DALIAN UNIVERSITY OF TECHNOLOGY
1949
大连理工大学

Alpha-Refine: Boosting Tracking Performance by Precise Bounding Box Estimation

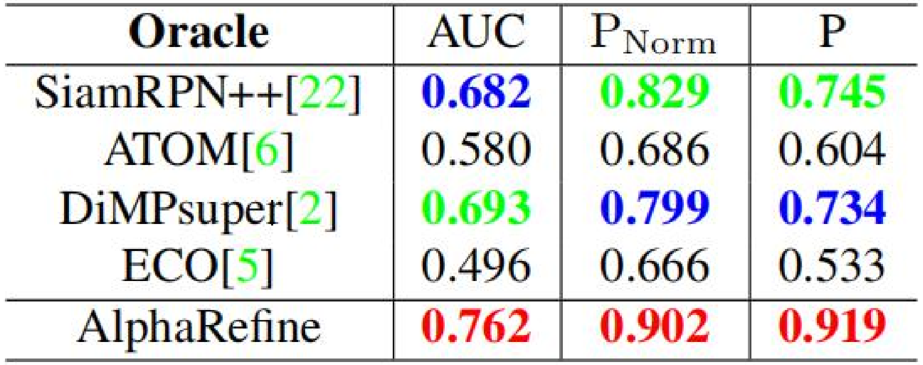
Oracle Experiment

Why wee need a refinement module?

Oracal Setting:

Reset the target center to ground truth center location at each frame.

The performances will be mainly determined by the box estimation capacity.

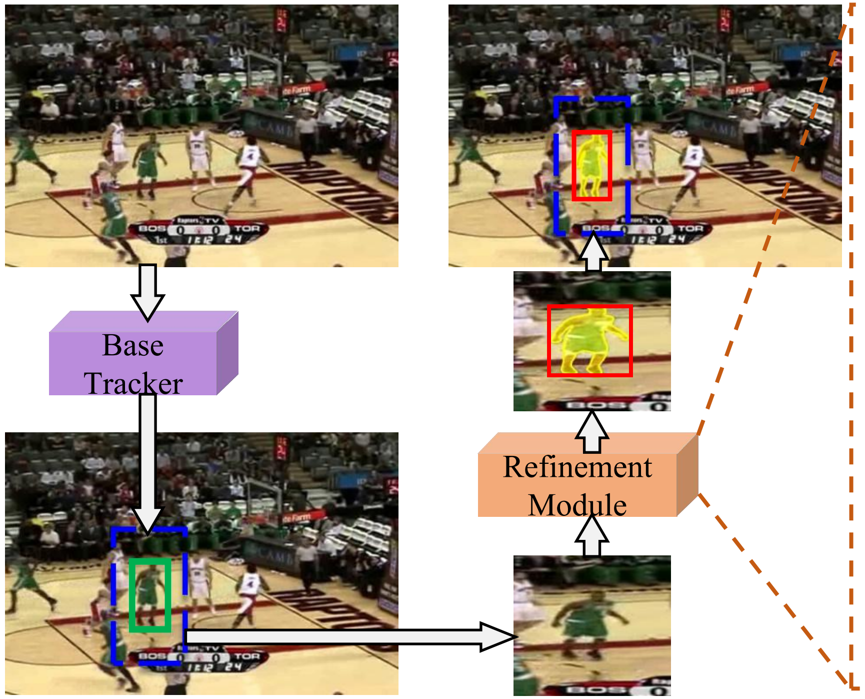


Oracle	AUC	P_{Norm}	P
SiamRPN++[22]	**0.682**	**0.829**	**0.745**
ATOM[6]	0.580	0.686	0.604
DiMPsuper[2]	**0.693**	**0.799**	**0.734**
ECO[5]	0.496	0.666	0.533
AlphaRefine	**0.762**	**0.902**	**0.919**

