

Distractor-aware Siamese Networks for Visual Object Tracking

Winning entry in VOT 2018 Real-time challenge

Speaker: Qiang Wang

Qiang Wang¹, Zheng Zhu¹, Bo Li², Wei Wu², Weiming Hu¹

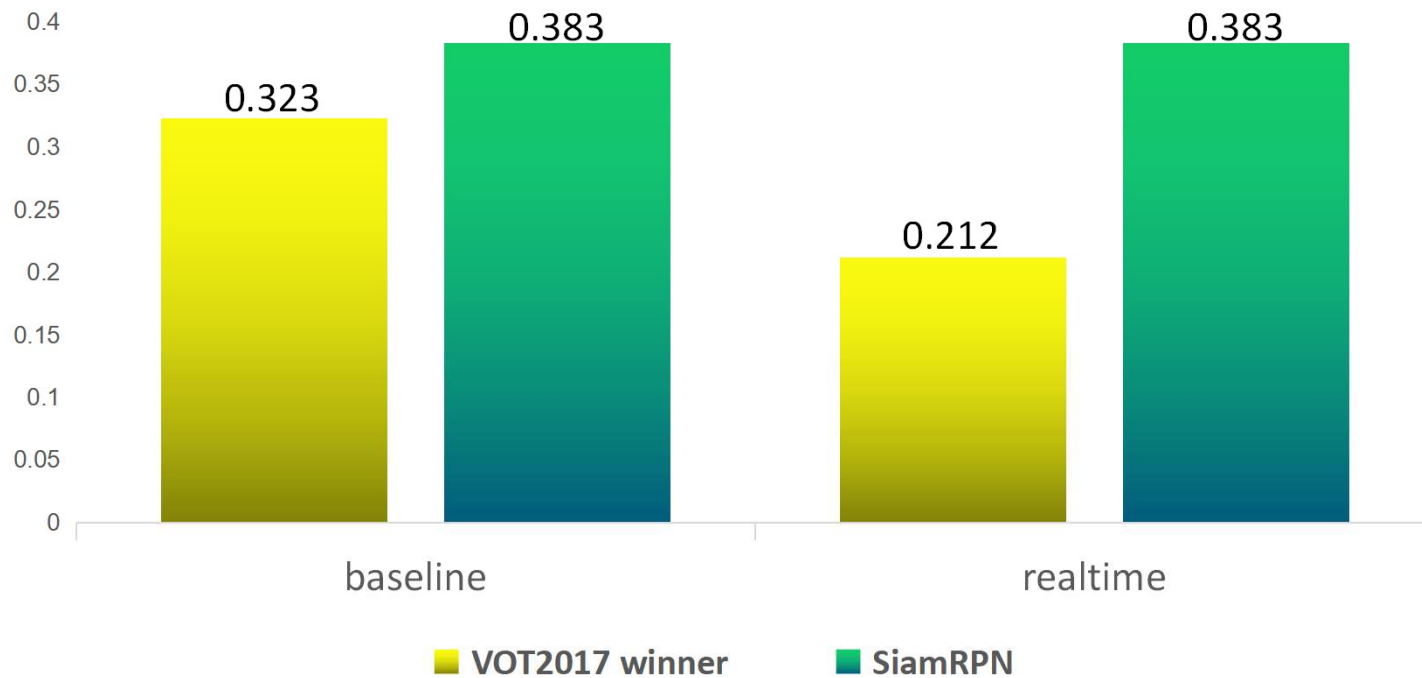
¹Institute of Automation, Chinese Academy of Sciences

²Sensetime Group Limited



中国科学院自动化研究所

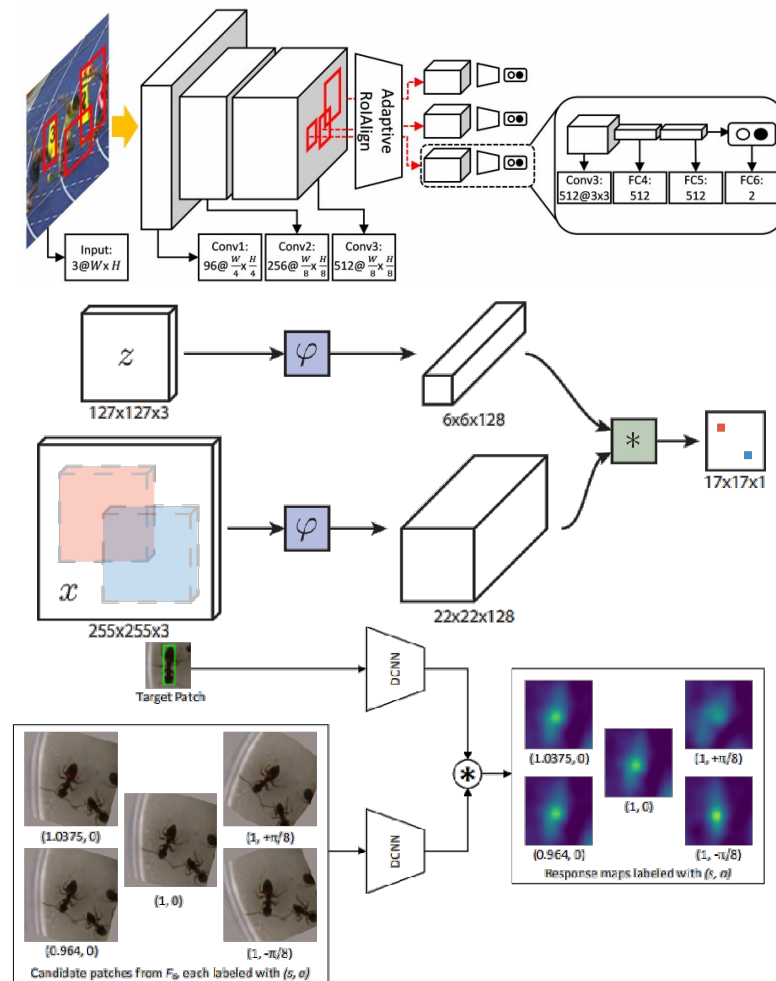
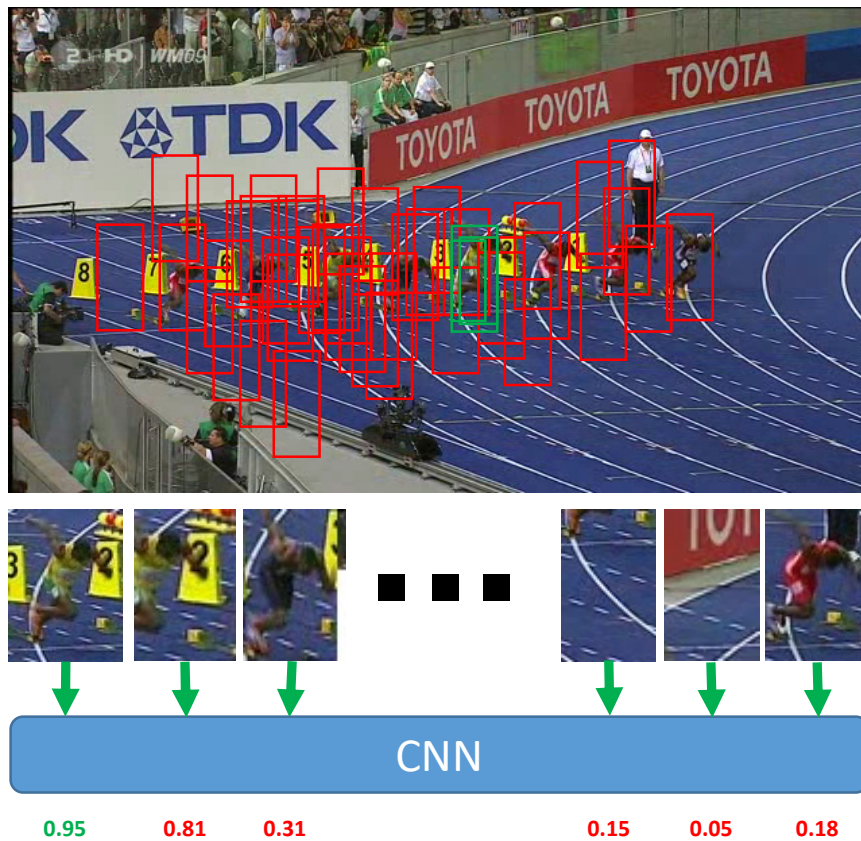
Institute of Automation, Chinese Academy of Sciences



