How to Aggregate Arbitrary Online Trackers [IS3-2-28]

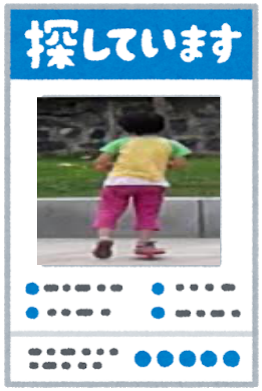
Heon Song (Kyushu Univ., RIKEN),
Daiki Suehiro (Kyushu Univ., RIKEN),
Seiichi Uchida (Kyushu Univ.)

KYUSHU
UNIVERSITY

INTRODUCTION

What is online Visual Object Tracking (VOT)?

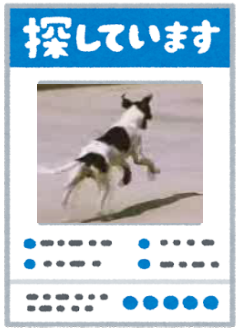
- Given a target location at the first frame,
  we should track the target and predict the location.



Frame 1 Frame 2 Frame 3

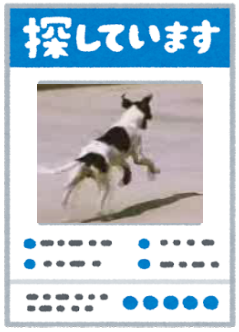
What makes online VOT difficult?

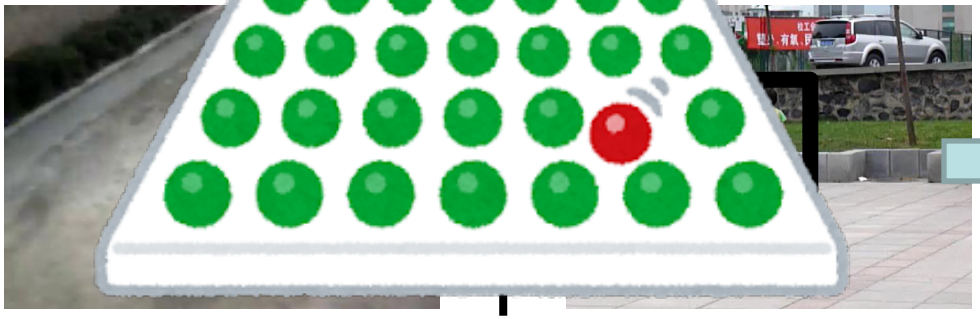
- Depending on videos,
  we may have to track completely different objects.
- There might be a heavy appearance change
  of the target or occlusion even in a video.



What makes online VOT difficult?

No almighty online tracker





Use multiple online trackers for robust tracking

I track animals well

I track plastics well

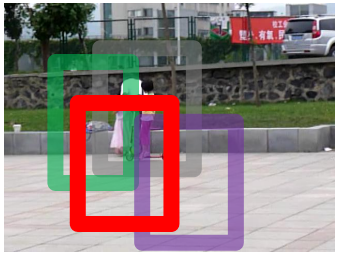
I track occluded object well

We call them “experts”

Use multiple online trackers for robust tracking

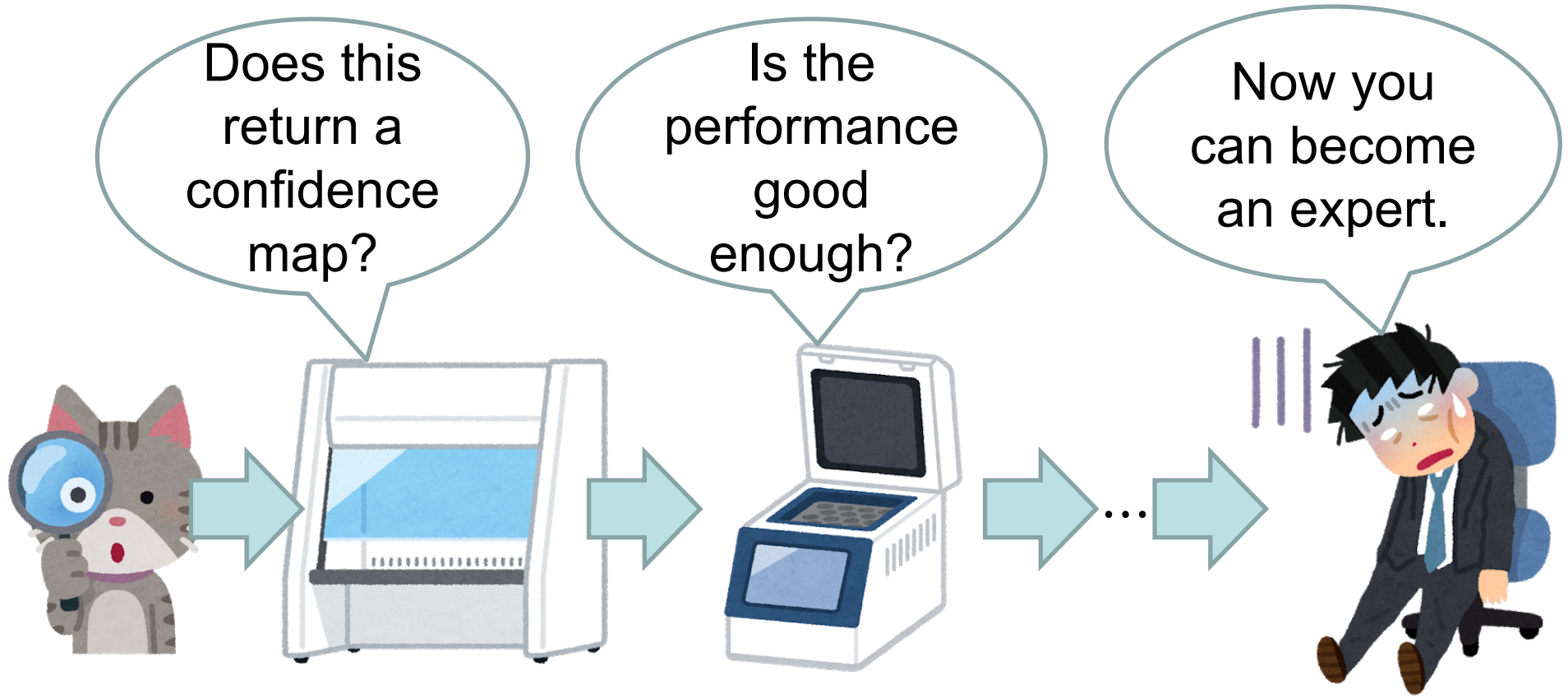
For example…

Simply average experts' estimation every frame



Problems with using experts

- Most algorithms using experts require a lot of restrictions to use online trackers as experts.