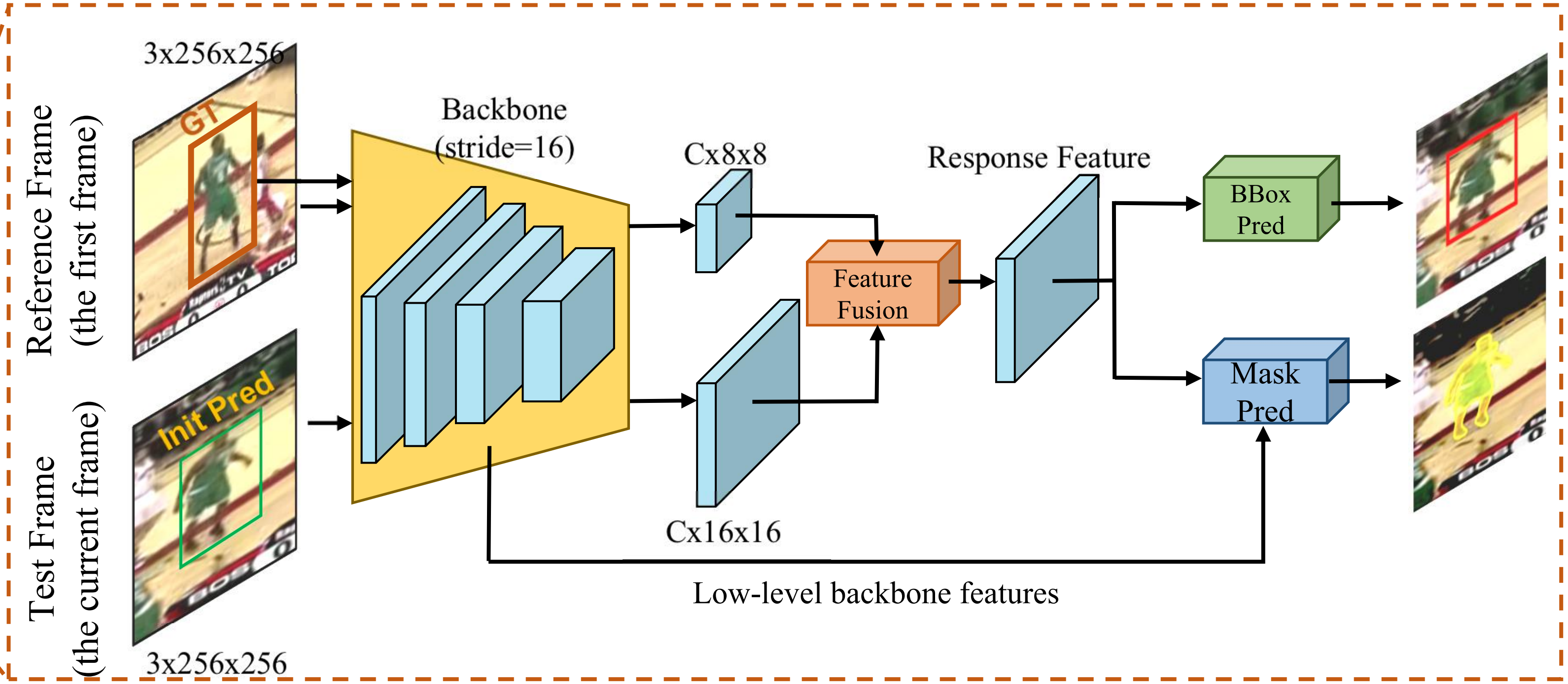
Observation:

1. Existing trackers suffer from low-quality box estimation.
2. A specially designed refinement module may be better at box estimation.

Architecture Overview



Tracking Pipeline



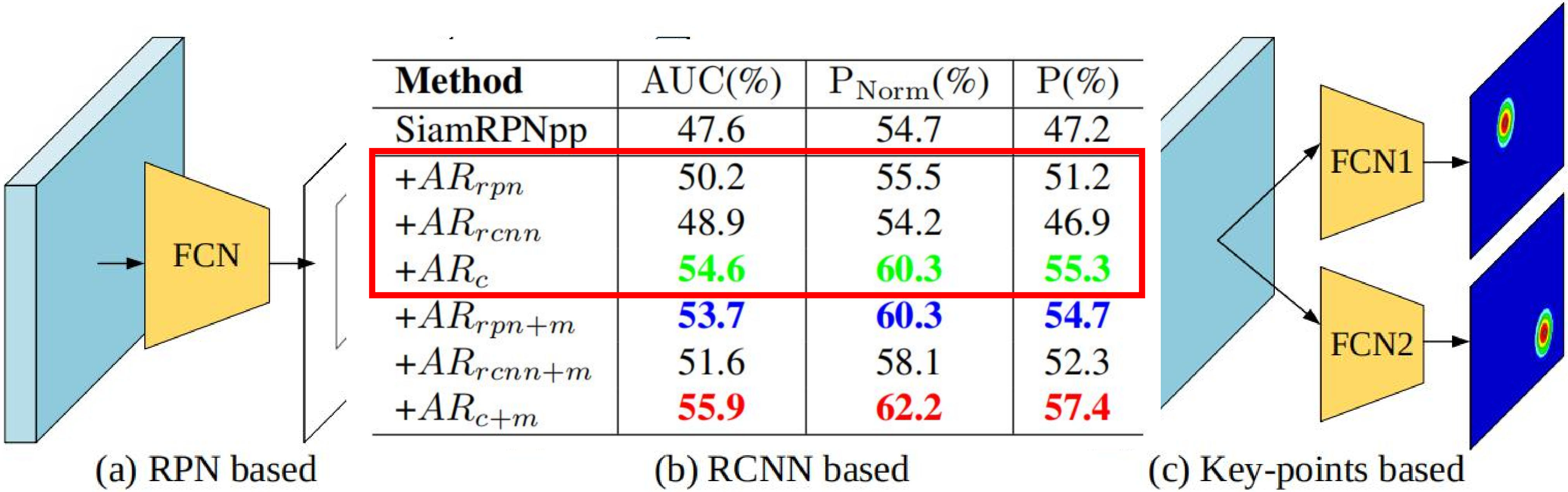
Inner Structure of the Refinement Pipeline

