

结构度量：一种新的评估前景图方法

程明明 · 范登平

Received: date / Accepted: date

摘要 前景图的度量对于物体分割算法的发展至关重要,特别是在显著物体检测领域,其目的是准确地检测和分割出场景中最显著的对象,现有的多种度量方法(例如 AUC、F1 分数、平均精度等)已经被用于评估前景图和 GT 图之间的相似性。现有的评估基于像素级误差,经常忽略结构相似性。然而,行为视觉研究表明,人类视觉系统对场景中的结构高度敏感。由此,本文提出了一种新颖、高效(每幅图像 0.005 秒)且易于计算的指标 **S-指标**(结构度量)来评估前景图。本文的新指标同时评估了前景图和 GT 图之间的区域信息和对象结构相似性。在 5 个广泛使用的基准数据集上使用 4 个元指标来证明该指标优于现有的指标。此外,本文对一个新数据库进行了行为评估研究。来自 45 名受试者的数据显示,一般情况下他们更喜欢由 S-指标选择得到的显著图,而不是由当前最先进的指标选择的显著图。本文的实验结果为前景图评估提供了新的思路,在前景图评估中,当前的指标无法真正评估模型的优缺点。代码:

<https://github.com/DengPingFan/S-measure>

Keywords S-指标 · 结构度量 · 前景图 · 评估 · 显著性目标检测

程明明和范登平就职于中国南开大学计算机科学学院。(Email: cmm@nankai.edu.cn; dengpfan@gmail.com)
这项工作的初始版本发表于 ICCV2017 [19]
本文为 IJCV2021 论文 [13] 的中译版。

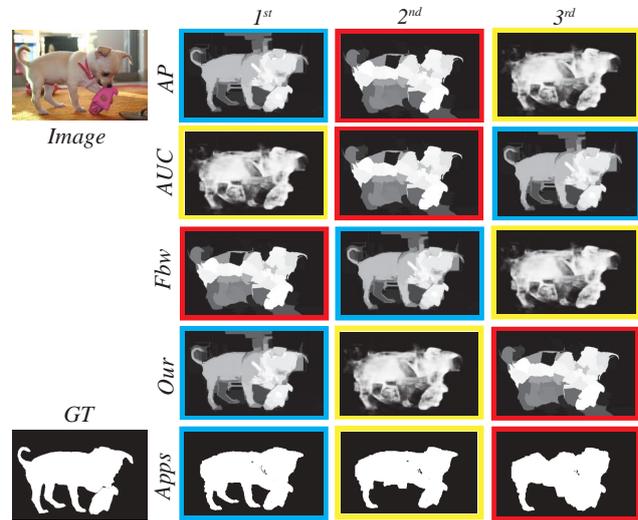


图 1 当前的指标评价不准确。比较 3 种最先进的显著物体检测算法产生的显著图的排序: DISC [15], MDF [44], 和 MC [84]。根据应用程序的排序(最后一行,在 Apps-章节. 5 查看细节),蓝色边框显著图排在第 1 位,其次是黄色和红色边框显著图。蓝色边框的显著图最准确地捕获了狗的结构。黄色边框的显著图看起来较模糊,虽然狗的整体轮廓仍然存在。红边显著图则几乎完全破坏了狗的结构。令人惊讶的是,所有基于像素误差(前 3 行)的指标都无法正确排序这些显著图,唯独本文新的指标(第 4 行)以正确的顺序排列了这三张显著图。

1 简介

在评估和比较各种计算机视觉算法的应用中,评估预测前景图(FM)与真值图(GT)的相似度是至关重要

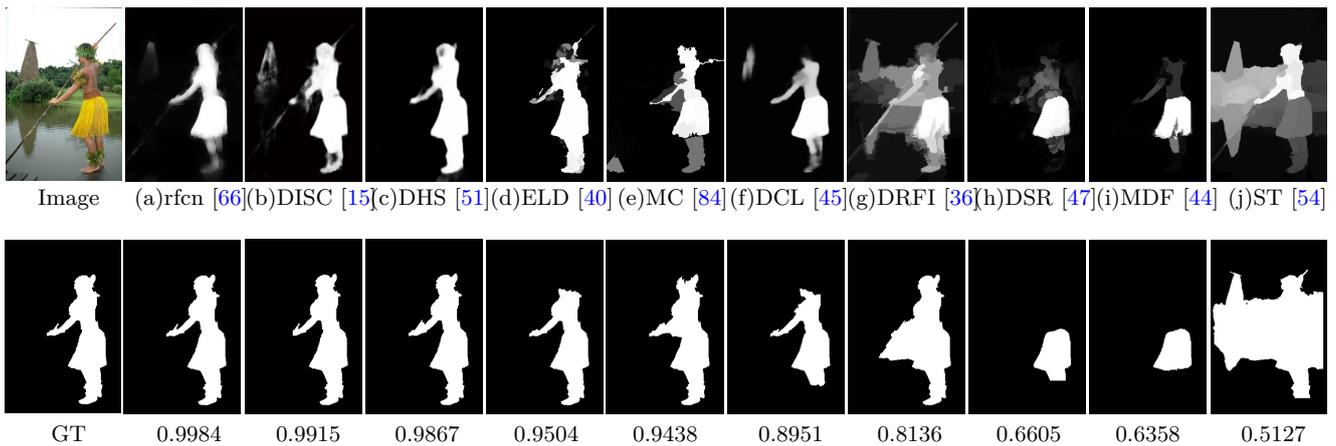


图 2 10 个显著性检测算法得到的结果图 (第 1 行) 作为 SalCut [17] 输入, 输出结果 (第 2 行) 用结构度量 ($\lambda = 0.5, K = 4$) 评估。排名结果清楚地表明, 本文的方法善于捕捉对象结构, 并可以提供一个可靠的排名。

的, 例如 AR COPY & PASTE¹ [63], 目标检测和识别 [6, 8, 32, 38, 64], 视频归纳 [28], 视频压缩 [31, 33], 图像分割 [70], 基于内容的图像检索 [46, 49, 50], 视觉跟踪, 照片合成 [14], 图像-文本匹配 [88], 图像采集浏览 [14] 等, 作为一个具体的例子, 这里重点关注显著目标检测 [5–7, 10, 12, 30, 43, 52, 65, 67, 71, 76, 77, 79, 82], 尽管方法是通用的, 可以用于其他目的, 但是值得注意的是, 显著对象不一定是前景对象 [9, 26].

GT 图通常是几个手动标注的平均值。因此, 它可以是二进制也可以是非二进制的。类似地, 预测的前景图也可以是二进制或非二进制的, 所以评估指标可以分为两种类型:

1. **二进制图评估:** 常见的方法包括 F_β -指标 [3, 17, 53] 和 PASCAL's VOC 分割指标 [18]。
2. **非二进制图的评估:** 三种传统的受欢迎的指标包括曲线下面积 (AUC), PR 曲线和平均精度 (AP) [18]. 加权 F_β -指标 (Fbw) [56] 作为一种新指标被用来弥补 AUC, PR 和 AP 的缺陷。

通常情况下我们希望前景图包含完整的对象结构。因此, 本文希望评价方法能够指导模型生成一个更完整的对象, 例如, 在图. 1(第一行) 蓝边的图比红色边框的图更好。而后面一个例子中, 狗的形状已经严重退化到难以从它的分割图猜测对象类别。令人惊讶的是, 所有

当前的评估指标都未能正确地对这些结果图进行排名 (就保存结构而言)。

本文使用了 10 个最先进的 (SOTA) 显著性检测算法, 获得 10 个显著性图 (图. 2 第一行), 然后将这些显著图提供给 SalCut² [17] 算法以生成相应的二进制显著图 (第 2 行)。最后, 使用本文的 **S-指标** 对这些图进行排名。评估值越低, 表示这个人的整体结构受到的破坏越大 (e 到 j 列)。这个实验清楚地表明, 新评估值强调了物体的整体结构。在这 10 个二进制图 (第二行) 中, 有 6 个图的得分低于 0.95, 比例为 60%。采用相同的阈值 (0.95), 可以发现 (*i.e.*, ECSSD [69], HKU-IS [44], PASCAL-S [48] 和 SOD [57]) 这 4 个流行显著性数据集的图像被破坏的比例分别为 66.80%、67.30%、81.82% 和 83.03%。用 F_β 指标评价二元图的比例分别为 63.76%、65.43%、78.32% 和 82.67%。这意味着本文的度量在保持对象结构上比 F_β 指标更严格。

为了解决现有评估方法的问题 (即对整个对象结构的灵敏度较低), 本文提出了一种基于两个观测的结构敏感相似性度量方法:

1. **区域角度:** 虽然很难描述前景显著图的物体结构, 但是注意到, 一个物体的整体结构可以通过组合物体各个部分 (区域) 的结构来很好地表达。

¹ <https://clipdrop.co/>

² <https://github.com/MCG-NKU/SalBenchmark/blob/master/Code/CmLib/Saliency/CmSalCut.cpp>

2. **物体角度:** 在高质量的前景图中, 前景区域与背景区域形成鲜明的对比。这些区域通常有近似均匀的对比度分布。

因此, 本文的结构测度包括区域感知结构相似性指标和对象感知结构相似性指标两部分。区域感知测度试图通过结合对象各部分的结构信息来捕获全局对象结构。图像质量评价 (IQA) 社区对区域的结构相似性进行了很好的研究 [68]。目标相似度量试图比较前景图和 GT 图中前景和背景区域的全局分布。

本文的度量方法与 AP、AUC、PASCAL、Fbw、 F_β -指标等多种现有度量方法进行了比较, 这些方法适用于一些广泛使用的显著目标检测基准, 包括 ASD [1], SOD [57], ECSSD [69], PASCAL-S [48] 和 HKU-IS [44]。大量的实证研究表明, 结构度量不仅提供了可靠的评价, 而且显著改善了现有指标的成绩。