- Improvements
 - Baseline experiments
 - **6.0%** absolute and **18.5%** relative improvement
 - Realtime experiments
 - **17.1%** absolute and **80.7%** relative improvement

How to speed up your tracker?

Only verify each candidate?

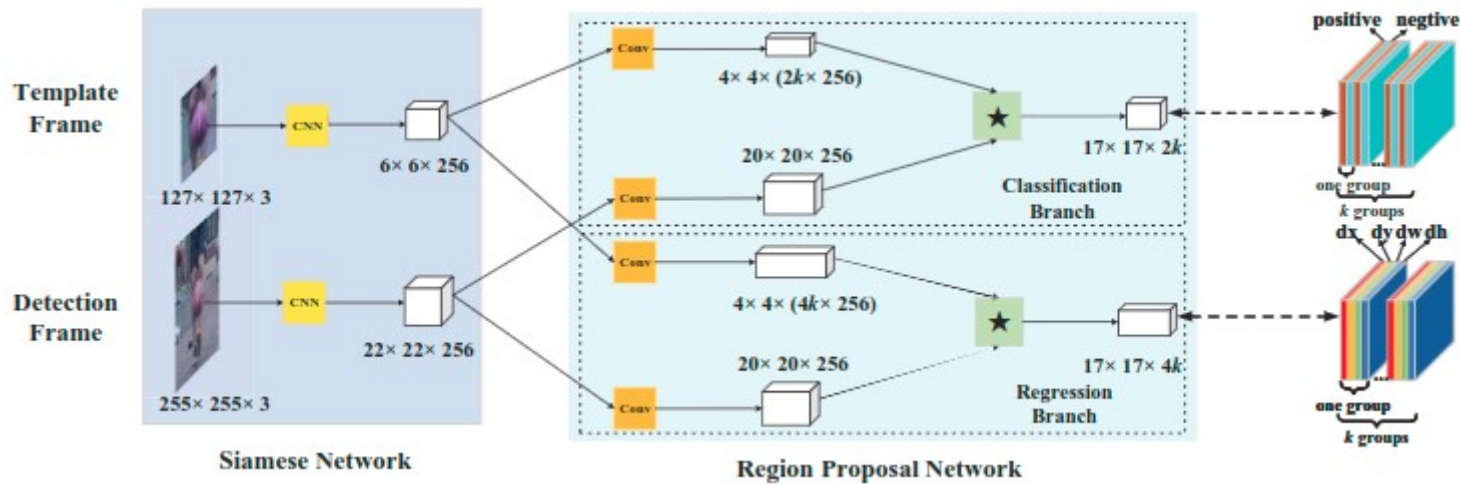


Jung, Ilchae and Son, Jeany and Baek, Mooyeol and Han, Bohyung. Real-Time MDNet. *ECCV*, 2018.

Bertinetto, Luca, et al. "Fully-convolutional siamese networks for object tracking." *ECCV workshop*, 2016.

Anfeng He, Chong Luo, Xinmei Tian, Wenjun Zeng. Towards a Better Match in Siamese Network Based Visual Object Tracker. *ECCV workshop*, 2018.

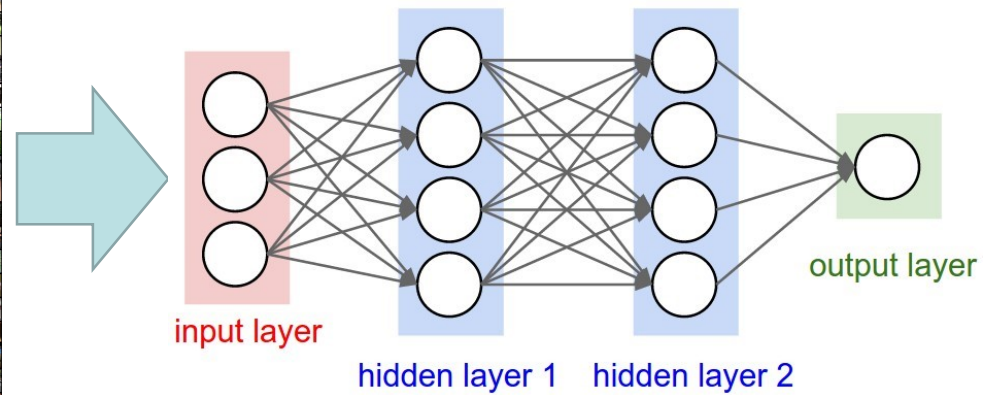
Baseline: SiamRPN



- Data-driven
- One-shot Detection
- Simultaneous Identification and Localization
- Anchor
- Fully Convolutional Networks
- Without Online Learning
- Without Multi-scale Detection