Pipeline:

1. The base tracker predict a preliminary result (Init Pred)
2. Alpha-Refine expand the preliminary prediction as its search region
3. Alpha-Refine predict a refined result in this search region, which is more precise than the preliminary prediction

Head Options

get the mass center of each heatmap

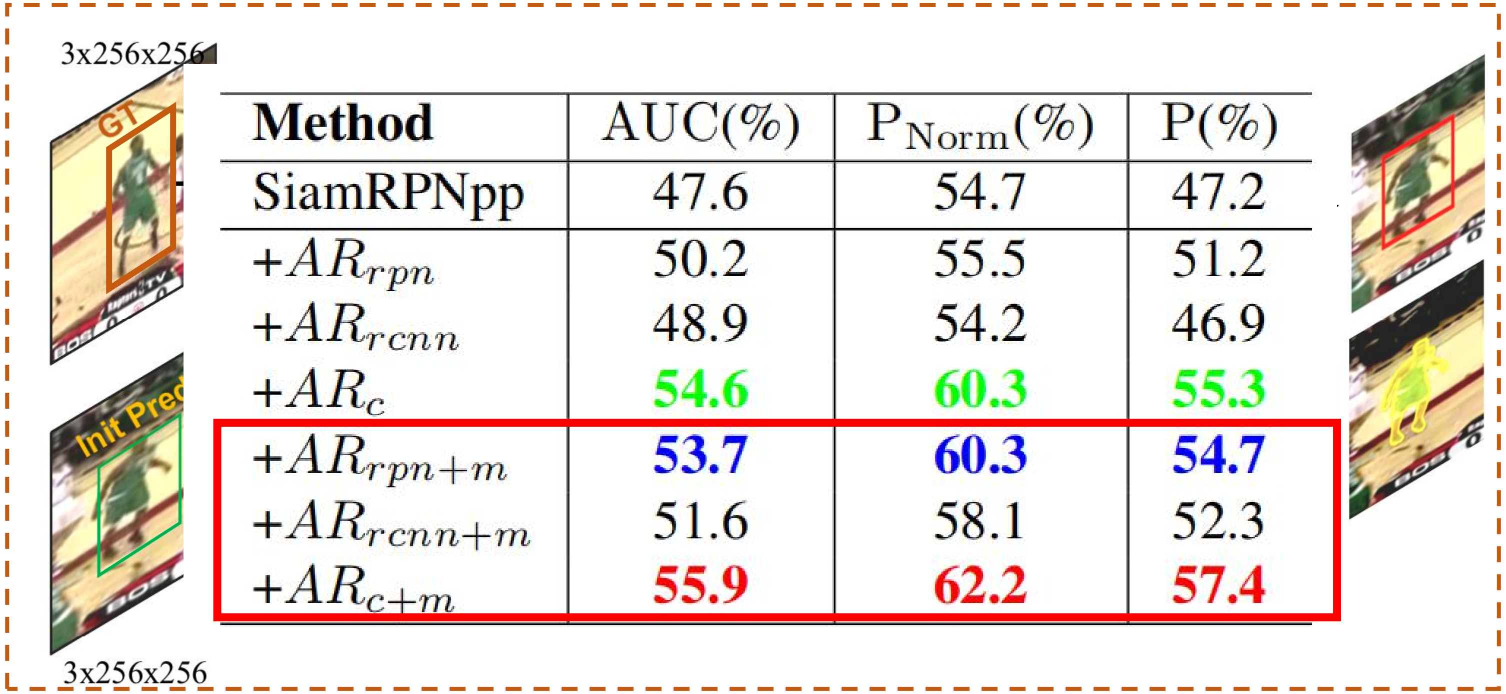


Method	AUC(%)	$\mathrm{P}_{\mathrm{Norm}}(\%)$	P(%)
SiamRPNpp	47.6	54.7	47.2
$+AR_{rpn}$	50.2	55.5	51.2
$+AR_{rcnn}$	48.9	54.2	46.9
$+AR_{c}$	**54.6**	**60.3**	**55.3**
$+AR_{rpn+m}$	**53.7**	**60.3**	**54.7**
$+AR_{rcnn+m}$	51.6	58.1	52.3
$+AR_{c+m}$	**55.9**	**62.2**	**57.4**

(a) RPN based (b) RCNN based (c) Key-points based

- RCNN-style flattens the feature tensor. Spatial information is lost.
- RPN maintain the spatial structure of the feature map, but the estimation of each box utilizes the information of one feature point.
