

Centermask 复现分析及总结

1. 数据集

COCO-Person 类别，训练集共有 64115 张图片 257253 个人体实例，验证集共有 2693 张图片 10777 个人体实例。

样本分布:大约 41% 的小目标 ($\text{area} < 32 \times 32$), 34% 的中等目标 ($32 \times 32 < \text{area} < 96 \times 96$), 和 24% 的大目标 ($\text{area} > 96 \times 96$)

2. 网络结构

训练采用网络结构为: centermask-lite-V-19-eSE-slim-dw-FPN-ms-4x.pth <其中: slim 代表通道减半, dw 代表使用深度可分离卷积,基于 VoVNet2>

3. 实验结果分析

①Recall

Average Recall	(AR) @[IoU=0.50:0.95 area= small maxDets=100]	= 0.331
Average Recall	(AR) @[IoU=0.50:0.95 area=medium maxDets=100]	= 0.551
Average Recall	(AR) @[IoU=0.50:0.95 area= large maxDets=100]	= 0.672

结论: 根据不同目标尺寸大小的召回率对比情况, 可知检测器对于小目标的检测效果还是不好。

②AP

[07/01 08:34:11 d2.evaluation.testing]:	copypaste: Task: bbox
[07/01 08:34:11 d2.evaluation.testing]:	copypaste: AP, AP50, AP75, APs, APm, AP1
[07/01 08:34:11 d2.evaluation.testing]:	copypaste: 50.4554, 78.9691, 53.3882, 32.6193, 56.5379, 69.3345
[07/01 08:34:11 d2.evaluation.testing]:	copypaste: Task: segm
[07/01 08:34:11 d2.evaluation.testing]:	copypaste: AP, AP50, AP75, APs, APm, AP1
[07/01 08:34:11 d2.evaluation.testing]:	copypaste: 42.4175, 75.0290, 42.9795, 22.9157, 47.2288, 62.4050

Backbone	Params.	AP _{mask}	AP _S ^{mask}	AP _M ^{mask}	AP _L ^{mask}	AP _S ^{box}	AP _S ^{box}	AP _M ^{box}	AP _L ^{box}	Time (ms)
MobileNetV2 [31]	28.7M	29.5	12.0	31.4	43.8	32.6	17.8	35.2	43.2	56
VoVNetV2-19 [19]	37.6M	32.2	14.1	34.8	48.1	35.9	20.8	39.2	47.6	59

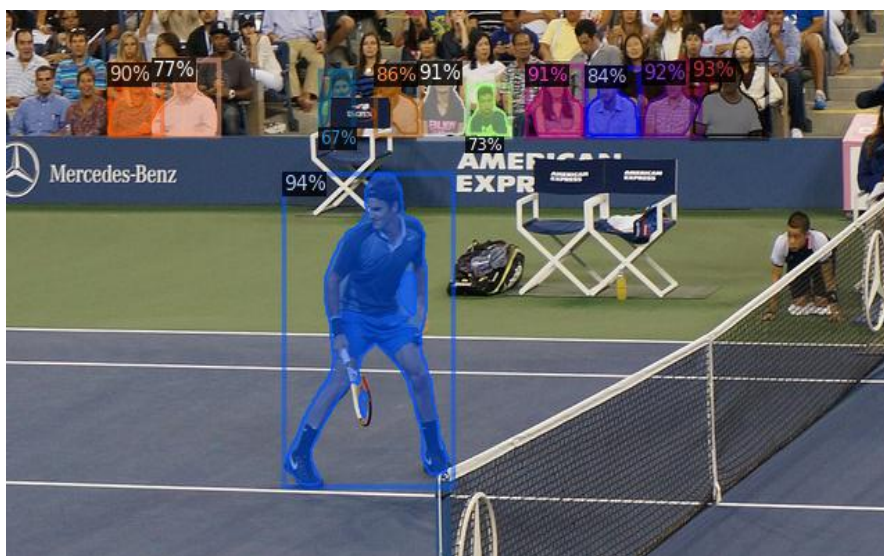
结论: AP_{mask} 为 42.4%, AP_{box} 为 50.4%, 其中 AP50 可以达到 75.0%, 与论文中的结果 (AP_{box}:35.9%, AP_{mask}:32.2%) 相比, 单独训练 Person 这一类别明显高于其 80 个类别的平均, 从而证明, 只要针对 Person 类别去训练优化, 肯定会有较好的结果。

③FPS

Inference done 2643/2693. 0.0260 s / img. ETA=0:00:02
Total inference time: 0:01:52.923636 (0.042010 s / img per device, on 1 devices)
Total inference pure compute time: 0:01:10 (0.026043 s / img per device, on 1 devices)

结论: 可以达到实时性要求

4. Badcase



5. Good



总结：整体而言，CenterMask 在速度与精度上都有了很大的提升，但通过分析 Badcase，发现对于密集型多目标 ($>5/6$) 分割效果较差，而同样的问题在 TensorMask 中得到了解决，后续的工作可以多一方面这样的思考。基于 centermask 可以有更多的尝试。

附：TensorMask 分割图

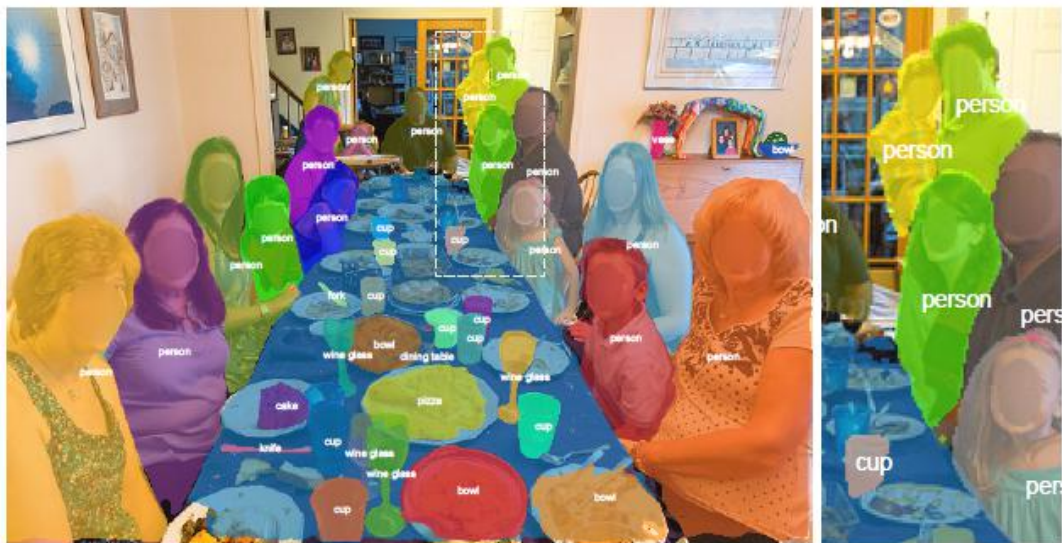


Figure 1. Selected output of *TensorMask*, our proposed framework