Classification Training

ImageNet



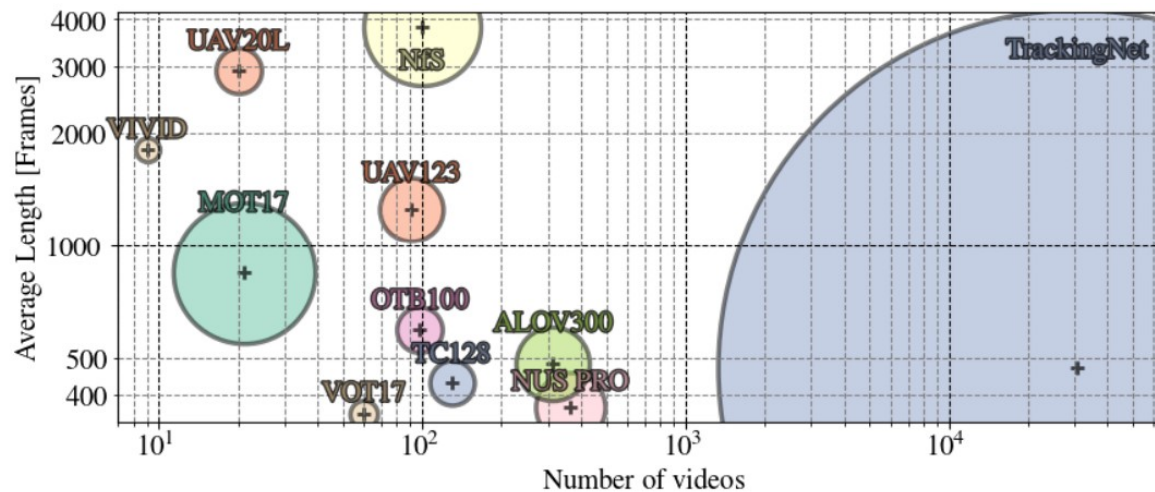
Classification v.s. Tracking



Classification	Tracking
• Predefined Categories(1000)	• General Object
• Recognition	• Metric Learning +Bounding box Regression
• Image Input	• Video Input

Training Datasets for Visual Tracking

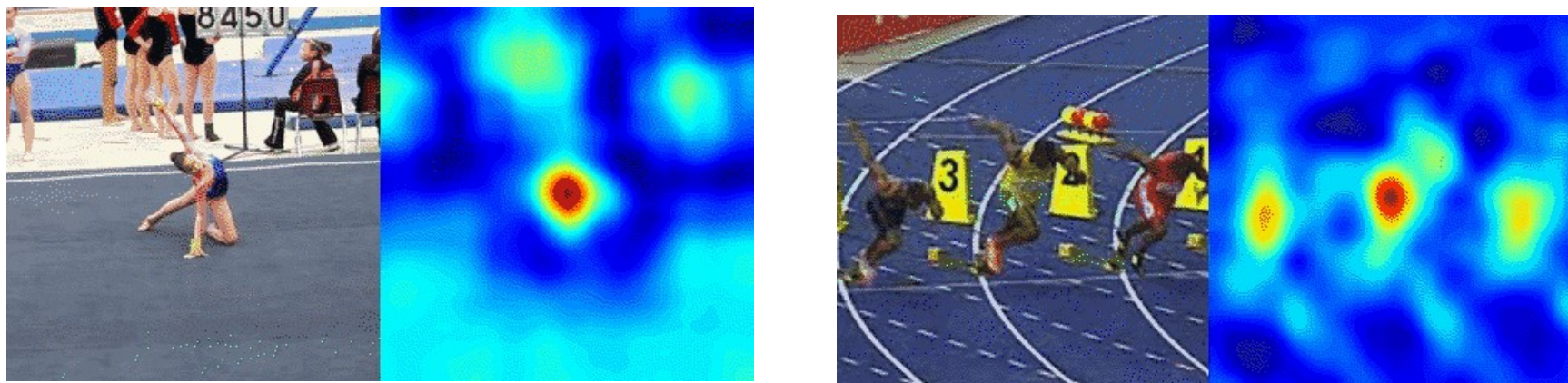
OxUvA long-term dataset & TrackingNet



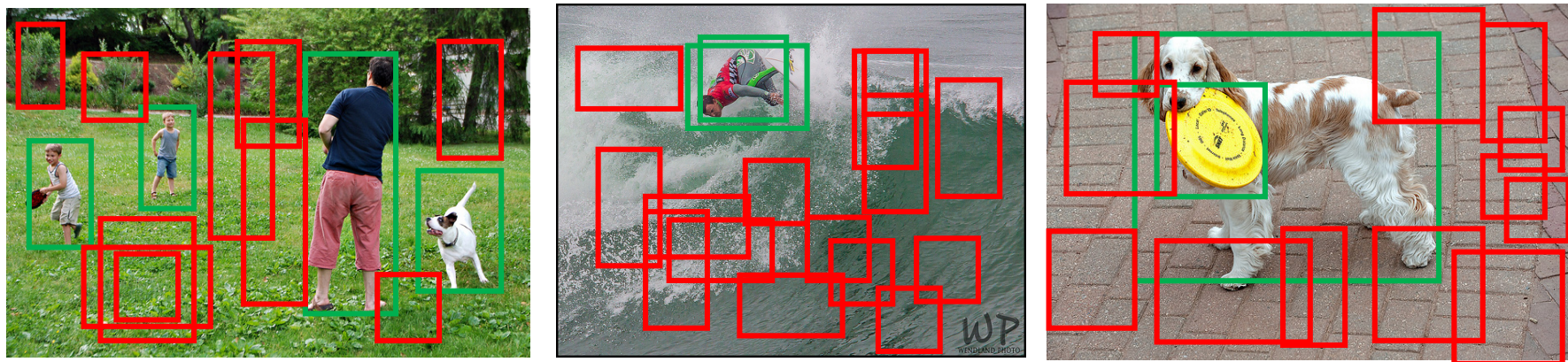
	OxUvA 2018	OTB-100 2015 [35]	VOT 2017 [14]	UAV123 2016 [22]	DTB 2017 [16]	NUS-PRO 2016 [15]	TC 2015 [17]	ALOV 2013 [29]	NfS 2017 [11]
Frames	1.55M	59k	21k	113k	15k	135k	60k	152k	380k
Tracks	366	100	60	123	70	365	128	314	100
... w/ absent labels	52%	0%	0%	0%	0%	0%	0%	0%	0%
Avg length (min)	2.36	0.33	0.20	0.51	0.12	0.21	0.26	0.27	0.26
Median length (min)	1.46	0.22	0.17	0.49	0.10	0.17	0.22	0.15	0.17
Max length (min)	20.80	2.15	0.83	1.71	0.35	2.8	2.15	3.32	1.44
Min length (min)	0.50	0.04	0.02	0.06	0.04	0.08	0.04	0.01	0.01
Avg absent labels	2.2	0	0	0	0	0	0	0	0
Object classes	22	16	24	9	15	8	27	—	—

