



基于孪生中心感知网络的实时视觉跟踪 Real-Time Visual Tracking Based on Siamese Center-Aware Network

周文璋², 文龙银³, 张立波^{1,*}, 独大为^{4,*}, 罗铁坚², 武延军¹

¹智能软件研究中心, 中国科学院软件研究所, 中国 ²中国科学院大学, 中国
³字节跳动, 美国 ⁴纽约州立大学奥尔巴尼分校, 美国

期刊: IEEE Transactions on Image Processing
通讯作者: 张立波
联系方式: 18655882017
libo@iscas.ac.cn

问题: 利用孪生中心感知分支捕捉和定位目标中心区域, 增加多尺度不同移动速度目标跟踪鲁棒性
应用: 视频和摄像头中物体的实时跟踪

动机

项目	代表性方法	不足
相关滤波器	ECO: Efficient convolution operators for tracking	这类方法需要花大量时间实时收集样本和更新模型, 尤其随着序列长度增加, 因此具有较慢的速度
	Learning spatial-temporal regularized correlation filters for visual tracking	
	Correlation tracking via joint discrimination and reliability learning	
深度孪生网络	Fully-convolutional siamese networks for object tracking	这些方法依赖预先设定的anchor类型, 不能很好自适应不同移动速度和尺寸变化, 尤其是快速移动和遮挡
	High performance visual tracking with siamese region proposal network	
	SiamRPN++: Evolution of siamese visual tracking with very deep networks	
循环神经网络	Antidecay LSTM for Siamese Tracking With Adversarial Learning	这类方法并不能有效捕捉物体结构变化趋势, 且导致物体模板信息发生偏差, 不能很好学习物体具有判别性特征
	Visual Tracking via Dynamic Memory Networks	
	DeepMTT: A deep learning maneuvering target-tracking algorithm based on bidirectional LSTM network	

挑战及贡献

1. 挑战

- 跟踪的物体具有不断变化的移动速度、尺度、光照环境和视角等, 尤其是物体快速移动和形态发生巨大变化, 影响跟踪目标的判断
- 跟踪的物体会随机出现丢失和遮挡现象, 导致跟踪失败

2. 贡献

- 提出了一种孪生中心感知的实时跟踪模型, 利用localization分支定位跟踪物体区域, 增加classification和regression分支识别和大小获取的精度
- 在localization分支, 利用全局上下文模块捕捉目标区域各部分之间的依赖关系, 增加目标整体的判断性



SiamCAN实时跟踪网络

1. 特征提取

模板图像和当前搜索图像通过共享的孪生ResNet50网络, 分别提取Conv5_3的特征图
特征图经过1*1的卷积进行特征降维, 模板图像需要从特征图中心裁剪7*7大小的特征区域

2. 分类分支

分类分支主要将从孪生网络提取的多尺度模板特征和当前检测特征分别对应进行深度相关卷积, 得到预测得分热图

$$F_{h \times w \times 2k}^{cls}(m) = \phi_m^{cls}(\alpha) * m^{cls}(\beta)$$

然后通过两个步长stride=2的3*3卷积和平均池化和全连层预测每层特征所占权重, 最后利用这些权重融合多尺度得分热图

$$O_{h \times w \times 2k}^{cls} = \sum_{m=1}^L \gamma_m^{cls} F_{h \times w \times 2k}^{cls}(m)$$

3. 回归分支

类似于分类分支, 回归分支主要将从孪生网络提取的多尺度模板特征和当前检测特征分别对应进行深度相关卷积, 得到物体大小预测图

$$F_{h \times w \times 2k}^{reg}(m) = \phi_m^{reg}(\alpha) * m^{reg}(\beta)$$

$$O_{h \times w \times 2k}^{reg} = \sum_{m=1}^L \gamma_m^{reg} F_{h \times w \times 2k}^{reg}(m)$$

4. 定位分支

基于anchor-based的分类分支和回归分支很难精准预测物体位置和大小。我们提出一种定位分支, 直接预测物体中心位置。定位分支主要将从孪生网络提取的多尺度模板特征和当前检测特征分别对应进行通道点乘, 得到预测物体中心热图

$$F_{h \times w \times 2k}^{loc}(m) = S[\phi_m^{loc}(\alpha)] * m^{loc}(\beta)$$

$$O_{h \times w \times 2k}^{loc} = \sum_{m=1}^L \gamma_m^{loc} F_{h \times w \times 2k}^{loc}(m)$$

