

[首页](#)[资讯](#)[深度资源](#)[产业视频](#)[GMIS峰会](#)[AI 商用搜索](#)[登录/注册](#)

SEARCH

何恺明等人提出新型半监督实例分割方法：学习分割Every Thing

By 路雪 2017年12月02日 14:57

伯克利和 FAIR 研究人员提出一种新型实例分割模型，该模型能利用目标检测边界框学习对应的分割掩码，因此大大加强了实例分割的目标数量。这种将目标检测的知识迁移到实例分割任务的方法可能是我们以后需要关注的优秀方法，机器之心对这篇论文进行了简要地介绍，更详细的内容请查看原文。

近来目标检测器准确度显著提升，获得了很多重要的新能力。其中最令人兴奋的一项是每个检测目标预测前景分割掩码，即实例分割（instance segmentation）。在实践中，典型的实例分割系统只能关注小部分视觉信息，一般约为 100 个目标类别。

该限制的主要原因是顶级的实例分割算法需要强大的监督系统，而此类监督数据很难收集新的类别，且比较昂贵。相比之下，边界框标注更丰富，也没有那么昂贵。这就引出了一个问题：在不对所有类别提供完整的实例分割标注情况下，我们还可以训练高质量实例分割模型吗？该论文以此为动机，引入了一个新型部分监督实例分割任务，提出了一种用于解决该问题的新型迁移学习方法。

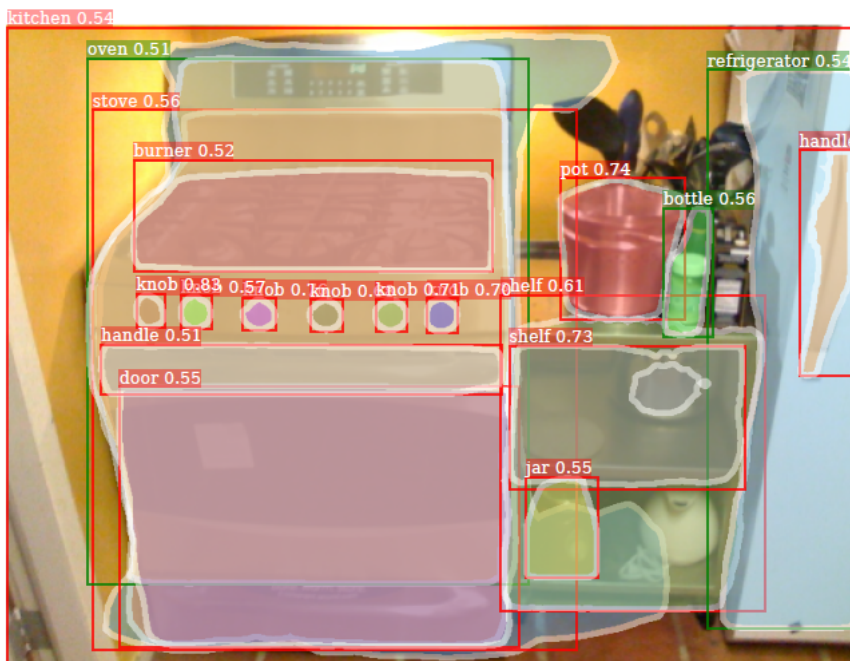


图 1. 训练部分监督的实例分割模型：类别子集（绿色框）在训练过程中有实例掩码标注；其他类别（红色框）只有边界框标注。上图显示的是使用 COCO 中 80 个类别的掩码标注对来自 Visual Genome 的 3000 个类别进行训练后的输出。

研究者按照下列方式构建部分监督的实例分割任务：（1）给定一个类别集，其中一小部分具备实例掩码标注，其他类别仅具备边界框标注；（2）实例分割算法利用该数据构建一个能够分割该集合中所有目标类别实例的模型。由于训练数据是强标注样本（带有掩码）和弱标注样本（只有边界框标注）的混合，该任务被认为是部分监督式的。

部分监督方法的主要优势是允许我们利用现有数据集的两种类型：大量类别具有边界框标注的数据集如 Visual Genome [19] 和少量类别具备实例掩码标注的数据集如 COCO [22]，构建大规模实例分割模型。这促使研究者将顶尖的实例分割方法扩展至数千个类别，这对现实世界应用部署至关重要。

