New work and A tutorial on SOT

Zhipeng Zhang

NLPR, CASIA

Contents

- Which paper should a fresher read?
- Which Github-repo should you fork?
- Advances in Siamese Tracking
- Our new paper: Ocean/Ocean+
- Challenges and future study
- Q&A

Papers to read

[survey] Marvasti-Zadeh S M, et al. Deep learning for visual tracking: A comprehensive survey.

[Siam 开山] Luca Bertinetto, et.al Fully-Convolutional Siamese Networks for Object Tracking.

[Siam 突破] Li, Bo, et al. High performance visual tracking with siamese region proposal network.

[Siam 突破] Zhang Z, et al. Deeper and wider siamese networks for real-time visual tracking.

[Siam 突破] Li B, Siamrpn++: Evolution of siamese visual tracking with very deep networks.

[Siam 突破] Wang Q et al, Fast online object tracking and segmentation: A unifying approach.

[Siam 突破] Zhang Z, Ocean: Object-aware anchor-free tracking.

[Siam 开山] Luca Bertinetto, et.al Fully-Convolutional Siamese Networks for Object Tracking.

[CF 开山] David S. Bolme et al, Visual Object Tracking using Adaptive Correlation Filters.

[CF 突破] J. F. Henriques, et al, High-speed tracking with kernelized correlation filters.

[CF 突破] Martin Danelljan, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking.

[CF 突破] Martin Danelljan, et al. ECO: Efficient Convolution Operators for Tracking.

[CF 突破] Martin Danelljan, et al. ATOM: Accurate Tracking by Overlap Maximization.

Github to Fork

[Results Comparison] https://github.com/JudasDie/Comparison

[Papers Collection] https://github.com/foolwood/benchmark_results

[TracKit] https://github.com/researchmm/TracKit [SiamDW/Ocean/Ocean+]

[SiamFC++] https://github.com/MegviiDetection/video_analyst

[SiamRPN++] https://github.com/STVIR/pysot

[SiamMask] https://github.com/foolwood/SiamMask

[Pytracking] https://github.com/visionml/pytracking [ATOM/DIMP/PrDiMP]

[VOT] https://github.com/votchallenge

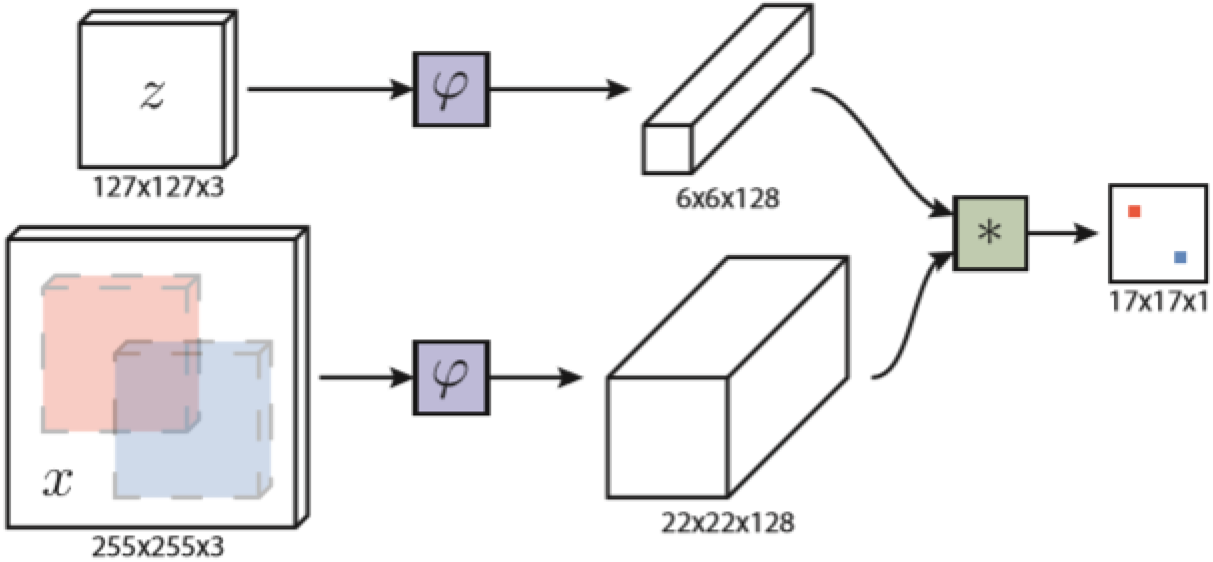
[GOT10K] https://github.com/got-10k/toolkit

[FairMOT] https://github.com/ifzhang/FairMOT

[TnesorRT] https://github.com/NVIDIA/TensorRT

Advances in Siamese Tracking

- SiamFC: Time is life



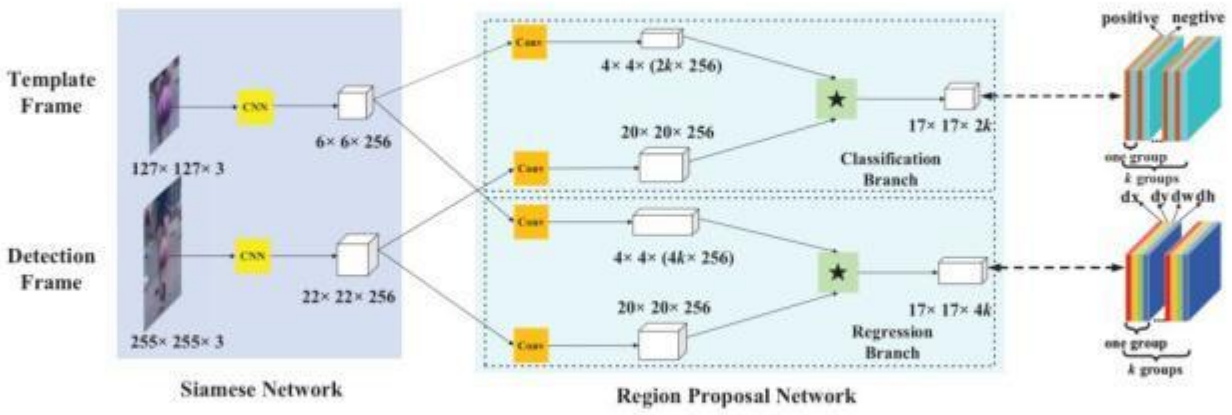
❑ Siamese network

❑ All in matching

❑ Fast! Fast! Fast!