Motivation

Response Maps predicted by SiamFC

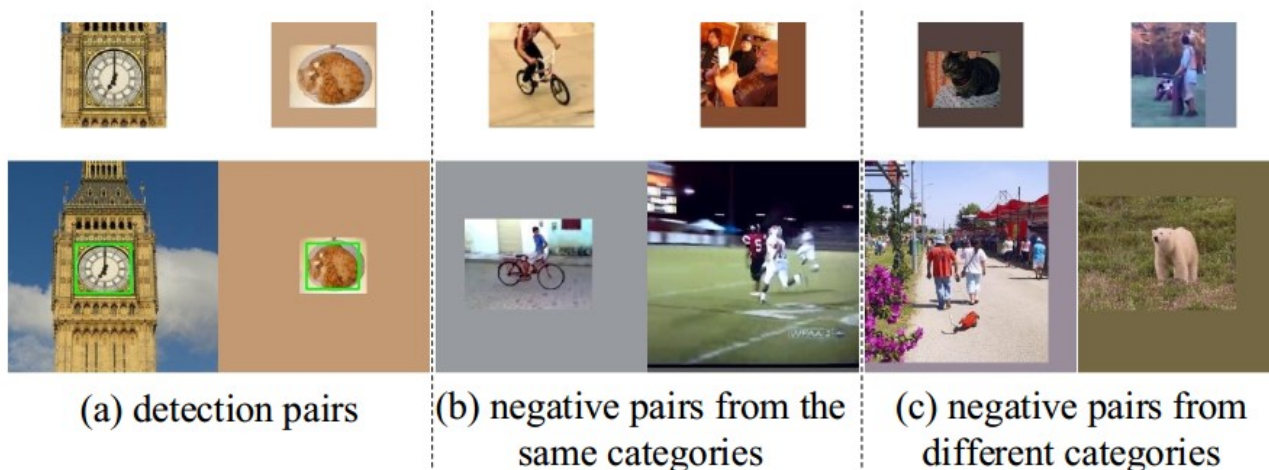


Non-semantic >> Semantic



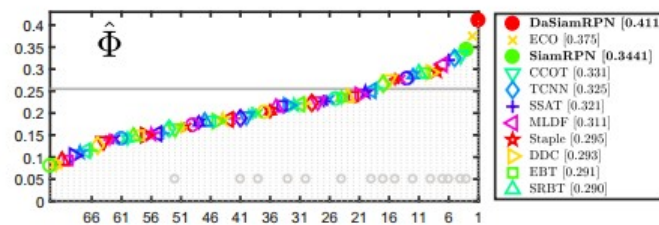
Distractor-aware Training

- Diverse categories of positive pairs can promote the generalization ability.
- Semantic negative pairs can improve the discriminative ability.
- Customizing effective data augmentation for visual tracking

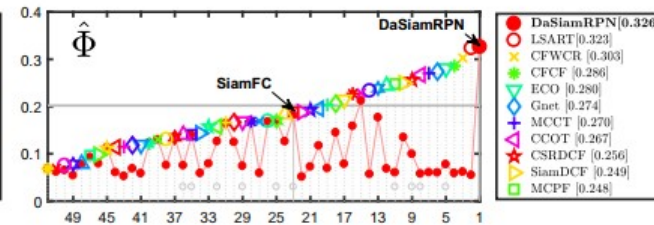


DaSiamRPN (ECCV2018)

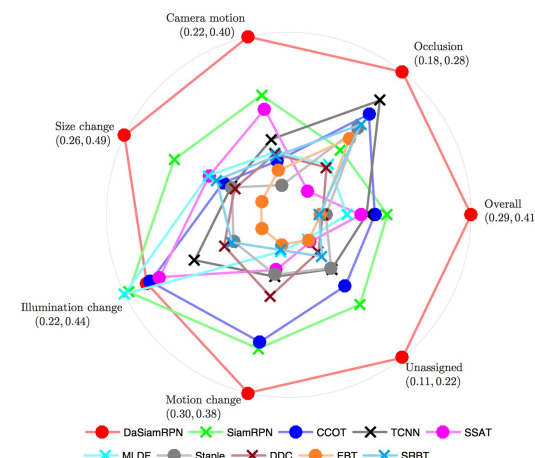
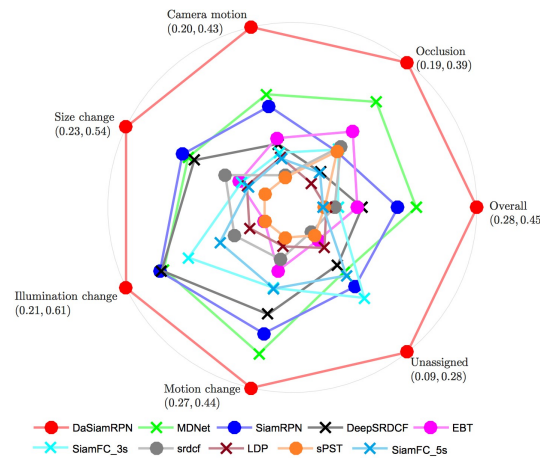
- State-of-the-art comparisons on VOT