为了解决部分监督实例分割问题，研究者提出了一种基于 Mask R-CNN [15] 的新型迁移学习方法，Mask R-CNN 完美适应研究者的任务，它将实例分割问题分解为边界框目标检测和掩码预测的子任务。这些子任务由联合训练的专门网络「heads」来处理。该方法背后的理念是经过训练后，边界框头部的参数对每个物体类别的嵌入进行编码，该嵌入表征使类别的视觉信息迁移至部分监督的掩码头部。

研究者将一个用来预测类别的实例分割参数的参数化权重迁移函数设计为边界框检测参数函数，从而实现该理念的具像化。权重迁移函数可以使用带有掩码标注的类别作为监督在 Mask R-CNN 中执行端到端的训练。在推断阶段，权重迁移函数用于预测每个类别的实例分割参数，从而使模型分割所有物体类别，包括训练阶段没有掩码标注的类别。

研究者在两种设置中对该方法进行了评估。首先，使用 COCO 数据集 [22] 将部分监督实例分割任务模拟成一种在具备高质量标注和评估指标的数据集上构建量化结果的方式。具体来说，将 COCO 类别分割成带有掩码标注的子集和系统只能获取边界框标注的余子集。由于 COCO 数据集仅包含少量语义分割类别（80 个），因此定量评估是精确可信的。实验结果证明该方法在强基线上改善了结果，在没有训练掩码的类别上的 mask AP 实现了 40% 的增长。

在第二种设置中，研究者使用 Visual Genome (VG) 数据集在 3000 个类别上训练大规模实例分割模型。VG 包含大量目标类别的边界框标注，但是定量评估很有难度，因为很多类别存在语义重叠（如近义词）、标注不够详尽，造成难以度量其查准率和查全率。此外，VG 不使用实例掩码进行标注。相反，研究者使用 VG 提供大规模实例分割模型的量化输出。模型输出如图 1 和图 5 所示。