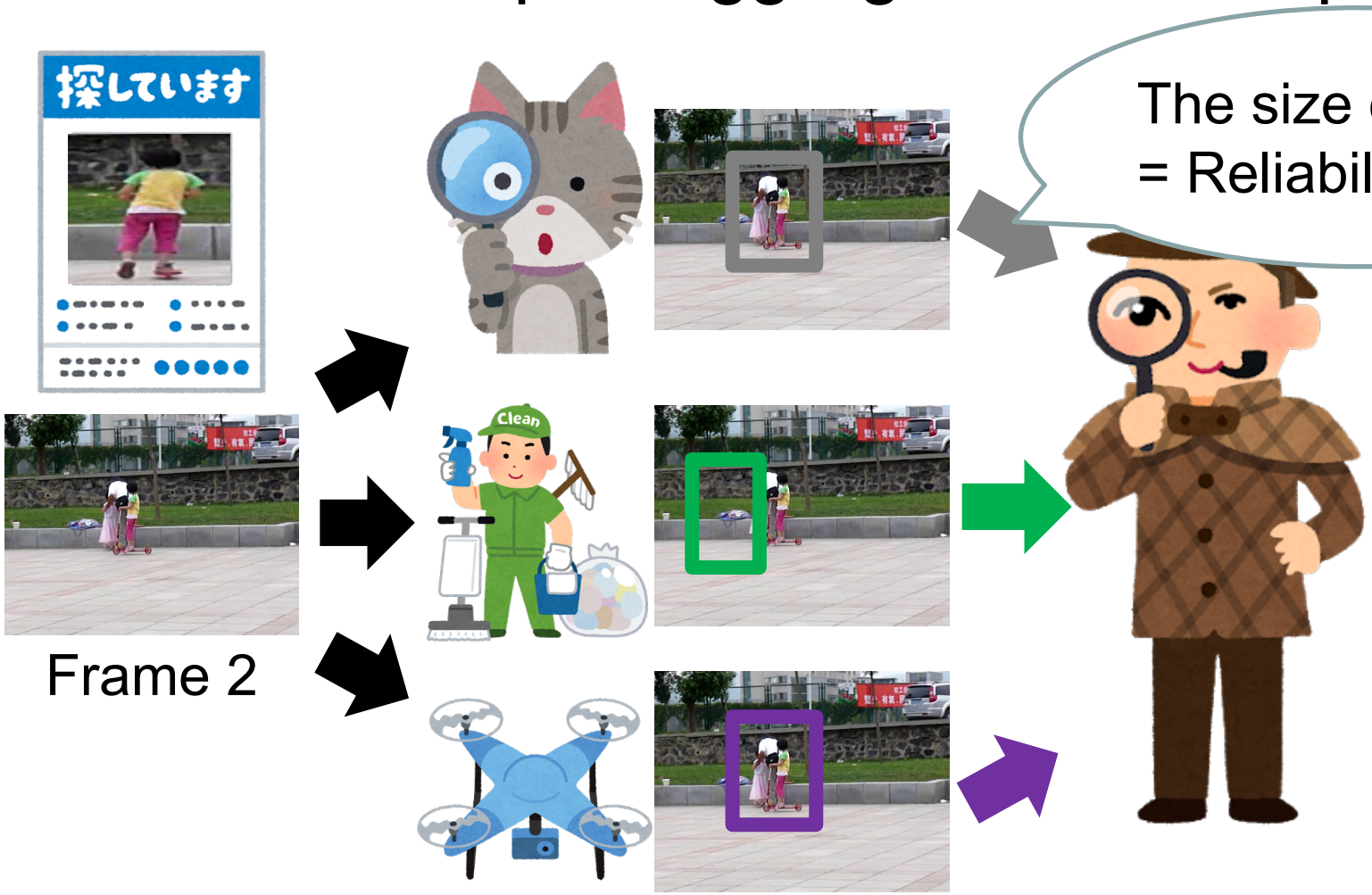
In our work

- The proposed method can aggregate arbitrary trackers.

- Moreover, The performance of the proposed method is theoretically guaranteed.

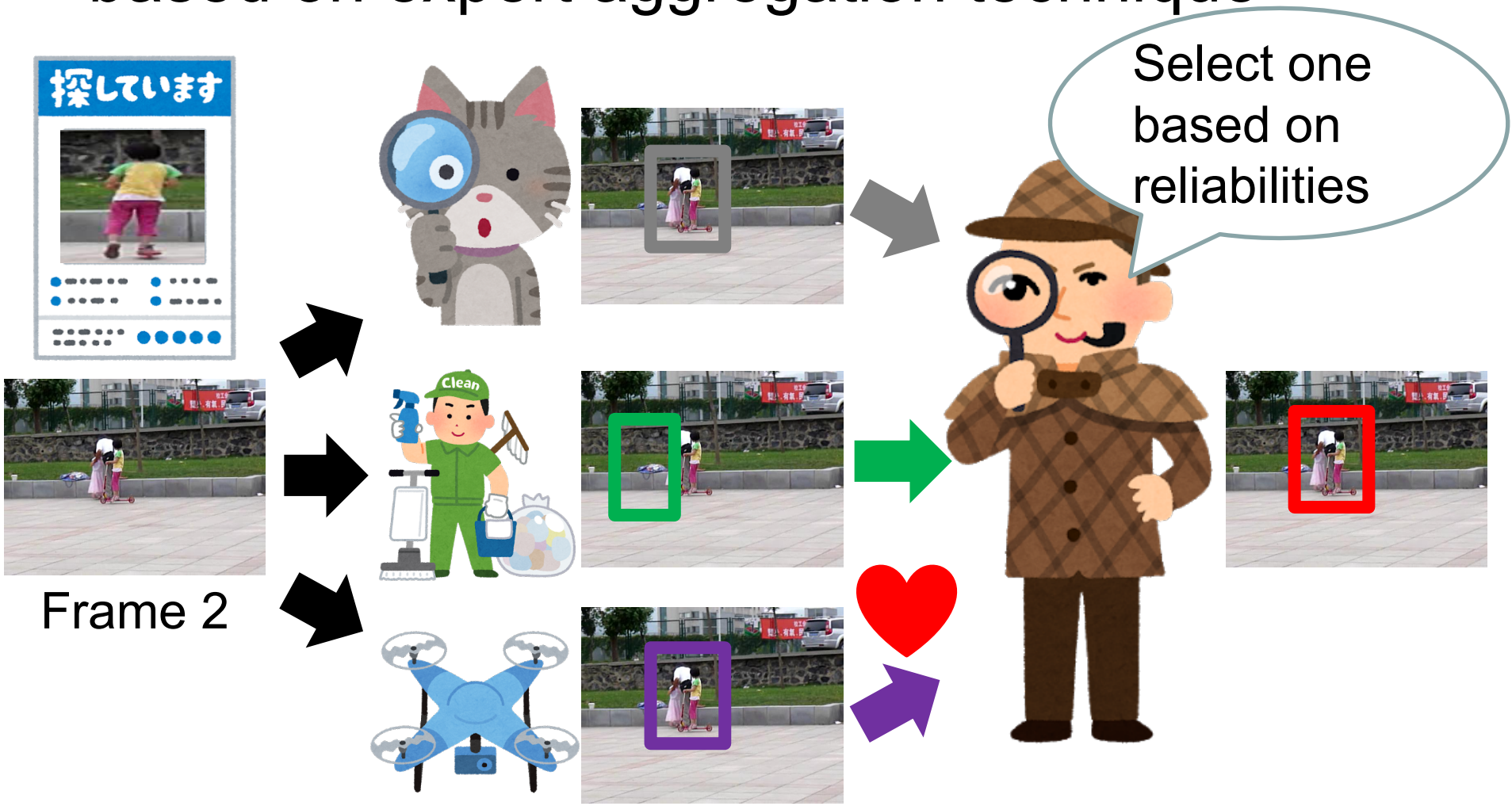
TRACKING WITH MULTIPLE ONLINE TRACKER

Aggregate multiple online trackers based on expert aggregation technique[1]



