级 SeaNet 仍有改进空间。另一方面，SeaNet 在准确率方面的优势十分明显，例如，在两个数据集上，SeaNet 的表现分别优于它们 3.50%~8.44%（ S_α）、4.87%~13.76%（F_β^{mean}）、5.51%~10.05%（ E_ξ^{mean}）以及 0.0037~0.0120（ $\mathcal{M}$）。

我们的 SeaNet 最初旨在减少现有轻量级光学遥感图像显著性目标检测方法 CorrNet 的参数量和浮点运算次数，[15], 尤其是后者。表II中的结果表明，SeaNet 在准确率没有显著下降的情况下实现了这一目标。具体而言，在计算复杂度方面，SeaNet 的参数量比 CorrNet 少 1.3× ，浮点运算次数少 12.4× ，且推理速度与之相当。在准确率方面，SeaNet 在 ORSSD 数据集上的 E_ξ^{adp} 略低于 CorrNet（ 0.9670 对比 0.9721），而在 EORSSD 数据集上的 $\mathcal{M}$则略高（0.0073 对比 0.0083）。

总体而言，SeaNet 在定量评价指标方面取得了一个第一名、两个第二名和五个第三名，同时

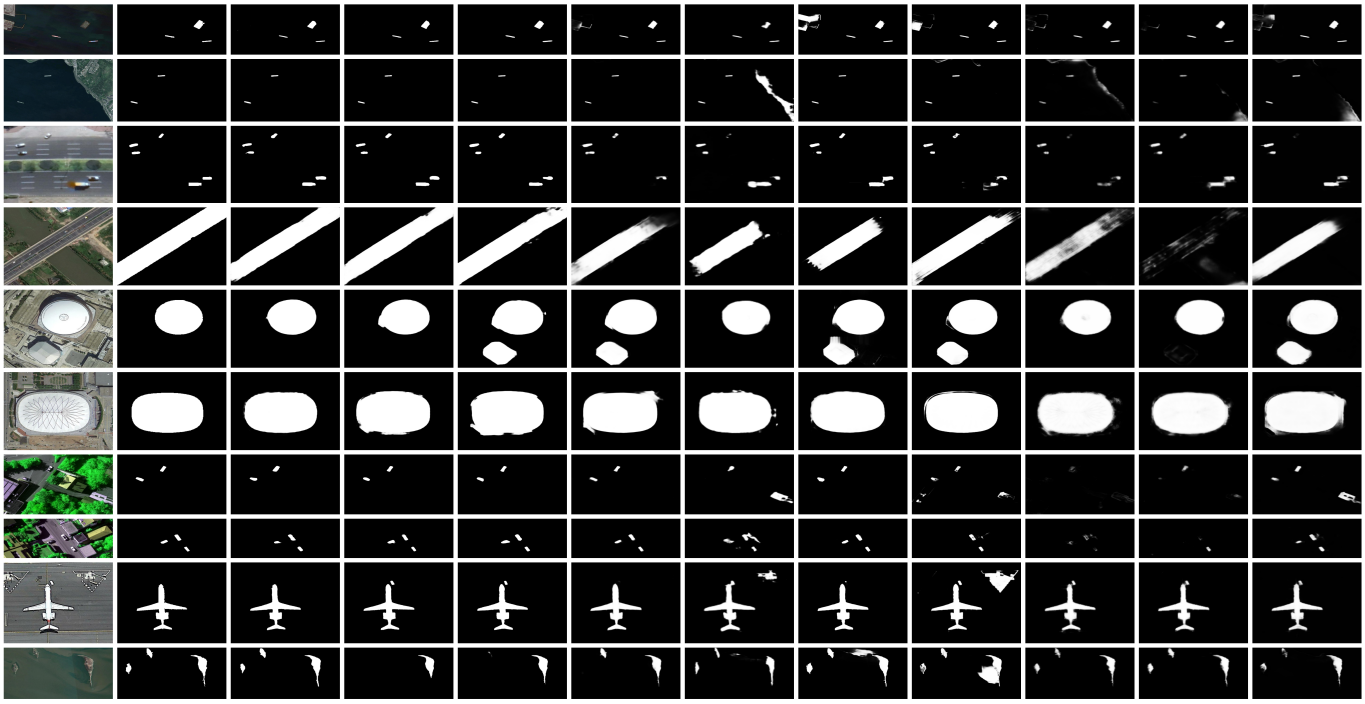


图4。与四类方法的定性比较，包括九种具有代表性的最先进方法。

具有更加理想的计算复杂度，在有效性和效率之间实现了平衡。这表明 SeaNet 是一种有望应用于实际场景的方法。

3）定性比较。在这里，我们在图4中展示了我们的 SeaNet 与四类方法的定性比较，这些方法包括九种具有代表性的最先进方法，涵盖四个独特且具有挑战性的 ORSI 场景。第一个场景包含多个目标，甚至多个微小目标，如图4的前三个案例所示。我们可以观察到，我们的 SeaNet 生成的显著图与专用光学遥感图像显著性目标检测方法（如 CorrNet 和 MCCNet）的显著图相似，能够准确突出显示所有显著目标，并且明显优于传统的轻量级 NSI-SOD 方法。这得益于 DSMM 的多尺度语义匹配。第二个场景包含一个较大的目标，如图4的第4至第6个案例所示。大多数方法都能定位该大目标，但无法完整地突出显示它；而我们的 SeaNet 能够突出显示整个大目标，并保留完整的边缘。这得益于 ESAM 基于边缘的细节增强。第三个场景包含杂乱的背景，如图4的第7和第8个案例所示。