



认知规律启发的物体分割评价标准及损失函数

范登平¹, 季葛鹏², 秦雪彬³, 程明明^{1*}

1. 南开大学计算机学院, 天津 300350, 中国

2. 武汉大学计算机学院, 武汉 430070, 中国

3. Department of Computing Science, Faculty of Science, University of Alberta, Edmonton T6G 2R3, Canada

* 通信作者. E-mail: cmm@nankai.edu.cn

收稿日期: 2020-11-30; 修回日期: 2021-01-24; 接受日期: 2021-02-22; 网络出版日期: 2021-09-17

新一代人工智能重大项目 (批准号: 2018AAA0100400)、国家自然科学基金优秀青年科学基金项目 (批准号: 61922046)、教育部指导高校科技创新规划项目和南开大学中央高校基本科研业务费专项资金项目 (批准号: 63201169) 资助

摘要 物体分割技术是计算机视觉中的研究热点, 在多个领域都有广泛的应用. 本文从人类视觉系统对场景中的全局信息和局部细节非常敏感的角度出发, 设计了一种新颖、高效且易于计算的增强匹配标准 (E_{ξ}) 来评估物体分割模型的性能. E_{ξ} 将局部像素值与全局平均值有机结合, 以便评估分割结果与标准结果在图像级和像素级的相似度. 在国际主流的 4 个公开数据集上的大量实验表明, E_{ξ} 在多个方面, 如应用关联度、随机偏好度、噪声偏好度、感知度上相比现有广泛采纳的评价标准 (IoU 和 F_{β}) 均有大幅相对提升. 通过利用加权二值交叉熵损失函数、本文的增强匹配损失函数以及加权交并比损失函数, 本文进一步设计了一套组合损失函数 (Hybrid- E_{loss}) 来促进网络学习到像素级、对象级和图像级的分割特征. 定性和定量的结果表明, 在 3 个不同领域的分割任务中使用这一组合损失函数能够进一步提高物体分割的精度.

关键词 物体分割技术, 评价标准, 视觉感知, 增强匹配标准, 损失函数

1 引言

物体分割技术在计算机领域有着广泛的应用基础, 如图像分割^[1]、显著目标分割^[2~6]、多模态显著物体分割^[7]、协同显著目标分割^[8]、前背景提取^[9]、视频目标分割^[10]以及目标检测与识别^[11,12]等. 在这些应用当中, 为了评估模型算法的优劣, 将物体分割模型输出结果 (foreground map, FM) 与标准结果 (ground-truth, GT) 进行定量比较是必不可少的, 通常需要引入评价标准去衡量两者之间的相似度. 本文主要研究二类 (目标类被标记为白色, 非目标类则为黑色) 物体分割¹⁾评价标准. 目前广

1) 本文在后续的章节将二值前景图等同于二类物体分割结果图.

引用格式: 范登平, 季葛鹏, 秦雪彬, 等. 认知规律启发的物体分割评价标准及损失函数. 中国科学: 信息科学, 2021, 51: 1475–1489, doi: 10.1360/SSI-2020-0370
Fan D-P, Ji G-P, Qin X B, et al. Cognitive vision inspired object segmentation metric and loss function (in Chinese). Sci Sin Inform, 2021, 51: 1475–1489, doi: 10.1360/SSI-2020-0370

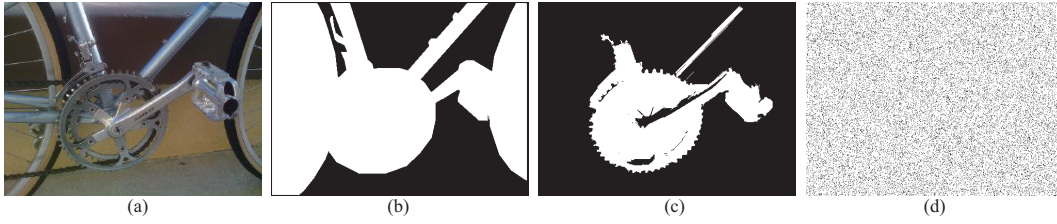


图 1 (网络版彩图) 现有评价标准评估结果不合理的示例. 合理的评价标准应赋予由先进算法生成的分割结果 (c) 更高的得分而不是随机高斯噪声结果 (d). 然而, 目前常用的标准如 IoU^[15], F_1/JI ^[14], Fbw^[17], CM^[19] 和 VQ^[18] 都更倾向于选择噪声结果. 唯独本文的标准能正确地将 (c) 排在 (d) 之前

Figure 1 (Color online) Inaccuracy of current evaluation measures. A measure should score the FM (c) generated by a state-of-the-art algorithm a higher value than the random Gaussian noise map (d). However, current common measures including IoU^[15], F_1/JI ^[14], Fbw^[17], CM^[19], and VQ^[18] prefer the noise map. Only our measure correctly ranked (c) higher than (d). (a) Image; (b) GT; (c) FM; (d) noise map

泛使用的 3 个评价标准为 F_β (F_1 , $\beta^2=1$)^[13], JI (Jaccard index)^[14] 和 IoU (intersection over union)^[15]. 此外, 基于 F_β 的评价标准^[16~18] 及其他设计思路的标准^[19~21] 也在某些特定的应用中相继被使用. 这些评价标准都采用了像素级方式来评价分割结果与标准结果的相似性. 图 1 分别展示了一幅测试图像 (图 1(a)) 的分割模型预测结果 (图 1(c)) 和随机高斯噪声结果 (图 1(d)). 相比随机噪声 (图 1(d)), 虽然图 1(c) 与标准结果图 1(b) 看起来更相似, 但迄今为止, 最常用的评价标准如 IoU^[15], F_1/JI ^[14] 以及最近提出的 Fbw^[17] 和 VQ^[18] 评价标准都认为图 1(d) 比图 1(c) 更好, 这显然与人类视觉评价结果相背离.

本文设计了一个新的评价标准, 称为增强匹配标准 (enhanced-alignment measure, E_ξ)^{[22]2)}. 它由一个同时考虑了像素级别和图像级属性的表达式组成. 通过实验, 本文证明了 E_ξ 是物体分割领域中一种简单、高效的评价标准. 如图 2^[23~25] 所示, 在 3 个用不同颜色 (红色、蓝色和黄色, 分别表示人工排序的最好、次最好与最坏的预测结果) 框标记的分割结果与标准结果的评估中, 与 3 个最近的评价标准 Fbw^[17], VQ^[18] 和 CM^[19] 比较时, 只有本文的 E_ξ 标准同时考虑了局部细节与全局信息, 因此能够正确地对 3 个分割结果进行排序. 本文的部分工作已被国际人工智能联合会议 (International Joint Conference on Artificial Intelligence, IJCAI)^[22] 收录. 具体来说, 会议版本简要地描述了 E_ξ 评价标准. 而本文则提供了 E_ξ 评价标准具体的实验细节. 此外, 本文基于 E_ξ 评价标准的思想进一步设计了一个组合损失函数来提高物体分割模型的性能. 最后, 本文在显著目标分割、伪装物体分割以及医疗图像分割 3 大领域上验证了组合损失函数的有效性. 本文的主要学术贡献归纳如下.

- 本文提出了一个简单高效的评价标准 E_ξ , 它通过一个紧凑项就能够同时捕捉图像级统计信息和像素级匹配信息. 实验结果表明, 在 4 个国际主流的数据集上 E_ξ 在多个元度量上优于最常用的评价标准 IoU^[15], F_1/JI ^[14], CM^[19] 以及最新的评价标准 VQ^[18] 和 Fbw^[17].
- 为了评测这些评价标准的性能, 本文提出了一个新的元度量 (噪声偏好度). 即, 优秀的评价标准应该赋予最先进算法生成的分割结果比随机高斯噪声结果更高的分数.
- 为了验证评价标准与人类感知一致性的程度, 本文还构建了一个新的 FMD 数据集. 该数据集合计 555 张图像, 并由人类受试者对分割结果进行主观排序. 这也是认知规律启发的物体分割评价标准领域第一个开源的数据集.
- 通过利用加权二值交叉熵损失函数 ($\mathcal{L}_{\text{ce}}^w$)、增强匹配损失函数 (\mathcal{L}_e) 以及加权交并比损失函数

2) 本文为 IJCAI2018 (E-measure)^[22] 工作的扩展.