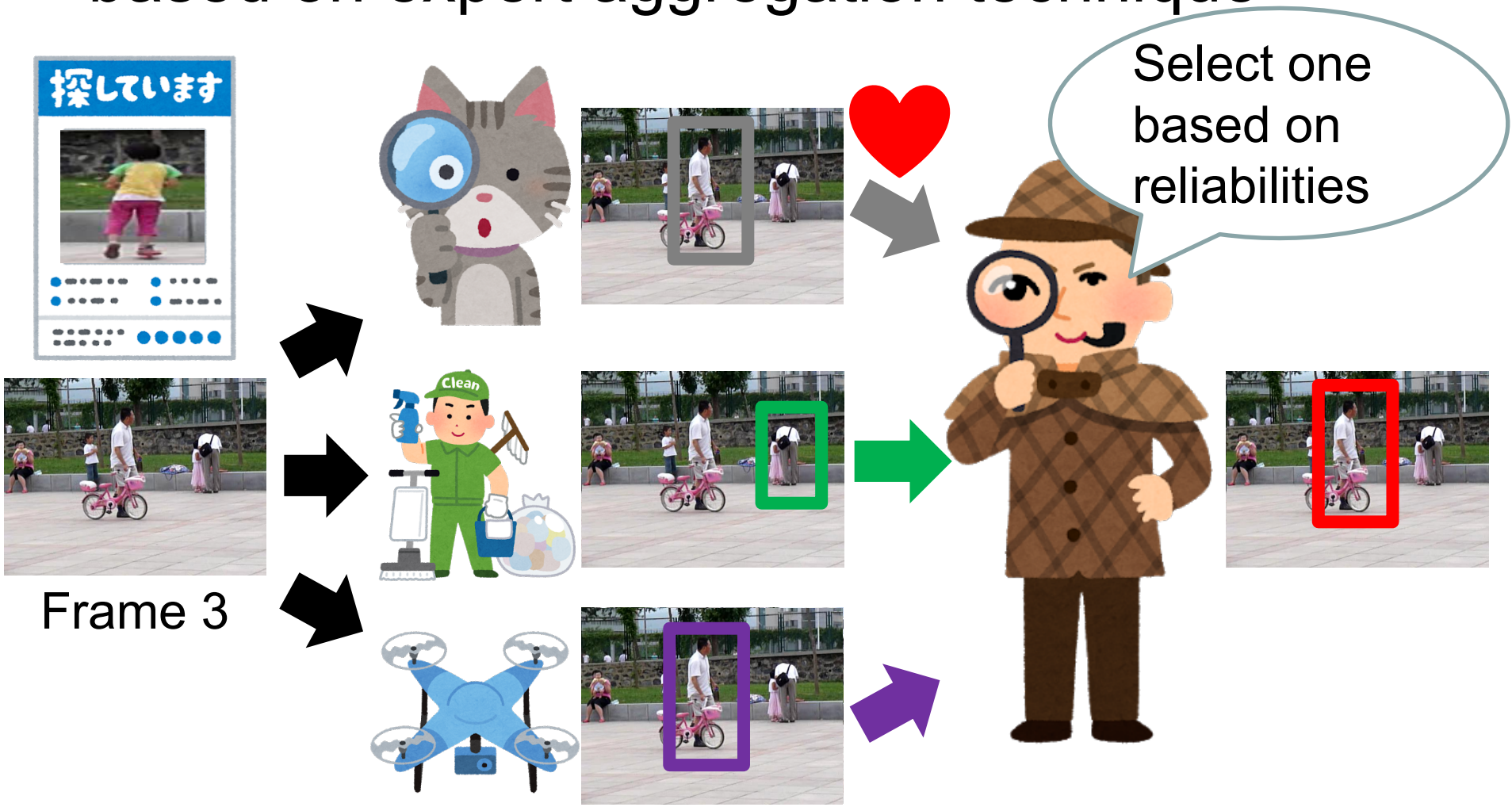
[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

Aggregate multiple online trackers based on expert aggregation technique[1]



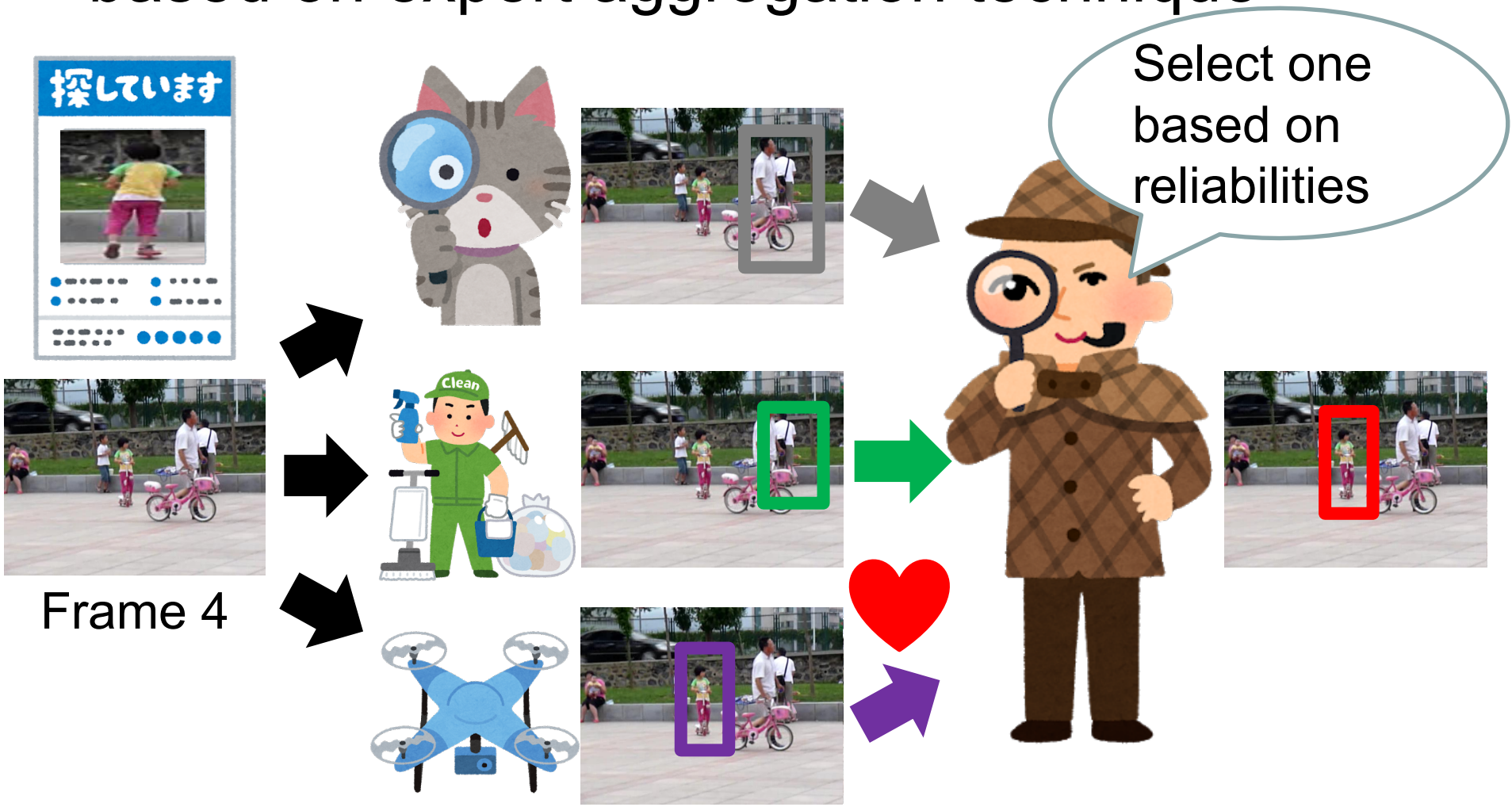
[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