这个工作是本文之前的 ICCV2017 版本 [19] 的扩展, 这两个版本的主要区别在于: (1) 本文将初始版本扩展到二进制前景图评估, 提供了一个既适用于二进制前景图也适用于非二进制前景图的统一评估。本文的工作为前景图评估提供了新的方式, 目前的度量方法未能充分评估模型的优缺点。(2) 本文提供了一组新的实验来验证本文的指标的有效性、鲁棒性和可扩展性, 这些扩展关注于非二进制显著图的评估。此外, 还在附录中给出了应用排名框架的更多细节。(3) 构建了具有代表性的显著性检测的在线基准和模型库, 将各种公开可用的具有统一输入/输出格式的显著性数据集 (GT 图像为 JPEG; PNG 格式)。(4) 本文还为现有的基准测试工作提供了 python³和 Matlab⁴ 版本代码, 这有利于许多相关任务和计算机视觉社区。(5) 本文做了很多尝试来改进论文的陈述和组织。首先, 添加或重构几个新数据, 以更好地说明元指标和这项工作的关键结论。其次, 增加了几个新的章节来描述更多关于现有方法缺陷的细节, 为 S-指标提供更多的理论支持和更多的理论细节, 包括区域感知 (章节. 4.1) 和对象感知 (章节. 4.2) 结构

相似度评估。新的内容将让读者更好地理解本文的评估方法。

2 现有的评估方法

前景图可以由各种算法生成 (例如, 显著性检测或目标分割)。显著性检测算法通常生成非二值图, 而目标分割算法通常生成二值图。因此, 前景图可以分为值在 [0,1] 范围内的非二进制图或值为 0 或 1 的二进制图, 每个值表示特定像素属于前景的概率 [56, 60]。

2.1 二进制图的评估

为了评估二元显著图, 从预测混淆矩阵中计算出四个值: 真阳性 (TP)、真阴性 (TN)、假阳性 (FP) 和假阴性 (FN)。然后使用这些值计算三个比率: 真阳性率 (TPR) 或召回率、假阳性率 (FPR) 和精度:

$$Recall = TPR = \frac{TP}{TP + FN} \quad (1)$$

$$FPR = \frac{FP}{TN + FP} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

精度和召回率组合可以计算经典的 F_β -指标。

$$\begin{aligned} F_\beta &= \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall} \\ &= \frac{(1 + \beta^2)TP}{(1 + \beta^2)TP + \beta^2FN + FP} \end{aligned} \quad (4)$$

其中 β 是一个用来平衡准确性和召回率的参数 (通常 $\beta = 1$ 为调和平均数)。另一个常用的二进制图评价指标是 PASCAL 指标:

$$PASCAL = \frac{TP}{TP + FN + FP} \quad (5)$$

³ <https://github.com/DengPingFan/eval-co-sod>

⁴ <https://dpfan.net/One-KeyEvaluation/>

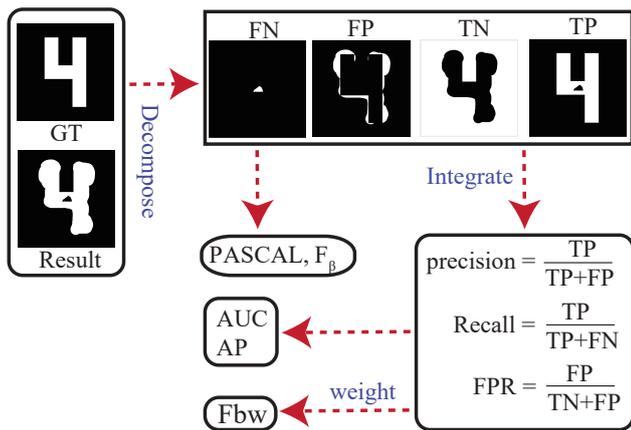


图 3 当前评估指标的框架。AP、AUC 和 Fbw 评价指标的计算方法类似，它们都以像素方式计算，并且忽略了结构上的相似性。

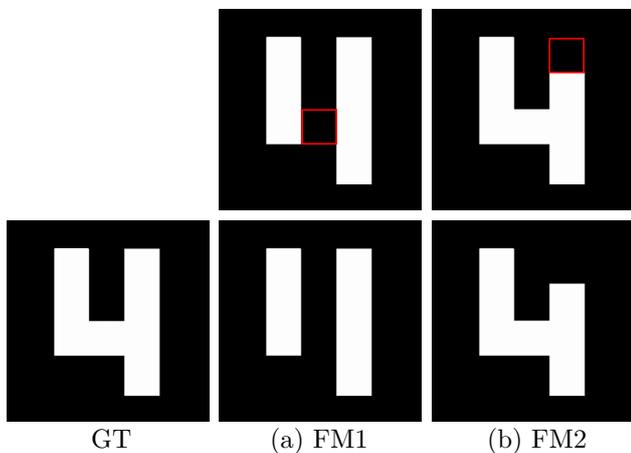


图 4 一个基于像素的计算示例。在 FM1 中，一个黑色方块落在数字的内部，而在 FM2 中它触及边界（第一行）。它们是两个具有相同 TP、TN、FP 和 FN 值的二进制显著图（第二行）。从视觉上看，FM2 比 FM1 更受欢迎，因为 FM1 破坏前景图的结构更剧烈，以至于数字很难识别。当前评估平均绝对误差 (MAE) 是按像素计算的方式，并独立处理像素。因此，它忽略了前景图的结构，因此它更偏好 FM1 而不是 FM2。

2.2 非二进制图的评估

AUC 和 AP 是两种公认的评价指标，生成非二进制图的算法采用三个步骤来评估模型预测（非二进制图）和标注 (GT) 之间的一致性。首先，对非二进制显著图应用多个阈值，得到多个二进制显著图；其次，将这些二进制显著图与真值进行比较，得到一组 TPR（参见式. (1)）和 FPR（参见式. (2)）的值，这些值被绘制在一个 2D 图（即 ROC 曲线），AUC 从中提取曲线下的面积。

AP 的计算方法与此类似。可以通过把召回率 r 看作精度的一个函数 $p(r)$ 得到一个精度（参见式. (3)）&

召回（参见式. (1)）曲线。AP 指标 [18] 是 $p(r)$ 在 $r = 0$ 到 $r = 1$ 的等间距 x 轴点上的平均值。

最近，一种称为 Fbw [56] 的指标给出了 F_β 指标的直观概括。Fbw 的作者指出了 AP 和 AUC 度量不准确的三个原因。为了改善这些缺陷，他们 1) 将 TP、TN、FP 和 FN 四个基本量扩展为非二进制值，2) 根据不同的位置和邻域信息，对不同的误差赋予不同的计算权重 (w)。

$$F_\beta^\omega = \frac{(1 + \beta^2) \text{Precision}^\omega \times \text{Recall}^\omega}{\beta^2 \cdot \text{Precision}^\omega + \text{Recall}^\omega} \quad (6)$$

虽然这种方法在其他方法上有所改进，但有时它不能正确地对前景图进行排序（见图. 1 的第三行）。在下一节中，本文将分析当前指标未能正确地对这些图进行排名的原因。

3 目前基于像素的评估

传统的度量方法 (AP、AUC、PASCAL、 F_β 和 Fbw) 依赖于两种类型的基本误差: FN、FP。由于这些基本误差是按像素计算的（见图. 3），不能完全捕获前景图的结构信息。然而，具有精细结构细节的前景图在一些应用（如图像检索、目标检测和分割）中往往是需要的。因此，对前景结构敏感的评价指标受到人们青睐。不过，上述指标未能满足这一期望。

图. 4 显示了一个虚构的例子，其中包含两种不同类型的前景图。在 FM1 中，一个黑色方块落在数字的内部，而在 FM2 中，它触及数字的边界。在人类看来，FM1 不如 FM2 受欢迎，因为它更彻底地破坏前景图。然而，目前的像素级 MAE（平均绝对误差）度量更倾向于 FM1 而不是 FM2。这似乎与常识相矛盾。

一个更现实的例子如图. 5 所示，这里的蓝边图比红边图更好地捕捉了金字塔，因为后者提供了一个模糊的检测图，主要突出了金字塔的顶部，而忽略了其余部分。从应用的角度来看（第三行 SalCut 算法的输出由显著图提供；第二行根据本文的指标），蓝边图提供了金字塔的完整形状。实际上，这种情况很常见。因此，如

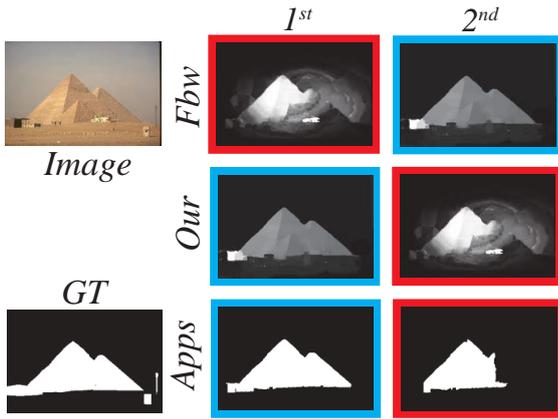


图 5 一个基于像素的示例。利用 DSR [47] 和 ST [54] 两种显著性检测算法生成两幅前景图。根据应用程序的排名和用户研究 (Apps-章节: 5; 最后一行), 蓝边图表现最好, 其次是红边图。由于 Fbw 指标没有考虑结构相似性, 所以结果是完全相反的。本文的指标 (第二行) 正确地将蓝边图排在更高的位置。

果评价指标不能捕获结构对象的信息, 就不能为应用中的模型选择提供可靠的信息。

4 本文的指标

本节将介绍评估前景图的新方法, 在图像质量评估 (IQA) 领域, 一种被称为结构相似度指标 (SSIM) [68] 的指标被广泛用于捕获原始图像和测试图像的结构相似度。

令 $x = \{x_i, i = 1, 2, \dots, N\}$ 以及 $y = \{y_i, i = 1, 2, \dots, N\}$ 分别为 FM 和 GT 像素值, 其中 $\bar{x}, \bar{y}, \sigma_x, \sigma_y$ 分别为 x 和 y 的均值和标准差, σ_{xy} 是两者之间的协方差 SSIM 由三个指标亮度、对比度和结构决定:

$$ssim = \frac{2\bar{x}\bar{y} + C_1}{(\bar{x})^2 + (\bar{y})^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (7)$$