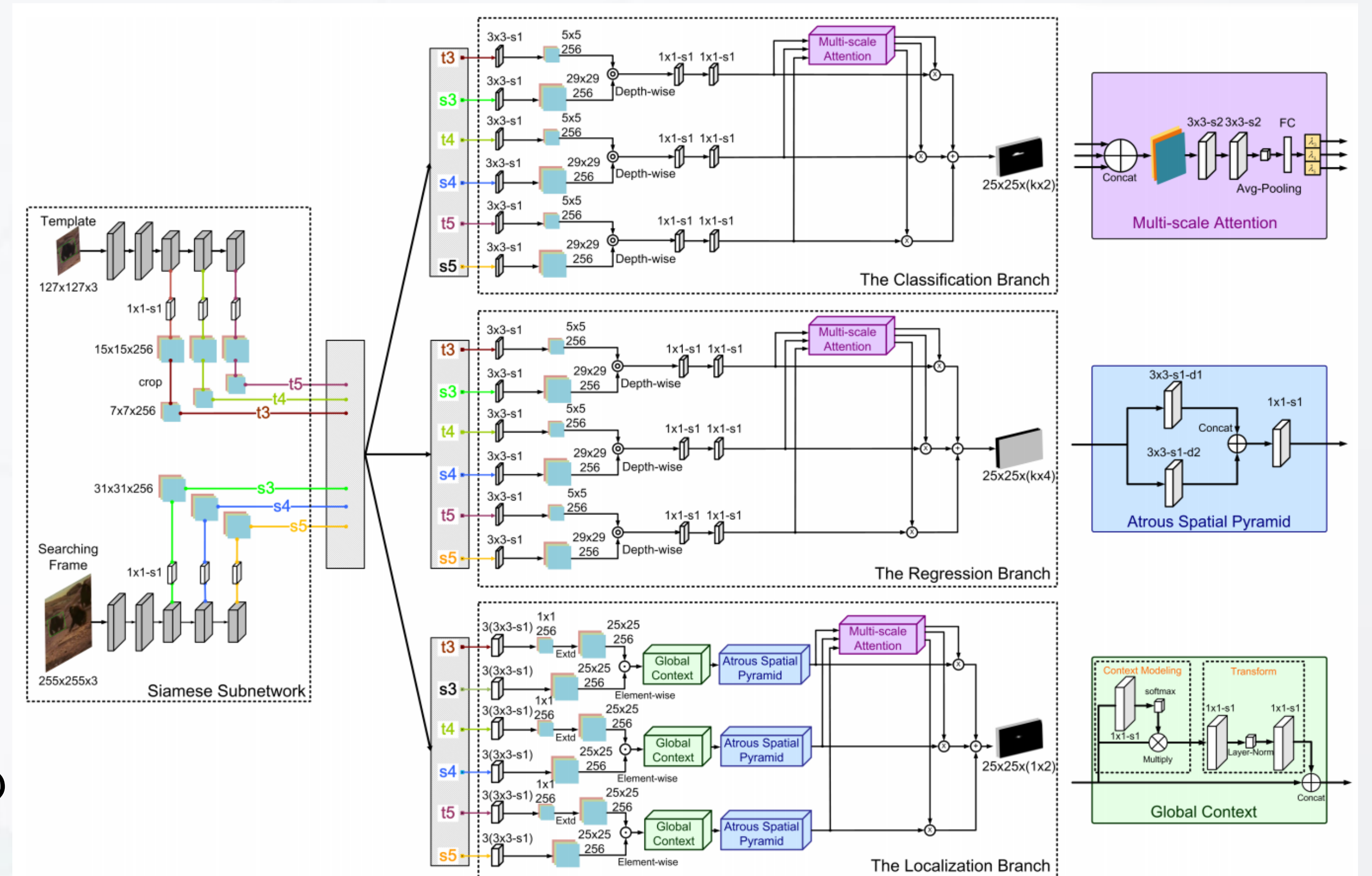
5. 全局上下文

我们利用全局上下文模块捕捉目标各部分以及目标和背景之间的长期依赖关系, 它由上下文模块和变换模块构成。上下文模块每个像素与其他所有像素之间存在依赖关系, 变换模块能够捕捉特征通道的依赖关系