我们的 SeaNet 能够在复杂背景中成功找到显著目标，而三种轻量级方法要么遗漏目标（HVPNet 和 SAMNet），要么引入背景区域（CSNet）。最后一个场景包含具有复杂几何形状的目标，如图4的最后两个案例所示。得益于 DSMM 与 ESAM 的协同作用，与这九种方法相比，我们的 SeaNet 具有明显优势：它不仅能够克服小目标的干扰，还能以精细的细节准确定位三个岛屿。总体而言，我们的 SeaNet 能够达到专用光学遥感图像显著性目标检测方法的水平，并优于 NSI-SOD 方法。

表 III 评估两个轻量级模块贡献的消融实验结果。每列中的最佳结果以**粗体**标示。

模型	#参数量 (M)↓	浮点运算次数 (G)↓	EORSSD [38] $S_\alpha\uparrow$	$F_\beta^{\text{mean}}\uparrow$	$E_\xi^{\text{mean}}\uparrow$	$\mathcal{M}\downarrow$
SeaNet（我们的方法）	2.76	1.66	**0.9208**	**0.8519**	**0.9651**	**0.0073**
无 *DSMM*	2.75 -0.01	1.56 -0.10	0.9158	0.8440	0.9580	0.0093
无 *ESAM*	2.70 -0.06	1.59 -0.07	0.9180	0.8404	0.9588	0.0084

C. 消融研究

为评估 SeaNet 各组成部分的有效性，我们在 EORSSD 数据集上进行了全面的消融研究。具体而言，我们分析了：1）两个轻量级模块的贡献；2）ESAM 各组成部分的有效性；以及 3）DSMM 各组成部分的有效性。各变体的参数设置和数据集与第IV-A节相同。

1）两个轻量级模块的贡献。为分析两个轻量级模块的贡献，我们提供了两个变体：1）移除 DSMM 和 SKC（即。，不含 *DSMM*）和 2）移除 ESAM（即，不含 *ESAM*）。定量结果和计算复杂度见表III。

我们发现，这两个模块都很轻量，即，DSMM 与 SKC 的组合仅有 0.01M 个参数和 0.10G 浮点运算次数，而 ESAM 有 0.06M 个参数和 0.07G 浮点运算次数。由于 DSMM 能够确定显著目标的位置，不使用 *DSMM* 会降低目标定位的准确性，从而导致像素级评估指标大幅下降，即，$\mathcal{M}$: 0.0073→0.0093。不使用 *EASM* 只会影响显著目标的细节，因此性能下降通常