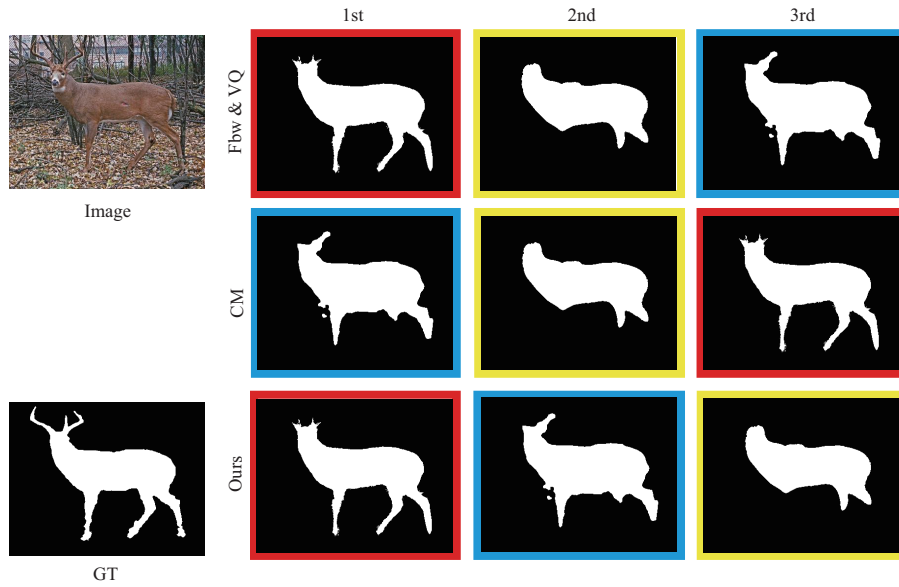


图 2 (网络版彩图) 证明本文标准有效性的示意图. 由 3 个最先进的显著目标分割模型生成的二值前景图 (阈值化后) 的排名, 生成模型包括 DCL [23], RFCN [24] 以及 DHS [25]. 3 种常用的、不同类型的标准 (CM, Fbw 和 VQ) 都无法对这些分割结果图给出完全正确的排序, 而只有本文的 E_ξ 给出了正确的排序

Figure 2 (Color online) Demonstration of effectiveness of our measure. The ranking of binary foreground maps (after thresholding) are generated by three state-of-the-art salient object detection models including DCL [23], RFCN [24] and DHS [25]. All of the three different types of popular measures (CM, Fbw and VQ) fail to rank these maps correctly. However, our E_ξ gives the right order

(\mathcal{L}_{iou}^w), 本文进一步设计了一套组合损失函数 (Hybrid- E_{loss}) 来促进网络学习到像素级、对象级和图像级分割特征. 定性和定量的结果表明, 在 3 个不同领域的分割任务 (显著目标分割、伪装物体分割和医学图像分割) 中使用本文的组合损失函数能够进一步提高物体分割的精度³⁾.

2 相关研究现状

F_β [13, 26, 27] 是一个常用的评价标准, 它同时考虑召回率 $recall = \frac{TP}{TP+FN}$ 和精确率 $precision = \frac{TP}{TP+FP}$:

$$F_\beta = \frac{(1 + \beta^2)precision \cdot recall}{\beta^2 \cdot precision + recall}, \tag{1}$$

其中 β 是一个调和参数, 用来权衡 recall 和 precision. 真正类 (true positives, TP), 真负类 (true negatives, TN), 假正类 (false positives, FP) 和假负类 (false negatives, FN) 是 4 个基本量. 设置 $\beta^2 = 1$ 即为经典的 F_1 评价标准: 另一个广泛使用的、基于 F_1 的标准是 Jaccard 指数 (JI [14]), 也称为 IoU 标准:

$$JI = IoU = \frac{TP}{TP + FN + FP}. \tag{2}$$

3) 代码. <https://github.com/GewelsJI/Hybrid-Eloss/>.

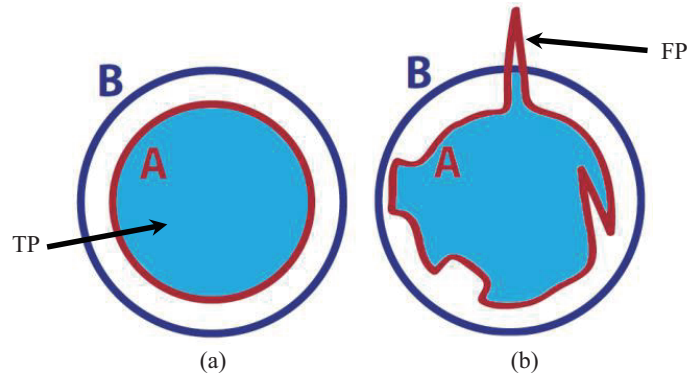


图 3 (网络版彩图) 基于区域的评价标准的局限性. 蓝色圆环表示标准结果的轮廓, 红色曲线为前景图的轮廓. IoU^[15] 和 F_1/JI ^[14] 标准的评价结果显示 (b) 与 (a) 的得分几乎一致, 即使 (b) 的形状与 (a) 有明显差异, 如钉子状突起和随机的凹槽^[19]

Figure 3 (Color online) Limitations of region-based measures. The blue circle represents GT and the red curve denotes FM. Based on IoU^[15], F_1/JI ^[14] measures, the intersection in (b) is almost equal to the intersection in (a) when compared with GT circle (blue circle curve), although it has spikes, wiggles and shape differences^[19]

表 1 目前常用的评价标准
Table 1 Existing popular evaluation measures

Measure	Publication	Pros	Cons
IoU/ F_1/JI ^[14]	BSVSN	Easy to calculate	Losing image level statistics
CM ^[19]	CVPRW	Considering both region and contour	Noise sensitive
Fbw ^[17]	CVPR	Assigning different weights for errors	Error location sensitive, complicated
VQ ^[18]	TIP	Weighting errors by psychological function	Subjective measure

F_1 和 IoU 的关系为 $Jl = \frac{F_1}{2 - F_1}$. Shi 等^[18] 提出了另一种主观的对象分割评价标准, 核心思想也是基于 F_1 标准的. 随后, Margolin 等^[17] 又设计了一个名为加权 F_β (Fbw) 的复杂评价标准:

$$Fbw = \frac{(1 + \beta^2) \text{precision}^\omega \cdot \text{recall}^\omega}{\beta^2 \cdot \text{precision}^\omega + \text{recall}^\omega}. \quad (3)$$

该标准为不同位置的错误分配不同的权重.

上述所有方法都与 F_β 密切相关. 它们独立地考虑每个像素位置来评价分割结果, 却忽略了重要的图像级别的信息, 这导致了在评估带有噪声 (如图 1 所示)、具有结构错误 (如图 2 所示)、形状不同 (如图 3 所示) 的分割结果时表现欠佳. 为此, Movahedi 等^[19] 提出了轮廓映射 (contour mapping, CM) 标准. 然而, 后续的元度量 3 实验表明, 这一标准对噪声非常敏感 (如图 1 所示), 导致性能不佳. 表 1 总结了当前常用的二类物体分割评价标准的优缺点.

3 认知规律启发的物体分割评价标准

本节将解释评估二值前景图的新方法的细节. 本文方法的一个重要优势是其简单性, 因为它仅包含一个能同时捕获全局统计信息和局部像素匹配信息的紧凑项. 因此, 本文的方法比现有最好的方法表现更优异. 图 4 展示了本文算法的过程.