Advances in Siamese Tracking

- SiamRPN: Detection to Tracking



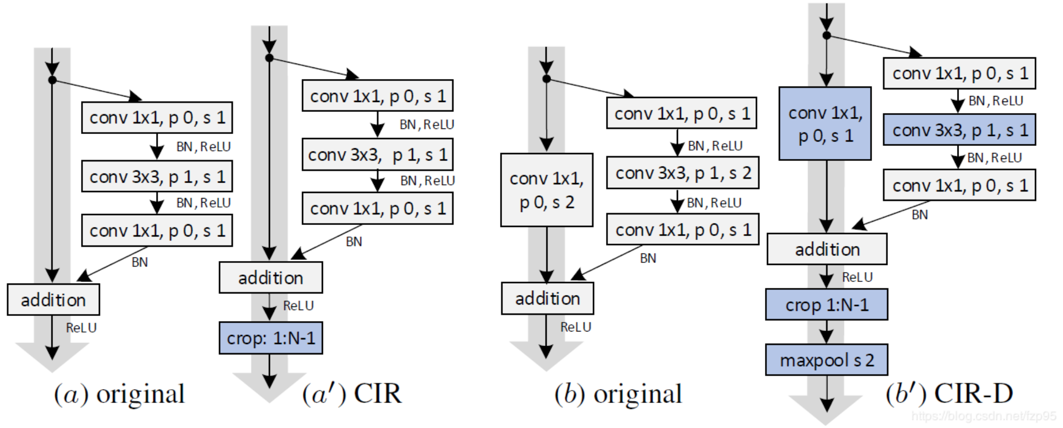
□ Region Proposal

□ Detection matters

□ Acc.! Acc.! Acc.!

Advances in Siamese Tracking

- SiamDW/SiamRPN++: Going deeper and wider



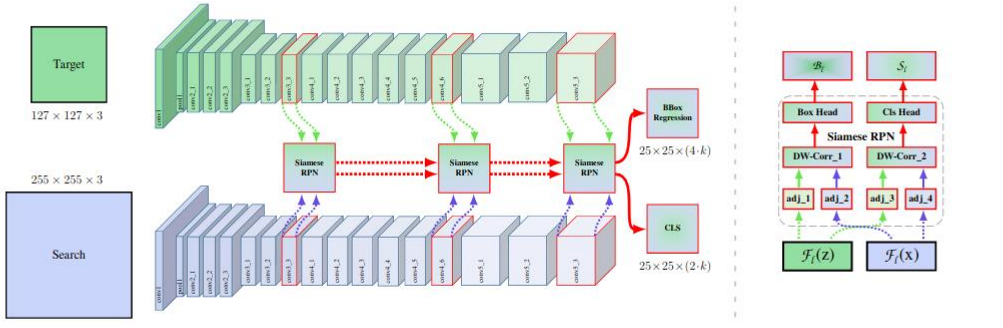
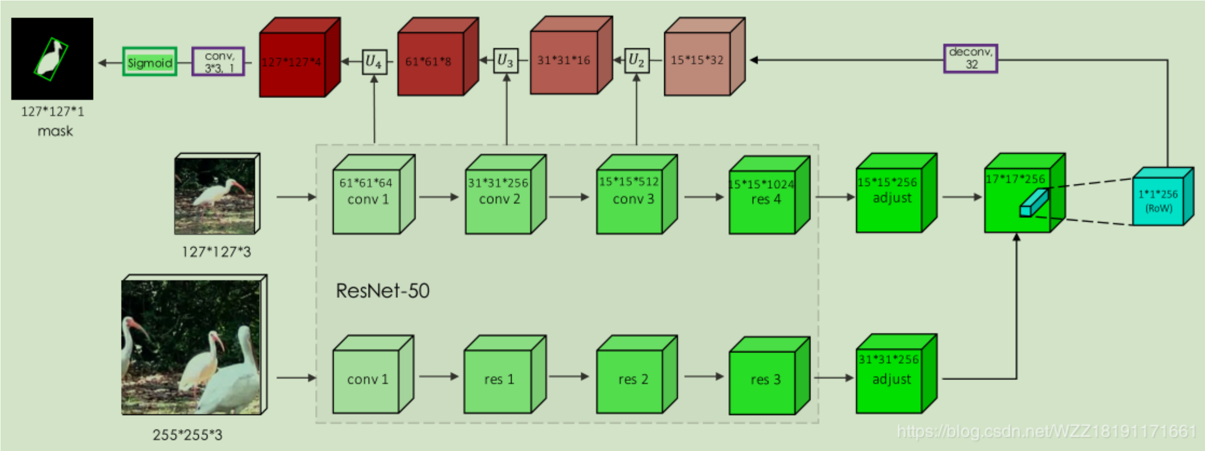


Figure 3. Illustration of our proposed framework. Given a target template and search region, the network ouputs a dense prediction by fusion the outputs from multiple Siamese Region Proposal (SiamRPN) blocks. Each SiamRPN block is shown on right.