- Corner Representation also maintain the spatial structure of feature map. In addition, the whole feature map work together to estiamte the box, fully utilize the spatial information.

Auxiliary Mask Head


3x256x256
GT
Init Pred
3x256x256

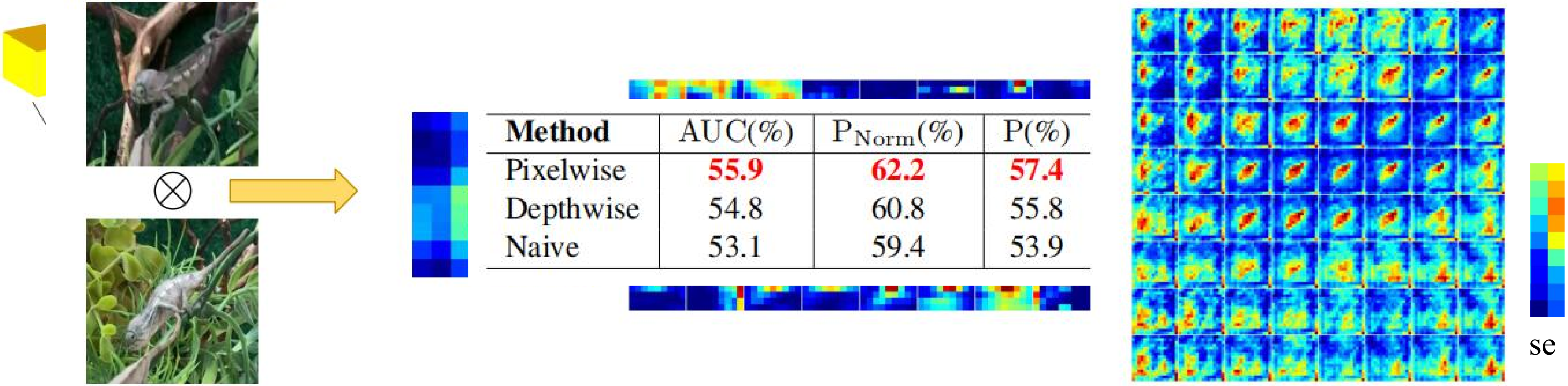
Method	AUC(%)	$P_{\mathrm{Norm}}(\%)$	P(%)
SiamRPNpp	47.6	54.7	47.2
$+AR_{rpn}$	50.2	55.5	51.2
$+AR_{rcnn}$	48.9	54.2	46.9
$+AR_{c}$	54.6	60.3	55.3
$+AR_{rpn+m}$	53.7	60.3	54.7
$+AR_{rcnn+m}$	51.6	58.1	52.3
$+AR_{c+m}$	55.9	62.2	57.4

Multi-task Training:

- Pixel-level supervision encourage refinement module to maintain more detailed spatial information, benefit box estimation
- Pixel-level supervision teach the network to discriminate foreground and background. More discriminative.
- Enable the box-based base tracker to make pixel level prediction, broaden the base tracker’s application.