表 IV 将两个轻量级模块嵌入 HVPNet 和 SAMNet 的消融实验结果。

模型	#参数量 (M)↓	浮点运算次数 (G)↓	EORSSD [38] $S_\alpha\uparrow$	$F_\beta^{\text{mean}}\uparrow$	$E_\xi^{\text{mean}}\uparrow$	$\mathcal{M}\downarrow$
SeaNet（我们的方法）	2.76	1.7	0.9208	0.8519	0.9651	0.0073
SAMNet	1.33	0.5	0.8622	0.7214	0.8700	0.0132
SeaNet-SAM	1.51	1.4	0.9173	0.8507	0.9631	0.0074
HVPNet	1.23	1.1	0.8734	0.7377	0.8721	0.0110
SeaNet-HVP	1.48	1.8	0.9202	0.8486	0.9646	0.0069

不如不使用 *DSMM*严重，例如， $\mathcal{M}$:0.0073→0.0084 和 E_ξ^{mean}: 0.9651→0.9588。这两个轻量级模块与 MobileNet-V2 协同工作，使 SeaNet 无需大量参数和较高的计算成本即可实现良好性能。

特别是，为了进一步说明这两个轻量级模块的有效性、灵活性和鲁棒性，我们将这两个轻量级模块及解码器嵌入 SAMNet 和 HVPNet 提出的特征提取骨干网络中，分别构建了两个变体，命名为 SeaNet-SAM 和 SeaNet-HVP。值得注意的是，我们的解码器包含 0.47M 个参数，浮点运算次数为 0.95G。如表IV所示，SeaNet-SAM 比原始 SeaNet 更轻量。SeaNet-HVP 的参数量显著少于原始 SeaNet，但浮点运算次数略高。两个变体的性能均与原始 SeaNet 相当，这表明这两个轻量级模块能够适配不同的骨干网络，且具有良好的鲁棒性。此外，SeaNet-SAM/SeaNet-HVP 的参数量与 SAMNet/HVPNet 相近，但浮点运算次数更高，主要是由解码器造成的。SeaNet-SAM/SeaNet-HVP 的性能显著优于 SAMNet/HVPNet，例如在F_β^{mean}上领先超过 10%。值得注意的是，SeaNet-SAM（1.51M 和 1.4G）与 HVPNet（1.23M 和 1.1G）的参数量和浮点运算次数相近，但 SeaNet-SAM 的表现远胜于 HVPNet。这一情况证明，当 SeaNet 变体的参数量和浮点运算次数与其他轻量级方法相当或相近时，SeaNet 变体仍显著优于它们。

2）DSMM 各组件的有效性。为分析 DSMM 各组件的有效性，我们在表 V中设计了 DSMM 的三种变体，并将其嵌入网络：1）移除空间语义匹配（*i.e., w/oSM*）；2）将空洞 DD 卷积改为常规 DD 卷积（*i.e., w/odilation*）；3）移除 DSMM 的通道相关性（*i.e.,w/o CCorr1*）。

根据表格III顶部的定量性能，我们发现 DSMM 的每个组件都是必不可少的。作为 DSMM 的关键部分，去除 *SM* 会削弱 DSMM 的目标定位能力，在这三个变体中表现最差，例如， E_ξ^{mean}: 0。9651→09595。去除膨胀卷积 会损害 DSMM 感知不同尺寸显著目标的能力，导致性能略有下降，

表 V 评估 DSMM 和 ESAM 各组件有效性的消融实验结果。每列中的最佳结果以**粗体**显示。

	模型	EORSSD [38] $S_\alpha\uparrow$	$F_\beta^{\text{均值}}\uparrow$	$E_\xi^{\text{mean}}\uparrow$	$\mathcal{M}\downarrow$
	SeaNet（我们的方法）	**0.9208**	**0.8519**	**0.9651**	**0.0073**
DSMM	无语义匹配	0.9170	0.8459	0.9595	0.0078
DSMM	无膨胀	0.9194	0.8490	0.9626	0.0076
DSMM	无通道相关性*1*	0.9187	0.8471	0.9619	0.0080
EASM	无边缘增强单元	0.9183	0.8441	0.9597	0.0078
EASM	无对齐	0.9189	0.8476	0.9617	0.0078
EASM	无通道相关性*2*	0.9197	0.8494	0.9627	0.0076