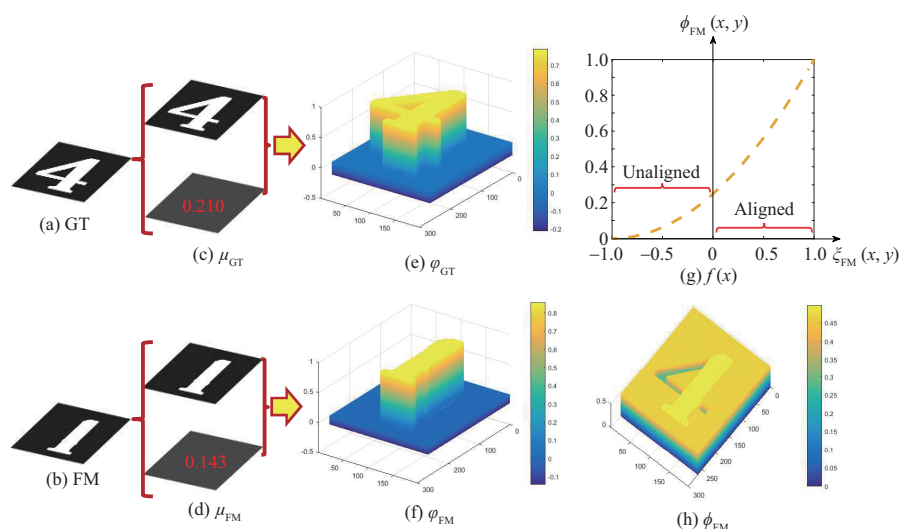


图 4 (网络版彩图) 本文 E_ξ 的流程图. (a) 标准结果图 GT; (b) 二值前景图 FM; (c) 和 (d) 为 GT 和 FM 的均值; (e) 和 (f) 为偏差矩阵由式 (4) 计算得到; (g) 是映射函数; (h) 为增强对齐矩阵, 可根据式 (6) 计算. “匹配” 和 “失配” 分别表示 $GT(x, y) = FM(x, y)$ 和 $GT(x, y) \neq FM(x, y)$ 时的位置

Figure 4 (Color online) The pipeline of our E_ξ . (a) Ground-truth (GT) map; (b) the binary foreground map (FM); (c) and (d) are the mean values maps of GT and FM, respectively; (e) and (f) are the bias matrices calculated by (4); (g) is the mapping function; (h) is the enhanced alignment matrix computed by (6). “Aligned” and “unaligned” denote those points whose $GT(x, y) = FM(x, y)$ and $GT(x, y) \neq FM(x, y)$, respectively

3.1 研究动机

认知视觉研究表明, 人类视觉系统对场景中的结构 (例如, 全局信息和局部细节) 高度敏感. 因此, 在评估 FM 和 GT 之间的相似性时, 同时考虑局部信息和全局信息至关重要. 基于上述观察, 通过一个紧凑的表达式同时获取图像统计信息和像素级匹配信息, 本文设计了一种新的二类 (目标类和背景类) 物体分割评价标准.

3.2 匹配项

为了设计一个能同时捕捉全局统计量和局部像素匹配信息的紧凑项, 本文定义了一个偏差矩阵 φ 来衡量分割结果 I 的每个像素值与它的全局平均值 μ_I 之间的距离:

$$\varphi_I = I - \mu_I \cdot \mathbb{A}, \quad (4)$$

其中, $I \in \{GT, FM\}$, 而 \mathbb{A} 是一个元素值均为 1 的矩阵, 大小与 I 相同. 本文分别计算标准结果图 GT 和分割结果图 FM 的偏差矩阵 φ_{GT} 和 φ_{FM} . 这个过程也称为信号中心化, 即从信号中减去平均信号. 受到 Wang 等^[28] 在结构相似度评价标准中引入亮度对比的启发, 本文也采用 φ_{GT} 和 φ_{FM} 之间的 Hadamard 乘积来衡量两个偏差矩阵之间的相关性. 本文定义了一个对齐矩阵 ξ 来描述它:

$$\xi_{FM} = \frac{2\varphi_{GT} \circ \varphi_{FM}}{\varphi_{GT} \circ \varphi_{GT} + \varphi_{FM} \circ \varphi_{FM}}, \quad (5)$$

\circ 表示 Hadamard 乘积, 对齐矩阵 ξ_{FM} 具有如下属性: $\xi_{FM}(x, y) \geq 0$ 仅当 φ_{GT} 和 φ_{FM} 符号相同, 即两个输入在 (x, y) 的位置处对齐. 对齐矩阵的元素值依赖于全局平均值, 即纳入了图像级统计信息. 这些特征使式 (5) 符合本文的设计目标.

3.3 增强匹配项

$\xi_{FM}(x, y)$ 的绝对值取决于 μ_{FM} 和 μ_{GT} 的相似性. 经过分析发现, 当两幅分割结果高度相似时, μ_{FM} 和 μ_{GT} 之间的进一步相似性可能会引起对齐位置正值增大, 而在未对齐位置的负值将减小. 将 $\xi_{FM}(x, y)$ 每个位置的值叠加后得到的总数并不总是提升从而不符合期望 (期望提升). 因此, 需要一个映射函数来抑制负值区域 ($\xi_{FM}(x, y) \leq 0$) 的减少 (这意味着具有较小的微分值) 并且增强正值区域 ($\xi_{FM}(x, y) \geq 0$) 的增大幅度.

为了实现这一目标, 需要用到“凸函数”. 本文也测试了其他形式的映射函数, 如高阶多项式或三角函数, 但发现二次型 ($f(x) = \frac{1}{4}(1+x)^2$ 如图 4(g) 所示) 是一个简单的且有效的函数, 并且实验效果最好. 因此, 本文的增强的对齐矩阵 ϕ 被定义为

$$\phi_{FM} = f(\xi_{FM}). \quad (6)$$

3.4 增强匹配标准

利用增强的对齐矩阵 ϕ 来捕获二值映射的两个属性 (像素级匹配和图像级统计), 本文的增强匹配标准 E_{ξ} 被定义为

$$Q_{FM} = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h \phi_{FM}(x, y), \quad (7)$$

其中 h 和 w 分别是分割结果图的高度和宽度. 利用 E_{ξ} 来评估图 1 中的分割结果 (图 1(c)) 与噪声结果 (图 1(d)), 发现本文新设计的标准能给出正确的评估结果.

4 元度量实验

为了测试评价标准的性能, 需要设计一套评估标准, 这套标准被称为元度量. 其基本思想是定义一些关于评价标准预期结果的理想状态, 然后测试评价标准满足这些状态的程度^[29]. 和已有的工作 [17, 29, 30] 保持一致, 本文采用了 4 个常用元度量: 应用关联度、随机偏好度、感知度、保真度, 以及本文引入的一个新的元度量噪声偏好度 (4.3 小节) 来进行后续实验.

数据集和模型. 为了和 Fbw^[17] 进行公平的比较, 后续元度量实验所采用的数据集也和 Fbw 保持一致: PASCAL-S^[31], ECSSD^[32], HKU-IS^[2] 以及 SOD^[33]. 为了验证评价标准对不同分割结果的评价性能, 需要选择具有不同特性的物体分割结果图. 本文选用了 3 个传统方法 (ST^[34], DRFI^[35] 和 DSR^[36]) 以及 7 个深度模型 (DCL^[23], RFCN^[24], MC^[37], MDF^[2], DISC^[38], DHS^[25] 和 ELD^[39]) 方法⁴⁾来生成二值分割结果图.

4.1 元度量 1: 应用关联度

元度量 1 中的应用关联度由 Margolin 等^[17] 首次提出, 它描述的是评价标准对物体分割结果的排序与应用程序对物体分割结果排序之间的一致性程度. 如图 5 所示, 首先将原图与标准分割结果一起输入到应用程序并得到最优输出. 然后, 将由不同的物体分割算法得到的分割结果与原图输入应用程序得到一组输出结果. 通过比较这组输出结果与最优输出就可得到物体分割算法的应用程序排名, 待评估的分割结果与标准分割结果越相似, 应用程序就会赋予分割结果优先的排名.