其中, C_1, C_2 和 C_3 被设置为非常小的值, 以避免分母 (例如, $\bar{x}^2 + \bar{y}^2$) 非常接近零时的不稳定性。

在式 (7) 中, 前两项分别表示亮度比较和对比度比较, 两者越接近 (如 \bar{x} 和 \bar{y} , 或 σ_x 和 σ_y), 亮度或对比度的比较值就越接近于 1。图像中物体的结构与受光照和反射率影响的亮度无关。因此, 结构比较公式的设

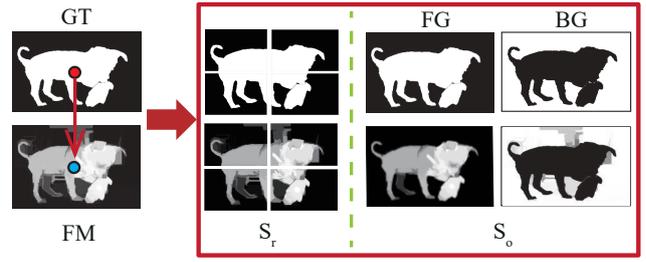


图 6 S-指标的框架。

计应独立于亮度和对比度。SSIM [68] 将两个单位向量 $(x - \bar{x})/\sigma_x$ 和 $(y - \bar{y})/\sigma_y$ 与这两个图像的结构联系起来, 由于这两个向量之间的相关性等价于 x 和 y 之间的相关系数, 因此结构比较公式由式 (7) 式的第三项表示。

在构建显著性目标检测或目标分割算法时, 研究人员往往更关注前景目标的结构。因此, 本文的结构度量结合了区域感知和对象感知的结构相似性。区域感知的结构相似性执行类似于 [68] 的操作, 其目的是捕获“目标部分”结构信息, 而不特别关注前景的完整性。另一方面, 对象感知的结构相似性主要是为了捕捉“目标整体”的结构信息, 该结构信息以整个对象为中心。

4.1 区域感知结构相似度

这一节中解释了如何评估区域感知的相似性。设计区域感知相似度来评估 FM 中目标-局部结构与 GT 图的相似度。首先, 使用水平和垂直的切线将 FM 和 GT 图分成四个区域, 这些切线在真值前景的中心相交。然后, 递归划分子图 [39]。块的总数为 K 。示例如图 6 所示。每个块的区域感知相似度 $ssim(k)$ 由式 (7) 单独计算, 为每个块分配不同的权重 (w_k), 与该块覆盖的真值前景区域成比例, 由此, 区域感知的结构相似性指标可以表示为:

$$S_r = \sum_{k=1}^K w_k \times ssim(k). \quad (8)$$

研究表明, 提出的 S_r 可以很好地描述 FM 和 GT 图之间的目标部分相似度。本文还尝试直接使用 SSIM 来评估 FM 和 GT 之间在图像级别或滑动窗口方式的相似性, 如 [68] 中所述, 这些方法无法捕捉到区域结构的相似性。

4.2 目标感知结构相似度

将前景图划分为块有助于评估目标部分结构的相似性。然而, 区域感知测度 (S_r) 并不能很好地从整体上解释物体相似性。对于显著性目标检测等高级视觉任务, 目标的相似度的评价至关重要。为了实现这一目标, 本文提出了一种对前景和背景分别进行评估的新方法。由于 GT 图通常具有**清晰的前景-背景对比和均匀的分布**等重要特征, 预测的 FM 有望具有这些特性。这有助于很容易地区分前景和背景。本文根据这些特征来设计对象结构相似度。

4.2.1 清晰的前背景对比度

首先发现, GT 图的前景区域通常与背景区域形成鲜明对比。使用一个类似于 SSIM 的亮度分量的公式, 来衡量 FM 前景区域和真值前景区域之间的平均概率的接近程度, 令 x_{FG} 和 y_{FG} 分别表示 FM 前景区域和真值前景区域的概率值, \bar{x}_{FG} 和 \bar{y}_{FG} 表示 x_{FG} 和 y_{FG} 的均值, 前景比较可以表示为:

$$O_{FG} = \frac{2\bar{x}_{FG}\bar{y}_{FG}}{(\bar{x}_{FG})^2 + (\bar{y}_{FG})^2}. \quad (9)$$

式 (9) 有几个不错的特性:

- 交换 \bar{x}_{FG} and \bar{y}_{FG} , O_{FG} 不会改变结果。
- O_{FG} 的范围是 $[0, 1]$ 。
- 当且仅当 $\bar{x}_{FG} = \bar{y}_{FG}$ 时, $O_{FG} = 1$ 。
- 两个图越接近, O_{FG} 越接近于 1 (最重要的属性)。

这些特性使式 (9) 能够达到目的。

4.2.2 均匀分布

本文的第二个观察是, GT 图的前景和背景区域通常分布均匀。因此, 重要的是给一个目标被一致高亮显示的 FM 分配一个较高的分数 (即, 在整个目标中相似的值; 如图. 5 所示)。如果前景值在 FM 中的变异性较大, 则其分布将不均匀。

在概率论和统计学中, 变异系数定义为标准差与均值之比 (σ_x/\bar{x}), 它是对概率分布离散度的标准度量。在这里, 用它来表示 FM 的离散度。换句话说, 利用变异系数来计算 FM 和 GT 分布之间的不相似性。根据式 (9), FM 和 GT 在目标级的总不相似度为:

$$D_{FG} = \frac{(\bar{x}_{FG})^2 + (\bar{y}_{FG})^2}{2\bar{x}_{FG}\bar{y}_{FG}} + \lambda \times \frac{\sigma_{x_{FG}}}{\bar{x}_{FG}}, \quad (10)$$

λ 是平衡这两项的常数。则 FM 和 GT 在目标上的相似度可以表示为:

$$S'_{FG} = \frac{1}{D_{FG}} = \frac{2\bar{x}_{FG}\bar{y}_{FG}}{(\bar{x}_{FG})^2 + (\bar{y}_{FG})^2 + 2\lambda \times \bar{y}_{FG} \times \sigma_{x_{FG}}}. \quad (11)$$

由于实际中 GT 前景的平均概率恰好为 1 ($\bar{y}_{FG} = 1$), FM 和 GT 在目标层面上的相似度可以重写为:

$$S_{FG} = \frac{2\bar{x}_{FG}}{(\bar{x}_{FG})^2 + 1 + 2\lambda \times \sigma_{x_{FG}}}. \quad (12)$$

为了计算背景对比 S_{BG} , 本文将背景作为前景的互补分量, 从 1 中减去 FM 和 GT 图 (将 1 改为 GT 为非二进制图时 GT 的最大值), 如图. 6 所示。那么, S_{BG} 可以类似地定义为:

$$S_{BG} = \frac{2\bar{x}_{BG}}{(\bar{x}_{BG})^2 + 1 + 2\lambda \times \sigma_{x_{BG}}}. \quad (13)$$

设 μ 为 GT 中前景区域与图像区域的比值 (宽 \times 高)。最后的对象感知结构相似度可以写成:

$$S_o = \mu \times S_{FG} + (1 - \mu) \times S_{BG}. \quad (14)$$

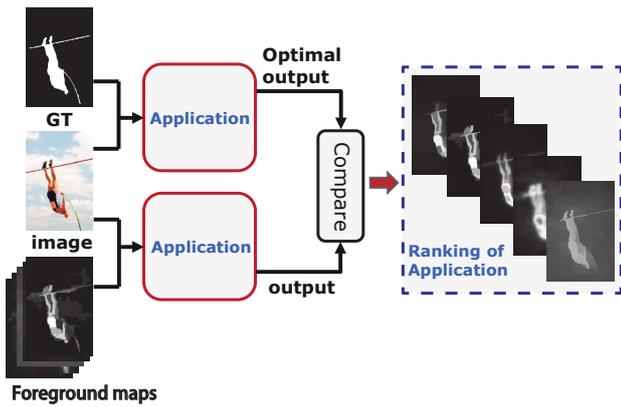


图 7 元指标 1: 程序排名。为了根据程序对前景图进行排名, 本文将使用 GT 时获得的输出与使用 FM 时的输出进行比较。FM 与 GT 图越相似, 其程序的输出应该与 GT 输出越接近。

4.3 结构度量

有了区域感知和对象感知的结构相似度评价定义, 可以将最终的指标表述为:

$$S = \alpha \times S_o + (1 - \alpha) \times S_r, \quad (15)$$

$\alpha \in [0,1]$ 。在本文的实现中, 设置 $\alpha = 0.5$, 为区域相似度和目标相似度分配相同的权重。利用该指标对图 1 中的三幅前景图进行评价, 可以正确地像人类排序一样对显著图进行排序。

5 实验

为了评估提出的新指标的质量, 本文使用了 Margolin *et al.* [56] 本文的 4 个元指标和本文的 1 个元指标 (人为判断)。这些元指标用于衡量评估指标的质量 [62]。为了进行公平的比较, 这 4 个元指标是在 ASD (又称 ASD1000) 数据集 [1] 上计算的, 使用 CA [29]、CB [35]、RC [17]、PCA [55] 和 SVO [11] 五种显著性检测算法生成 (共 5000 张图) 非二进制前景图 (二进制图是通过向 SalCut [17] 提供非二进制图生成的)。

配置和运行时间。 本文在所有实验中设定 $\lambda = 0.5$ 以及 $K = 4$ [19], 还使用一台 CPU 机器在 ASD1000 数据集上测试了 S 指标。单个图像的平均运行时间为 0.0053 秒。