- □ Perceptual inconsistency
- □ Position Bias
- □ Deeper! Deeper! Deeper!

Advances in Siamese Tracking

- SiamMask: Segmentation to Tracking



❑ Segmentation Matters

❑ Rotated Box

❑ Pixel! Pixel! Pixel!

Advances in Siamese Tracking

- SiamRCNN: rethinking where should we go

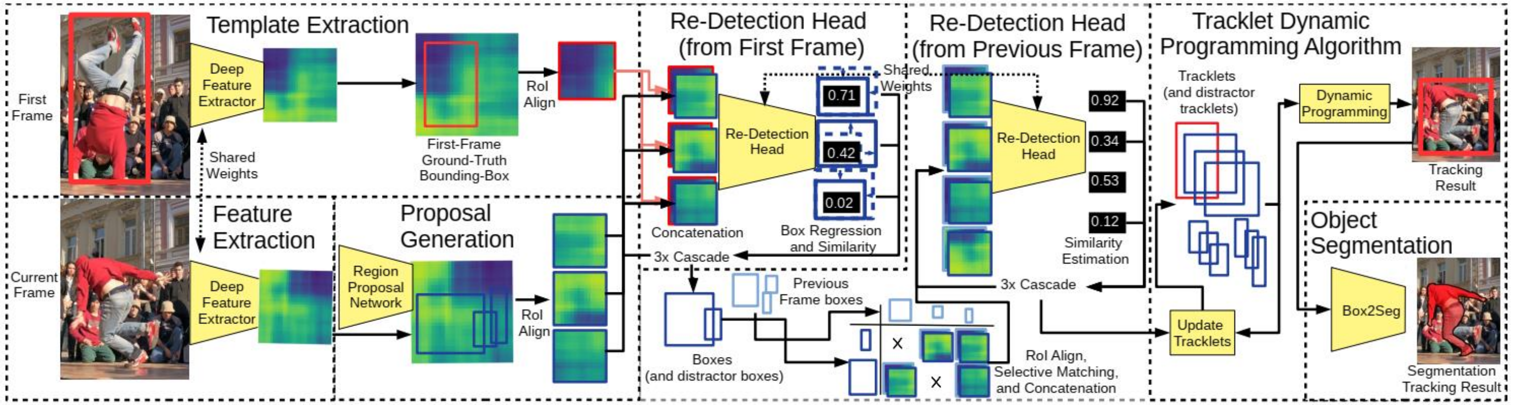


Figure 2: Overview of Siam R-CNN. A Siamese R-CNN provides re-detections of the object given in the first-frame bounding box, which are used by our Tracklet Dynamic Programming Algorithm along with re-detections from the previous frame. The results are bounding box level tracks which can be converted to segmentation masks by the Box2Seg network.

□ ReID Matters

□ Sequential Reasoning

□ Lost! Lost! Lost!