6. 损失函数

总的损失函数由分类分支, 回归分支和定位分支构成, 分别优化三个分支网络, 共同优化孪生特征提取网络

$$L = \lambda_{cls} L_{cls}(u, u^*) + \lambda_{reg} L_{reg}(p, p^*) + \lambda_{loc} L_{loc}(c, c^*)$$



实验

➢ VOT数据集(M. Kristan et al, 2016 and 2018)

➢ 60个视频序列

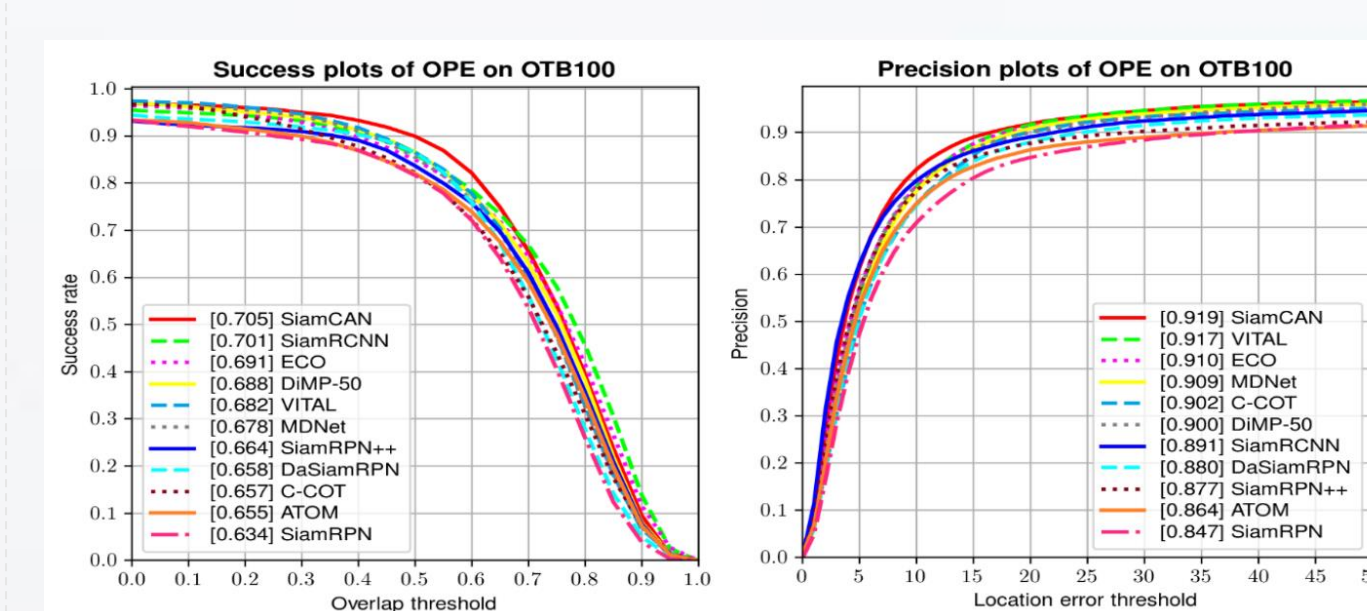
Tracker	A	R	EAO
SiamFC [35]	0.530	0.460	0.235
C-COT [9]	0.539	0.238	0.331
CSRF [58]	0.510	0.240	0.338
SiamRPN [11]	0.560	1.080	0.344
FCF [42]	0.581	1.020	0.356
C-RPN [39]	0.594	0.950	0.363
SiamRPN++ [41]	0.580	0.240	0.370
ECO [1]	0.550	0.200	0.375
ASRCF [59]	0.565	0.187	0.391
DaSiamRPN [40]	0.610	0.220	0.411
ROAM++ [60]	0.599	0.174	0.441
SiamMask+ [7]	0.643	0.219	0.455
SiamRPN++ [6]	0.642	0.196	0.464
SiamMask_E [12]	0.677	0.224	0.466
PTS [61]	0.642	0.144	0.471
SiamCAN*	0.636	0.149	0.513