例。如，在 F_β^{mean} 上下降了 0.29%。去除通道相关性*1* 仅关注空间层面的特征交互，却忽略了通道层面的特征交互，导致特征交互不完整，性能下降。因此，上述分析证明，我们的 DSMM 设计合理且有效。

3）EASM 各组件的有效性。为分析 EASM 各组件的有效性，我们在表 V 中设计了 EASM 的三个变体，并将其嵌入网络中：1）移除两个 EEU（即，不含 *EEU*）；2）移除边缘对齐（即，移除式 10中的边缘对齐损失，记为不含对齐）；3）移除 EASM 的通道级相关性（即，不含 *CCorr2*）。

根据表 III底部的定量结果，我们观察到 EASM 的各个组件均有助于提升最终性能。两个 EEU 负责增强特征中的边缘区域，因此不含 *EEU* 时无法取得令人满意的性能，例如， F_β^{mean}: 0。8519→0.8441。所提出的边缘对齐是一种有趣的机制，能够在不增加参数量和浮点运算次数的情况下提高边缘信息的准确性。不含对齐 会导致四项指标全部下降。与不含 *CCorr1*类似，不含 *CCorr2* 也舍弃了 EASM 中的通道级特征交互，从而影响性能。结合不含 *CCorr1* 和不含 *CCorr2*的结果可以得出，通道级特征交互对 SeaNet 至关重要，不可舍弃。以上分析表明，ESAM 的这些组件缺一不可。

V. 结论

本文旨在对低层特征和高层特征进行差异化处理，从而提出一种名为 SeaNet 的高效轻量级 ORSI-SOD 解决方案。针对高层特征，我们提出轻量级 DSMM，通过空间语义匹配探索目标位置，并考虑通道级特征交互。空间语义匹配利用空洞DD卷积感知多个显著目标以及不同尺寸的目标，从而更好地适应光学遥感图像中的复杂场景。针对低层特征，我们提出轻量级 ESAM，基于边缘信息增强显著目标的细节，并以创新的自对齐方式对边缘信息进行校正。通过这两个模块的紧密协作