Aggregate multiple online trackers based on expert aggregation technique[1]



[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

Aggregate multiple online trackers based on expert aggregation technique[1]

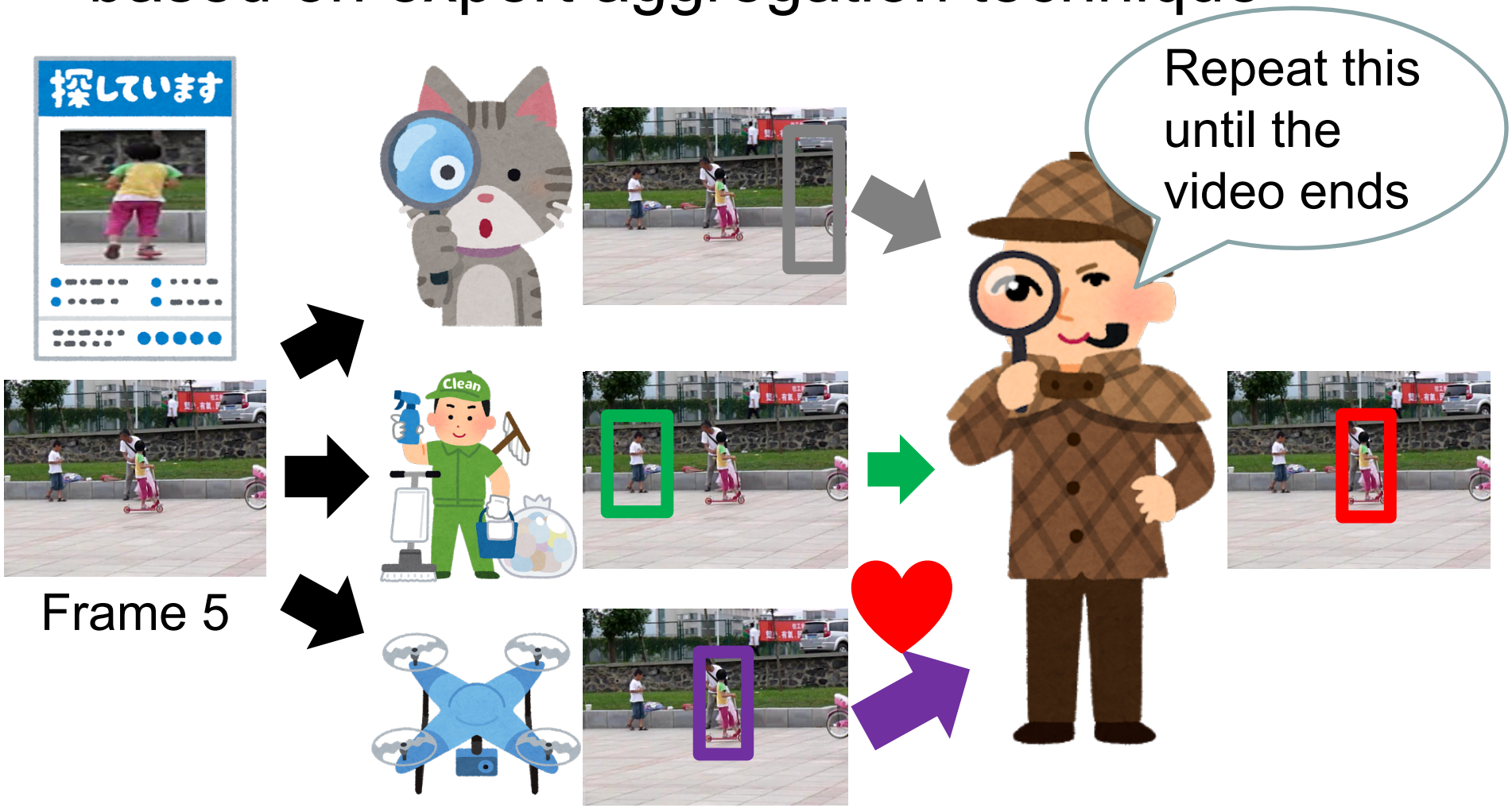


[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

Aggregate multiple online trackers based on expert aggregation technique[1]

[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

Aggregate multiple online trackers based on expert aggregation technique[1]



[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

How can we get feedback?



Where is she?







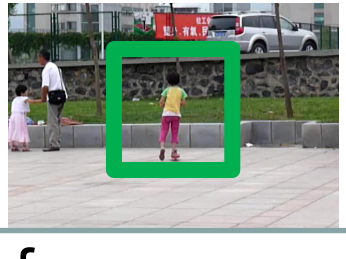
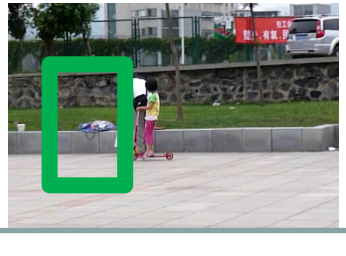
Frame 1 → Frame 2 → Frame 3

How can we get feedback?







Here she is!

Anchor frame =
When the target location can be predicted with high confidence


Frame 1 ⇨ Frame 2 ⇨ Frame 3 ⇨ Frame 4

How can we get feedback?