Feature Fusion Options

(a) Reference image



Method	AUC(%)	P_{Norm}(%)	P(%)
Pixelwise	**55.9**	**62.2**	**57.4**
Depthwise	54.8	60.8	55.8
Naive	53.1	59.4	53.9

(b) Test image

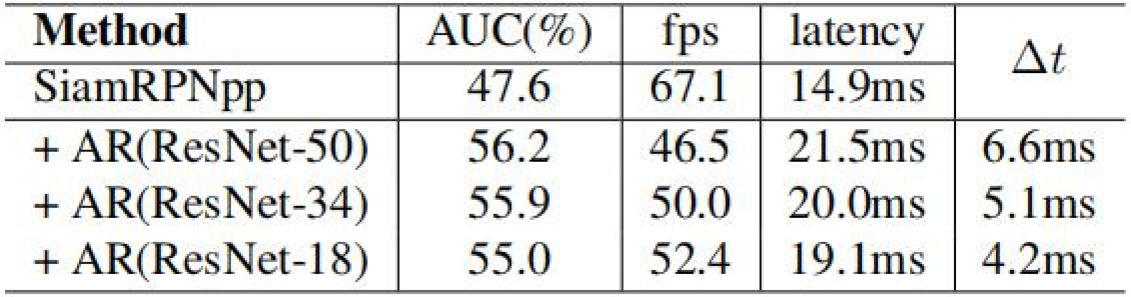
(c) Naive correlation responses

(d) Depth-wise correlation responses (the first 32 channels)

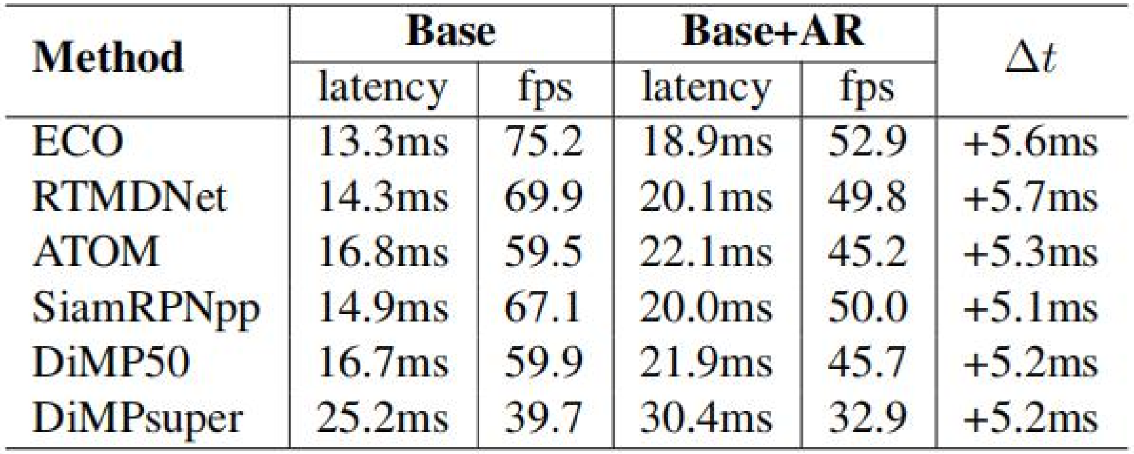
(e) Pixel-wise correlation responses (of all 64 pixel kernels)

- Naive Correlation or Depth-wise Correlation blur the response.
- Pixel-wise Correaltion maintain more spatial information, helpful to box estimation.

Computation Load



Method	AUC(%)	fps	latency	Δt
SiamRPNpp	47.6	67.1	14.9ms	
+ AR(ResNet-50)	56.2	46.5	21.5ms	6.6ms
+ AR(ResNet-34)	55.9	50.0	20.0ms	5.1ms
+ AR(ResNet-18)	55.0	52.4	19.1ms	4.2ms