VOT2016

VOT2018

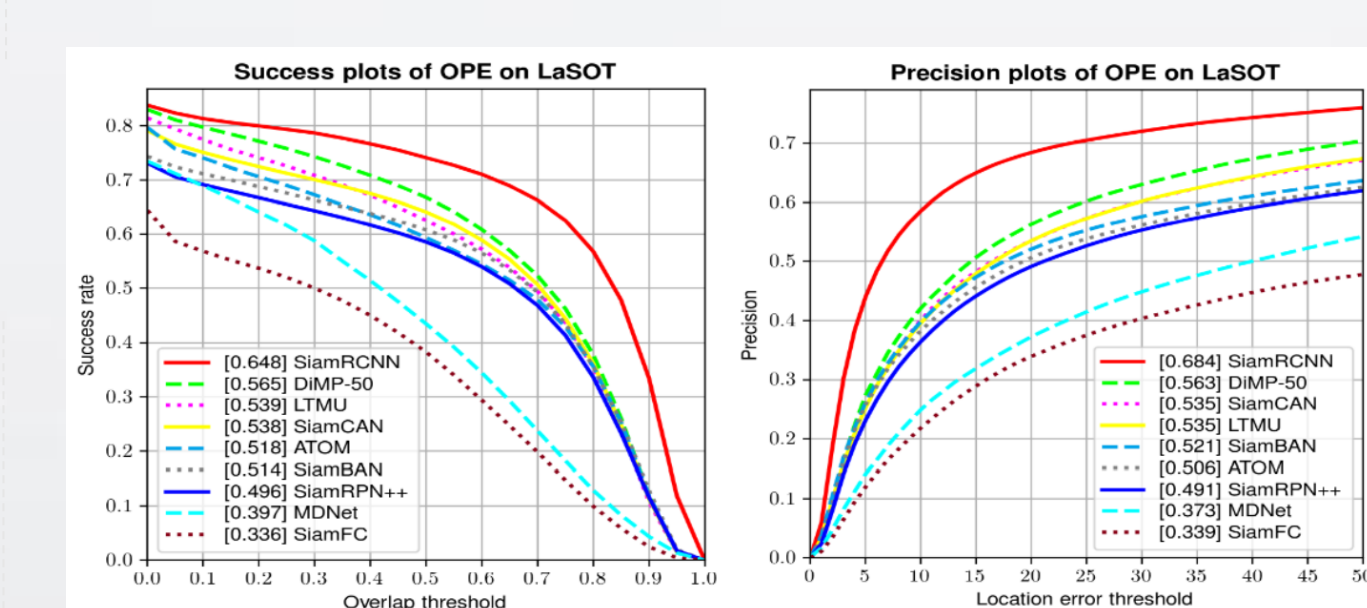
➢ OTB数据集(Y.Wu et al, 2015)