4) 这些算法得到的结果图为非二值分割结果图, 本文采用 2 倍均值作为阈值得到最终的二值分割结果图.

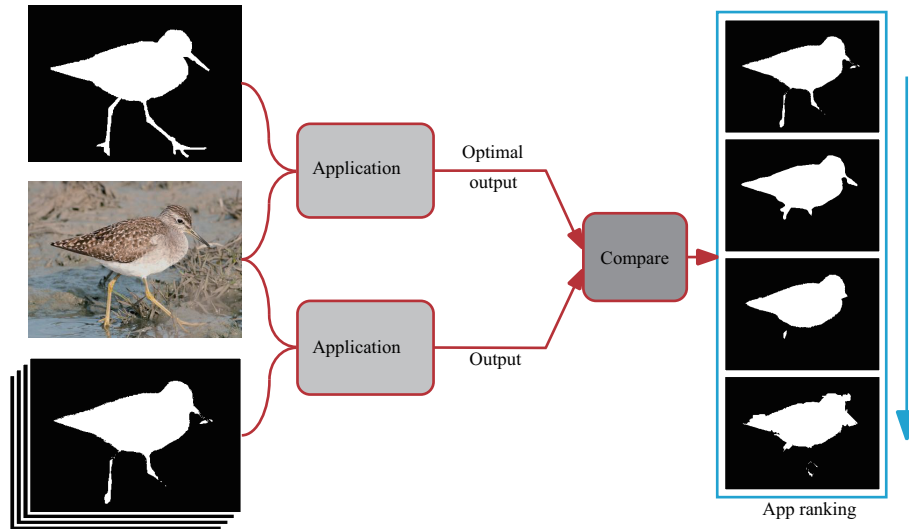


图 5 (网络版彩图) 应用程序排名. 为了得到应用程序对二值前景图的排名, 本文对比应用程序输入标准分割结果 GT 以及输入二值前景结果 FM 时的输出结果, FM 前景图与 GT 越相似, 其应用程序输出将越接近 GT 输出
 Figure 5 (Color online) Application ranking. To rank foreground maps according to an application, we compare the output when using the GT, to the output when using the FM foreground map. The more similar an FM foreground map is to a GT map, the closer its application’s output should be to the GT output

表 2 E_{ξ} 与现有评价标准在 4 个元度量上的定量比较

Table 2 Quantitative comparison between the E_{ξ} and current measures on four meta-measures^{a)}

Measure	PASCAL-S ^[31]			ECSSD ^[32]			SOD ^[33]			HKU-IS ^[2]			FMD (ours)
	MM1↓	MM2↓ (%)	MM3↓ (%)	MM1↓	MM2↓ (%)	MM3↓ (%)	MM1↓	MM2↓ (%)	MM3↓ (%)	MM1↓	MM2↓ (%)	MM3↓ (%)	MM4↓
CM ^[19]	0.610	49.78	100.0	0.504	34.62	100.0	0.723	29.89	56.22	0.613	25.26	100.0	1.492
VQ ^[18]	0.339	17.97	15.32	0.294	7.445	6.162	<u>0.335</u>	9.143	14.05	0.331	3.067	1.800	0.161
IoU/ F_1 / J_I ^[14]	<u>0.307</u>	9.426	5.597	<u>0.272</u>	4.097	1.921	0.342	<u>4.571</u>	6.857	<u>0.303</u>	0.900	0.197	<u>0.124</u>
Fbw ^[17]	0.308	<u>5.147</u>	<u>4.265</u>	0.280	<u>2.945</u>	<u>1.152</u>	0.361	6.286	<u>5.714</u>	0.312	<u>0.535</u>	<u>0.083</u>	0.149
E_{ξ} (ours)	0.247	3.093	0.001	0.247	0.641	0.001	0.273	0.571	0.001	0.274	0.084	0.001	0.121

a) MM: meta-measure. These differences are all statistically significant at the $\alpha < 0.05$ level.

为了公平比较, 本文和文献 [17] 一样也采用基于上下文的图像检索⁵⁾来执行应用关联度实验. 并且采用文献 [17] 推荐的 $\theta = 1 - \rho$ ^[40] 得分来检验评价标准给出的排序结果与应用程序给出的排序结果之间的相关性. 其中, θ 的浮动范围为 [0, 2], 0 分意味着评价标准得到的排序和应用程序得到的排序一致, 而 2 分表示完全相反的排序结果. 从表 2 元度量 1 实验发现本文的 E_{ξ} 评价标准相对于当前流行的评价标准在性能上有很大提升, 在 PASCAL-S^[31], ECSSD^[32], SOD^[33] 和 HKU-IS^[2] 数据集上, 比现有最好的标准相对提高了 19.54%, 9.19%, 18.51% 和 9.57%.

4.2 元度量 2: 随机偏好度

随机偏好度用来衡量评价标准赋予一幅没有考虑图像内容的随机结果图得分大于先进算法分割结果图得分的程度. 如图 6 所示, 本文使用一个贴近图像边缘的中心圆表示随机结果图, 并期望前景图 (图 6(c)) 相对图 6(d) 会得到更高的分数.

5) Margolin 等^[17] 证实图像检索、对象检测和物体分割等应用程序具有相似的结果. 图像检索详情见文献 [22].



图 6 (网络版彩图) 元度量 2: 性能最好的模型结果图 vs. 随机结果图. 评价标准应该为从最新模型获得的结果图赋予更高的分数, 而不是那些没有考虑内容的随机分割结果图

Figure 6 (Color online) Meta-measure 2: SOTA vs. generic. An evaluation measure should give FM (c) generated by SOTA method a higher score than the generic map (d) that does not consider contents of the image. (a) Image; (b) GT; (c) FM; (d) generic map

通过统计一张随机结果图得分高于上述 10 个模型生成结果的平均分的次数, 就可以得到错误排序比率. 考虑到偶然情况下, 随机的圆形结果得分可能会高于某个最好模型得到的糟糕结果^[17]. 因此引入 10 个先进模型的结果的均值可以增加元度量 2 的鲁棒性. 为了提高实验的可靠性, 需要对不同算法生成的 10 个结果图进行初步筛选. 如果 10 张结果图的评分都高于一个阈值, 则视为“好的结果图”. 本文选择了数据集中约 80% 的“好的结果图”来进行元度量 2 实验. 如果错误排序比率得分越低, 评价标准表现的效果就越好. 实验结果表明 E_{ξ} 的性能在 4 个数据集 PASCAL-S, ECSSD, SOD 以及 HKU-IS 上都胜过其他评价标准.

4.3 元度量 3: 噪声偏好度

第 3 个元度量是本文新提出的一个度量, 它依据的性质是评价标准应该偏好于最先进模型生成的分割结果图而不是随机噪声图.

本文采用与元度量 2 相似的实验来进行元度量 3 的实验, 但是这里使用高斯 (Gauss) 随机噪声图替代了中心圆结果图. 从表 2 实验结果来看, 本文的方法性能最好, 这归功于 E_{ξ} 考虑到局部像素匹配和全局统计信息.

4.4 元度量 4: 感知度

第 4 个度量方法是考察评价标准与人为排序之间的相关性, 称为感知度. 据我们所知, 在此之前没有开源的、经过人为排序的二类分割结果图数据集. 为此, 需要构建一个新的数据集以辅助感知度实验. 首先, 从 4 个数据集中随机挑选一些已经通过元度量 1 中应用程序排序后的分割结果图. 其次, 邀请 10 名受试者对这些经过应用程序排序过的结果图进行人为排序. 在某些情况下, 不同受试者的排序结果有所不同, 本文仅仅保留所有受试者都一致接受的排序. 如图 7 所示, 最终, 得到一个 555 张图像的 FMD 数据集⁶⁾, 它包括了 185 张 GT 图像, 以及对应的 3 种排序结果图.