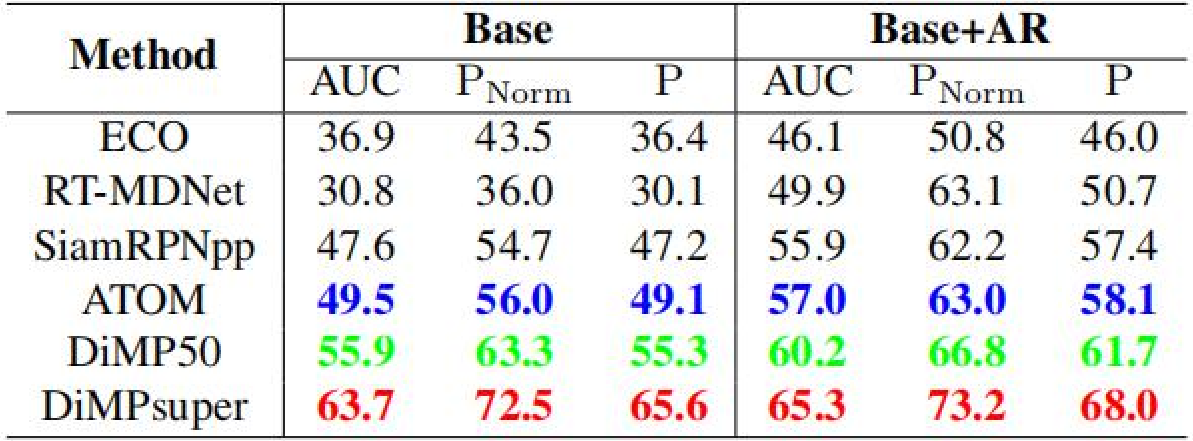


Method	Base		Base+AR		Δt
	latency	fps	latency	fps	
ECO	13.3ms	75.2	18.9ms	52.9	+5.6ms
RTMDNet	14.3ms	69.9	20.1ms	49.8	+5.7ms
ATOM	16.8ms	59.5	22.1ms	45.2	+5.3ms
SiamRPNpp	14.9ms	67.1	20.0ms	50.0	+5.1ms
DiMP50	16.7ms	59.9	21.9ms	45.7	+5.2ms
DiMPsuper	25.2ms	39.7	30.4ms	32.9	+5.2ms

- Alpha-Refine module introduces few computation loads (merely about 5-6ms every frame), while significantly improving the tracking accuracies

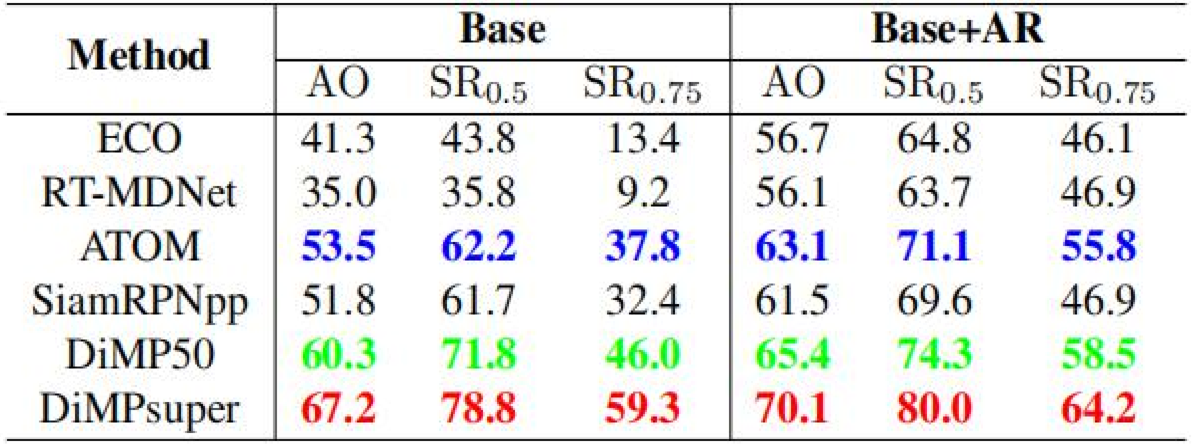
Experiment on More Datasets

Results on LaSOT



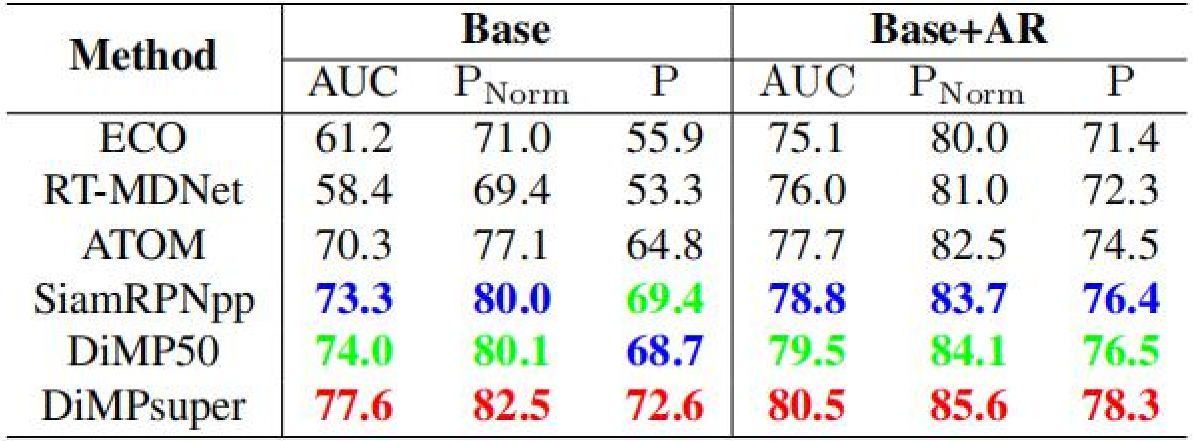
Method	Base			Base+AR		
	AUC	P_{Norm}	P	AUC	P_{Norm}	P
ECO	36.9	43.5	36.4	46.1	50.8	46.0
RT-MDNet	30.8	36.0	30.1	49.9	63.1	50.7
SiamRPNpp	47.6	54.7	47.2	55.9	62.2	57.4
ATOM	**49.5**	**56.0**	**49.1**	**57.0**	**63.0**	**58.1**
DiMP50	**55.9**	**63.3**	**55.3**	**60.2**	**66.8**	**61.7**
DiMPsuper	**63.7**	**72.5**	**65.6**	**65.3**	**73.2**	**68.0**