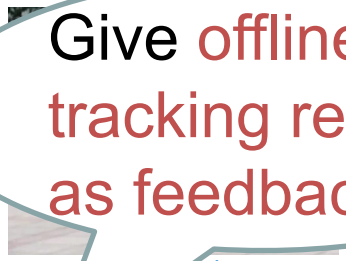
















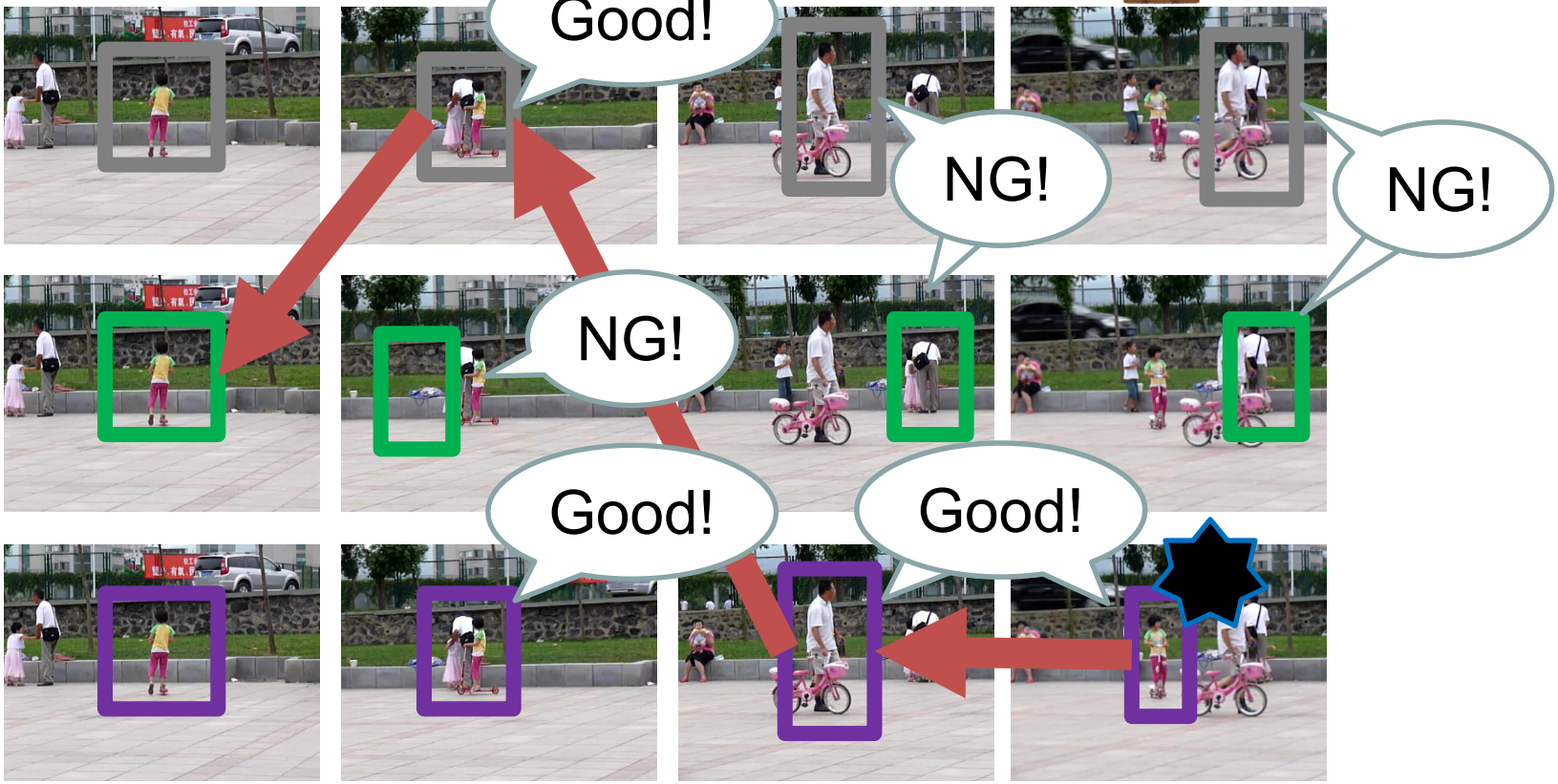
Give offline tracking results as feedback

Anchor frame

Anchor frame

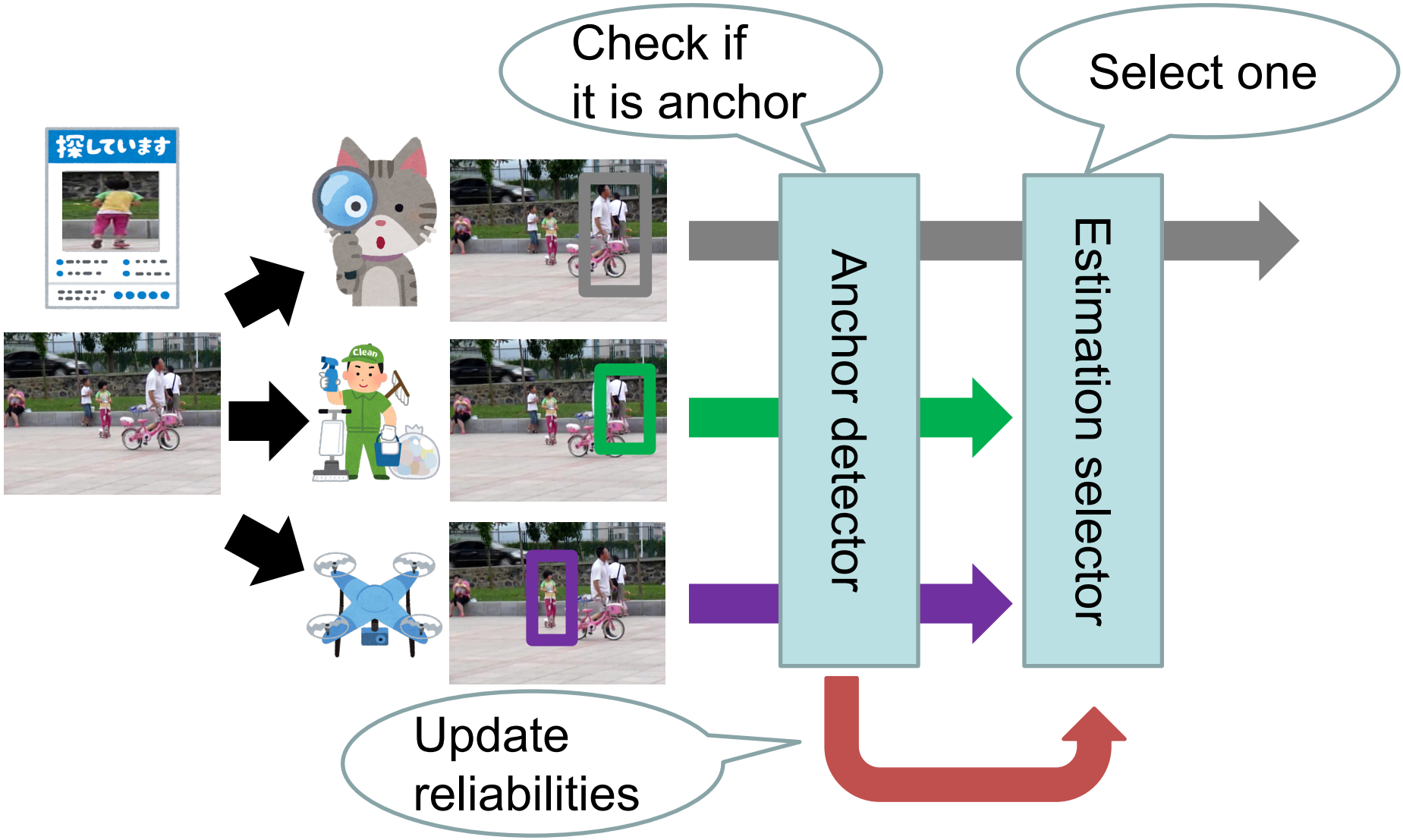
Frame 1 ⇨ Frame 2 ⇨ Frame 3 ⇨ Frame 4

How can we get feedback?



Frame 1 ⇨ Frame 2 ⇨ Frame 3 ⇨ Frame 4

Overview of the proposed method



Why should we use online prediction?

Total error L_1^T

Total error L_2^T

Total error L_3^T

minimum

Defined only at the end of the video

$L_2^T (= L_{best}^T)$

Difference
=Regret
$= L_{Ours}^T - L_{best}^T$

Total error L_{Ours}^T

L_{Ours}^T

Why should we use online prediction?

- Surprisingly, the regret of the proposed method is bounded without any assumption.

- We do not need to know who will be the best expert, which data will be given, and how long the video is.

The length of the video

$$Regret = L^T_{Ours} - L^T_{best} \leq O\left(\sqrt{T \ln N}\right)$$

The number of the experts

EXPERIMENTS & RESULTS

Experts and Benchmarks

- I would like to demonstrate that the proposed method is robust for arbitrary videos and experts.
- For arbitrary video, I evaluated the proposed method on six different benchmarks.
- For arbitrary experts, I prepared twelve experts and divide into 2 groups according to their performance.

Comparision with High group

- High group included six high-performance experts.
- Ours outperformed the experts on almost benchmarks.