为了定量评估感知度, 本文继续使用 θ 分数 (见元度量 1) 来检验这个元度量. 得分越低, 说明评价标准与人的排序结果越一致. 从表 2 和图 2 可见, 本文的标准胜过其他标准.

4.5 元度量 5: 保真度

如图 8 所示, 第 5 个元度量表达的是, 当使用错误的 GT 图时, “好的分割结果图” 的分数应该减少. 为了筛选出 “好的分割结果图”, 本文仔细分析了 4 个流行的数据集 (PASCAL-S^[31], SOD^[33],

6) FMD 数据集的分割结果从第 4 节中 10 个方法生成的结果中挑选, 我们仅仅留下差异最大的 2 个方法, 这有利于防止受试者由于多张结果图太近似进行随机选择. https://drive.google.com/file/d/18tV4Fn8SZrVA5GunPpGbM_uabzDEEayZ/view?usp=sharing.

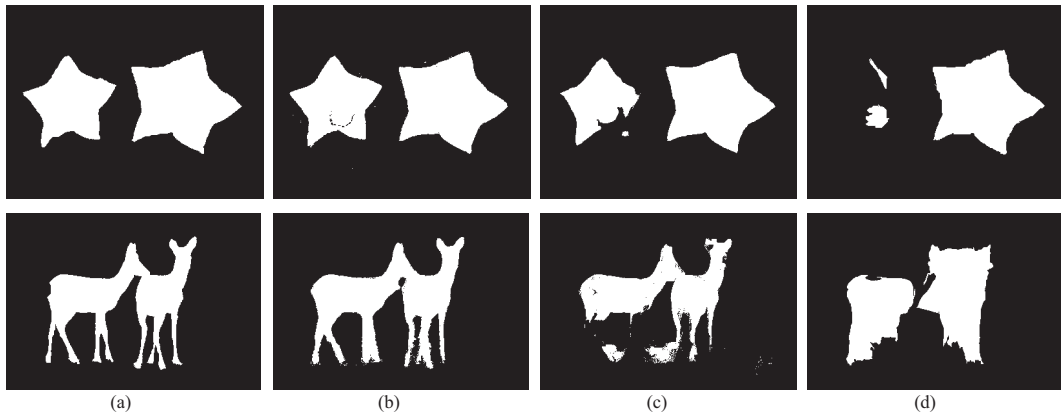


图 7 元度量 4: 感知度. 人工排序结果图来自本文构建的 FMD 数据集

Figure 7 Meta-measure 4: human ranking. Ranked examples are from FMD database newly created by us. (a) GT; (b) Rank 1; (c) Rank 2; (d) Rank 3

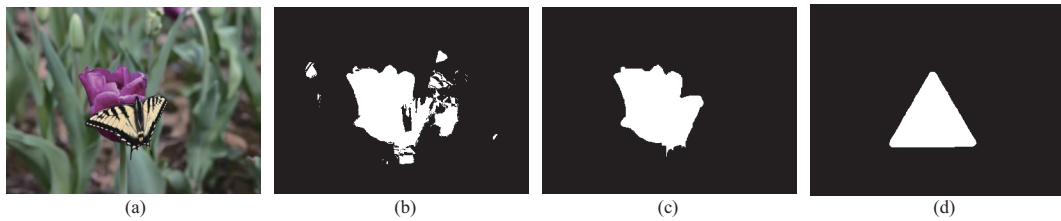


图 8 (网络版彩图) 元度量 5: 保真度. 评价标准应该给那些用正确的分割结果图 (c) 作为参考的“好”结果 (b) 更高的得分, 而不是那些使用随机选择的分割结果图 (d) 作为参考的“好”结果更高的得分

Figure 8 (Color online) Meta-measure 5: ground-truth switch. An evaluation measure should assign the “good” map (b) a higher score when using the right GT (c) as the reference than using the randomly switched GT (d). (a) Image; (b) FM; (c) GT; (d) switched GT

ECSSD^[32] 和 HKU-IS^[2]), 经验发现 $F_1 \geq 0.8$ 普遍上是较好的结果图. 为了和 Fbw^[17] 进行公平比较, 本文同样采用一个评价标准在错误使用 GT 的情况下将“好的分割结果图”赋予更高得分的频次作为评估分数. 实验结果表明, 当前所有的评价标准不相上下 (4 个数据集的平均结果为: VQ^[18] 为 0.000925%, CM^[19] 为 0.001675%, IoU/JI/ F_1 ^[14] 为 0.0014% 和本文 E_ξ 为 0.0523%). 本文的评价方法相对于其他方法有 0.05% 的差距, 这也是 E_ξ 评价标准的一个局限性.

5 认知规律启发的物体分割损失函数

为了促进物体分割技术的发展, 本文扩展了 E_ξ 评价标准的应用范围. 进一步设计了一套组合损失函数 Hybrid- E_{loss} 来促使深度学习学习到像素级、对象级和图像级的分割特征. 本文将组合损失函数解耦成 3 个损失函数, 包括加权二值交叉熵损失函数 \mathcal{L}_{ce}^w ^[41]、增强匹配损失函数 \mathcal{L}_e ^[22] 以及加权交叉并比损失函数 (\mathcal{L}_{iou}^w)^[41], 并在 3 个图像分割任务 (显著目标分割、伪装物体分割和医疗图像分割) 上验证了其通用性. 在显著目标分割和伪装物体分割实验中, 本文采用 4 个常用评价标准: 结构评价标准 (S_α)^[30]、F-measure ($F_\beta^{max}, \beta^2 = 0.3$)^[42]、增强匹配标准 (E_ξ^{max})^[22] 以及平均绝对误差 (\mathcal{M})^[43]. 对于医疗图像分割任务, 我们进一步补充了 Dice 和 IoU 评价标准.

表 3 通过解耦本文提出的组合损失函数 Hybrid- E_{loss} 进行的 SCRN 剥离实验结果Table 3 Ablation studies for SCRN via decoupling the proposed Hybrid- E_{loss} function

SCRN ^[44]	DUT-OMRON ^[46]				DUTS-Test ^[45]				HKU-IS ^[47]			
	$S_{\alpha} \uparrow$	$F_{\beta}^{\max} \uparrow$	$E_{\xi}^{\max} \uparrow$	$\mathcal{M} \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta}^{\max} \uparrow$	$E_{\xi}^{\max} \uparrow$	$\mathcal{M} \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta}^{\max} \uparrow$	$E_{\xi}^{\max} \uparrow$	$\mathcal{M} \downarrow$
$\mathcal{L}_{\text{ce}}^{\text{w}}$	0.829	0.764	0.870	0.062	0.880	0.859	0.928	0.044	0.917	0.921	0.959	0.036
$\mathcal{L}_{\text{iou}}^{\text{w}}$	0.824	0.761	0.864	0.053	0.881	0.868	0.929	0.035	0.911	0.920	0.957	0.028
\mathcal{L}_{e}	0.814	0.748	0.866	0.056	0.866	0.849	0.927	0.039	0.900	0.908	0.954	0.034
Hybrid- E_{loss}	0.833	0.770	0.868	0.055	0.887	0.872	0.930	0.036	0.918	0.925	0.959	0.028

5.1 损失函数

在二值分割模型的训练中, 最常用的是二值交叉熵和交并比损失函数. 与文献 [41] 中定义一致, 针对常用的二值交叉熵和交并比损失函数各增加了加权因子项, 给予网络对难像素点 (如边界区域或者易错部分) 更多的关注度. 因此, 我们引入加权二值交叉熵损失函数 $\mathcal{L}_{\text{ce}}^{\text{w}}$ 和加权交并比损失函数 $\mathcal{L}_{\text{iou}}^{\text{w}}$ 分别捕获像素级别和图像级别特征. 此外, 我们进一步采用本文的增强匹配损失函数 \mathcal{L}_{e} 来补足网络对于对象级别特征的表征能力. 具体而言, \mathcal{L}_{e} 损失函数定义为

$$\mathcal{L}_{\text{e}}(I_{\text{FM}}, I_{\text{GT}}) = 1 - \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h \phi_{\text{FM}}(x, y), \quad (8)$$

