

高精度图像二类分割

秦雪彬¹, 戴航¹, 胡晓彬², 范登平^{*3}, 邵岭⁴, Luc Van Gool³

¹ 穆罕默德·本·扎耶德人工智能大学, 阿布扎比, 阿拉伯联合酋长国

² 腾讯优图实验室, 上海, 中国

³ 苏黎世联邦理工学院, 苏黎世, 瑞士

⁴ 特斯联科技集团, 中国

⁵ xuebin@ualberta.ca, hang.dai@mbzuai.ac.ae, xiaobin.hu@tum.de, dengpfan@gmail.com, ling.shao@ieee.org, vangool@vision.ee.ethz.ch

摘要 本文系统地研究了一项新任务, 称为图像二类分割 (DIS), 旨在从自然图像中分割出高精度的物体。为此, 本文收集了第一个大规模 DIS 数据集, 称为 DIS5K, 其中包含 5470 张高分辨率 (如 2K、4K 或更大) 图像, 覆盖了各种背景中伪装的、显著的或结构稠密的物体。DIS 使用非常精细化的标签进行注释。此外, 本文引入了一个简单的中间监督基线 (IS-Net), 使用特征级和掩码级指导来进行 DIS 模型训练。IS-Net 在 DIS5K 上的表现优于各种前沿基线, 这使其成为一个通用的自学习监督网络, 可以促进 DIS 的未来研究。此外, 本文设计了一种称为人工矫正量 (HCE) 的新指标, 它近似于校正假阳性和假阴性所需的鼠标点击操作次数。HCE 被用来衡量模型和现实应用之间的差距, 因此是现有指标的补充。最后, 本文进行了最大规模的基准测试, 评估了 16 个代表性的分割模型, 提供了关于对象复杂性的深入讨论, 并展示了几个潜在的应用 (例如背景去除、艺术设计和 3D 重建)。希望这些努力可以为学术界和工业界开辟有前景的发展方向。项目主页: <https://xuebinqin.github.io/dis/index.html>。

Keywords: 图像二类分割, 高分辨率, 衡量标准

1 引言

目前, 驱动大量人工智能 (AI) 模型的计算机视觉数据集的标注精度在一定程度上满足了机器感知系统的要求。然而, 人工智能已经步入了一个要求计算机视觉算法输出高度精确从而支持精细的人机交互时代。相比于分类 [15, 39, 74] 和检测 [29, 30, 69], 分割可以为广泛的应用提供更精确的几何目标描述, 如图像编辑 [31]、AR/VR [65]、医学图像分析 [71] 和机器人操作 [7] 等。

* 通信作者: 范登平。本文为 ECCV2022 [64] 的中文翻译版。由俞珍妮翻译, 范登平、胡晓彬校稿。