	OTB2015 [14]		TCo128 [15]		UAC123 [16]		NFS [17]		LaSOT [18]		VOT2018 [19]
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
ATOM [2]	0.67	0.87	0.6	0.81	0.62	0.82	0.58	0.69	0.51	0.51	0.52
DaSiamRPN [3]	0.65	0.88	0.53	0.75	0.57	0.78	0.55	0.67	0.43	0.42	0.43
SiamMCF [4]	0.65	0.85	0.57	0.78	0.54	0.77	0.57	0.7	0.44	0.45	0.45
SiamRPN++ [5]	0.69	0.9	0.58	0.77	0.6	0.8	0.6	0.74	0.49	0.51	0.5
SPM [6]	0.67	0.87	0.58	0.79	0.59	0.77	0.57	0.67	0.47	0.48	0.48
THOR [7]	0.64	0.85	0.52	0.72	0.57	0.77	0.57	0.68	0.4	0.41	0.47
Ours	**0.7**	**0.91**	**0.62**	**0.84**	**0.62**	**0.83**	**0.61**	**0.75**	**0.53**	**0.55**	**0.52**

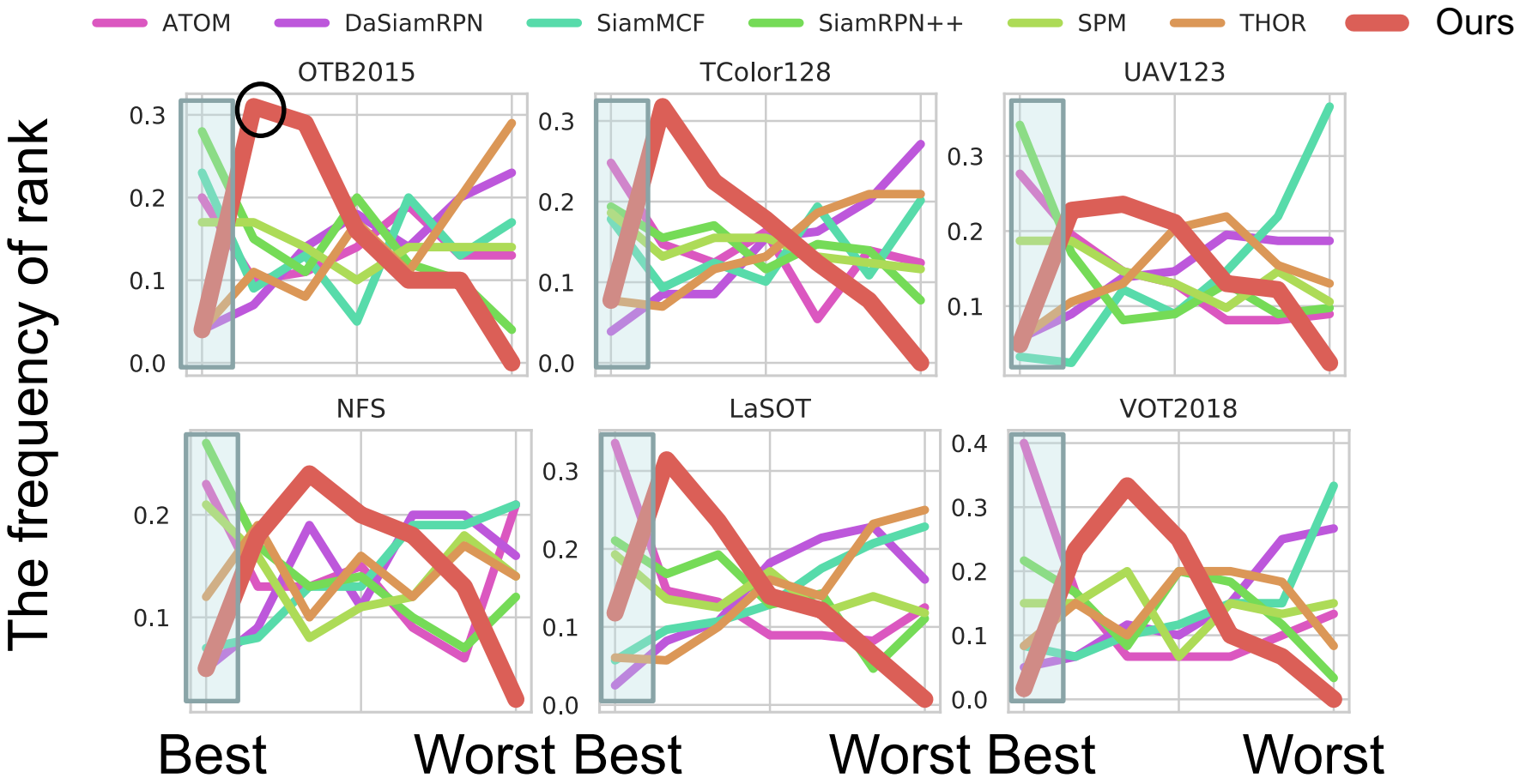
Experts

2nd

1st

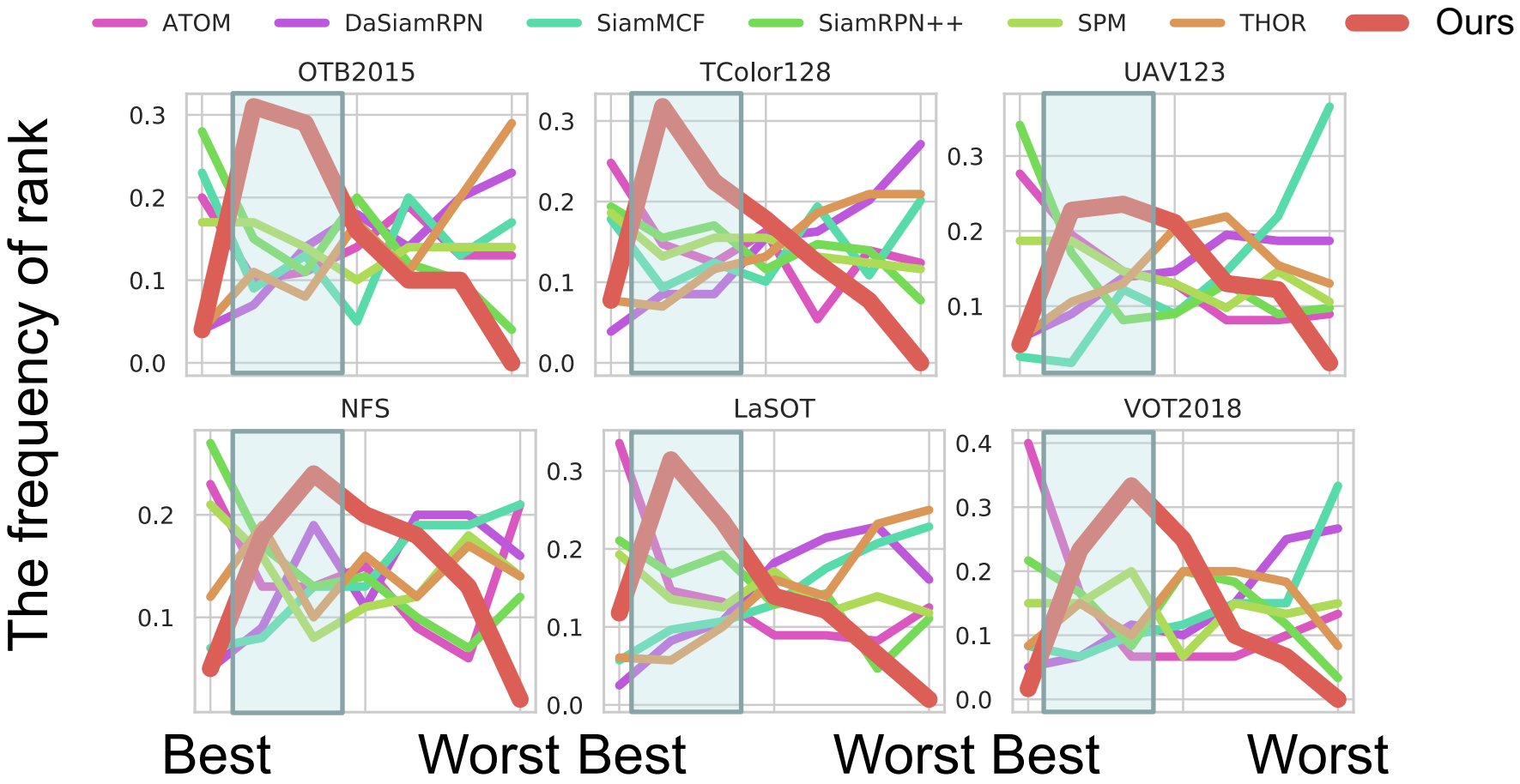
Comparision with High group

- No tracker was always the best, and the best expert drastically changed over the videos.



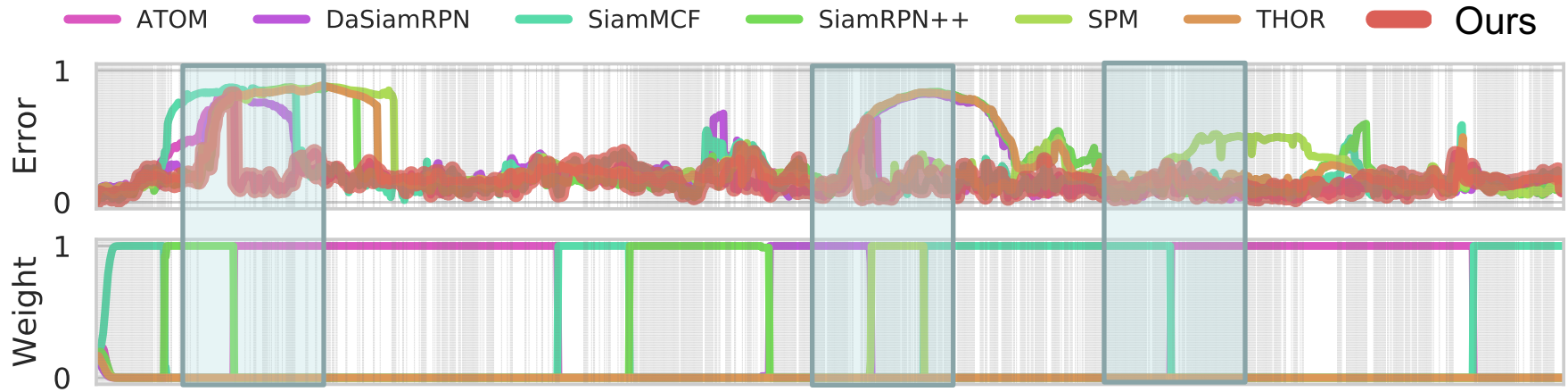
Comparision with High group

- Ours achieved the second or third best performance for most videos.



Tracking example

- Ours(AAA) can follow the best expert by adaptively aggregating the trackers.



Comparision with Low group

- Low group included six low-performance experts.
- Ours achieved at least the second best performance for all benchmarks.

	OTB2015 [14]		TCo128 [15]		UAC123 [16]		NFS [17]		LaSOT [18]		VOT2018 [19]
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
GradNet [8]	0.63	0.85	0.56	0.76	0.51	0.74	0.51	0.64	0.36	0.38	0.4
MemTrack [9]	0.63	0.82	0.54	0.74	0.49	0.7	0.5	0.61	0.34	0.35	0.39
SiamDW [10]	0.67	0.91	0.53	0.75	0.46	0.69	0.5	0.63	0.35	0.34	0.37
SiamFC [11]	0.59	0.78	0.52	0.7	0.51	0.74	0.51	0.6	0.35	0.36	0.33
SiamRPN [12]	0.63	0.83	0.52	0.71	0.58	0.77	0.56	0.66	0.45	0.45	0.48
Staple [13]	0.6	0.79	0.51	0.68	0.45	0.64	0.41	0.48	0.24	0.23	0.3
Ours	**0.66**	**0.87**	**0.59**	**0.82**	**0.56**	**0.78**	**0.58**	**0.69**	**0.45**	**0.46**	**0.45**

Experts

1st

2nd

Reference

[1] Quanrud, Kent and Daniel Khashabi, “Online Learning with Adversarial Delays,” in Proc. NIPS, 2014.

[2] Danelljan, Martin, et al, "Atom: Accurate tracking by overlap maximization," in Proc. CVPR, 2019.