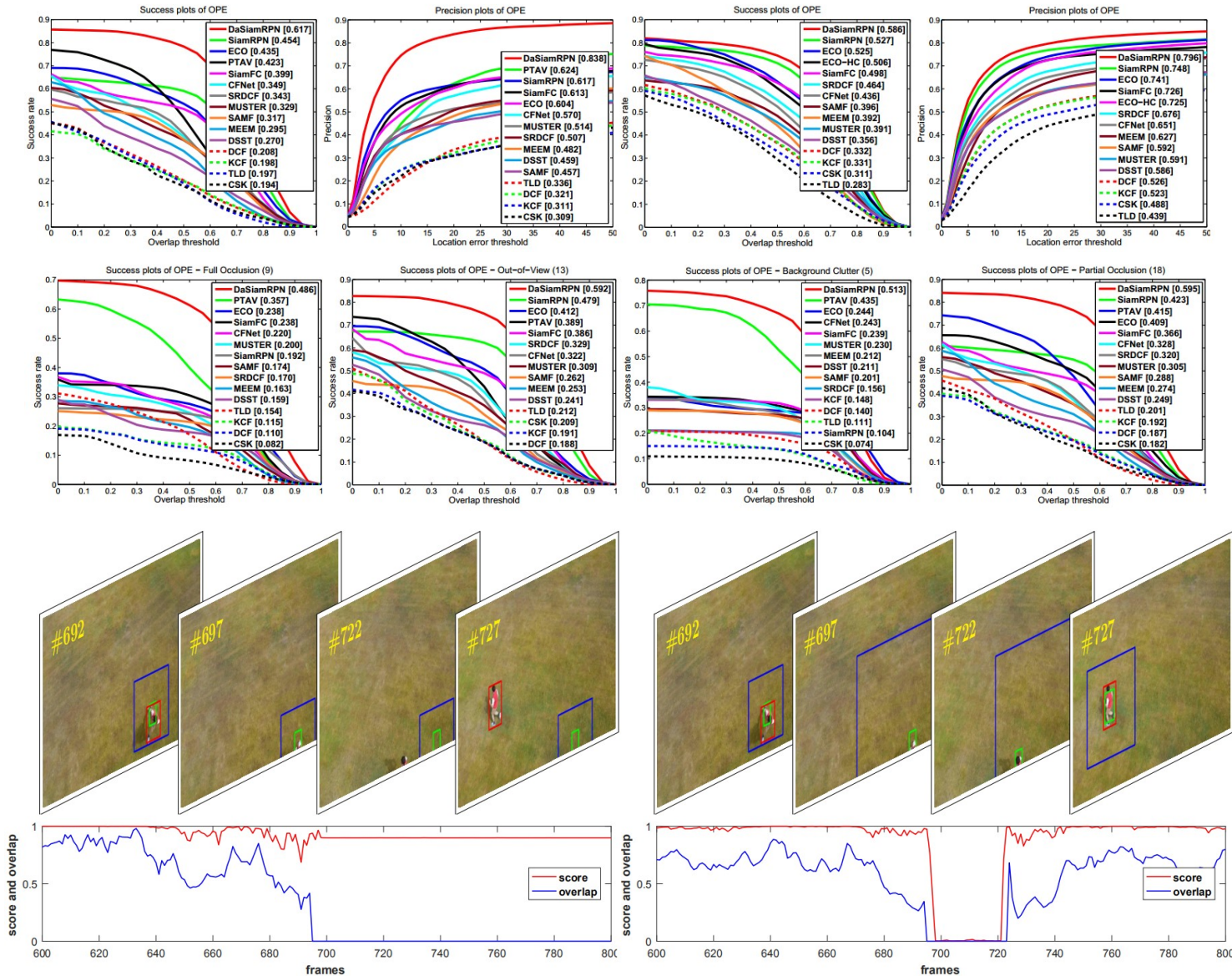
(a) EAO on VOT2016



(b) EAO on VOT2017

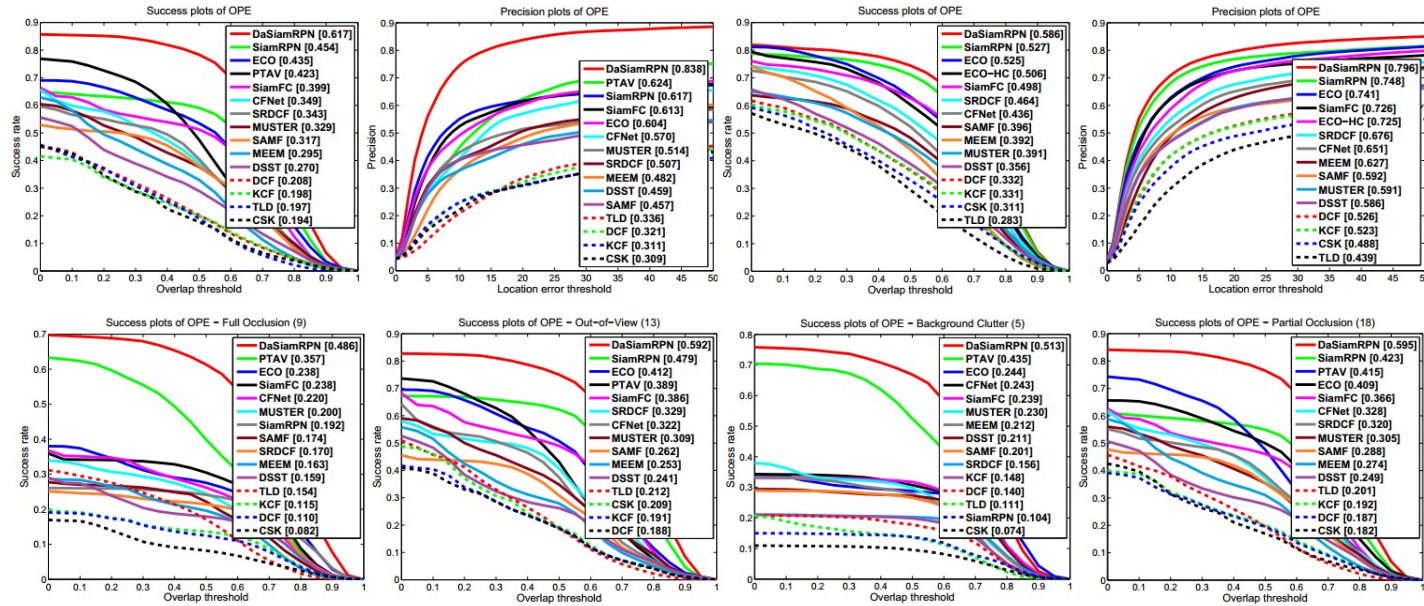


Second Prize in VOT2018-LT



(a) scores and overlaps in SiamRPN (b) scores and overlaps in DaSiamRPN

Zheng Zhu*, Qiang Wang*, Bo Li*, Wei Wu, Junjie Yan, and Weiming Hu. Distractor-aware Siamese Networks for Visual Object Tracking. *ECCV*, 2018



Trackers	OTB-2013		OTB-2015		VOT2015			VOT2016			VOT2017			FPS
	OP	DP	OP	DP	A	R	EAO	A	R	EAO	A	R	EAO	
SiamFC	77.8	80.9	73.0	77.0	0.533	0.88	0.289	0.53	0.46	0.235	0.50	0.59	0.188	86
CFNet	76.8	80.7	69.9	74.7	-	-	-	-	-	-	-	-	-	75
Staple	75.3	79.2	70.9	78.4	0.57	1.39	0.300	0.54	0.38	0.295	0.52	0.69	0.169	80
CSRDCF	75.3	80.0	70.7	78.7	0.56	0.86	0.320	0.51	0.24	0.338	0.49	0.36	0.256	13
BACF	82.3	84.4	76.7	81.5	0.59	1.56	-	-	-	-	-	-	-	35
ECO-HC	80.9	87.4	78.4	85.6	-	-	-	0.54	0.30	0.322	0.49	0.44	0.238	60
CREST	85.9	88.8	77.5	83.7	-	-	-	0.51	0.25	0.283	-	-	-	1
MDNet	91.1	94.8	85.4	90.9	0.60	0.69	0.378	0.54	0.34	0.257	-	-	-	1
C-COT	83.1	89.9	82.0	89.8	0.54	0.82	0.303	0.54	0.24	0.331	0.49	0.32	0.267	0.3
ECO	88.7	93.0	84.9	91.0	-	-	-	0.55	0.20	0.375	0.48	0.27	0.280	8
SiamRPN	85.7	88.3	81.9	85.0	0.58	1.13	0.349	0.56	0.26	0.344	0.49	0.46	0.244	200
DaSiamRPN	89.4	90.7	86.5	88.0	0.63	0.66	0.446	0.61	0.22	0.411	0.56	0.34	0.326	160