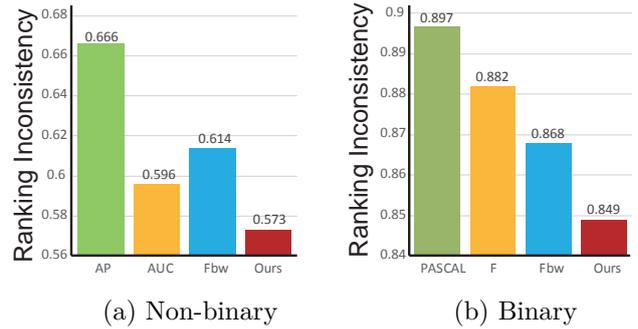


图 8 元指标 1-结果。评价指标与 SalCut 分割 (非二值评价) 和图像检索 (二值评价) 应用程序给出的评价指标的排序相关性。本文使用 1-spearman's ρ 的结果展示。分数越低, 在预测应用程序的偏好方面, 评价就越好。本文的评估方法优于其他评估方法。

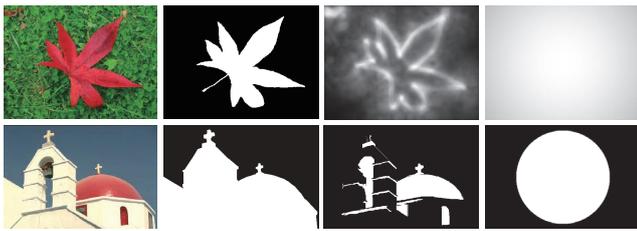
5.1 元指标 1: 应用排序

第一个元指标检查评估度量与使用前景图 [56] 的程序的排序相关性。假设 GT 图是应用程序的最优输入 (图 7 中的顶部路径)。然后给定前景图, 本文比较程序的输出 (图 7 底部路径) 和 GT 的输出。显著图越接近 GT, 其应用程序输出就越接近 GT 输出。本文将 AP、AUC、Fbw、PASCAL、 F_β 等二元和非二元评价指标的排序结果与应用程序的排序结果进行了比较。

[56] 的工作包括三个方面的应用: 目标检测、分割和图像检索。在这里, 本文使用 SalCut [17] 方法 (对于非二进制) 和图像检索 (对于二进制) 应用程序 (参见章节 7) 来计算这个元指标⁵。

本文利用 1-Spearman ρ 指标 [4] 来评价指标的排序精度, 其值越低表示排序一致性越好。得分为 0 表示评价对显著图的排名与应用程序的排名相同。2 分表示评价指标以完全相反的顺序对前景图进行排序。不同指标 (AP、AUC、Fbw、Ours) 的比较如图 8(a) 所示, 这表明结构度量在其他备选方法中排名一致性最好。根据图 1 所示的例子, 所有的非二进制度量方法不能正确地对前景图进行排序。本文的度量方法对这些图进行了正确的排序。在二进制图的情况下, S-measure 也获得了 5.35%、3.74%、2.19% 的提升, 优于 PASCAL、 F_β 和 Fbw 指标, 后者的得分各自为 0.897、0.882 和 0.868, 而 S 指标为 0.849。

⁵ 本文使用 Fbw [56] 进行相同的实验设置以进行公平比较。Fbw 只提供了检索应用程序, 从而无法实现其他两个应用程序。



(a) Image (b) GT (c) FM (d) Generic

图 9 元指标 2: SOTA vs. 通用图。评估指标应给予由 SOTA 方法 (c) 生成的 FM 比不考虑图像内容的通用图 (d) 更高的分数。不幸的是, 目前所有的评估方法给 (d) 中的图的分数的分数都高于 (c)。只有本文的方法正确地将 SOTA 结果排在更高的位置。

5.2 元指标 2: SOTA vs. 通用图

第二个元指标是, 指标应该更喜欢由 SOTA 算法实现的输出, 而不是丢弃图像内容的通用基线图 (例如, 有中心的高斯映射图像, 见图. 9(d))。一个好的评价方法应该将 SOTA 模型生成的 SM 排在比通用图更高的位置。

本文统计 5 个 SOTA 模型 (CA [29], CB [35], RC [17], PCA [55], SVO [11]) 生成的通用图得分高于平均得分的次数。平均分表明了模型的鲁棒性。结果如图. 10 所示。这里的值越低, 度量准确度就越好。在超过 1000 张图片中, 本文的度量方法在非二进制图只有 11 个错误 (即, 通用图超过 SOTA)。与此同时, AP 和 AUC 的衡量标准非常差, 排序错误明显更多。本文的度量方法也相对 PASCAL、 F_β 和 Fbw 取得了很大的提高。

5.3 元指标 3: 真值切换

第三项元指标表明, 当切换到错误的 GT 图时, 优秀的 SM 不应该获得更高的分数。在 Margolin *et al.* [56] 中, 当一个 SM 得分在 0.5 分以上 (与原始 GT 图相比) 时, 它被认为是“好的”。使用这个阈值 (0.5), 前 41.8% 的 5000 张图被认为是“优秀”图。为了公平比较, 本文遵循 Margolin *et al.* 选择相同比例的“好”图。对于 1000 张图片中的每一张, 随机测试了 100 个 GT 切换。然后计算当使用不正确的 GT 图时, 度量增加显著图得分的次数百分比。

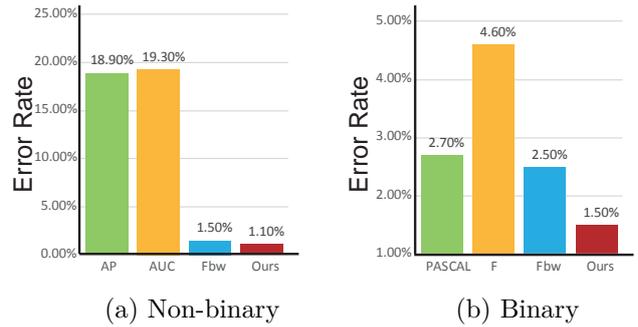
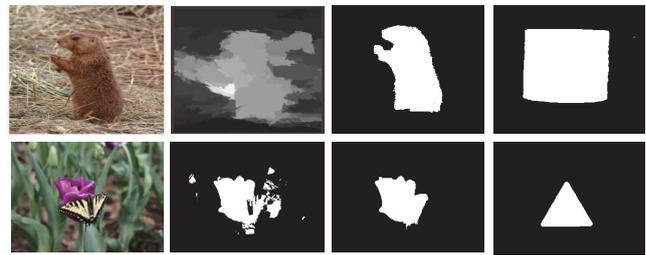


图 10 元指标 2 - 结果评估度量将一个通用图像 (非二进制圆或二进制中心高斯图) 排名高于由 SOTA 模型生成的 FM 的百分比。分数越低, 评价指标越好。本文的标准达到了最好的效果。



(a) Image (b) FM (c) GT (d) Switched GT

图 11 元指标 3: 真值切换。当使用错误的切换 GT 作为参考时, FM 的分数会降低。然而, 当使用 (d) 而不是 (c) 作为参考 GT 图时, AUC 和 AP 都给了 (b) 中的图更高的分数。使用 S 指标方法, 当切换到随机真值 (d) 时, (b) 的得分适当降低。

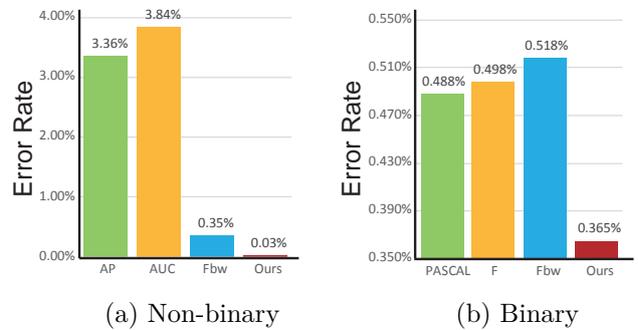


图 12 元指标 3 - 结果。当使用不正确的 GT 图时, 评估度量值分配更高分数的次数百分比 (在含有 1000 张图片的 ASD 数据集上测试)。分数越低, 衡量标准越好。本文的方法在非二进制和二进制显著图方面都比其他方法取得了显著的改进。

结果如图. 12 所示。分数越低, 匹配正确 GT 的能力就越强。与第二好的方法相比, 本文的方法表现最好, 大约好 10 倍。这是因为本文的度量方法捕获了 FM 和 GT 图之间的对象结构相似性。当使用随机选择的 GT 时, 本文的度量会给“好的”FM 分配一个较低的值, 因为在随机 GT 中物体结构已经改变。

5.4 元指标 4: 标注敏感性

第四项元指标规定, 评估度量不对 GT 边界手工标注中的微小错误/不准确敏感。为了执行这个元指标, 使用形态学操作制作了一个稍微修改过的 GT 图。示例如图. 13所示。虽然 (b) 和 (c) 中的两幅 GT 图略有不同, 但当使用 (b) 或 (c) 作为参考时, 一个好的度量方法不应在两幅前景图 (d) 和 (e) 之间切换排名。

本文使用 1-Spearman ρ 指标来检验标注错误前后的排名相关性。分数越低, 对注释错误 [56] 的评估指标就越健壮。结果如图. 14所示。本文的指标优于 AP 和 AUC, 但不是最好的。检查这个原因, 意识到并不总是分数越低, 评价方法就越好。问题在于, 有时候“轻微的”不准确的手工注释可能会改变 GT 图的结构, 而这反过来又会改变排名。

仔细检查结构变化后的影响, 发现主要的结构变化往往对应于 GT 与其形态变化版本之间的差异图中连续的大区域。本文使用了腐蚀版本的差分图的总和作为主要结构变化的度量, 并对所有 GT 图进行排序。

在变化最小的前 10% 的图像中, S 指标和 Fbw 具有相同的 MM4 分数 (均为 0)。当 GT 的拓扑没有变化时, S 指标和 Fbw 保持原有的排名。这可以在图. 15的例子中看到, 虽然 GT 图 (GT 和 Morphologic GT) 略有不同, 但根据使用的 GT 图, Fbw 和本人的度量方法都维持了两个显著图的排名顺序。

对于变化最多的前 10% 的图像, 让 3 个用户判断 GT 图像是否有主要的结构变化, 其中 100 张 GT 图像中有 95 张被认为有主要的结构变化 (例如, 每组的小条、细节、小部分和细线; 如图. 16所示), 对此本文认为保持相同的排名是不好的。图. 17是一个典型的例子。当使用 GT 图作为参考时, Fbw 和 S 指标对这两个图进行了适当的排序。但是, 当使用 Morphologic GT 作为参考时, 排名结果是不同的。显然, 蓝边 SM 在视觉和结构上比红边 SM 更类似于 Morphologic GT 图。一个好的衡量标准应该是蓝边 SM 高于红边 SM。所以