Results on GOT-10k



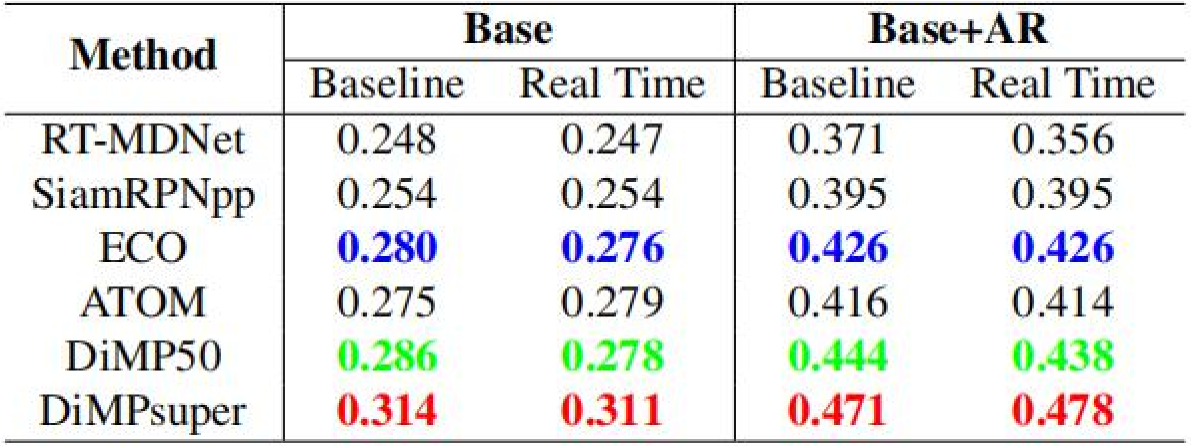
Method	Base			Base+AR		
	AO	$SR_{0.5}$	$SR_{0.75}$	AO	$SR_{0.5}$	$SR_{0.75}$
ECO	41.3	43.8	13.4	56.7	64.8	46.1
RT-MDNet	35.0	35.8	9.2	56.1	63.7	46.9
ATOM	**53.5**	**62.2**	**37.8**	**63.1**	**71.1**	**55.8**
SiamRPNpp	51.8	61.7	32.4	61.5	69.6	46.9
DiMP50	**60.3**	**71.8**	**46.0**	**65.4**	**74.3**	**58.5**
DiMPsuper	**67.2**	**78.8**	**59.3**	**70.1**	**80.0**	**64.2**

Results on TrackingNet



Method	Base			Base+AR		
	AUC	P_{Norm}	P	AUC	P_{Norm}	P
ECO	61.2	71.0	55.9	75.1	80.0	71.4
RT-MDNet	58.4	69.4	53.3	76.0	81.0	72.3
ATOM	70.3	77.1	64.8	77.7	82.5	74.5
SiamRPNpp	**73.3**	**80.0**	**69.4**	**78.8**	**83.7**	**76.4**
DiMP50	**74.0**	**80.1**	**68.7**	**79.5**	**84.1**	**76.5**
DiMPsuper	**77.6**	**82.5**	**72.6**	**80.5**	**85.6**	**78.3**