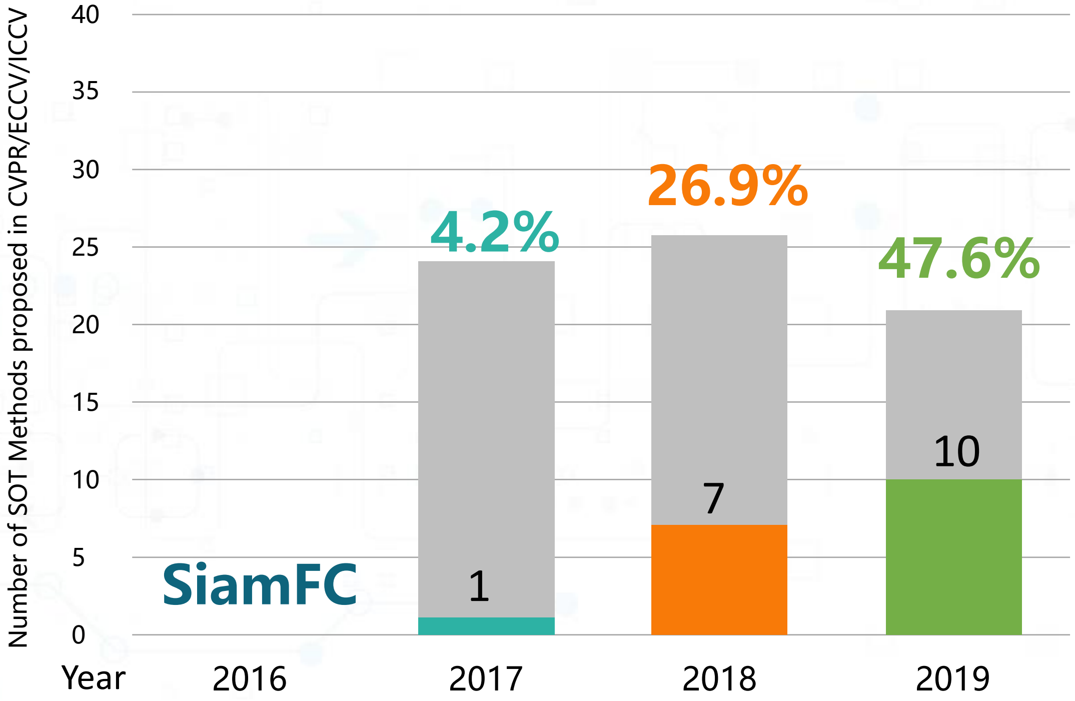
ECCV'20 ONLINE 23-28 AUGUST 2020

Ocean: Object-aware Anchor-Free Tracking

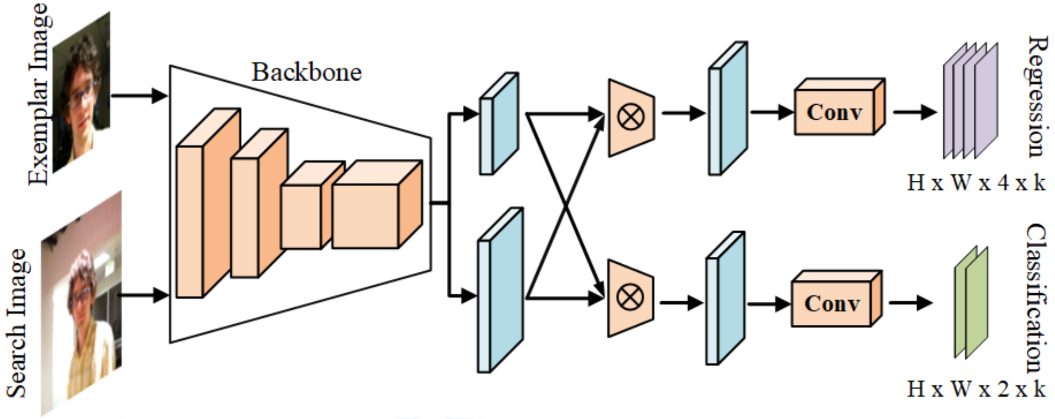
Microsoft

Background

- **Booming of Siamese Tracking**



- **RPN based approaches**

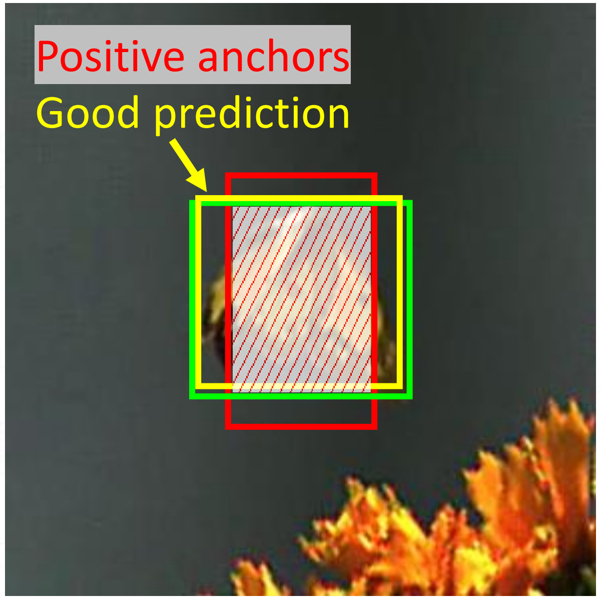


[CVPR'18] **SiamRPN**
[ECCV'18] **DaSiam**
[CVPR'19] **SiamDW, SiamRPN++, SiamMask, C-RPN**
[CVPR'20] **SiamAtt**

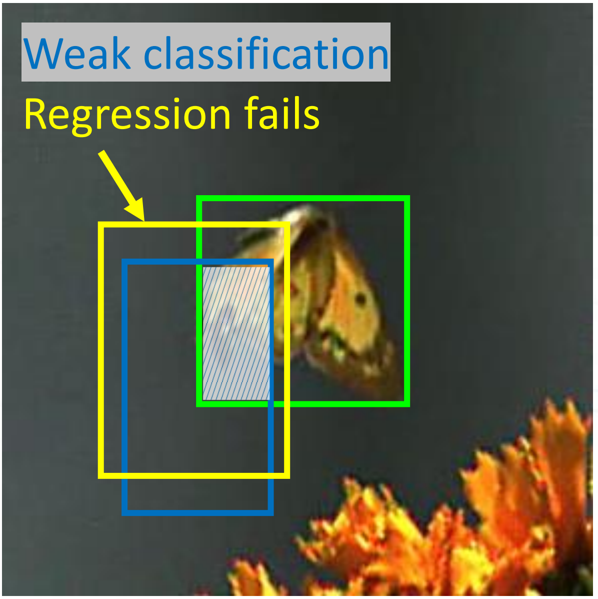
[SiamRPN] B Li, et.al. High Performance Visual Tracking with Siamese Region Proposal Network [DaSiam] Z Zhu, et.al . Distractor-aware Siamese Networks for Visual Object Tracking
[SiamRPN++] B Li, et.al. SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks [SiamDW] Z Zhang, et.al. Deeper and Wider Siamese Networks for Real-Time Visual Tracking
[SiamMask] Q Wang, et.al Fast Online Object Tracking and Segmentation: A Unifying Approach [C-RPN] Siamese Cascaded Region Proposal Networks for Real-Time Visual Tracking
[SiamFC] Fully-Convolutional Siamese Networks for Object Tracking [SiamAtt] Deformable Siamese Attention Networks for Visual Object Tracking

Motivation

- **How does anchor work?**



- **Why anchor regression fails?**



Positive anchors : large overlap between Anchor (red) and GT (green)

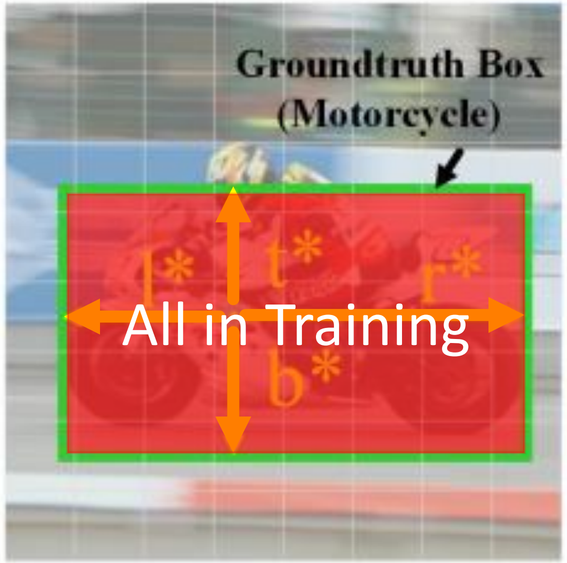
Weak classification: small overlap between Anchor (red) and GT (green)