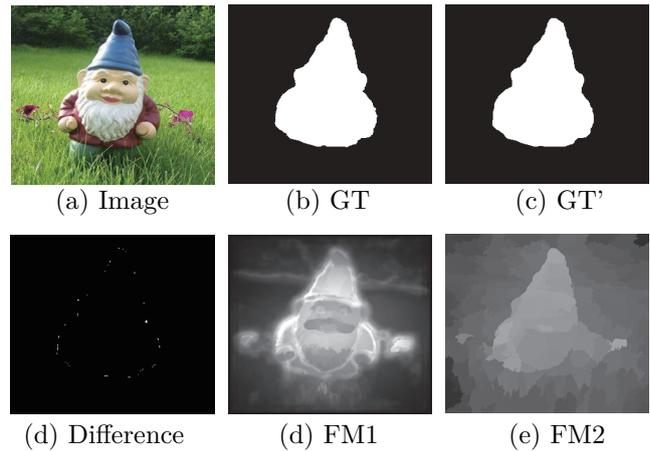


图 13 元指标 4: 标注敏感性。评估指标不对 GT 边界的手工标注的微小变化敏感。虽然 GT (b) 和 GT' (c) 几乎是相同的, 但一些度量方法改变了两个前景图 (d) 和 (e) 的排名顺序, 这取决于使用的不同 GT, S 指标的度量结果始终排名 (d) 高于 (e), 彩色浏览效果最佳。

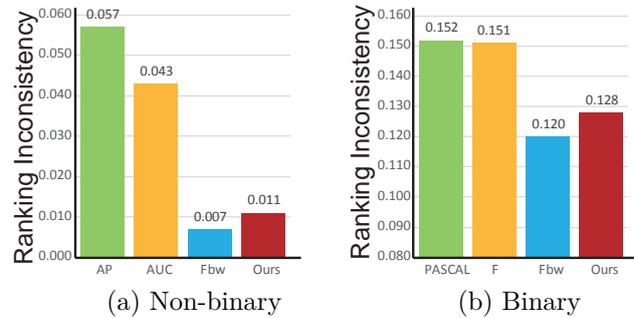


图 14 元指标 4 - 结果。在人工标注不准确的情况下, 评价指标的排序相关性。本文使用 1-Spearman’s ρ 度量来展示结果。分数越低越好。

这两张图的排名应该改变。Fbw 方法未能达到这一目的, 但本文的方法给出了正确的顺序。

上述分析表明, 这种度量方法不是很可靠。因此, 不把它包括在比较中。

5.5 进一步比较

图. 8, 图. 10, 和 图. 12的结果表明, 在 ASD1000 数据集上, 使用 3 个评估方法, 本文的方法获得了最佳性能。然而, 一个好的评估方法应该在几乎所有的数据集上表现良好。为了证明 S 指标的方法的鲁棒性, 进一步使用 10 种 SOTA 的显著目标检测算法在另外 4 个广泛使用的基准数据集上进行实验。

前景图的采集。 本文使用 10 种 SOTA 算法, 包括 3 种传统模型 (如, ST [54], DRFI [36] 和 DSR [47]) 和 7

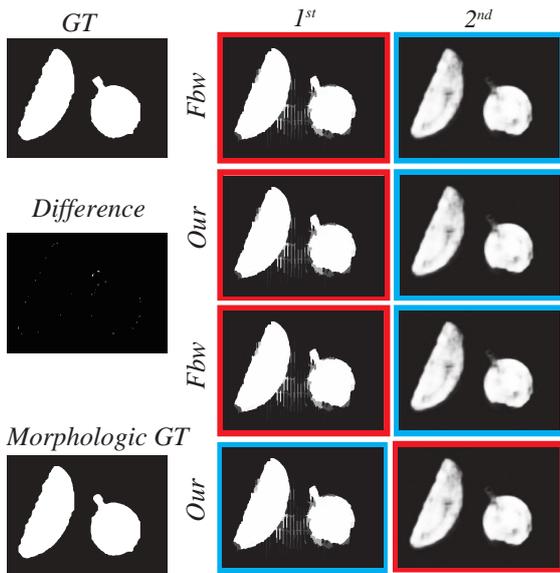


图 15 结构不变的情况下。S 指标和 Fbw 度量对 GT 边界手工标注的不准确性 (结构不变) 都不敏感。

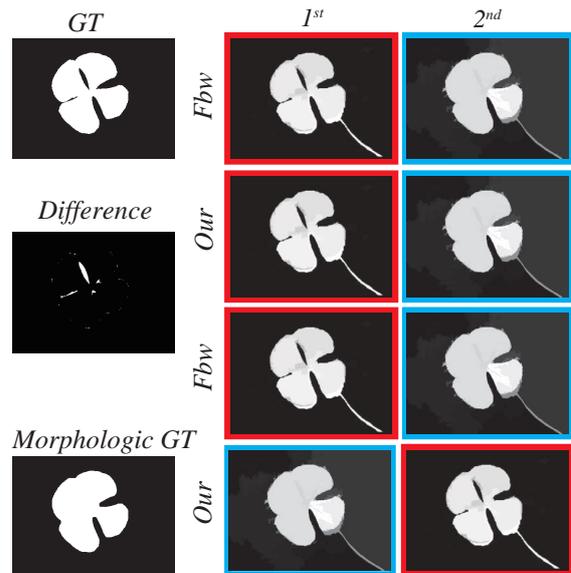
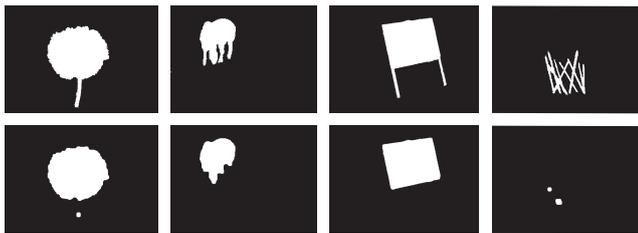


图 17 结构变化的情况。评估指标的排名应该对结构变化敏感。令人惊讶的是, 目前的最佳衡量标准 (Fbw) 并没有考虑到结构变化。根据本文的评估方法, 对图进行了正确的排名。建议在屏幕上观看获得最佳的视觉效果。



(a) Group1 (b) Group2 (c) Group3 (d) Group4

图 16 结构改变实例。第一行展示了 GT 图。第二行展示了形态改变后的图像, 可以观察到极大的结构变化。

种基于深度学习的模型 (DCL [45], rfcn [66], MC [84], MDF [44], DISC [15], DHS [51] 与 ELD [40]) 来生成二进制和非二进制前景图。在 [1] 中, 采用图像相关自适应阈值法来对非二值显著图进行阈值化, 从而得到二值显著图。

基准数据集. 被广泛使用的 4 个数据集包括 PASCAL-S [48], ECSSD [69], HKU-IS [44] 和 SOD [57]。PASCAL-S 包含 850 幅具有挑战性的图像, 这些图像在高度复杂的背景中有多个目标。ECSSD 包含 1000 个语义上有意但结构上复杂的图像。HKU-IS 是另一个包含 4445 幅大比例图像的大型数据集, 该数据集集中的大部分图像包含多个低对比度的显著目标。最后, 还在 SOD 数据集 (BSDS 数据集的子集) 上评估 S 指标。它包含相对较少的图像 (300 张), 但包含多个复杂的对象。

结果. 非二值图与二值图的定量比较结果如表 1 所示。根据第一个元指标, S 指标对于二进制和非二进制显著图的评估都表现得最好。这表明本文的方法比其他方法更接近于人。

对于 MM2 中的评价结果 (二值和非二值), S 指标优于现有的四种指标 (AP、AUC、F、PASCAL) 且有较大的差距。在两个较简单的数据集 (ECSSD 和 HKU-IS) 上的结果显示, 本文的度量和 Fbw 在二元和非二元图上的表现相当。

根据元指标 3, 与排名第二的指标相比, 本文的指标在 PASCAL、ECSSD、SOD 和 HKU-IS 上的非二值错误率分别降低了 67.62%、44.05%、17.81% 和 69.23%。对于二值显著图, 本文的方法也在 PASCAL、ECSSD、SOD 和 HKU-IS 上分别降低了 62.86%、52.38%、10.96% 和 61.54% 的错误率, 排名第二的得分分别为 62.86%、52.38%、10.96% 和 61.54%。这表明, S 指标具有更高的能力来捕获 FM 和 GT 图之间的结构相似性。

总的来说, S 指标在大多数情况下胜出, 这表明它比其他指标更鲁棒。

Table 1 非二值 (N-binary) & 二值显著图与 3 个元指标的当前指标的定量比较。最好的结果用**粗体**突出显示。MM: 元指标。

| Type | Measure | PASCAL-S [48] | | | ECSSD [69] | | | SOD [57] | | | HKU-IS [44] | | |
|----------|---------|---------------|--------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|
| | | MM1 | MM2(%) | MM3(%) | MM1 | MM2(%) | MM3(%) | MM1 | MM2(%) | MM3(%) | MM1 | MM2(%) | MM3(%) |
| N-binary | AP | 0.452 | 12.1 | 5.50 | 0.449 | 9.70 | 3.32 | 0.504 | 9.67 | 7.69 | 0.518 | 3.76 | 1.25 |
| | AUC | 0.449 | 15.8 | 8.21 | 0.436 | 12.1 | 4.18 | 0.547 | 14.0 | 8.27 | 0.519 | 7.02 | 2.12 |
| | Fbw | 0.365 | 7.06 | 1.05 | 0.401 | 3.00 | 0.84 | 0.384 | 16.3 | 0.73 | 0.498 | 0.36 | 0.26 |
| | Ours | 0.320 | 4.59 | 0.34 | 0.312 | 3.30 | 0.47 | 0.349 | 9.67 | 0.60 | 0.424 | 0.34 | 0.08 |
| Binary | F | 0.757 | 21.65 | 1.62 | 0.619 | 12.5 | 1.53 | 0.760 | 25.3 | 1.81 | 0.593 | 4.54 | 0.41 |
| | PASCAL | 0.905 | 19.41 | 1.59 | 0.787 | 11.0 | 1.54 | 0.911 | 24.0 | 1.81 | 0.786 | 3.62 | 0.41 |
| | Fbw | 0.802 | 13.76 | 1.72 | 0.675 | 7.50 | 1.54 | 0.814 | 19.0 | 1.42 | 0.665 | 2.16 | 0.58 |
| | Ours | 0.655 | 12.50 | 1.11 | 0.523 | 7.30 | 1.29 | 0.707 | 15.0 | 1.56 | 0.483 | 2.27 | 0.32 |