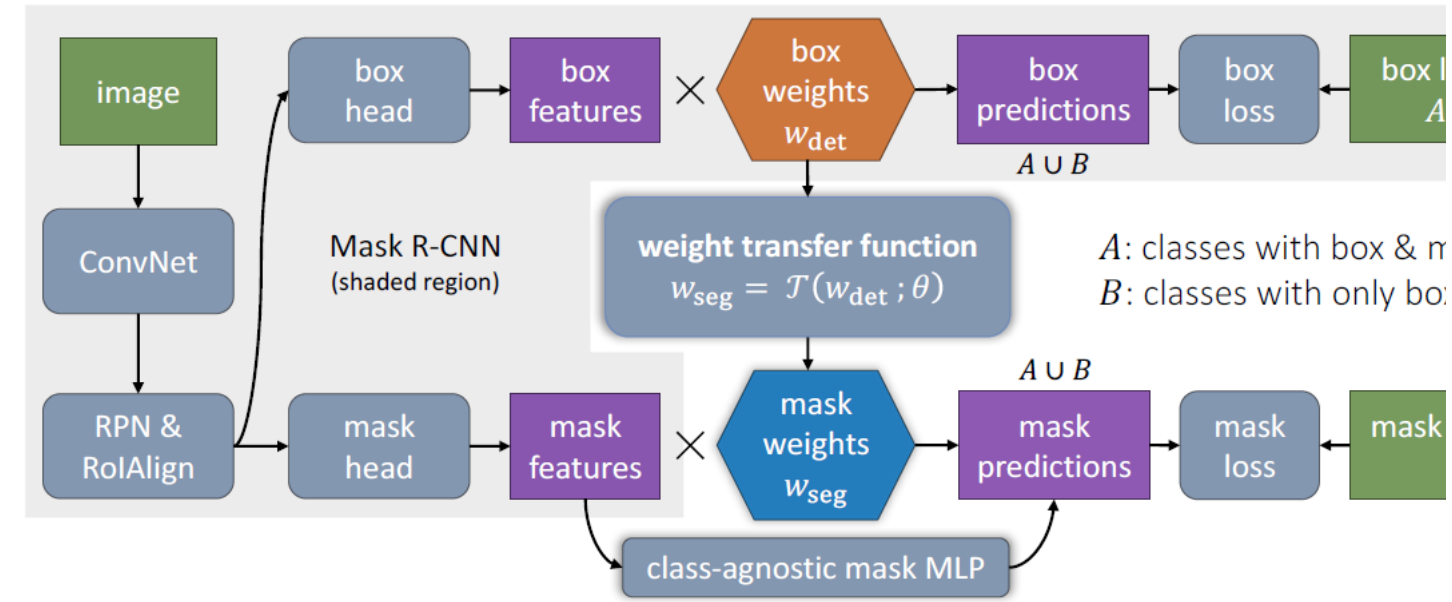


图 2. Mask R-CNN 方法的详细描述。Mask R-CNN 没有直接学习掩码预测参数 w_{seg} ，而是使用学得的权重迁移函数 T ，利用对应的检测参数 w_{det} 预测类别的分割参数 w_{seg} 。在训练中， T 仅需要集 A 中类别的掩码数据，测试阶段中它可应用于集 $A \cup B$ 中所有类别。研究者还使用补充性 (complementary) 全连接多层感知机 (MLP) 增强掩码头部标签。

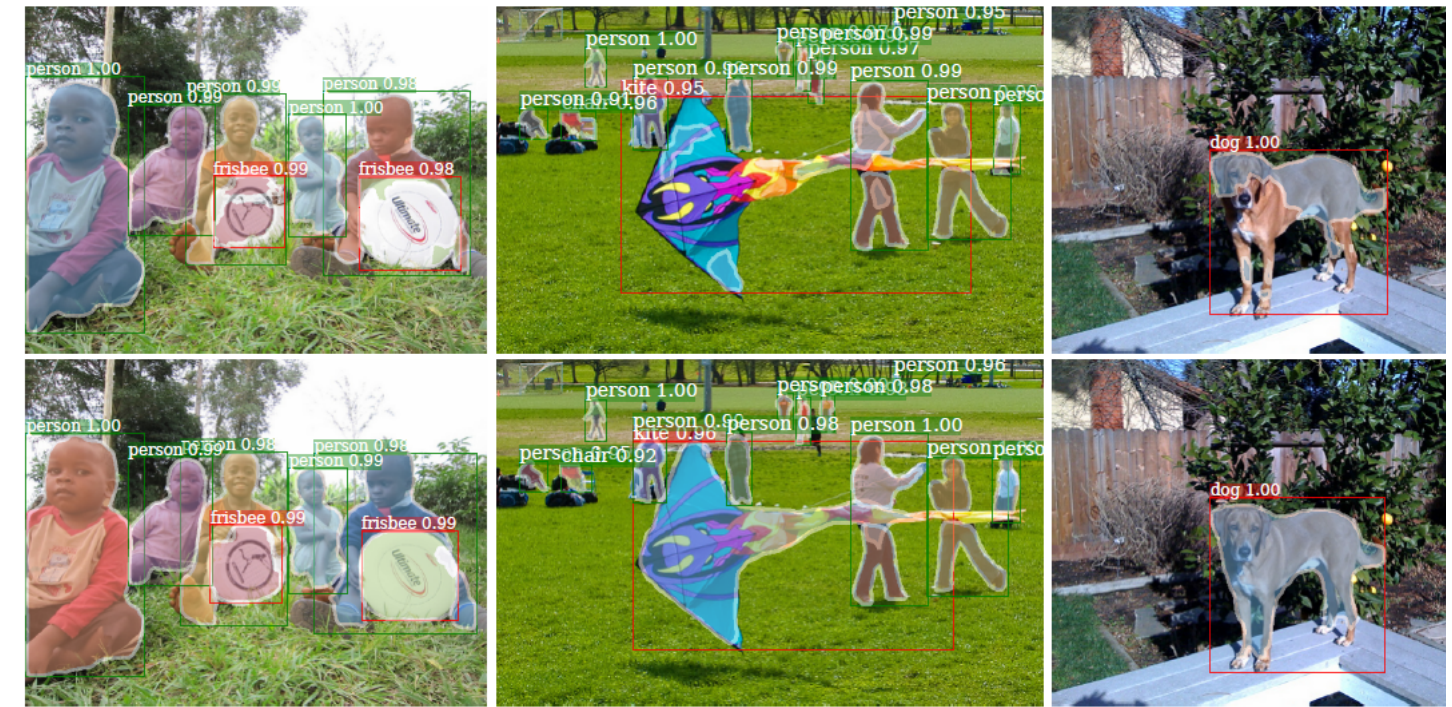


图 4. 类别不可知基线上的掩码预测（第一行）vs. Mask R-CNN 方法（第二行）。绿色框是集 A 中的类别，红色框是集 B 中的类别。左面两列是 $A = \{voc\}$ ，右面两列是 $A = \{non-voc\}$ 。