Positive anchors → Good regression (yellow)

Weak classification → Regression Fails (yellow)

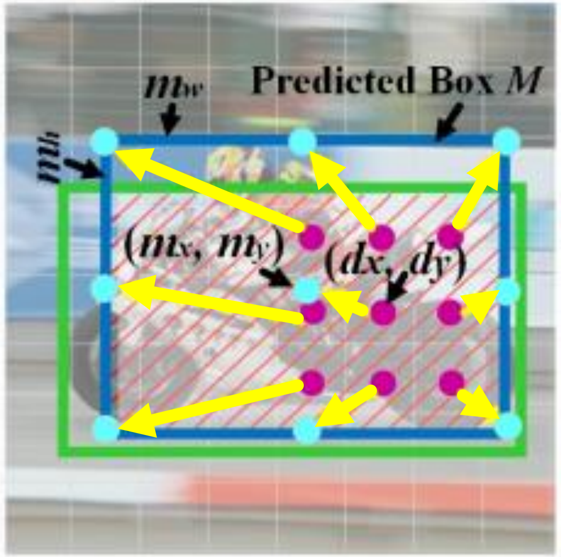
Method

- **Anchor-free regression**
- **Object-aware classification**



Help to learn better feature

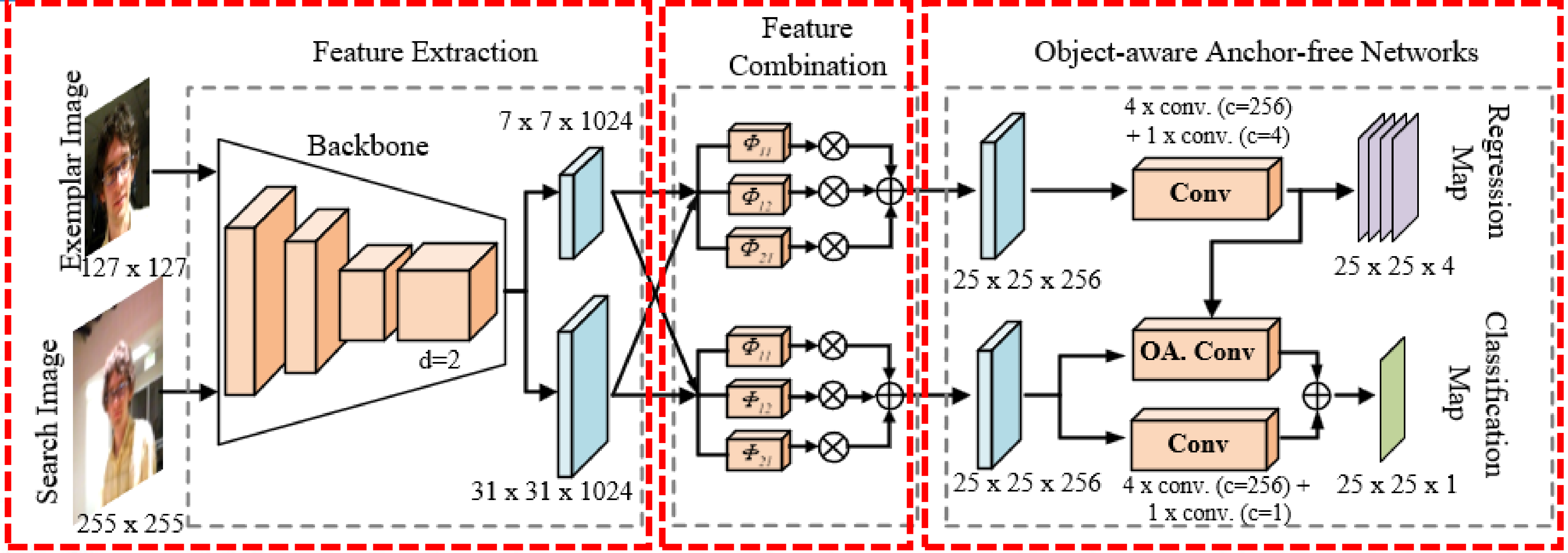
Select better bounding box



Consider more samples in the training of regression network.

Learn object-aware feature with the predicted bounding boxes.