➢ 100个视频序列



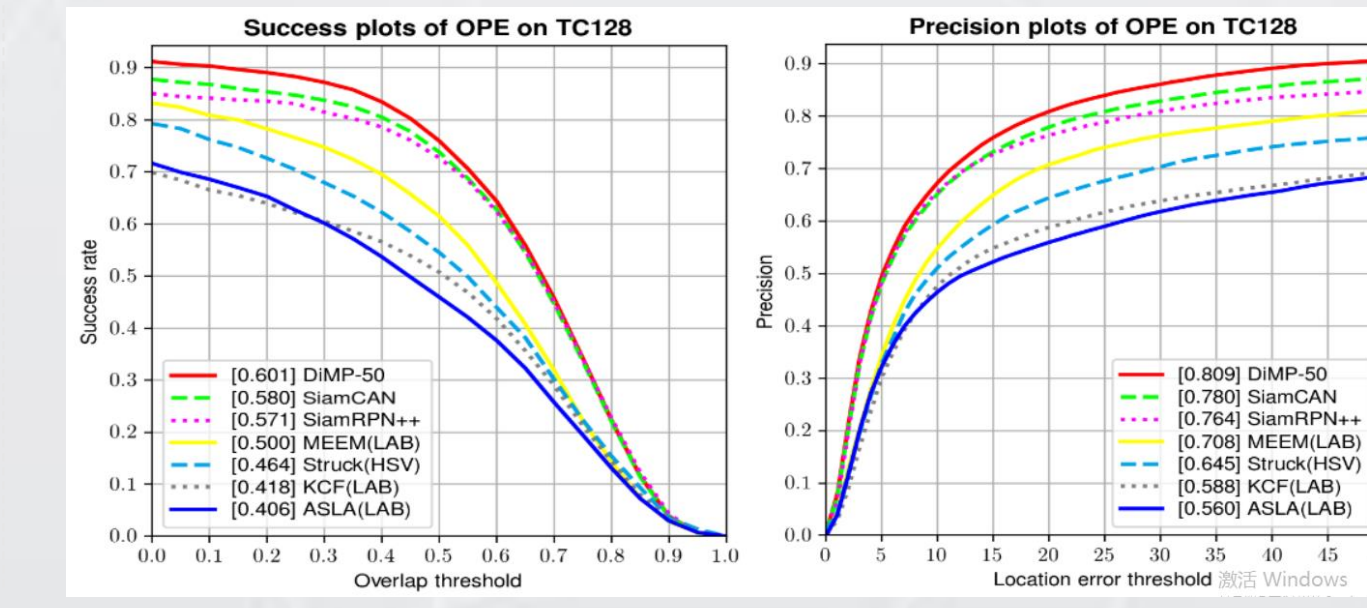
➢ LASOT数据集(H. Fan et al, 2019)

➢ 280个视频序列, 平均视频2500帧



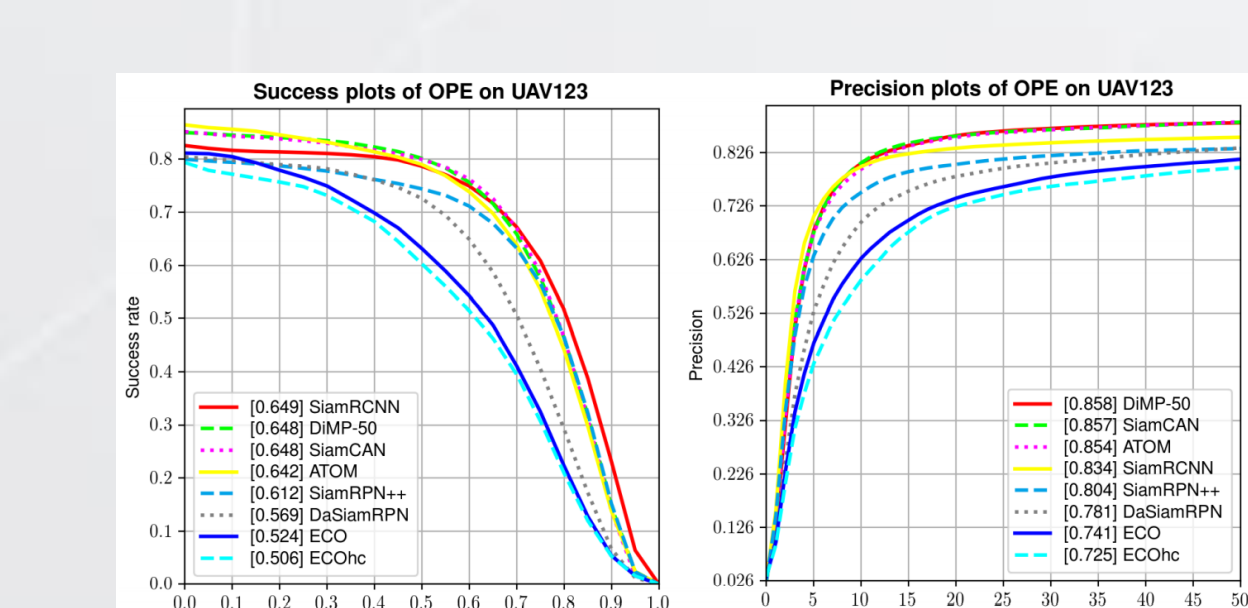
➢ TC-128数据集(P. Liang et al, 2015)