其中 $\phi_{\text{FM}}(x, y)$ 见式 (6), 它代表网络训练过程中的预测图 FM 在像素点 (x, y) 所对应的损失值. 因此, 我们的组合损失函数 Hybrid- E_{loss} 被定义为

$$\text{Hybrid-}E_{\text{loss}}(I_{\text{FM}}, I_{\text{GT}}) = \lambda_1 \mathcal{L}_{\text{ce}}^{\text{w}}(I_{\text{FM}}, I_{\text{GT}}) + \lambda_2 \mathcal{L}_{\text{iou}}^{\text{w}}(I_{\text{FM}}, I_{\text{GT}}) + \lambda_3 \mathcal{L}_{\text{e}}(I_{\text{FM}}, I_{\text{GT}}), \quad (9)$$

其中, 系数 $\{\lambda_1, \lambda_2, \lambda_3\}$ 用于调节组合损失函数各个部分的重要性. 本文实验设置中均默认设置为 1.

5.2 显著目标分割

实现细节. 采用与显著目标分割模型 SCRN^[44] 中一致的训练标准, 本文在 DUTS-Train^[45] (含有 10553 张训练样本) 训练集上, 使用上述 4 种损失函数对 SCRN 模型重新进行训练. 所有实验除了用于优化的损失函数外, 其训练过程的超参数均保持一致: 批大小设置为 36, 初始学习率为 $1\text{E}-4$, 总训练周期为 100 次, 每隔 60 次训练周期学习率衰减 10%, 训练与测试阶段的输入图片统一重采样到 352×352 大小. 并对训练数据使用 3 种数据增强技术以增强模型的泛化能力, 包括随机翻转、随机剪裁以及随机旋转.

实验结果. 从表 3 的结果来看, 本文的组合损失函数 Hybrid- E_{loss} 在 DUT-OMRON^[46] (包含 5168 张测试样本), DUTS-Test^[45] (包含 5019 张测试样本) 和 HKU-IS^[47] (包含 4445 张测试样本) 3 个基准测试集上获得了总体最优的性能. 如图 9 可视化结果所示, 加权交并比损失函数 $\mathcal{L}_{\text{iou}}^{\text{w}}$ 学习到了更大范围的物体分割结果, 加权二值交叉熵损失函数 $\mathcal{L}_{\text{ce}}^{\text{w}}$ 学习到了更加细节的结果, 而增强匹配损失函数 \mathcal{L}_{e} 围绕物体周围学习. 三者相辅相成, 共同学习到更加准确的分割结果.

5.3 伪装物体分割

实现细节. 采用与伪装目标分割模型 SINet^[48] 中一致的训练标准, 本文在由 CAMO^[49] (含有 1000 张训练样本) 和 COD10K^[48] (含有 3040 张训练样本) 组成的训练集上, 使用上述 4 种损失函数对 SINet 模型重新进行训练. 所有实验除了用于优化的损失函数外, 其训练过程的超参数均保持一致:

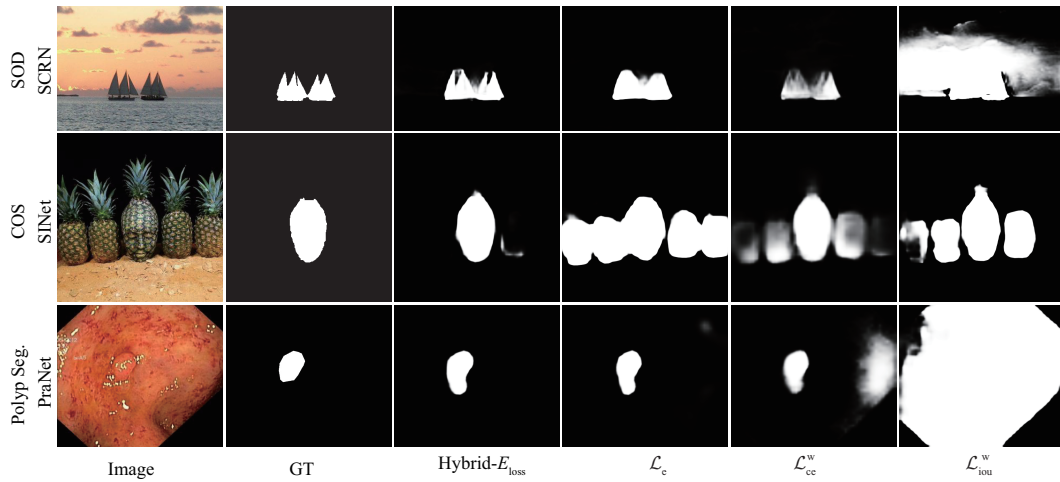


图 9 (网络版彩图) 组合损失函数结果展示对比图. 通过解耦组合损失函数 ($\text{Hybrid-}E_{\text{loss}} = \mathcal{L}_{\text{ce}}^w + \mathcal{L}_{\text{iou}}^w + \mathcal{L}_e$), 本文在 3 个常见的目标分割任务上验证了组合损失函数的有效性, 包括: 显著目标分割 (SOD, 如 SCRNet^[44])、伪装物体分割 (COS, 如 SINet^[48]) 与息肉分割 (Polyp Seg., 如 PraNet^[51])

Figure 9 (Color online) Visual comparison of the proposed Hybrid- E_{loss} function via decoupling it, i.e., Hybrid- $E_{\text{loss}} = \mathcal{L}_{\text{ce}}^w + \mathcal{L}_{\text{iou}}^w + \mathcal{L}_e$. We demonstrate the effectiveness of three sub-variants and our Hybrid- E_{loss} on three typical object segmentation tasks, including salient object detection (SOD, i.e., SCRNet^[44]), camouflaged object segmentation (COS, i.e., SINet^[48]) and polyp segmentation (Polyp Seg., i.e., PraNet^[51])

表 4 通过解耦本文提出的组合损失函数 Hybrid- E_{loss} 进行的 SINet 剥离实验结果

Table 4 Ablation studies for SINet via decoupling the proposed Hybrid- E_{loss} function

SINet ^[48]	CHAMELEON ^[50]				CAMO-Test ^[49]				COD10K-Test ^[48]			
	$S_{\alpha} \uparrow$	$F_{\beta}^{\max} \uparrow$	$E_{\xi}^{\max} \uparrow$	$\mathcal{M} \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta}^{\max} \uparrow$	$E_{\xi}^{\max} \uparrow$	$\mathcal{M} \downarrow$	$S_{\alpha} \uparrow$	$F_{\beta}^{\max} \uparrow$	$E_{\xi}^{\max} \uparrow$	$\mathcal{M} \downarrow$
$\mathcal{L}_{\text{ce}}^w$	0.876	0.843	0.947	0.040	0.772	0.743	0.854	0.090	0.791	0.706	0.882	0.044
$\mathcal{L}_{\text{iou}}^w$	0.855	0.821	0.927	0.041	0.774	0.748	0.852	0.081	0.787	0.713	0.885	0.040
\mathcal{L}_e	0.814	0.771	0.935	0.046	0.746	0.722	0.844	0.088	0.750	0.658	0.872	0.045
Hybrid- E_{loss}	0.888	0.862	0.960	0.029	0.796	0.777	0.870	0.076	0.799	0.721	0.893	0.038

批大小设置为 36, 初始学习率为 $1\text{E}-4$, 每隔 60 个周期学习率衰减 10%, 训练与测试阶段的输入图片统一重采样到 352×352 大小, 并对训练数据进行数据增强以增强模型的泛化能力, 包括随机翻转、随机剪裁、随机旋转.

