

目标跟踪调研报告

视觉跟踪 (Visual Tracking)

经典的方法都与相关滤波 (KCF, SRDCF, CF2, CCOT, ECO... 等等)。视觉跟踪领域的方法的选择, 一般基于三点: 足够高效、开源、基于特定场景效果较好。相关滤波的发展已经经过了几代迭代, 深度学习方向也在不断积攒 (广度高而深度浅)。目前来看, 应该是以 SiamFC 为代表的 Siamese Tracker 脱颖而出。

早期的跟踪算法都是坐标轴对齐的矩形框。但随着跟踪精度的不断提升, 数据集的难度在不断提升, 在 VOT2015 时即提出使用旋转矩形框来作为标记。在 VOT2016 的时候提出自动的通过 mask 来生成旋转框的方法。

将 2013 年以后的跟踪分为几类:

- 第一类是预测 score 的方法, 这类算法以相关滤波和 SiameFC 为代表。通过预测候选区域的 score map 来得到物体的位置, 物体的尺度大小通常是通过图像金字塔得到。同时无法得到物体的长宽比变化。
- 第二类就是以 GOTURN 和 SiamRPN 为代表的做 boundingbox regression 的方法。这也是 SiamRPN 取得当前最好结果的核心所在, 充分挖取精度方向的红利。实际上并不是 SiamRPN 预测的有多稳定, 而是在预测正确的时候, 会给出更为精确的 box。利用网络预测长宽比可以调整 box

目标分割

视频目标分割 (VOS) 领域, 之前普遍流行的算法是利用语义分割网络在线进行一个二分类的训练, 然后再后续帧进行预测。这种方法在训练过程中一般都会花费数分钟, 给人一种电脑假死的感觉。最近越来越多的不需要在线 finetune 的算法被提出。但其速度仍然无法到达令人满意的状态, 例如 FAVOS 和 OSMN 分别需要 1s / 帧, 120ms / 帧。这距离真正的实时运行还是有一定差异。另一方面, VOS 算法的第一帧需要给定目标的 mask, 这在人机交互的场景中很难时间, 这个 mask 获取成本过高。

所以我们提出了对视觉目标跟踪 (VOT) 和视频目标分割 (VOS) 的统一框架 SiamMask。我们将初始化简化为视频跟踪的 box 输入即可, 同时得到 box 和 mask 两个输出

SiamMask

结果

1. 实验评价

- 在视频跟踪领域 (VOT), VOT2016 和 VOT2018 数据集上的性能, 该方法达到比以前更好的结果, 同时保持了 55fps 的超实时的性能表现。

	SiamMask	SiamMask-2B	DaSiamRPN [72]	SiamRPN [31]	SA_Siam_R [17]	CSRDCF [37]	STRCF [32]	LSART [56]	ECO [15]
EAO \uparrow	0.347	0.334	0.326	0.244	0.337	0.263	0.345	0.323	0.280
Accuracy \uparrow	0.602	0.575	0.569	0.490	0.566	0.466	0.523	0.495	0.484
Robustness \downarrow	0.288	0.306	0.337	0.460	0.258	0.318	0.215	0.218	0.276
Speed (fps) \uparrow	35	40	160	200	32.4	48.9	2.9	1.7	3.7

- 在视频目标分割领域（VOS），我们取得了当前最快的速度。在 DAVIS2017 和 Youtube-VOS 上，我们和最近发表的较为快速的算法对比，我们的算法可以取得可比较的分割精度，同时速度快了近一个数量级。对比经典的 OSVOS，我们的算法快了近三个数量级，使得视频目标分割可以得到实际使用。

	FT	M	$\mathcal{I}_M \uparrow$	$\mathcal{I}_O \uparrow$	$\mathcal{I}_D \downarrow$	$\mathcal{F}_M \uparrow$	$\mathcal{F}_O \uparrow$	$\mathcal{F}_D \downarrow$	Speed
OnAVOS [61]	✓	✓	61.6	67.4	27.9	69.1	75.4	26.6	0.1
OSVOS [7]	✓	✓	56.6	63.8	26.1	63.9	73.8	27.0	0.1
FAVOS [11]	✗	✓	54.6	61.1	14.1	61.8	72.3	18.0	0.8
OSMN [67]	✗	✓	52.5	60.9	21.5	57.1	66.1	24.3	8.0
SiamMask	✗	✗	51.1	60.5	-1.1	55.0	64.3	1.9	35

1. 主观评测