SeaNet 能够在解码器中快速、准确地推断显著目标。性能分析和消融研究表明，与最先进方法相比，SeaNet 具有良好的有效性和效率。

参考文献

[1] R. Cong、J. Lei、H. Fu、M.-M. Cheng、W. Lin 和 Q. Huang，“综合信息视觉显著性检测综述”，IEEE电路与系统视频技术汇刊，第29卷，第10期，第2941–2959页，2019年10月。[2] K. Gu、S. Wang、H. Yang、W. Lin、G. Zhai、X. Yang 和 W. Zhang，“显著性引导的屏幕内容图像质量评估”，IEEE多媒体汇刊，第18卷，第6期，第1098–1110页，2016年6月。[3] S. Yang、Q. Jiang、W. Lin 和 Y. Wang，“SGDNet：用于无参考图像质量评估的端到端显著性引导深度神经网络”，载于 ACM多媒体会议论文集，2019年10月，第1383–1391页。[4] G. Li、Z. Liu、R. Shi 和 W. Wei，“基于深度神经网络的约束注视点分割”，神经计算，第368卷，第180–187页，2019年11月。[5] G. Li、Z. Liu、R. Shi、Z. Hu、W. Wei、Y. Wu、M. Huang 和 H. Ling，“结合目标定位与边界保持的基于个人注视的目标分割”，IEEE图像处理汇刊，第30卷，第1461–1475页，2021年1月。[6] P. Zhang、W. Liu、D. Wang、Y. Lei、H. Wang 和 H. Lu，“通过深度多尺度时空判别显著图进行非刚性目标跟踪”，模式识别，第100卷，第107130页，2020年4月。[7] W. Wang、Q. Lai、H. Fu、J. Shen、H. Ling 和 R. Yang，“深度学习时代的显著目标检测：深入综述”，IEEE模式分析与机器智能汇刊，第44卷，第6期，第3239–3259页，2022年6月。[8] G. Li、Z. Liu 和 H. Ling，“ICNet：用于RGB-D显著目标检测的信息转换网络”，IEEE图像处理汇刊，第29卷，第4873–4884页，2020年3月。[9] G. Li、Z. Liu、L. Ye、Y. Wang 和 H. Ling，“用于RGB-D显著目标检测的跨模态加权网络”，载于 欧洲计算机视觉会议论文集，2020年8月，第665–681页。[10] W. Wang、J. Shen、J. Xie、M.-M. Cheng、H. Ling 和 A. Borji，“重新审视深度学习时代的视频显著性预测”，IEEE模式分析与机器智能汇刊，第43卷，第1期，第220–237页，2021年1月。[11] Y. LeCun、B. Boser、J. S. Denker、D. Henderson、R. E. Howard、W. Hubbard 和 L. D. Jackel，“应用于手写邮政编码识别的反向传播”，神经计算，第1卷，第4期，第541–551页，1989年12月。[12] B. Xu、H. Liang、R. Liang 和 P. Chen，“全局定位、局部分割：一种用于显著目标检测的带知识回顾网络的渐进式架构”，载于 人工智能促进协会年会论文集，2021年2月，第3004–3012页。[13] Y. Liu、Y.-C. Gu、X.-Y. Zhang、W. Wang 和 M.-M. Cheng，“通过层次化视觉感知学习实现轻量级显著目标检测”，IEEE控制论汇刊，第51卷，第9期，第4439–4449页，2021年9月。[14] G. Li、Z. Liu、W. Lin 和 H. Ling，“用于光学遥感图像显著目标检测的多内容互补网络”，IEEE地球科学与遥感汇刊，第60卷，第1–13页，2022年。[15] G. Li、Z. Liu、Z. Bai、W. Lin 和 H. Ling，“通过特征相关性实现光学遥感图像中的轻量级显著目标检测”，IEEE地球科学与遥感汇刊，第60卷，第1–12页，2022年。[16] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于 医学图像计算与计算机辅助干预会议论文集，2015年10月，第234–241页。[17] Q. Hou、M.-M. Cheng、X. Hu、A. Borji、Z. Tu 和 P. Torr，“通过短连接进行深度监督显著目标检测”，载于 IEEE计算机视觉与模式识别会议论文集，2017年7月，第5300–5309页。[18] 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 欧洲计算机视觉会议论文集，2020年8月，第35–51页。[19] J. Zhao、J.-J. Liu、D.-P. Fan、Y. Cao、J. Yang 和 M.-M. Cheng，“EGNet：显著目标检测的边缘引导网络”，载于 IEEE国际计算机视觉大会论文集，2019年10月，第8779–8788页。[20] H. Zhou、X. Xie、J.-H. 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 IEEE计算机视觉与模式识别会议论文集，2020年6月，第9138–9147页。[21] C. Li、R. Cong、J. Hou、S. Zhang、Y. Qian 和 S. Kwong，“用于光学遥感图像显著目标检测的双流金字塔嵌套网络”，IEEE地球科学与遥感汇刊，第57卷，第11期，第 9156–9166 页，2019 年 11 月。[22] X. Zhou、K. Shen、Z. Liu、C. Gong、J. Zhang 和 C. Yan，“用于光学遥感图像显著目标检测的边缘感知多尺度特征集成网络”，IEEE地球科学与遥感汇刊，第60卷，第1–15页，2022年。