实验结果. 本文在 CHAMELEON^[50], CAMO-Test^[49] 和 COD10K-Test^[48] 3 个基准测试集上验证了损失函数的有效性, 详见表 4. 图 9 第 2 行展示了伪装物体分割结果. 单独的损失函数都不能很好地学习到伪装对象的本质特征, 唯独组合损失函数学习到了完整的对象.

5.4 医疗图像分割

实现细节. 遵循与医疗图像处理领域息肉分割模型 PraNet^[51] 中一致的训练标准, 本文在 Kvasir-Train^[52] (含有 900 张训练样本) 和 CVC-ClinicDB-Train^[53] (含有 550 张训练样本) 组成的训练集上 (含有 10553 张训练样本) 使用不同的损失函数 (即 $\mathcal{L}_{\text{ce}}^w$, $\mathcal{L}_{\text{iou}}^w$, \mathcal{L}_e 和 Hybrid- E_{loss}) 对 PraNet 模型重新进行训练. 训练超参数设定与数据增强规则与 5.3 小节相同.

实验结果. 本文在 EndoScene 的测试集^[54] (又称为 CVC-300, 包含 60 张测试样本), CVC-ClinicDB^[53] (包含 62 张测试样本), 和 Kvasir^[52] (包含 100 张测试样本) 3 个基准测试集上

表 5 通过解耦本文提出的组合损失函数 Hybrid- E_{loss} 进行的 PraNet 剥离实验结果Table 5 Ablation studies for PraNet via decoupling the proposed Hybrid- E_{loss} function

PraNet [51]	CVC-300 [54]					CVC-ClinicDB [53]					Kvasir [52]				
	Dice \uparrow	IoU \uparrow	Fbw \uparrow	$E_{\xi}^{\max}\uparrow$	$\mathcal{M}\downarrow$	Dice \uparrow	IoU \uparrow	Fbw \uparrow	$E_{\xi}^{\max}\uparrow$	$\mathcal{M}\downarrow$	Dice \uparrow	IoU \uparrow	Fbw \uparrow	$E_{\xi}^{\max}\uparrow$	$\mathcal{M}\downarrow$
$\mathcal{L}_{\text{ce}}^{\text{w}}$	0.872	0.795	0.829	0.974	0.008	0.870	0.810	0.853	0.977	0.013	0.869	0.807	0.850	0.940	0.029
$\mathcal{L}_{\text{iou}}^{\text{w}}$	0.875	0.807	0.853	0.957	0.009	0.901	0.850	0.896	0.961	0.012	0.894	0.842	0.885	0.950	0.033
\mathcal{L}_{e}	0.866	0.788	0.825	0.963	0.009	0.880	0.817	0.853	0.970	0.014	0.882	0.821	0.855	0.950	0.031
Hybrid- E_{loss}	0.886	0.816	0.860	0.975	0.009	0.912	0.858	0.903	0.987	0.008	0.895	0.837	0.881	0.951	0.028

验证了所提出的组合损失函数 Hybrid- E_{loss} 的有效性, 详见表 5. 可视化结果如图 9 第 3 行所示.

6 总结与讨论

本文从人类认知规律出发, 广泛地探索了物体分割技术的评价标准并设计了一个简单、高效的增强匹配标准 E_{ξ} 来评估物体分割模型的性能. E_{ξ} 首次将局部像素和全局像素有机结合以进行评价物体分割结果, 并在多个元度量实验中取得了显著提高的性能. 同时, 本文构建了第一个开源的经过人工标注的物体分割排序数据集 FMD, 为社区进一步研究评价标准提供了便利. 最后, 在显著目标分割、伪装物体分割和医疗图像分割领域, 本文设计了一套组合损失函数 Hybrid- E_{loss} . 一系列消融实验证明了本文提出的组合损失函数的通用性. 未来我们计划将损失函数进一步应用到更广泛的图像分割领域尤其是医疗领域中的脑肿瘤分割 [55, 56] 和胰腺分割 [57] 等.

参考文献

- 1 Qin C, Zhang G, Zhou Y, et al. Integration of the saliency-based seed extraction and random walks for image segmentation. *Neurocomputing*, 2014, 129: 378–391
- 2 Li G, Yu Y. Visual saliency based on multiscale deep features. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 5455–5463
- 3 Zhang D, Meng D, Han J. Co-saliency detection via a self-paced multiple-instance learning framework. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39: 865–878
- 4 Zhang P, Wang D, Lu H, et al. Amulet: aggregating multi-level convolutional features for salient object detection. In: *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2017. 202–211
- 5 Li C, Cong R, Piao Y, et al. RGB-D salient object detection with cross-modality modulation and selection. In: *Proceedings of European Conference on Computer Vision*, 2020. 225–241
- 6 Zhang D, Tian H, Han J. Few-cost salient object detection with adversarial-paced learning. *Adv Neural Inform Process Syst*, 2020, 33: 12236–12247
- 7 Fan D P, Lin Z, Zhang Z, et al. Rethinking RGB-D salient object detection: models, data sets, and large-scale benchmarks. *IEEE Trans Neural Netw Learn Syst*, 2021, 32: 2075–2089
- 8 Fan D P, Li T, Lin Z, et al. Re-thinking co-salient object detection. *IEEE Trans Pattern Anal Mach Intell*, 2021. doi: 10.1109/TPAMI.2021.3060412
- 9 Tian Y, Li J, Yu S, et al. Learning complementary saliency priors for foreground object segmentation in complex scenes. *Int J Comput Vis*, 2015, 111: 153–170
- 10 Chen C, Li S, Wang Y, et al. Video saliency detection via spatial-temporal fusion and low-rank coherency diffusion. *IEEE Trans Image Process*, 2017, 26: 3156–3170
- 11 Kanan C, Cottrell G. Robust classification of objects, faces, and flowers using natural image statistics. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. 2472–2479
- 12 Rutishauser U, Walther D, Koch C, et al. Is bottom-up attention useful for object recognition? In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004

- 13 Arbeláez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell*, 2011, 33: 898–916
- 14 Jaccard P. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 1901, 37: 547–579
- 15 Everingham M, van Gool L, Williams C K I, et al. The Pascal visual object classes (VOC) challenge. *Int J Comput Vis*, 2010, 88: 303–338
- 16 Csurka G, Larlus D, Perronnin F, et al. What is a good evaluation measure for semantic segmentation? In: *Proceedings of British Machine Vision Conference*, 2013
- 17 Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps? In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014. 248–255
- 18 Shi R, Ngan K N, Li S, et al. Visual quality evaluation of image object segmentation: subjective assessment and objective measure. *IEEE Trans Image Process*, 2015, 24: 5033–5045
- 19 Movahedi V, Elder J H. Design and perceptual validation of performance measures for salient object segmentation. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. 49–56
- 20 Villegas P, Marichal X. Perceptually-weighted evaluation criteria for segmentation masks in video sequences. *IEEE Trans Image Process*, 2004, 13: 1092–1103
- 21 McGuinness K, O'Connor N E. A comparative evaluation of interactive segmentation algorithms. *Pattern Recogn*, 2010, 43: 434–444
- 22 Fan D P, Gong C, Cao Y, et al. Enhanced-alignment measure for binary foreground map evaluation. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018. 698–704
- 23 Li G, Yu Y. Deep contrast learning for salient object detection. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. 478–487
- 24 Wang L, Wang L, Lu H, et al. Saliency detection with recurrent fully convolutional networks. In: *Proceedings of European Conference on Computer Vision*, 2016. 825–841
- 25 Liu N, Han J. DHSNet: deep hierarchical saliency network for salient object detection. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. 678–686
- 26 Cheng M M, Mitra N J, Huang X, et al. Global contrast based salient region detection. *IEEE Trans Pattern Anal Mach Intell*, 2015, 37: 569–582
- 27 Liu T, Yuan Z J, Sun J, et al. Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell*, 2011, 33: 353–367
- 28 Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*, 2004, 13: 600–612
- 29 Pont-Tuset J, Marques F. Measures and meta-measures for the supervised evaluation of image segmentation. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013. 2131–2138
- 30 Fan D P, Cheng M M, Liu Y, et al. Structure-measure: a new way to evaluate foreground maps. In: *Proceedings of IEEE International Conference on Computer Vision*, 2017. 4548–4557
- 31 Li Y, Hou X, Koch C, et al. The secrets of salient object segmentation. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014. 280–287
- 32 Xie Y L, Lu H C, Yang M-H. Bayesian saliency via low and mid level cues. *IEEE Trans Image Process*, 2013, 22: 1689–1698
- 33 Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings of IEEE International Conference on Computer Vision*, 2001. 416–423
- 34 Liu Z, Zou W B, Meur O L. Saliency tree: a novel saliency detection framework. *IEEE Trans Image Process*, 2014, 23: 1937–1952
- 35 Wang J, Jiang H, Yuan Z, et al. Salient object detection: a discriminative regional feature integration approach. *Int J Comput Vis*, 2017, 123: 251–268
- 36 Li X, Lu H, Zhang L, et al. Saliency detection via dense and sparse reconstruction. In: *Proceedings of IEEE International Conference on Computer Vision*, 2013. 2976–2983
- 37 Zhao R, Ouyang W, Li H, et al. Saliency detection by multi-context deep learning. In: *Proceedings of IEEE Computer*