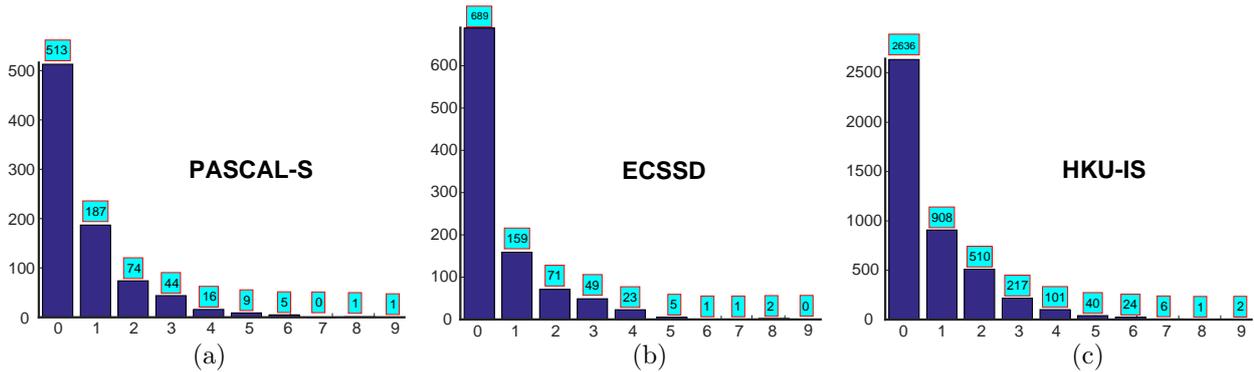


图 18 Fbw 和 S 指标之间的排名差距。(a)-(c) 是用来计算 Fbw 和 S 指标之间的排名差距的三个数据集。图中 y 轴是图像的数量。x 轴是排名差距。

5.6 元指标 5: 人为判断

在此, 本文提出了一种新的元指标来评价前景评价指标。这个元度量规定, 根据评估度量的图排名应该与人类排名高度一致, 有人认为“人是评价任何分割算法输出的最好的法官” [59]。然而, 由于时间和金钱成本, 对数据集的所有图像进行主观评价是不切实际的。据我们所知, 在目标分割领域还没有这样的视觉相似度评价数据库可以满足这些要求, 这里, 本文主要关注用于收集这样一个数据库的非二值显著图。

实验样本. 源前景图取自三个大尺度数据集: PASCAL-S、ECSSD 和 HKU-IS。如上所述, 本文使用 10 个 SOTA 显著性模型为每个数据集生成显著图。因此, 有 10 个前景图对应于每幅图像, 使用 Fbw 和本文的方法来评估这 10 张图, 然后根据每一种方法选出排名第一的图。如果两个指标选择了相同的图, 它们的排名差距为 0。如果一个指标将一个图排在第一位, 而另一个指标将同一图排在第 n 位, 那么它们的排名差距是 $|n-1|$ 。图. 18-20(a)、(b)、(c) 为两指标之间的排名差距直方图。蓝框表示每个排序距离对应的图像数量。

一些排名差距大于 0 的图被选为我们应用的用户研究的候选图。

用户调研. 本文从三个数据集中随机选择了 100 对图, 图. 21(b) 上方显示了一个示例试验, 左边是根据 S 指标得到的最佳图, 最右边是根据 Fbw 得到的最佳图。用户被要求选择他认为与 GT 图最相似的图。在本例中, 这两个显著图明显不同, 因此用户很容易做出决定。在另一个例子中 (图. 21(b) 的下方), 两张图过于相似, 使得使用者很难选择哪一个更接近 GT, 所以我们避免向受试者展示这类情况。最后, 留下了 50 组的实验样本, 本文开发了一款手机 app (如图. 21(a)) 来进行用户调研。出于实验目的收集了 45 个观众的数据, 受试者视力正常或已矫正。(年龄分布 19-29 岁; 从本科到博士的教育; 来自历史、医学、金融等 10 个不同专业; 男性 25 人, 女性 20 人)。

结果. 测试结果 (Fbw 与 S 指标) 如图. 22 所示, 受调者更喜欢 S 指标选择的图的试验百分比 (平均高于受试者) 是 63.69%。使用相同的程序来进行两个额外的

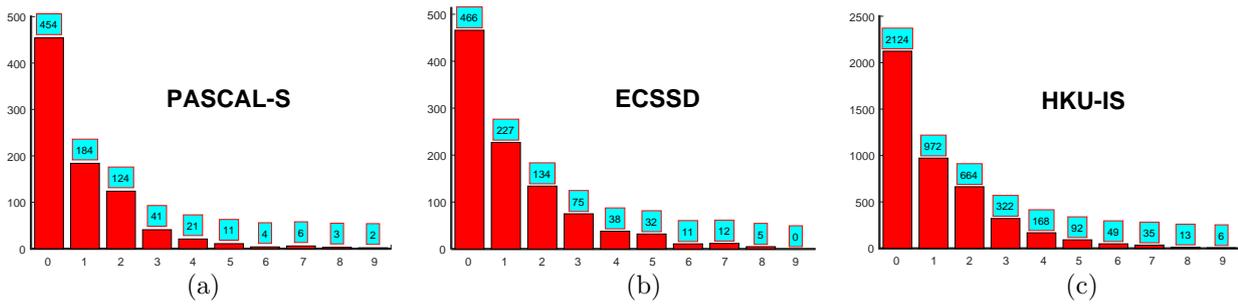


图 19 AP 和 S 指标之间的排名差距。(a)-(c) 是三个用于计算 AP 和本文的 S-指标之间的排名差距的数据集。图中 y 轴是图像的数量,x 轴是排名差距。

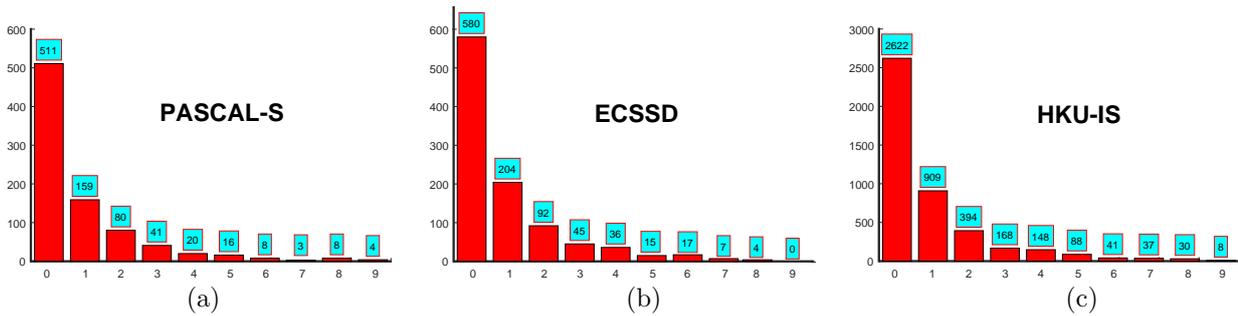
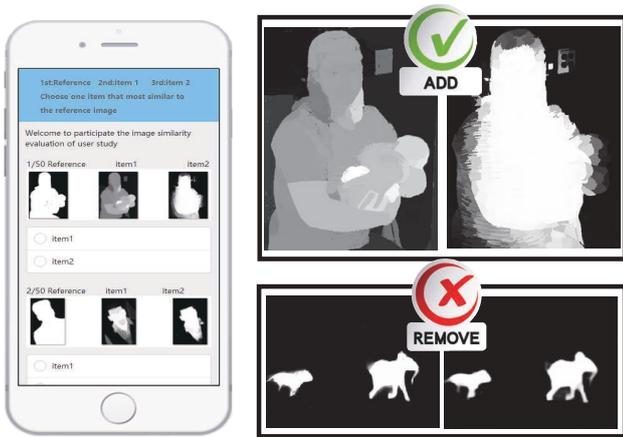


图 20 AUC 和本文的指标之间的排名差距。(a)-(c) 是用来计算 AUC 和 S-指标之间的排名差距的三个数据集。图中 y 轴是图像的数量,x 轴是排名差距。



(a) 手机 app (b) 一个实验案例

图 21 用户调研平台。

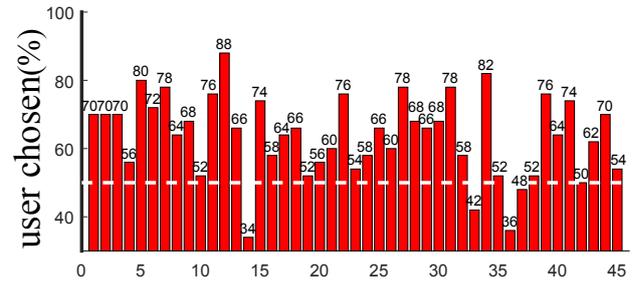


图 22 用户调研的结果 (Fbw & S-指标)。x 轴是用户 id,y 轴表示被测者更喜欢我们 S 指标选择的图的试验的百分比。

用户研究 (AP vs. S 指标, AUC vs. S 指标)。结果分别为 72.11% 和 73.56%, 这表明 S 指标更接近人为判断。

5.7 显著性模型比较

在此, 本文比较了 4 个数据集 (PASCAL-S、ECSSD、HKU-IS 和 SOD) 上的 10 个 SOTA 模型, 图. 23展

示了 10 个模型的排名。根据 S 指标, 排名前三的型号依次是 dhsnet、DCL 和 rfcn。此外, 我们还建立了许多具有代表性的在线基准 (1.<http://dpfan.net/socbenchmark>; 2.<http://dpfan.net/d3netbenchmark>; 3.<http://dpfan.net/cosod3k>; 4.<http://dpfan.net/camouflage/>) 来比较 S-指标和其他指标。