[23] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen，“MobileNetV2：倒置残差和线性瓶颈”，发表于 IEEE计算机视觉与模式识别会议论文集，2018年6月，第4510–4520页。[24] A. G. Howard、M. Zhu、B. Chen、D. Kalenichenko、W. Wang、T. Weyand、M. Andreetto 和 H. Adam，“MobileNets：面向移动视觉应用的高效卷积神经网络”，arXiv预印本 arXiv:1704.04861，2017年。[25] X. Hu、L. Zhu、J. Qin、C.-W. Fu 和 P.-A. Heng，“循环聚合深层特征以进行显著目标检测”，发表于 人工智能促进协会年会论文集，2018年2月，第6943–6950页。[26] Z. Deng、X. Hu、L. Zhu、X. Xu、J. Qin、G. Han 和 P.-A. Heng，“R^{3}Net：用于显著性检测的循环残差细化网络”，发表于 IJCAI论文集，2018年7月，第684–690页。[27] 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”，发表于 IEEE计算机视觉与模式识别会议论文集，2019年6月，第1623–1632页。[28] X. Qin、Z. Zhang、C. Huang、C. Gao、M. Dehghan 和 M. Jagersand，“BASNet：边界感知显著目标检测”，发表于 IEEE计算机视觉与模式识别会议论文集，2019年6月，第7479–7489页。[29] M. S. Lee1、W. Shin 和 S. W. Han，“TRACER：极致注意力引导的显著目标追踪网络”，发表于 人工智能促进协会年会论文集，2022年。[30] Y. Pang、X. Zhao、L. Zhang 和 H. Lu，“用于显著目标检测的多尺度交互网络”，发表于 IEEE计算机视觉与模式识别会议论文集，2020年6月，第9410–9419页。[31] 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”，发表于 人工智能促进协会年会论文集，2020年2月，第10 599–10 606页。[32] J. Li、Z. Pan、Q. Liu 和 Z. Wang，“具有通道注意力的堆叠式U形显著目标检测网络”，IEEE多媒体汇刊，第23卷，第1397–1409页，2021年。[33] Y. Y. Ke 和 T. Tsubono，“用于精准显著目标检测的递归轮廓显著性融合网络”，发表于 IEEE冬季计算机视觉应用会议论文集，2022年1月，第1360–1370页。[34] C. Li、R. Cong、C. Guo、H. Li、C. Zhang、F. Zheng 和 Y. Zhao，“用于光学遥感图像显著目标检测的并行下采样-上采样融合网络”，神经计算，第415卷，第411–120页，2020年11月。[35] Z. Huang、H. Chen、B. Liu 和 Z. Wang，“用于光学遥感图像显著目标检测的语义引导注意力细化网络”，遥感，第13卷，第11期，第2163页，2021年5月。[36] R. Cong、Y. Zhang、L. Fang、J. Li、Y. Zhao 和 S. Kwong，“RRNet：用于光学遥感图像显著目标检测的并行多尺度注意力关系推理网络”，IEEE地球科学与遥感汇刊，第60卷，第1–11页，2022年。[37] G. Li、Z. Liu、D. Z. Lin 和 H. Ling，“用于光学遥感图像显著目标检测的邻近上下文协调网络”，IEEE控制论汇刊，第53卷，第1期，第526–538页，2023年1月。[38] Q. Zhang、R. Cong、C. Li、M.-M. Cheng、Y. Fang、X. Cao、Y. Zhao 和 S. Kwong，“用于光学遥感图像显著目标检测的密集注意力流体网络”，IEEE图像处理汇刊，第30卷，第1305–1317页，2021年。[39] Z. Tu、C. Wang、C. Li、M. Fan、H. Zhao 和 B. Luo，“基于多尺度区域与边界联合模型的 ORSI 显著目标检测”，IEEE地球科学与遥感汇刊，第60卷，第1–13页，2022年。[40] S.-H. Gao、Y.-Q. Tan、M.-M. Cheng、C. Lu、Y. Chen 和 S. Yan，“仅用10万参数实现高效显著目标检测”，发表于 欧洲计算机视觉会议论文集，2020年8月，第702–721页。[41] Y. Liu、X.-Y. Zhang、J.-W. Bian、L. Zhang 和 M.-M. Cheng，“SAMNet：用于轻量级显著目标检测的立体注意力多尺度网络”，IEEE图像处理汇刊，第30卷，3814，2021年 3804页。– 。[42] K. Simonyan 和 A. Zisserman，“用于大规模图像识别的超深卷积网络”，发表于 ICLR论文集，2015年5月，第1–14页。[43] Y.-H. Wu、Y. Liu、J. Xu、J.-W. Bian、Y.-C. Gu 和 M.-M. Cheng，“MobileSal：极高效的RGB-D显著目标检测”，IEEE模式分析与机器智能汇刊，第44卷，第12期，第10 261–10 269页，2022年12月。[44] Y. Zhang、J. Zhang、Q. Wang 和 Z. Zhong，“DyNet：用于加速卷积神经网络的动态卷积”，arXiv预印本 arXiv:2004.10694，2020年。[45] J. Lu、J. Yang、D. Batra 和 D. Parikh，“用于视觉问答的层次化问题-图像协同注意力”，发表于 NeurIPS论文集，2016年12月，第289–297页。