Results Ablation

SiamRPN		A	R	EAO	FPS
AlexNet 256C	CVPR'18	0.490	0.460	0.244	250

Baseline:

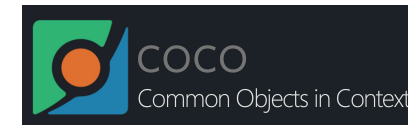
CVPR'18 Paper version of SiamRPN

Optimized Coding ⚡⚡⚡

Results Ablation (Training)

SiamRPN		A	R	EAO	FPS
AlexNet 256C	CVPR'18	0.490	0.460	0.244	250
	+ ImageNetDet&COCO	0.510	0.410	0.273	250
	+ train-pair aug.	0.560	0.340	0.326	250

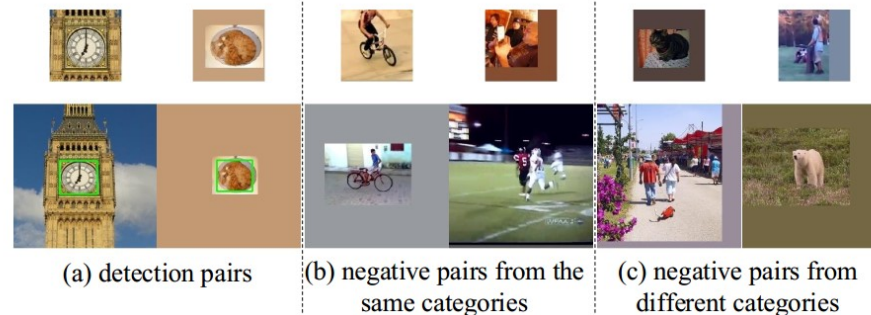
IMAGENET



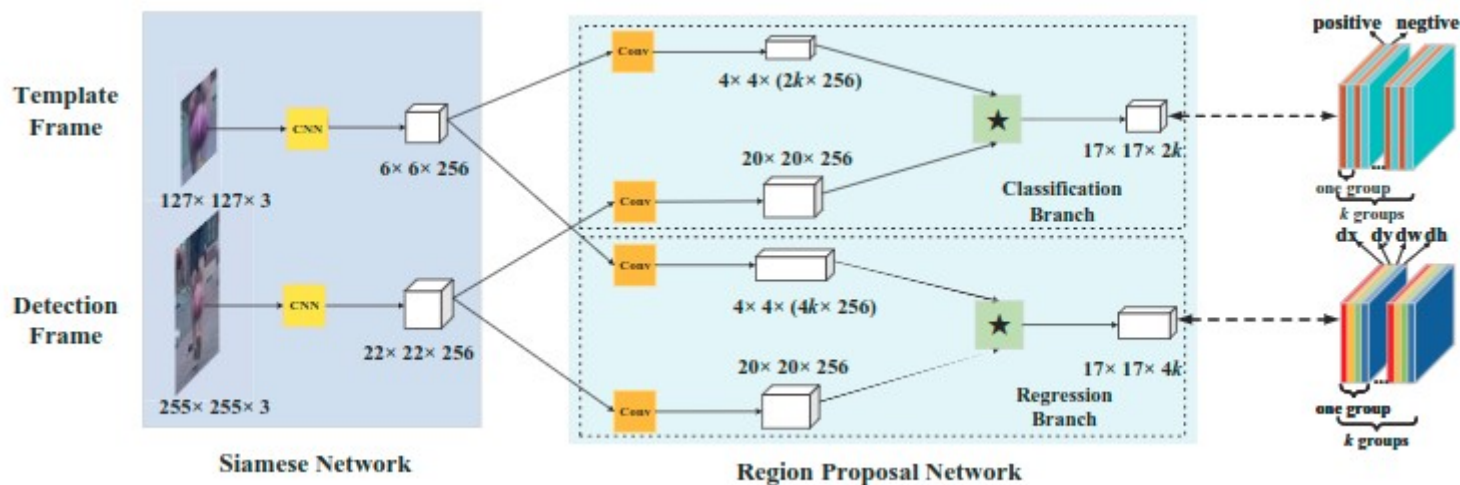
200 synsets for Object detection



DaSiamRPN
(ECCV2018)



Results Ablation (Backbone)



Kernel Size

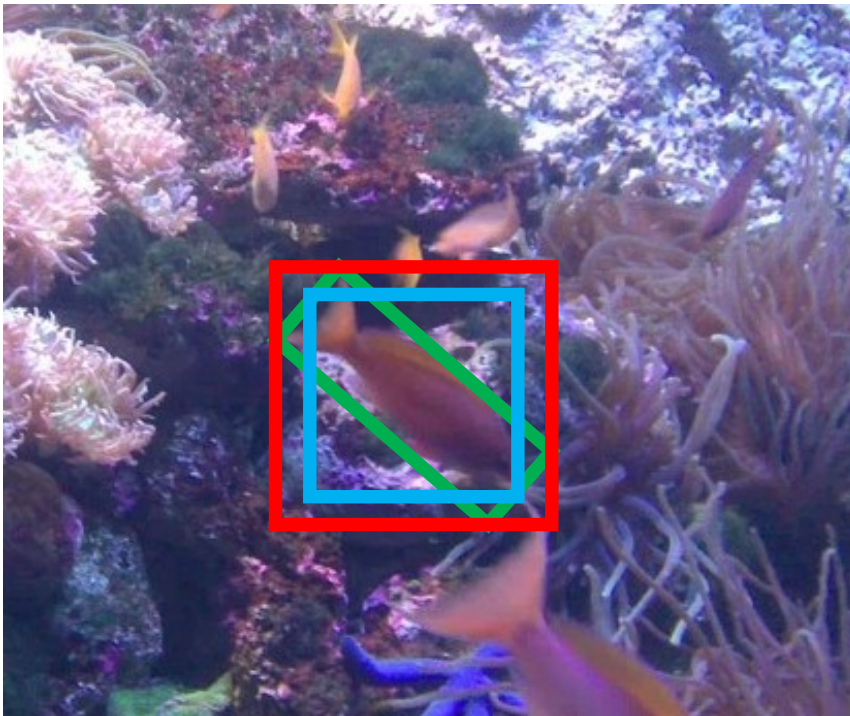
SiamRPN	Lost	EO
11+5+(3+3+3)	65	0.324
11+(3+3)+(3+3+3)	65	0.322
(9+3)+5+(3+3+3)	63	0.324
(7+5)+5+(3+3+3)	66	0.325
(7+3+3+3)+5+(3+3+3)	66	0.312
...	...	...
(3*5)+(3+3)+(3+3+3)	67	0.301