[3] Zhu, Zheng, et al. "Distractor-aware siamese networks for visual object tracking," in Proc. ECCV, 2018.

[4] Morimitsu, Henrique, "Multiple context features in Siamese networks for visual object tracking,” in Proc. ECCV, 2018.

[5] Li, Bo, et al, "Siamrpn++: Evolution of siamese visual tracking with very deep networks,” in Proc. CVPR, 2019.

[6] Wang, Guangting, et al, "Spm-tracker: Series-parallel matching for real-time visual object tracking," in Proc. CVPR, 2019.

[7] Sauer, Axel, Elie Aljalbout, and Sami Haddadin, "Tracking holistic object representations,” in Proc. BMVC, 2019.

[8] Li, Peixia, et al, "Gradnet: Gradient-guided network for visual object tracking,” in Proc. ICCV, 2019.

[9] Yang, Tianyu, and Antoni B. Chan, "Learning dynamic memory networks for object tracking,” in Proc. ECCV, 2018.

[10] Zhang, Zhipeng, and Houwen Peng, "Deeper and wider siamese networks for real-time visual tracking,” in Proc. CVPR, 2019.

[11] Bertinetto, Luca, et al, "Fully-convolutional siamese networks for object tracking.” in Proc. ECCV, 2016.

Reference

[12] Li, Bo, et al, "High performance visual tracking with siamese region proposal network," in Proc. CVPR, 2018.

[13] Bertinetto, Luca, et al, "Staple: Complementary learners for real-time tracking," in Proc. CVPR, 2016.

[14] Wu, Yi, Jongwoo Lim, and Ming-Hsuan Yang, "Online object tracking: A benchmark," in Proc. CVPR, 2013.

[15] Liang, Pengpeng, Erik Blasch, and Haibin Ling, "Encoding color information for visual tracking: Algorithms and benchmark," TIP 24.12 (2015): 5630-5644.

[16] Mueller, Matthias, Neil Smith, and Bernard Ghanem, "A benchmark and simulator for uav tracking," in Proc. ECCV, 2016.

[17] Kiani Galoogahi, Hamed, et al, "Need for speed: A benchmark for higher frame rate object tracking," in Proc. CVPR, 2017.

[18] Fan, Heng, et al, "Lasot: A high-quality benchmark for large-scale single object tracking," in Proc. CVPR, 2019.

[19] Kristan, Matej, et al, "A novel performance evaluation methodology for single-target trackers," TPAMI, 38.11 (2016): 2137-2155.