➢ 128个视频序列



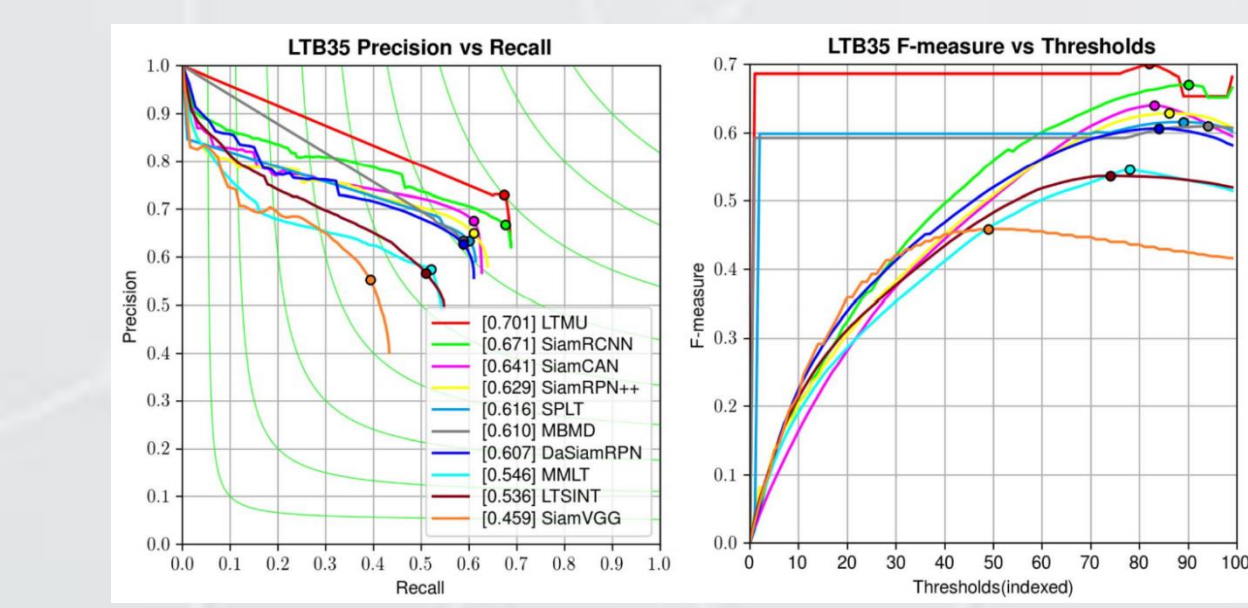
➢ UAV数据集(M. Mueller et al, 2016)

➢ 123个视频序列



➢ LTB35数据集(A. Lukezic, 2018)

➢ 35个视频序列



消融研究

1. 不同跟踪算法速度和网络复杂度对比

Trackers	Online Update	FPS	GFLOPs	# Params
SiamKPN [47]		28	76.07	90.75M
ATOM [8]	✓	28	3.97	17.29M
DIMP-50 [64]	✓	29	10.35	43.10M
Ocean [48]	✓	52	23.66	26.46M
SiamMask_E [12]		52	20.09	21.48M
SiamRPN++ [6]		53	59.56	53.95M
SiamBAN [45]		54	59.55	53.93M
SiamMask [12]		54	20.08	21.48M
DaSiamRPN [40]		96	21.85	90.44M
SiamCAN		45	64.62	61.14M