Framework

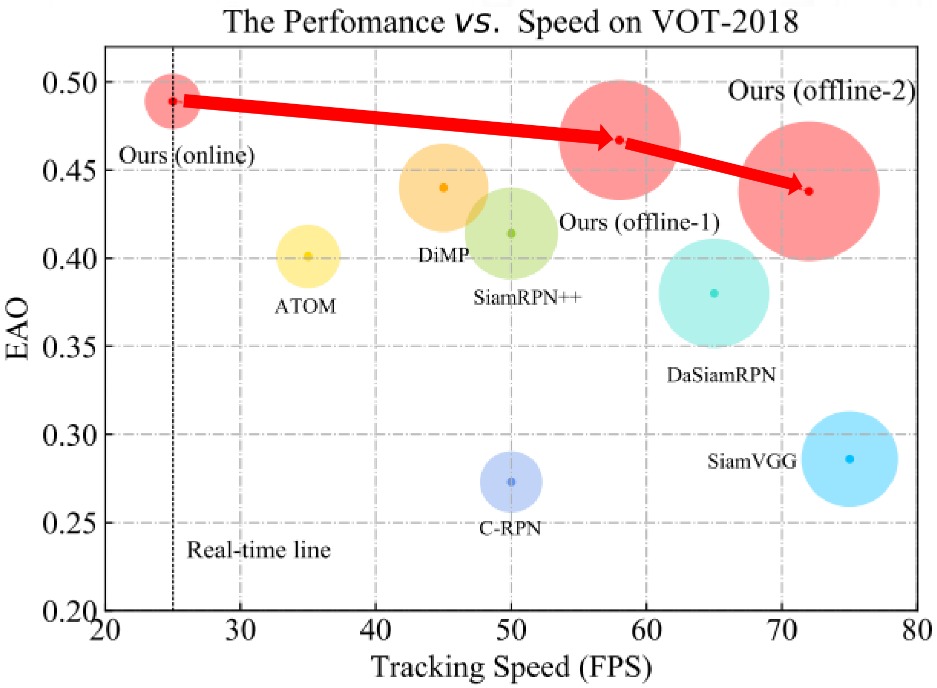


- Feature extraction: ResNet50 *conv1-conv4*
- Feature combination: Parallel layers with different dilated strides at *x* and *y* axis
- Target localization: anchor-free regression + object-aware classification

ECCV'20 ONLINE 23-28 AUGUST 2020

Results & Ablations

- **Results**



- **Ablations**

#Num	Components	EAO
①	baseline	0.358
②	+ centralized sampling	0.396
③	+ feature combination	0.438
④	+ object-aware classification	0.467
⑤	+ online update	0.489

4.2 points

2.9 points

#Num	Dilated Kernels	EAO
①	Φ_{11}	0.425
②	$\Phi_{11}\Phi_{11}$	0.433
③	$\Phi_{11}\Phi_{12}$	0.446
④	$\Phi_{11}\Phi_{21}$	0.443
⑤	$\Phi_{11}\Phi_{12}\Phi_{21}$	0.467

2.1 points

4.2 points

Microsoft

Towards Accurate Pixel-wise Object Tracking by Attention Retrieval

Motivation

- **Trend of SOT Community**

  Bounding Box --> Mask (VOT2020)

- **Weakness of existing Tracking-Segmentation Methods**

  Cascaded Structure: Box → Mask (Accurate)

  Too Slow!

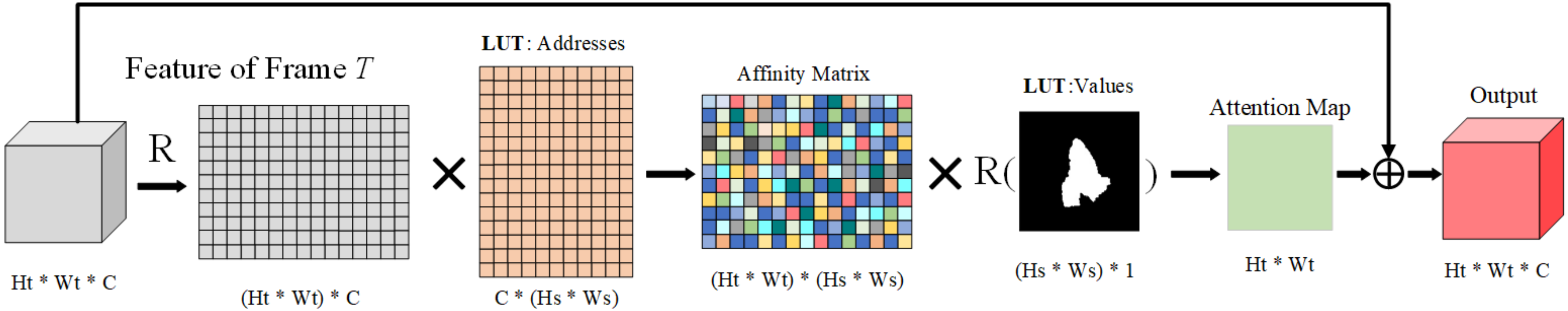
  Box error → Mask error

  Parallel Structure [SiamMask/D3S]: segmentation branch (Fast)