Channel Size

SiamRPN	Lost	EO
256F+256R	65	0.324
512F+256R	62	0.342
512F+1024R	56	0.341
512F+512R+SE	62	0.344
512F+1024R+SE	64	0.322
1024F+512R	74	0.318
1024F+1024R	73	0.312
512F+512R	60	0.345

Results Ablation (Trick1)

SiamRPN		A	R	EAO	FPS
AlexNet 256C	CVPR'18	0.490	0.460	0.244	250



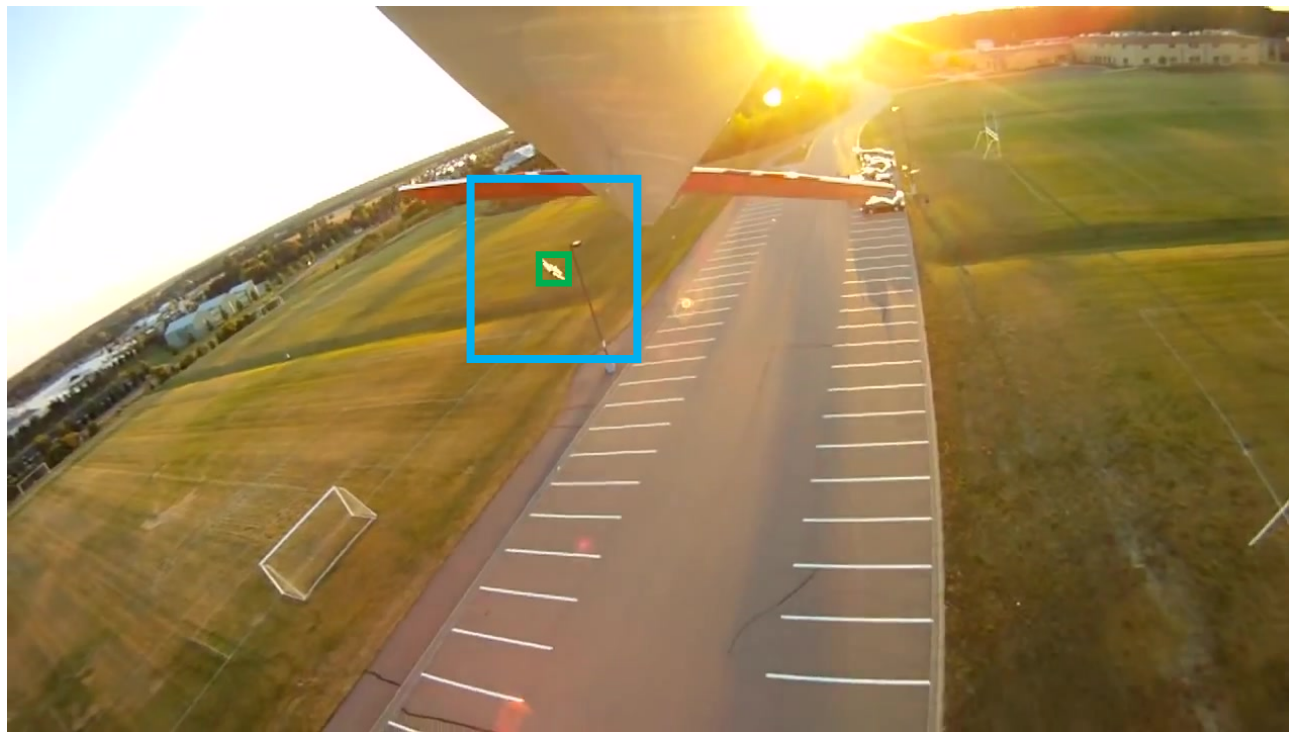
 Rotated Rect

 MinMax Rect

 Aligned Rect

Results Ablation (Trick2)

SiamRPN		A	R	EAO	FPS
AlexNet 256C	CVPR'18	0.490	0.460	0.244	250



Results Ablation (Trick2)

SiamRPN		A	R	EAO	FPS
AlexNet 256C	CVPR'18	0.490	0.460	0.244	250



Results Ablation

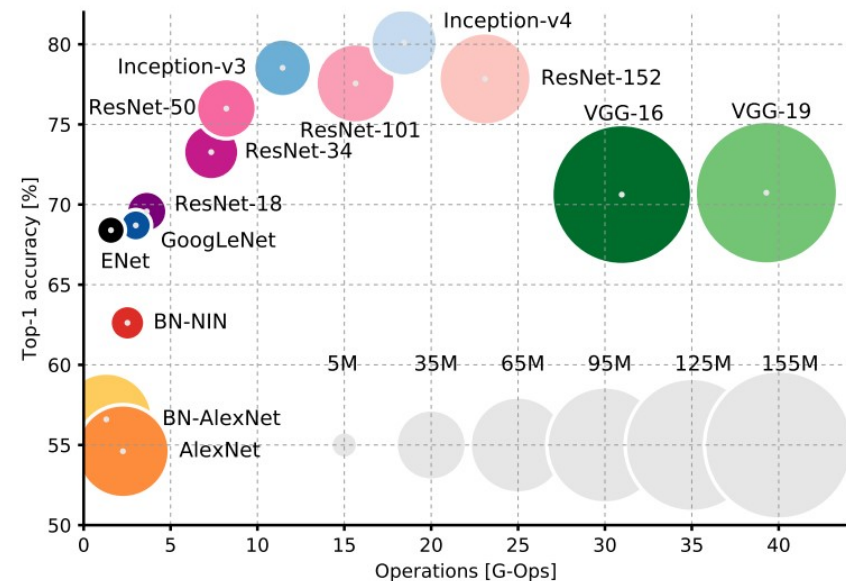
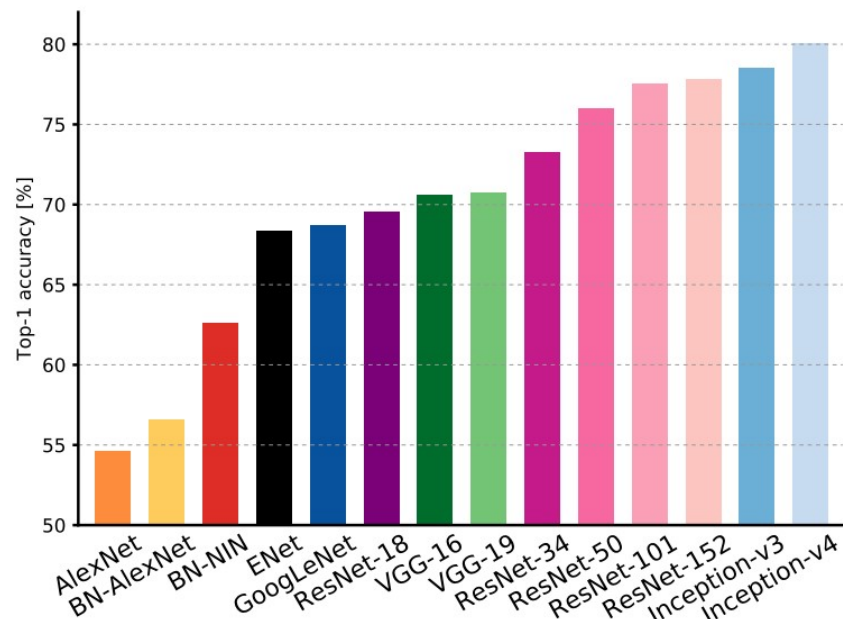
SiamRPN		A	R	EAO	FPS
AlexNet 256C	CVPR'18	0.490	0.460	0.244	250
	+ ImageNetDet&COCO	0.510	0.410	0.273	250
	+ train-pair aug.	0.560	0.340	0.326	250
AlexNet 512C	+ wider	0.570	0.330	0.345	100
	+ large search region	0.571	0.330	0.361	90
(final entry)	+ adaptive Search Region	0.588	0.320	0.383	90