[46] X. Lu、W. Wang、J. Shen、D. Crandall 和 J. Luo，“利用协同注意力孪生网络进行零样本视频目标分割”，IEEE模式分析与机器智能汇刊，第44卷，第4期，第2228–2242页，2022年4月。[47] G. Li、Z. Liu、M. Chen、Z. Bai、W. Lin 和 H. Ling，“用于RGB-D显著目标检测的层次交替交互网络”，IEEE图像处理汇刊，第30卷，第3528–3542页，2021年3月。[48] Z. Bai、Z. Liu、G. Li 和 Y. Wang，“用于协同显著性检测的自适应组间一致性网络”，IEEE多媒体汇刊，2021年。 [ ] 10 1109 2021 3138246 在线。可获取：。/ [49] TMM。。F. Yu 和 V. Koltun，“利用空洞卷积进行多尺度上下文聚合”，发表于ICLR论文集，2016年5月，第1–13页。[50] J.-J. Liu、Q. Hou、M.-M. Cheng、J. Feng 和 J. Jiang，“一种用于实时显著目标检测的简易平均池化设计”，发表于IEEE计算机视觉与模式识别会议论文集，2019年6月，第3912–3921页。[51] N. Srivastava、G. Hinton、A. Krizhevsky、I. Sutskever 和 R. Salakhutdinov，“Dropout：一种防止神经网络过拟合的简易方法”，机器学习研究杂志，第15卷，第1期，第1929–1958页，2014年6月。[52] K. He、X. Zhang、S. Ren 和 J. Sun，“深入探究整流器：超越 ImageNet分类的人类水平表现”，发表于 IEEE国际计算机视觉大会论文集，2015年12月，第1026–1034页。[53] D.-P. Fan、M.-M. Cheng、Y. Liu、T. Li 和 A. Borji，“结构度量：评估前景图的新方法”，发表于 IEEE国际计算机视觉大会论文集，2017年10月，第4548–4557页。[54] R. Achanta、S. Hemami、F. Estrada 和 S. Susstrunk，“频率调谐显著区域检测”，发表于 IEEE计算机视觉与模式识别会议论文集，2009年6月，第1597–1604页。[55] D.-P. Fan、C. Gong、Y. Cao、B. Ren、M.-M. Cheng 和 A. Borji，“用于二值前景图评估的增强对齐度量”，发表于 IJCAI论文集，2018年7月，第698–704页。[56] A. Paszke、S. Gross、F. Massa、A. Lerer、J. Bradbury、J. Chanan、T. Killeen、Z. Lin、N. Gimelshein、L. Antiga、A. Desmaison、A. Kopf、E. Yang、Z. DeVito、M. Raison、A. Tejani、S. Chilamkurthy、B. Steiner、L. Fang、J. Bai 和 S. Chintala，“PyTorch：一种命令式风格的高性能深度学习库”，发表于 NeurIPS论文集，2019年12月，第8024–8035页。[57] D. P. Kingma 和 J. L. Ba，“Adam：一种随机优化方法”，发表于 ICLR论文集，2015年5月，第1–15页。