  Background clutter → False positive predictions

Method

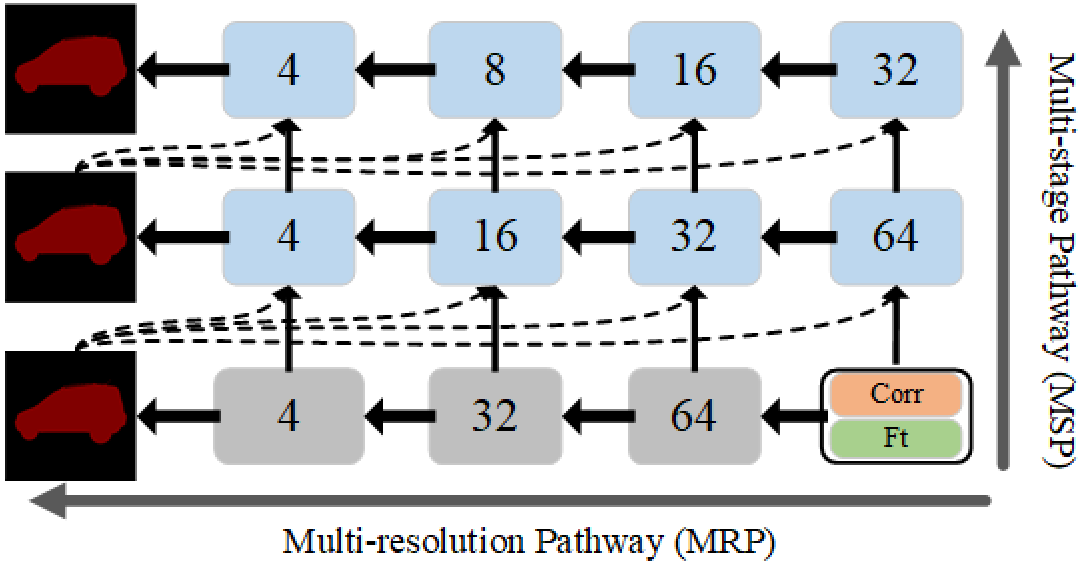
- **Attention Retrieval Network (ARN)**



- ➢ **Soft spatial constraints** → suppress negative influence of background clutter
- ➢ **Only matrix multiplication** → Very fast
- ➢ **Use initial mask** → infuse the information of the mask in the starting frame

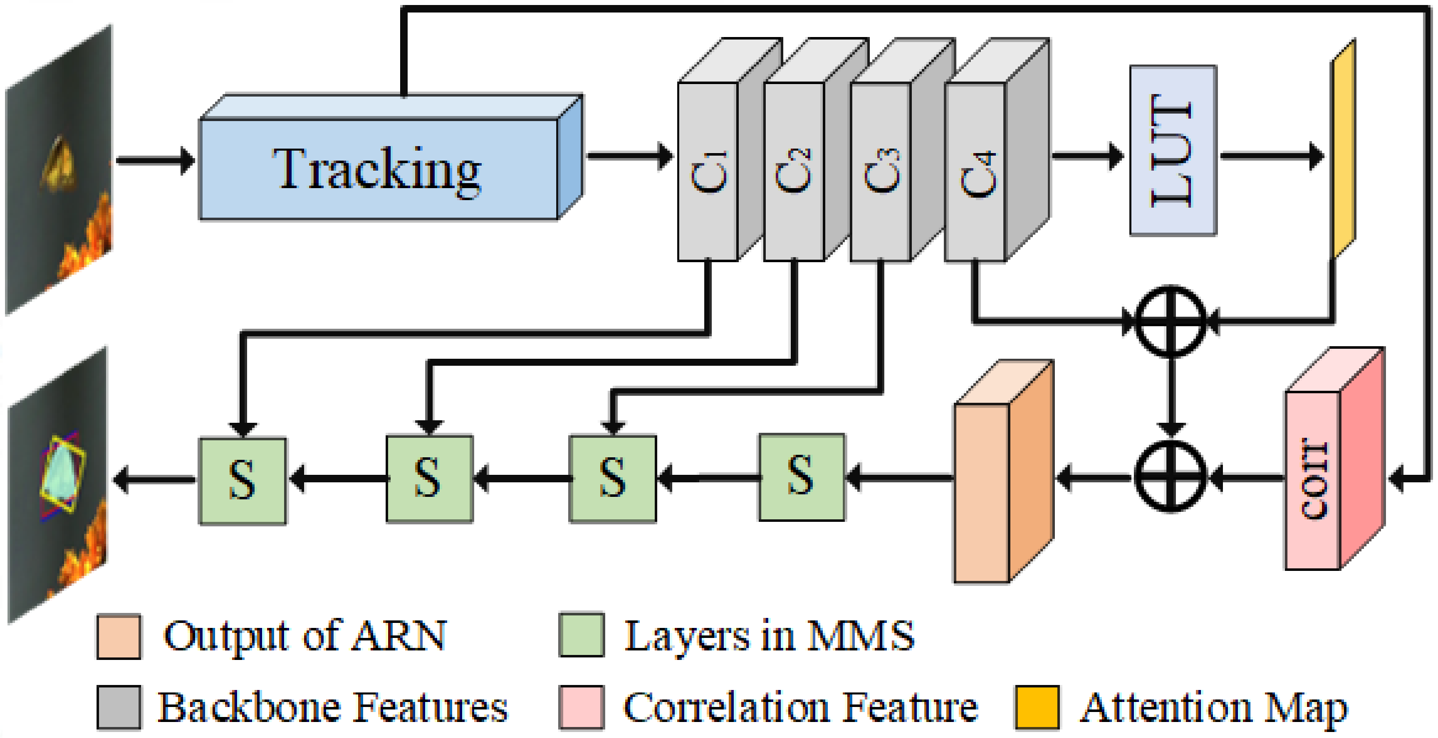
Method

- **Multi-resolution Multi-stage Segmentation (MMS)**



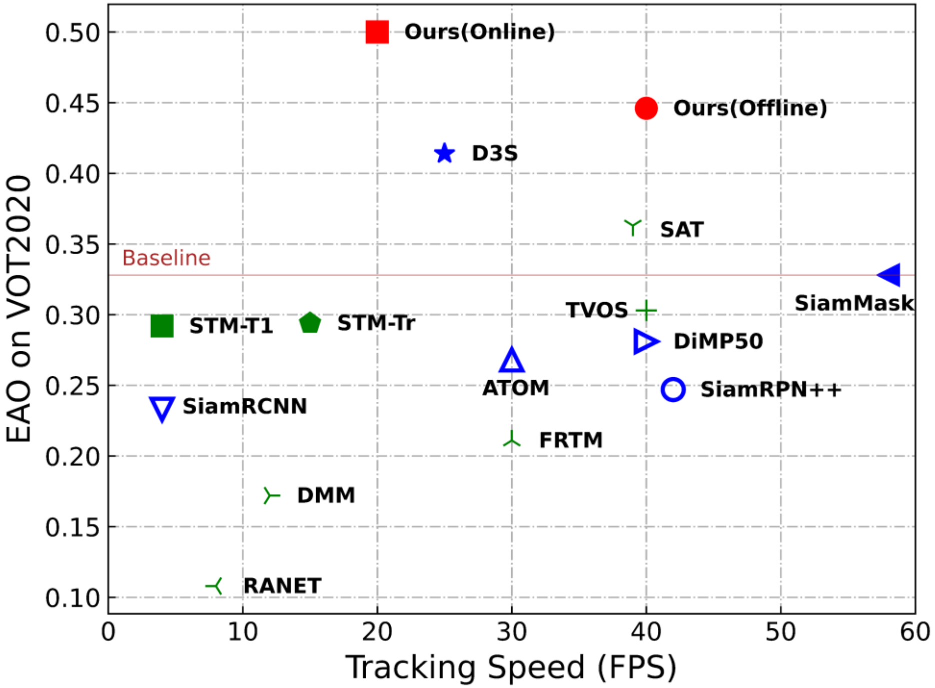
- ➢ **Reusing predicted mask →** further suppress background clutter
- ➢ **Small Channels in MSP →** fast