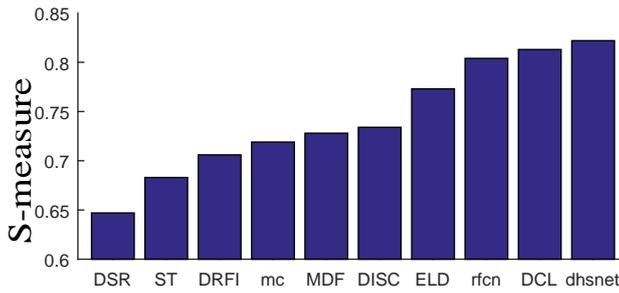


图 23 利用 S 指标对十个显著性模型的排名。y 轴表示每个数据集 (PASCAL-S [48], ECSSD [69], HKU-IS [44], SOD [57]) 上的平均得分。

6 消融实验

为了考查本文的 S-指标中每个部分的贡献, 进一步对一个由人类进行排名的数据集 [21] 进行消融研究, FMDatabase⁶由 185 张彩色图片和 555 张排序图组成。与元指标 1 类似, 本文也利用 1-Spearman's ρ 指标来评估排名性能。

Table 2 基于 1-spearman's 指标的人为排序消融实验 (使用 FMDatabase-IJCAI'18)

| Settings | 1 - Spearman's measure |
|------------------|------------------------|
| Object-aware | 0.195 |
| Region-aware | 0.142 |
| S-measure (Ours) | 0.140 |
| Fbw [56] | 0.149 |
| SSIM [68] | 0.223 |

如表. 2所示, 观察到本文的 S-指标在 FMDatabase 上优于其他设置 (例如目标感知、区域感知)。这清楚地表明, 仅区域级或目标级结构相似度无法提供稳定的评价指标, 由于目标感知结构相似性主要集中在前景-背景对比度的评价上, 它更像是一个全局评估。另一方面, 区域感知链接到基于窗口级统计的局部评估, 与现有的两个经典指标 (即 Fbw 和 SSIM) 相比, 我们也发现 S 指标达到了最好的结果。

7 讨论与结论

本文分析了目前基于像素误差的显著性评价方法, 表明它们忽略了结构相似性。然后, 本文提出了一种新的结构相似性测度, 即 **S-指标**, 该指标同时评估显著性图和



图 24 失败情况。由于这些预测图没有明显的结构, 本文的 S 指标排名 (d) 通用图高于 (c) rfcn。

GT 图之间的区域感知和目标感知的结构相似性。S 指标基于两个重要的特征: 1) 强烈的前景背景对比, 2) 均匀的显著性分布。此外, 本文的指标是有效且易于计算的。在 5 个数据集上的实验结果表明, 本文的度量方法优于现有方法, 包括 AP、AUC 和 Fbw。最后, 本文对 100 张显著图和 50 张 GT 图的数据库进行了用户调研。来自 45 名受试者的数据显示, 一般来讲, 相比 Fbw 他们更认可本文的 S 指标选择的显著图。

总之, 所有指标都是双刃剑, 很难说哪种方法是最好的。这些方法与实际应用背景是紧密耦合的, 例如, 有些应用可能倾向于重要区域的正确性, 而有些应用可能倾向于局部结构的连续性。我们观察到一些失败案例, 其中没有 (或更少) 对象结构的预测显著图将获得更高的 S-指标得分。例如, 如图. 24所示, 在 GT (b) 的对象没有清晰的结构的情况下, 本文的 S-指标就不能很好地工作。因此, 为了评估前景图, 我们需要评估它是否同时在多个指标上表现良好。

总而言之, 本文的方法为前景图评估提供了新的见解, 当前的方法未能充分检查模型的优缺点。目前, 已经发现显著性 (如 RGB SOD [16, 75, 83, 87], RGB-D SOD [24, 27, 74, 85, 86], RGB-T SOD [81], light field SOD [37, 61, 78], VSOD [34], 360 SOD, SID [42], Saliency Ranking [2], Co-SOD [23, 25, 80], and HR SOD [72]) 社区已经开始广泛采用该方法, 甚至在伪装目标检测 [20, 58, 73, 73] 和医学图像分割 [22] 领域也有应用。

⁶ <http://dpfan.net/e-measure/>

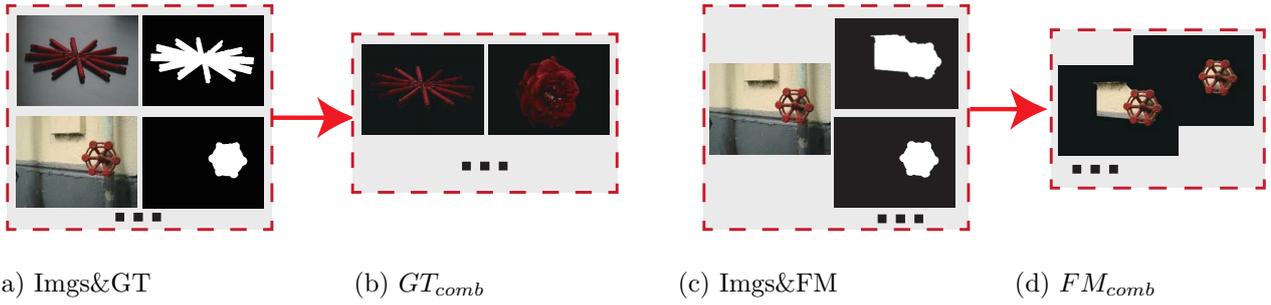


图 25 图像与前景图的整合。(a) 和 (c) 是图像 (Iimgs) 与 GT 和 FM。(b) 和 (d) 为组合图像

附录

图像评估程序

本文使用公开的基于内容的图像检索系统 LIRE [41] 作为应用的基础。首先,生成如图. 25(a)-(d) 所示的组合得到的图像。对于每一个组合的 GT 图像 (例如, GT_1, \dots, GT_n ; 其中 n 表示图像总数), 本文使用 LIRE 提取 CEDD 特征, 然后搜索 100 个最相似的图像列表 $GT_{lst-i} = \{G_{1-i}, \dots, G_{100-i}\}$ 。LIRE 还对 100 幅图像进行评分 ($GT_{score-i} = \{G_{s1-i}, \dots, G_{s100-i}\}$), 表示相似度值。因此, 对于每一幅 FM_{comb} 图像, 可以得到 100 幅相似度最高的图像 FM_{lst-i} 对应的评分 $FM_{score-i}$ 。最后, 令 $Q_i = \{GT_{lst-i} \cap FM_{lst-i}\}$ 。搜索 FM_k 等于 FM_{lst-i} 当中的 G_i , 如果 G_i 存在, 我们记录索引 k 和对应的 F_{sk-i} 。LIRE 分配的每个 FM 的相似度 S_i 是:

$$S_i = \begin{cases} F_{sk-i} + \frac{1}{k} + \frac{\|Q_i\|}{100}, & G_i \in Q_i \\ \frac{\|Q_i\|}{100}, & otherwise \end{cases} \quad (16)$$

References

- Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 1597–1604 (2009)
- Amirul Islam, M., Kalash, M., Bruce, N.D.: Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 7142–7150 (2018)
- Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 898–916 (2011)
- Best, D., Roberts, D.: Algorithm as 89: the upper tail probabilities of spearman’s rho. Journal of the Royal Statistical Society. Series C (Applied Statistics) **24**(3), 377–379 (1975)
- Borji, A.: What is a salient object? a dataset and a baseline model for salient object detection. IEEE Trans. Image Process. **24**(2), 742–756 (2015)
- Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: A benchmark. IEEE Trans. Image Process. **24**(12), 5706–5722 (2015)
- Borji, A., Itti, L.: State-of-the-art in visual attention modeling. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 185–207 (2013)
- Borji, A., Sihite, D., Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. IEEE Trans. Image Process. **22**(1), 55–69 (2013)
- Borji, A., Sihite, D.N., Itti, L.: What stands out in a scene? a study of human explicit saliency judgment. Vis. Res. **91**(0), 62–77 (2013)
- Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., Torralba, A.: Mit saliency benchmark (2015) (2015)
- Chang, K.Y., Liu, T.L., Chen, H.T., Lai, S.H.: Fusing generic objectness and visual saliency for salient object detection. In: Int. Conf. Comput. Vis., pp. 914–921 (2011)
- Chen, H., Li, Y.f.: Progressively complementarity-aware fusion network for rgb-d salient object detection. In: IEEE Conf. Comput. Vis. Pattern Recog. (2018)
- Chen, M.M., Fan, D.P.: Structure-measure: A new way to evaluate foreground maps. Int. J. Comput. Vis. **129**, 2622–2638 (2021)
- Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: internet image montage. ACM Trans. Graph. **28**(5), 124 (2009)
- Chen, T., Lin, L., Liu, L., Luo, X., Li, X.: Disc: Deep image saliency computing via progressive representation learning. IEEE T. Neural Netw. Learn. Syst. **27**(6), 1135–1149 (2016)
- Chen, Z., Xu, Q., Cong, R., Huang, Q.: Global context-aware progressive aggregation network for salient object detection. In: AAAI Conf. Art. Intell. (2020)
- Cheng, M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)
- Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
- Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Int. Conf. Comput. Vis., pp. 4548–4557 (2017)
- Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. arXiv preprint arXiv:2102.10274 (2021)
- Fan, D.P., Ji, G.P., Qin, X., Cheng, M.M.: Cognitive vision inspired object segmentation metric and loss function. SSI (2021). DOI 10.1360/SSI-2020-0370
- Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranel: Parallel reverse attention network for polyp segmentation. In: Med. Image. Comput. Comput. Assist. Interv., pp. 263–273 (2020)
- Fan, D.P., Li, T., Lin, Z., Ji, G.P., Zhang, D., Cheng, M.M., Fu, H., Shen, J.: Re-thinking co-salient object detection. IEEE Trans. Pattern Anal. Mach. Intell. (2021)
- Fan, D.P., Lin, Z., Zhang, Z., Zhu, M., Cheng, M.M.: Re-thinking rgb-d salient object detection: Models, data sets,