backbone	method	voc \rightarrow non-voc: test on $B = \{\text{non-voc}\}$						non-voc	
		AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L	AP	AP ₅₀
R-50-FPN	class-agnostic (baseline)	19.2	36.4	18.4	11.5	23.3	24.4	23.9	42.9
	Mask ^X R-CNN (ours)	23.7	43.1	23.5	12.4	27.6	32.9	28.9	52.2
	fully supervised (oracle)	33.0	53.7	35.0	15.1	37.0	49.9	37.5	63.1
R-101-FPN	class-agnostic (baseline)	18.5	34.8	18.1	11.3	23.4	21.7	24.7	43.5
	Mask ^X R-CNN (ours)	23.8	42.9	23.5	12.7	28.1	33.5	29.5	52.4
	fully supervised (oracle)	34.4	55.2	36.3	15.5	39.0	52.6	39.1	64.5

表 2. Mask^X R-CNN 的端到端训练。如表 1 所示，研究者使用 T 的『cls+box, 2-layer, LeakyReLU』实现，添加 MLP 掩码层 (『transfer+MLP』)，遵循相同的评估协议。研究者还报告 AP50 和 AP75 (在 0.5 和 0.75 IoU threshold 上评估的平均查准率)，以及小 (APS)、中 (APM) 和大 (APL) 目标上的 AP。对 ResNet-50-FPN 和 ResNet-101-FPN 主干网来说，在没有掩码训练数据的集 B 中的类别上，该方法显著优于基线方法。



图 5. Mask^X R-CNN 在 Visual Genom 中 3000 个类别上的掩码预测示例。绿色框是与 COCO 重叠的 80 个类别（集 A 有掩码训练数据），红色框有 COCO 中没有的 2920 个类别（集 B 没有掩码训练数据）。我们可以看到该模型在集 B 的很多类别上生成了合理的掩码预测。

Learning to Segment Every Thing

Ronghang Hu^{1,2,*} Piotr Dollár² Kaiming He² Trevor Darrell¹ Ro

¹BAIR, UC Berkeley ²Facebook AI Research (FAIR)

论文链接：<https://arxiv.org/abs/1711.10370>

目前，目标实例分割的已有方法需要用分割掩码标注所有的训练样本。这一需求使得新类别的手工标注异常昂贵，也限制了实例分割模型只能识别大约 100 个有较好注释的类别。该论文的目标是提出一种新的部分监督式训练范式，加上一种全新的权重迁移函数，能够在超大分类的数据集（所有目标都有边框注释）上训练实例分割模型，只有一小部分有掩码注释。这些能力使得我们能够在 Visual Genome 数据集上训练 Mask R-CNN，使用边框注释检测、分割 3000 多种视觉概念，且在 COCO 数据集上训练的模型能使用掩码注释检测分割 80 多类目标。我们在 COCO 数据集上认真评估了提出的该方法。该方法是实例分割模型向更为广阔的理解视觉世界所迈出的第一步。

声明：本文由机器之心编译出品，原文来自arXiv，作者Ronghang Hu等人，转载请查看要求，机器之心对于违规侵权者保有法律追诉权。

[何恺明FAIR理论论文Mask R-CNN](#)



[提交评论](#)

登录后参与评论[去登录](#)



[关于我们](#)[寻求报道](#)[商务合作](#)[服务条款](#)

©2017版权所有 机器之心（北京）科技有限公司

京 ICP 备 12027496

全球人工智能信息服务

友情链接

[Synced Global](#)[机器之心](#) [Medium](#) [博客PaperWeekly](#)[网易智能动脉网](#)[硬蛋网](#)



联系电话：+86 010-57150141

联系邮箱：contact@jiqizhixin.com