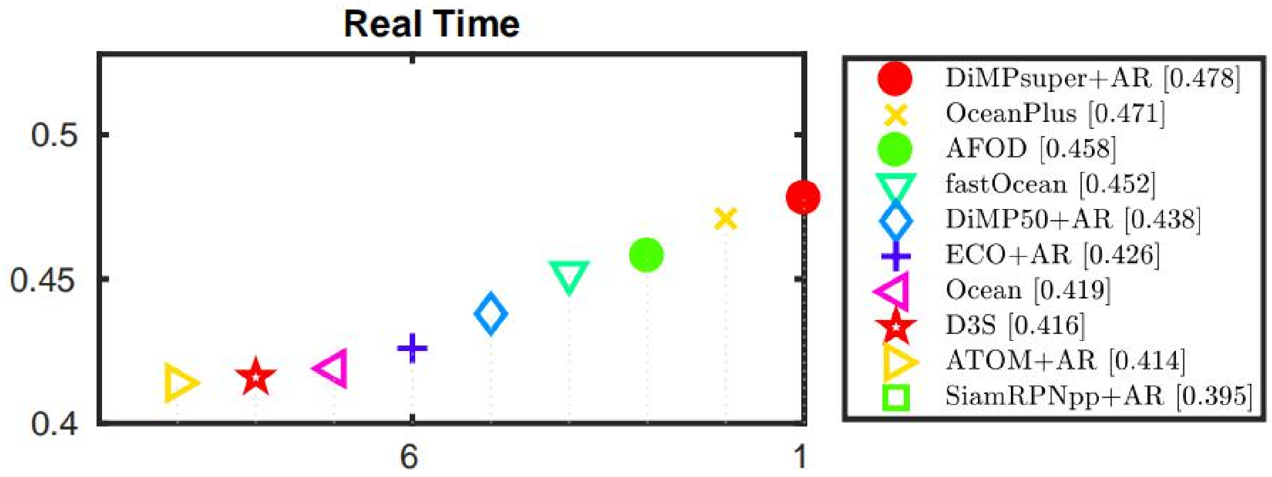
Results on VOT2020

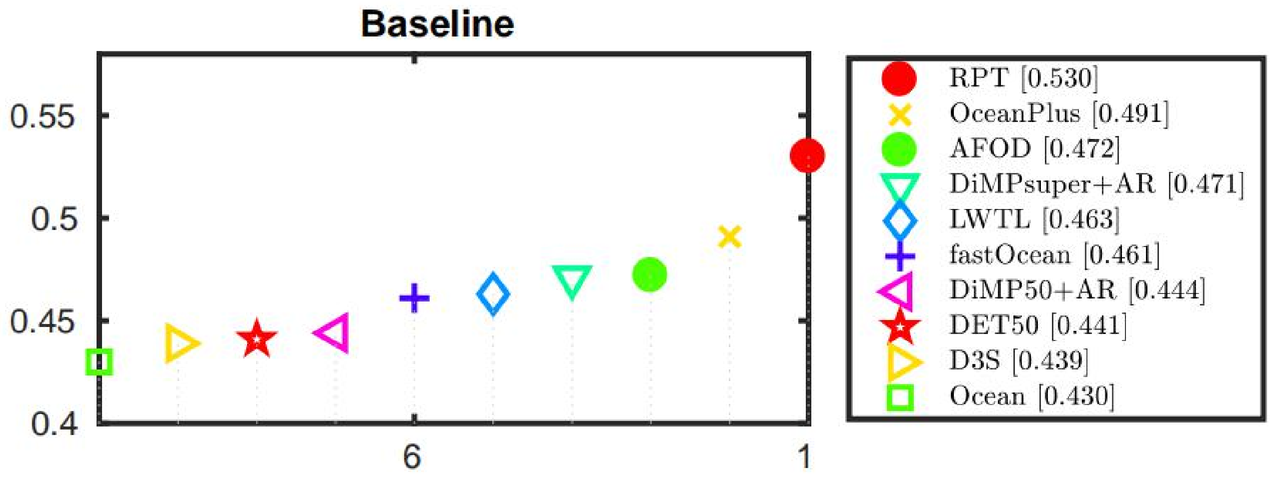


Method	Base		Base+AR	
	Baseline	Real Time	Baseline	Real Time
RT-MDNet	0.248	0.247	0.371	0.356
SiamRPNpp	0.254	0.254	0.395	0.395
ECO	**0.280**	**0.276**	**0.426**	**0.426**
ATOM	0.275	0.279	0.416	0.414
DiMP50	**0.286**	**0.278**	**0.444**	**0.438**
DiMPsuper	**0.314**	**0.311**	**0.471**	**0.478**

- Alpha-Refine significantly improve the base trackers under different benchmarks.

Winner of VOT2020 RealTime





- Alpha-Refine is the key component of the winner method in VOT2020 RealTime.