- and large-scale benchmarks. *IEEE T. Neural Netw. Learn. Syst.* (2021)
25. Fan, Q., Fan, D.P., Fu, H., Tang, C.K., Shao, L., Tai, Y.W.: Group collaborative learning for co-salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2021)
 26. Feng, D., Barnes, N., You, S., McCarthy, C.: Local background enclosure for rgb-d salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2343–2350 (2016)
 27. Fu, K., Fan, D.P., Ji, G.P., Zhao, Q., Shen, J., Zhu, C.: Siamese network for rgb-d salient object detection and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* (2021)
 28. Ghosh, J., Lee, Y.J., Grauman, K.: Discovering important people and objects for egocentric video summarization. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1346–1353 (2012)
 29. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(10), 1915–1926 (2012)
 30. Gorji, S., Clark, J.: Going from image to video saliency: Augmenting image salience with dynamic attentional push. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
 31. Guo, C., Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Process.* **19**(1), 185–198 (2010)
 32. Islam, M.A., Kalash, M., D. B. Bruce, N.: Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
 33. Itti, L.: Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans. Image Process.* **13**(10), 1304–1318 (2004)
 34. Ji, Y., Zhang, H., Jie, Z., Ma, L., Wu, Q.J.: Casnet: A cross-attention siamese network for video salient object detection. *IEEE T. Neural Netw. Learn. Syst.* (2020)
 35. Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., Li, S.: Automatic salient object segmentation based on context and shape prior. In: *Brit. Mach. Vis. Conf.*, p. 9 (2011)
 36. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: A discriminative regional feature integration approach. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2083–2090 (2013)
 37. Jiang, Y., Zhou, T., Ji, G.P., Fu, K., Zhao, Q., Fan, D.P.: Light field salient object detection: A review and benchmark. *arXiv preprint arXiv:2010.04968* (2020)
 38. Kanan, C., Cottrell, G.: Robust classification of objects, faces, and flowers using natural image statistics. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2472–2479 (2010)
 39. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, pp. 2169–2178 (2006)
 40. Lee, G., Tai, Y.W., Kim, J.: Deep saliency with encoded low level distance map and high level features. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 660–668 (2016)
 41. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. *ACM T Multimed. Comput.* **2**(1), 1–19 (2000)
 42. Li, G., Xie, Y., Lin, L., Yu, Y.: Instance-level salient object segmentation. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2386–2395 (2017)
 43. Li, G., Xie, Y., Wei, T., Lin, L.: Flow guided recurrent neural encoder for video salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
 44. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 5455–5463 (2015)
 45. Li, G., Yu, Y.: Deep contrast learning for salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 478–487 (2016)
 46. Li, L., Jiang, S., Zha, Z., Wu, Z., Huang, Q.: Partial-duplicate image retrieval via saliency-guided visually matching. *IEEE Trans. Multimedia* **20**(3), 13–23 (2013)
 47. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.H.: Saliency detection via dense and sparse reconstruction. In: *Int. Conf. Comput. Vis.*, pp. 2976–2983 (2013)
 48. Li, Y., Hou, X., Koch, C., Rehg, J.M., Yuille, A.L.: The secrets of salient object segmentation. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 280–287 (2014)
 49. Liu, G., Fan, D.: A model of visual attention for natural image retrieval. In: *IEEE Inf. Sci. Cloud Comput. Companion*, pp. 728–733 (2013)
 50. Liu, G.H., Yang, J.Y., Li, Z.: Content-based image retrieval using computational visual attention model. *Pattern Recogn.* **48**(8), 2554–2566 (2015)
 51. Liu, N., Han, J.: Dhsnet: Deep hierarchical saliency network for salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 678–686 (2016)
 52. Liu, N., Han, J., Yang, M.H.: Picanet: Learning pixel-wise contextual attention for saliency detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
 53. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 353–367 (2011)
 54. Liu, Z., Zou, W., Le Meur, O.: Saliency tree: A novel saliency detection framework. *IEEE Trans. Image Process.* **23**(5), 1937–1952 (2014)
 55. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1139–1146 (2013)
 56. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 248–255 (2014)
 57. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Int. Conf. Comput. Vis.*, vol. 2, pp. 416–423 (2001)
 58. Mei, H., Ji, G.P., Wei, Z., Yang, X., Wei, X., Fan, D.P.: Camouflaged object segmentation with distraction mining. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2021)
 59. Pal, N.R., Pal, S.K.: A review on image segmentation techniques. *Pattern Recogn.* **26**(9), 1277–1294 (1993)
 60. Peng, H., Li, B., Xiong, W., Hu, W., Ji, R.: Rgb-d salient object detection: a benchmark and algorithms. In: *Eur. Conf. Comput. Vis.*, pp. 92–109 (2014)
 61. Piao, Y., Rong, Z., Zhang, M., Lu, H.: Exploit and replace: An asymmetrical two-stream architecture for versatile light field saliency detection. In: *AAAI Conf. Art. Intell.*, pp. 11865–11873 (2020)
 62. Pont-Tuset, J., Marques, F.: Measures and meta-measures for the supervised evaluation of image segmentation. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2131–2138 (2013)
 63. Qin, X., Fan, D.P., Huang, C., Diagne, C., Zhang, Z., Sant’Anna, A.C., Suárez, A., Jagersand, M., Shao, L.: Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704* (2021)
 64. Rutishauser, U., Walther, D., Koch, C., Perona, P.: Is bottom-up attention useful for object recognition? In: *IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, pp. II–37 (2004)
 65. Tiantian, W., Zhang, L., Lu, H., Borji, A.: Detect globally, refine locally: A novel approach to saliency detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)

66. Wang, L., Wang, L., Lu, H., Zhang, P., Ruan, X.: Saliency detection with recurrent fully convolutional networks. In: *Eur. Conf. Comput. Vis.*, pp. 825–841 (2016)
67. Wang, W., Shen, J., Dong, X., Borji, A.: Salient object detection driven by fixation prediction. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
68. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
69. Xie, Y., Lu, H., Yang, M.H.: Bayesian saliency via low and mid level cues. *IEEE Trans. Image Process.* **22**(5), 1689–1698 (2013)
70. Yu, Q., Xie, L., Wang, Y., Zhou, Y., K. Fishman, E., L. Yuille, A.: Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
71. Zeng, Y., Lu, H., Zhang, L., Feng, M., Borji, A.: Learning to promote saliency detectors. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
72. Zeng, Y., Zhang, P., Zhang, J., Lin, Z., Lu, H.: Towards high-resolution salient object detection. In: *Int. Conf. Comput. Vis.*, pp. 7234–7243 (2019)
73. Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H., Fan, D.P.: Mutual graph learning for camouflaged object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2021)
74. Zhang, J., Fan, D.P., Dai, Y., Anwar, S., Saleh, F., Aliakbarian, S., Barnes, N.: Uncertainty inspired rgb-d saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* (2021)
75. Zhang, J., Xie, J., Barnes, N.: Learning noise-aware encoder-decoder from noisy labels by alternating back-propagation for saliency detection. In: *Eur. Conf. Comput. Vis.* (2020)
76. Zhang, J., Zhang, T., Dai, Y., Harandi, M., Hartley, R.: Deep unsupervised saliency detection: A multiple noisy labeling perspective. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
77. Zhang, L., Dai, J., Lu, H., He, Y., Wang, G.: A bi-directional message passing model for salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
78. Zhang, M., Li, J., Wei, J., Piao, Y., Lu, H.: Memory-oriented decoder for light field salient object detection. In: *Adv. Neural Inform. Process. Syst.*, pp. 898–908 (2019)
79. Zhang, P., Wang, D., Lu, H., Wang, H., Ruan, X.: Amulet: Aggregating multi-level convolutional features for salient object detection. In: *Int. Conf. Comput. Vis.* (2017)
80. Zhang, Q., Cong, R., Hou, J., Li, C., Zhao, Y.: Coadnet: Collaborative aggregation-and-distribution networks for co-salient object detection. In: *Adv. Neural Inform. Process. Syst.* (2020)
81. Zhang, Q., Huang, N., Yao, L., Zhang, D., Shan, C., Han, J.: Rgb-t salient object detection via fusing multi-level cnn features. *IEEE Trans. Image Process.* **29**, 3321–3335 (2019)
82. Zhang, X., Wang, T., Qi, J., Lu, H., Wang, G.: Progressive attention guided recurrent network for salient object detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2018)
83. Zhao, J.X., Liu, J.J., Fan, D.P., Cao, Y., Yang, J., Cheng, M.M.: Egnnet: Edge guidance network for salient object detection. In: *Int. Conf. Comput. Vis.*, pp. 8779–8788 (2019)
84. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1265–1274 (2015)
85. Zhao, X., Pang, Y., Zhang, L., Lu, H., Zhang, L.: Suppress and balance: A simple gated network for salient object detection. In: *Eur. Conf. Comput. Vis.* (2020)
86. Zhou, T., Fan, D.P., Cheng, M.M., Shen, J., Shao, L.: Rgb-d salient object detection: A survey. *Comput. Vis. Media* pp. 1–33 (2021)
87. Zhuge, M., Fan, D.P., Liu, N., Zhang, D., Xu, D., Shao, L.: Salient object detection via integrity learning. *arXiv preprint arXiv:2101.07663* (2021)
88. Zhuge, M., Gao, D., Fan, D.P., Jin, L., Chen, B., Zhou, H., Qiu, M., Shao, L.: Kaleido-bert: Vision-language pre-training on fashion domain. In: *IEEE Conf. Comput. Vis. Pattern Recog.* (2021)