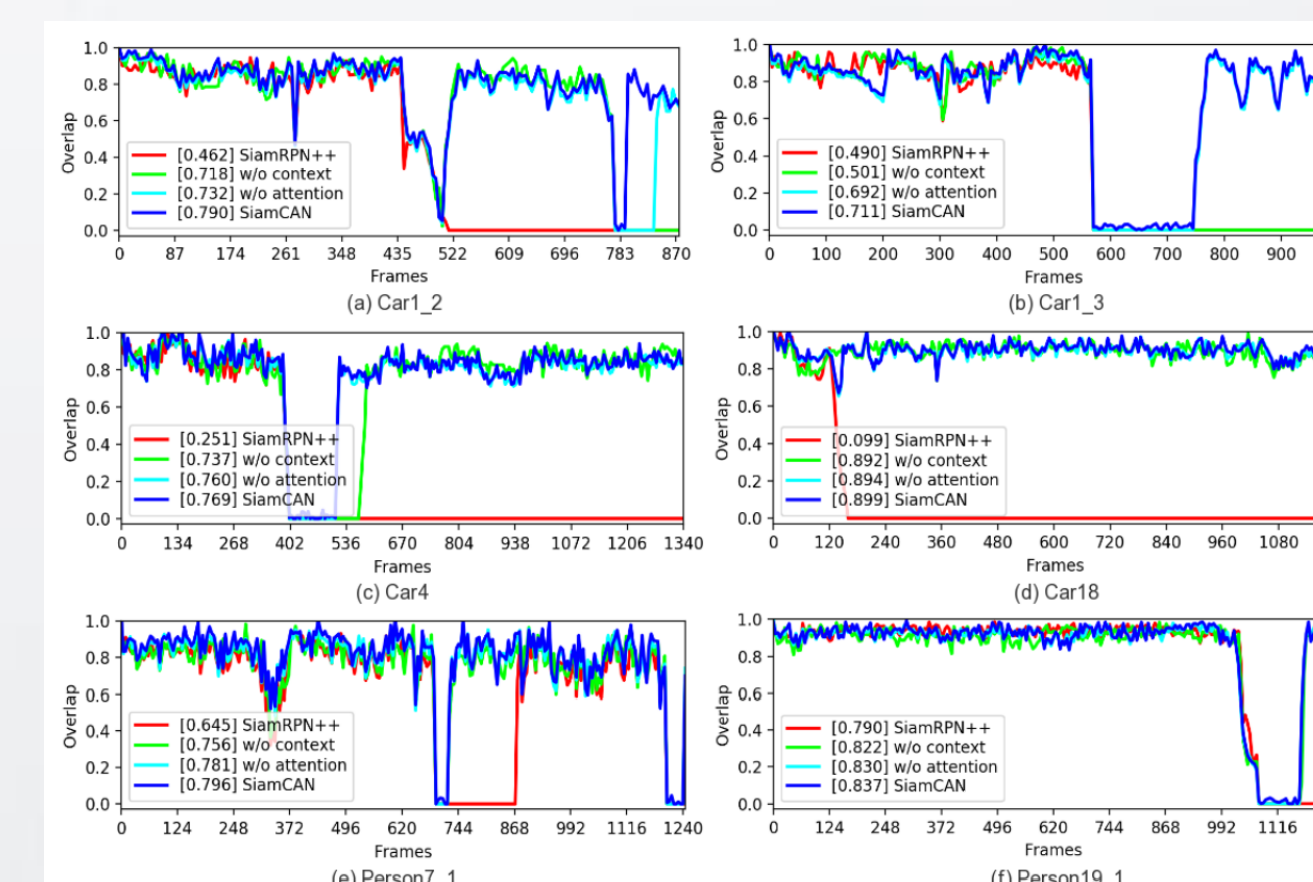
2. 主要模块的EAO指标对比

Component	SiamCAN				
	localization branch?	global context?	multi-scale attention?	✓	✓
VOT2016	0.464	0.472	0.488	0.494	0.504
VOT2018	0.415	0.421	0.432	0.446	0.447