Framework

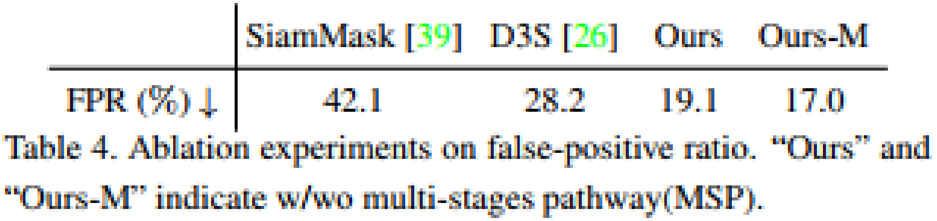

Tracking
C1
C2
C3
C4
LUT
corr
S
S
S
S
Output of ARN
Layers in MMS
Backbone Features
Correlation Feature
Attention Map

Results & Ablations

- **Results**

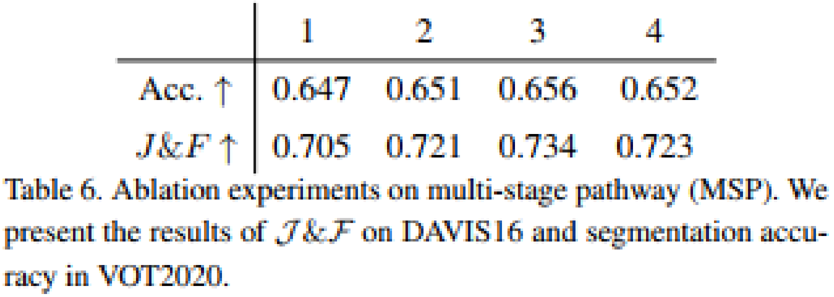


- **Ablations**



	SiamMask [39]	D3S [26]	Ours	Ours-M
FPR (%) ↓	42.1	28.2	19.1	17.0

Table 4. Ablation experiments on false-positive ratio. “Ours” and “Ours-M” indicate w/wo multi-stages pathway(MSP).

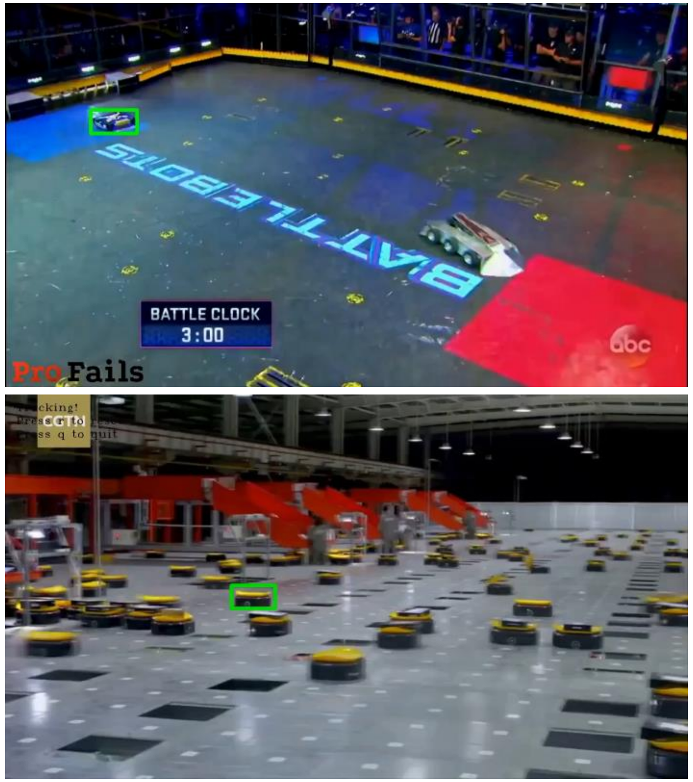


	1	2	3	4
Acc. ↑	0.647	0.651	0.656	0.652
$\mathcal{J}\&\mathcal{F}$ ↑	0.705	0.721	0.734	0.723

Table 6. Ablation experiments on multi-stage pathway (MSP). We present the results of $\mathcal{J}\&\mathcal{F}$ on DAVIS16 and segmentation accuracy in VOT2020.

Demo

- **Ocean**



- **Ocean+**

Challenges & Future Study

Challenges

- Siamese can’t go deeper
- Trackers are too Slow
- No essential novelty/improvement
- A new framework is required

Future Study

- Tacking and Segmentation
- Merging MOT and SOT
- Involve other learning method (e.g. Self-training)

https://github.com/researchmm/TracKit

https://github.com/JudasDie/Comparision