- CVPR'18 Paper version of SiamRPN
- ECCV'18 Paper version of DaSiamRPN

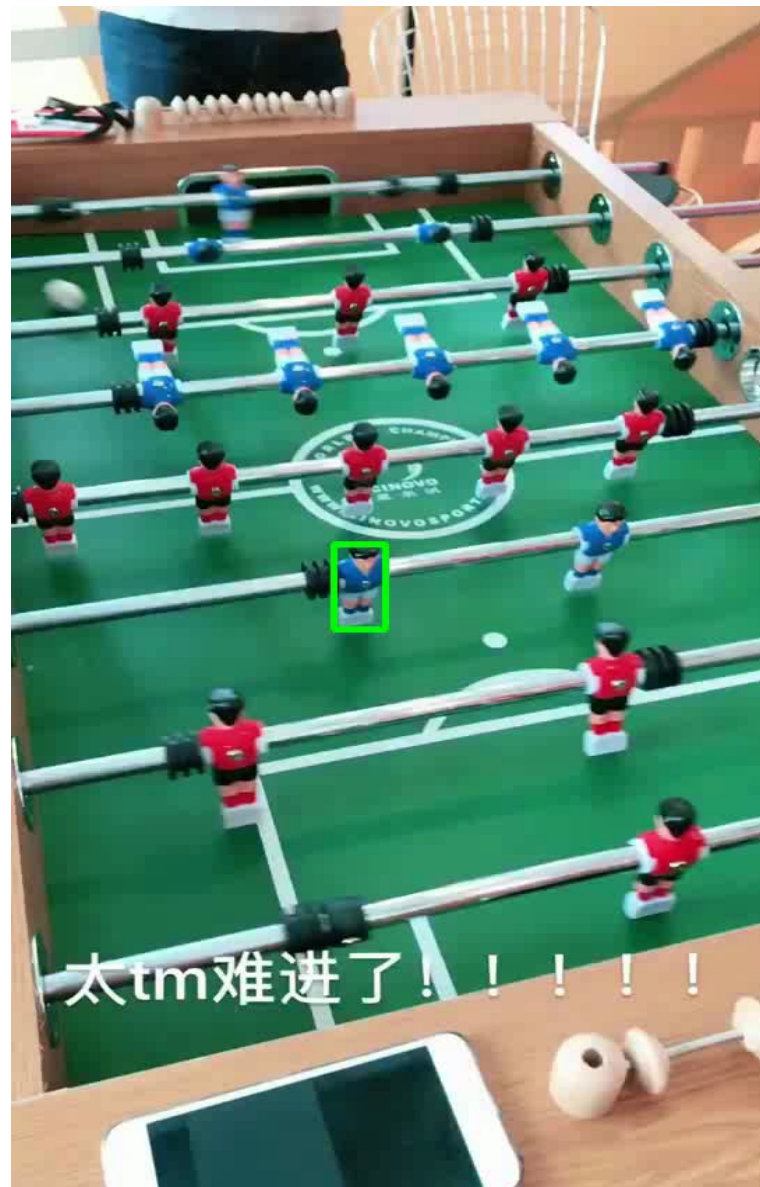
Performance Drop / No Effect

The backbone net is the engine of recognition!!!

- × Modern networks: VGG, Inception or ResNet (padding? or training methods?)
- × Attention Network
- × Ensemble (× Average × Max Win)
- × Focal Loss



Visualizations



Conclusion

SiamRPN

- ✓ Simultaneous Identification and Localization
- ✓ Training Strategies
- ✓ Widen Backbone
- ✓ Adaptive Search Area



	Baseline SiamRPN $\Rightarrow$	Training Strategy $\Rightarrow$	Network 512C $\Rightarrow$	Adaptive Searching
EAO	0.244	0.326	0.361	0.383
Accuracy	0.490	0.560	0.571	0.588
Robustness	0.460	0.340	0.330	0.320
FPS	250	250	100	90

Thanks! Questions?

foolwood / DaSiamRPN

Unwatch 13 Unstar 239 Fork 56

Code Issues 6 Pull requests 1 Projects 0 Wiki Insights Settings

Distractor-aware Siamese Networks for Visual Object Tracking (ECCV2018)

Manage topics

2 commits 1 branch 0 releases 1 contributor MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

foolwood DaSiamRPN VOT2018		Latest commit a2cf47a 23 days ago
code	DaSiamRPN VOT2018	23 days ago
LICENSE	DaSiamRPN VOT2018	23 days ago
README.md	DaSiamRPN VOT2018	23 days ago
run_install.sh	DaSiamRPN VOT2018	23 days ago
tracker_SiamRPN.m	DaSiamRPN VOT2018	23 days ago
votresult.png	DaSiamRPN VOT2018	23 days ago



Zheng Zhu*, Qiang Wang*, Bo Li*, Wei Wu, Junjie Yan, and Weiming Hu. Distractor-aware Siamese Networks for Visual Object Tracking. *ECCV*, 2018
<https://github.com/foolwood/DaSiamRPN>