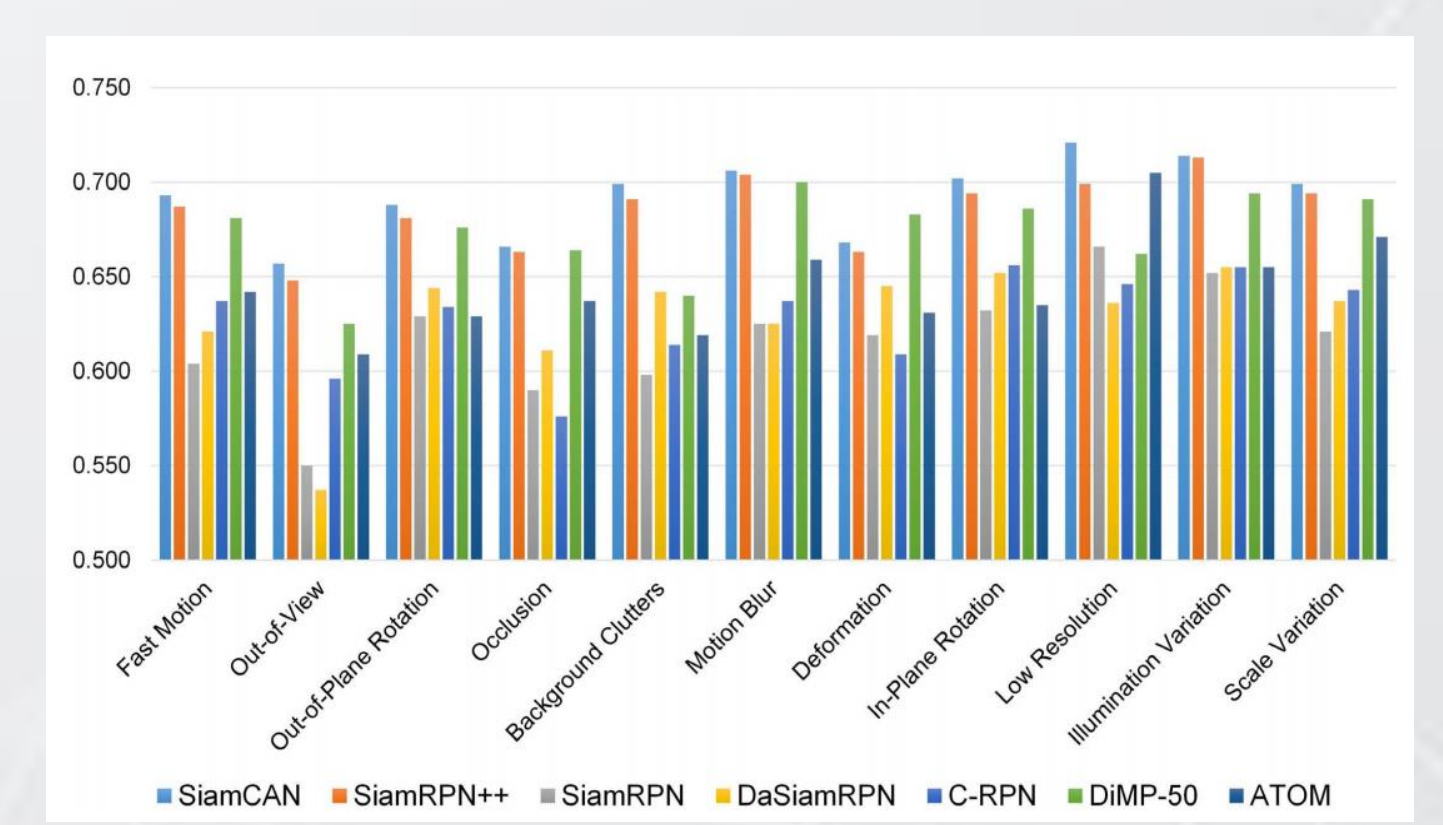
3. 多阶段训练优化的比较

Optimization	A	R	EAO
one-stage	0.615	0.243	0.410
two-stage	0.604	0.178	0.463
three-stage	0.605	0.183	0.462

4. 不同模块在6个视频序列跟踪结果



5. 在OTB数据集上不同属性的比较结果



6. SiamCAN变体和几种Siamese网络算法的比较

Trackers	Architecture	OTB100 [21]		UAV123 [25]	
		Success Score	Precision Score	Success Score	Precision Score
SiamBAN [45]	Anchor-free	0.696	0.910	0.631	0.833
SiamCAR [46]	Anchor-free	0.697	0.910	0.614	0.760
SiamRPN++ [6]	Anchor-based	0.664	0.877	0.612	0.804
SiamRCNN [52]	Anchor-based	0.701	0.891	0.649	0.834
SiamCAN	Fusion	0.701	0.916	0.640	0.843