- Society Conference on Computer Vision and Pattern Recognition, 2015. 1265–1274
- 38 Chen T, Lin L, Liu L, et al. DISC: deep image saliency computing via progressive representation learning. *IEEE Trans Neural Netw Learn Syst*, 2016, 27: 1135–1149
- 39 Lee G, Tai Y W, Kim J. Deep saliency with encoded low level distance map and high level features. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. 660–668
- 40 Best D, Roberts D. Algorithm as 89: the upper tail probabilities of Spearman’s rho. *J R Stat Soc Ser C App Stat*, 1975, 24: 377–379
- 41 Wei J, Wang S, Huang Q. F³net: fusion, feedback and focus for salient object detection. In: *Proceedings of the Association for the Advance of Artificial Intelligence*, 2020. 12321–12328
- 42 Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009. 1597–1604
- 43 Perazzi F, Krähenbühl P, Pritch Y, et al. Saliency filters: contrast based filtering for salient region detection. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012. 733–740
- 44 Wu Z, Su L, Huang Q. Stacked cross refinement network for edge-aware salient object detection. In: *Proceedings of IEEE International Conference on Computer Vision*, 2019. 7264–7273
- 45 Wang L, Lu H, Wang Y, et al. Learning to detect salient objects with image-level supervision. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2017. 136–145
- 46 Yang C, Zhang L, Lu H, et al. Saliency detection via graph-based manifold ranking. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013. 3166–3173
- 47 Li G, Yu Y. Visual saliency detection based on multiscale deep CNN features. *IEEE Trans Image Process*, 2016, 25: 5012–5024
- 48 Fan D P, Ji G P, Sun G, et al. Camouflaged object detection. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020. 2777–2787
- 49 Le T N, Nguyen T V, Nie Z, et al. Anabranh network for camouflaged object segmentation. *Comput Vision Image Understand*, 2019, 184: 45–56
- 50 Skurowski P, Abdulameer H, Błaszczak J, et al. Animal camouflage analysis: chameleon database. Unpublished Manuscript, 2018. <http://kgwisc.aei.polsl.pl/index.php/en/dataset/63-animal-camouflage-analysis>
- 51 Fan D P, Ji G P, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation. In: *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020. 263–273
- 52 Jha D, Smedsrud P H, Riegler M A, et al. Kvasir-SEG: a segmented polyp dataset. In: *Proceedings of International Conference on Multimedia Modeling*, 2020. 451–462
- 53 Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Computized Med Imag Graph*, 2015, 43: 99–111
- 54 Vázquez D, Bernal J, Sánchez F J, et al. A benchmark for endoluminal scene segmentation of colonoscopy images. *J Healthcare Eng*, 2017, 2017: 1–9
- 55 Zhang D, Huang G, Zhang Q, et al. Cross-modality deep feature learning for brain tumor segmentation. *Pattern Recogn*, 2021, 110: 107562
- 56 Zhang D, Huang G, Zhang Q, et al. Exploring task structure for brain tumor segmentation from multi-modality MR images. *IEEE Trans Image Process*, 2020, 29: 9032–9043
- 57 Zhang D, Zhang J, Zhang Q, et al. Automatic pancreas segmentation based on lightweight DCNN modules and spatial prior propagation. *Pattern Recogn*, 2021, 114: 107762

Cognitive vision inspired object segmentation metric and loss function

Deng-Ping FAN¹, Ge-Peng JI², Xuebin QIN³ & Ming-Ming CHENG^{1*}

1. College of Computer Science, Nankai University, Tianjin 300350, China;

2. School of Computer Science, Wuhan University, Wuhan 430070, China;

3. Department of Computing Science, Faculty of Science, University of Alberta, Edmonton T6G 2R3, Canada

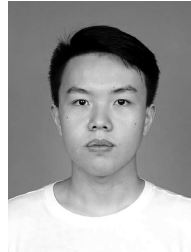
* Corresponding author. E-mail: cmm@nankai.edu.cn

Abstract Object segmentation (OS) technology is a research hotspot in computer vision, and it has a wide range of applications in many fields. Cognitive vision studies have shown that human vision is highly sensitive to both global information and local details in scenes. To this end, we design a novel, efficient, and easy-to-use enhanced-alignment measure (E_{ξ}) for evaluating the performance of the OS model. E_{ξ} combines local pixel values with the image-level mean value, jointly evaluates the image-/pixel-level similarity between a segmentation result and a ground-truth (GT) result. Extensive experiments on the four popular benchmarks via five meta-measures, i.e., application ranking, demoting generic, denying noise, human ranking, and recognizing GT, show significant relative improvement compared with existing widely-adopted evaluation metrics such as IoU and F_{β} . By using the weighted binary cross-entropy loss, the enhanced-alignment loss, and the weighted IoU loss, we further design a hybrid loss function (Hybrid- E_{loss}) to guide the network to learn pixel-, object- and image-level features. Qualitative and quantitative results show further improvement in terms of accuracy when using our hybrid loss function in three different OS tasks.

Keywords object segmentation, metric, cognitive vision, enhanced-alignment measure, loss function



Deng-Ping FAN was born in 1988. He received his Ph.D. degree from Nankai University in 2019. He joined the Inception Institute of Artificial Intelligence (IIAI) in 2019. His research interests include computer vision and visual attention, especially on RGB salient object detection (SOD), RGB-D SOD, video SOD, and Co-SOD.



Ge-Peng JI was born in 1996. He is currently pursuing his M.S. degree in communication and information system at the School of Computer Science, Wuhan University. His research interests lie in designing deep neural networks and applying deep learning in various fields of low-level vision, such as RGB salient object detection, RGB-D salient object detection, video salient object detection (VSOD), camouflaged object detection (COD), and medical image segmentation.



Xuebin QIN was born in 1990. He received his Ph.D. degree from the University of Alberta, Edmonton, Canada, in 2020. Since March 2020, he is a post-doctoral fellow in the Department of Computing Science and the Department of Radiology and Diagnostic Imaging, University of Alberta, Canada. His research interests include highly accurate image labeling, salient object detection, and image segmentation and detection.



Ming-Ming CHENG was born in 1985. He received his Ph.D. degree from Tsinghua University, Beijing, in 2012, and then worked with Prof. Philip Torr in the University of Oxford for 2 years. He is now a professor at Nankai University, Tianjin, leading the Media Computing Lab. His research interests include computer vision and computer graphics.