[20] Zhang, Li, Yuan Li, and Ramakant Nevatia, "Global data association for multi-object tracking using network flows," in Proc. CVPR, 2008

[21] He, Kaiming, et al, "Deep residual learning for image recognition," in Proc. CVPR, 2016.

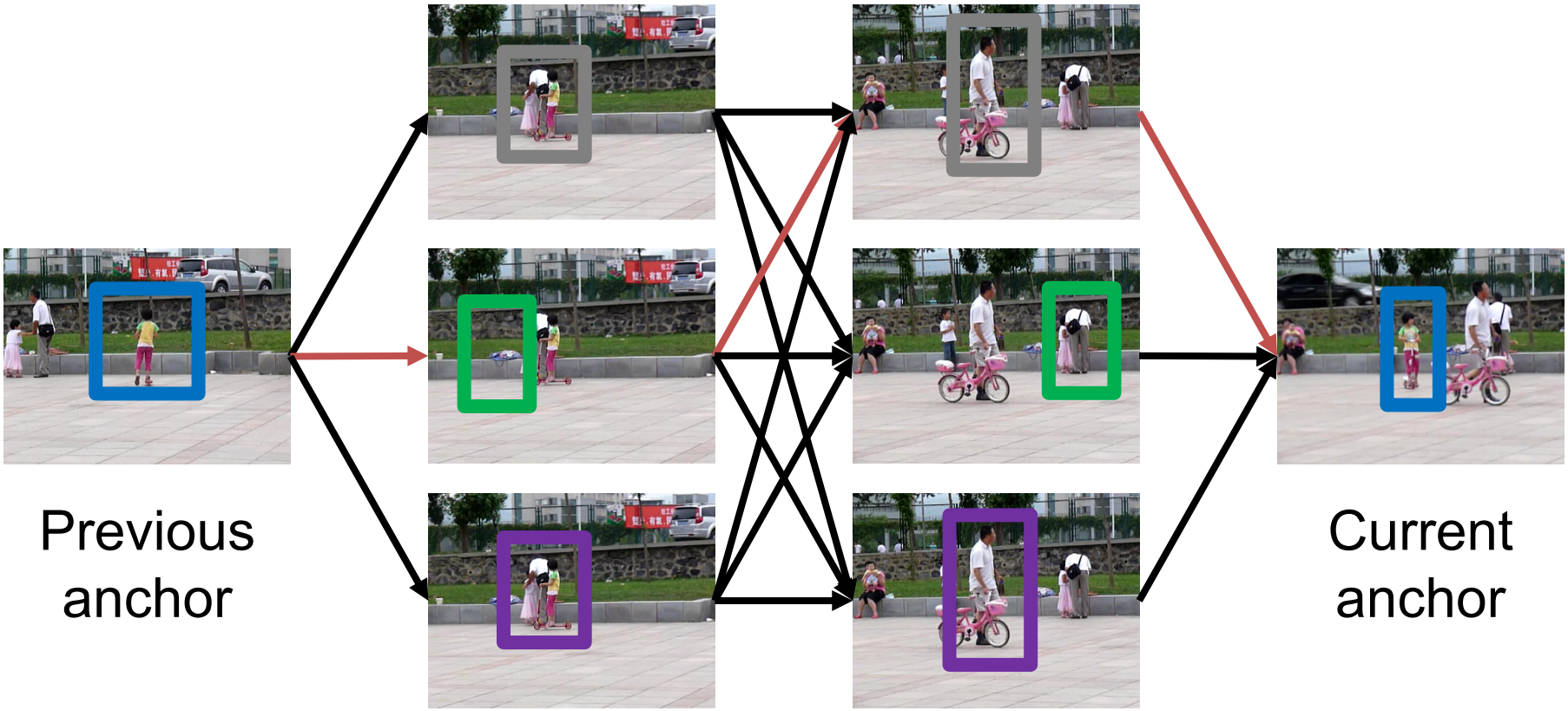
[22] Rezatofighi, Hamid, et al, "Generalized intersection over union: A metric and a loss for bounding box regression," in Proc. CVPR. 2019.

[23] Wang, Ning, et al, "Multi-cue correlation filters for robust visual tracking," in Proc. CVPR, 2018.

APPENDIX

How to track the target offline

- Create a graph with each tracker's estimation as a node as below and find minimum-cost flow[20].



[20] Zhang, Li, Yuan Li, and Ramakant Nevatia, "Global data association for multi-object tracking using network flows," in Proc. CVPR, 2008

How to detect anchor frame

- Extract a target feature vector from the target image which is cropped at the first frame by ResNet[21].
- In the same way, feature vectors are extracted from bounding boxes predicted by trackers every frame.
- If the cosine similarity between the target feature vector and each extracted feature vector by trackers is greater than a threshold, the frame is considered to be an anchor frame.

[21] He, Kaiming, et al, "Deep residual learning for image recognition," in Proc. CVPR, 2016.

How to calculate the loss of experts

The offline tracker gives the bounding box of the target object $y^{u+1}, \ldots, y^{t}$ between the previous anchor frame $u+1$ to current anchor frame t. Using the offline tracking results, the loss of tracker i is calculated using the following equation[22]:

$$L_i = \sum_{\tau=u+1}^{t} 1 - \mathrm{GIoU}(f_i^{\tau}, y^{\tau}),$$

where f_i^{τ} is the i-th tracker's estimation of the target location at frame τ.

[22] Rezatofighi, Hamid, et al, "Generalized intersection over union: A metric and a loss for bounding box regression," in Proc. CVPR. 2019.

How to update the reliability

Based on the loss, the reliability of tracker i is updated according to following equation:

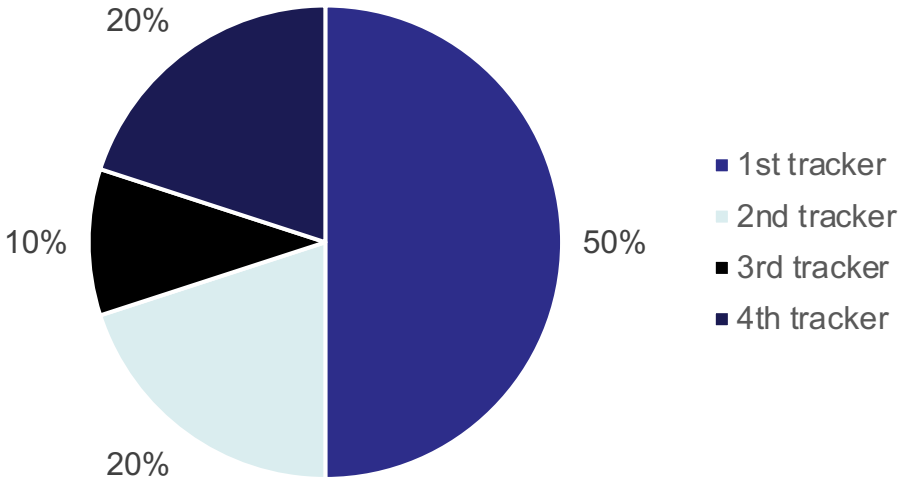
$$w_i^{t+1} = \frac{w_i^t \exp(-\eta L_i)}{\sum_{j=1}^{N} w_j^t \exp\left(-\eta L_j\right)},$$

where η is a learning rate and N is the number of aggregated trackers.

How to select one estimation

- For selecting one of experts' estimation, we use roulette wheel selection, known as fitness proportionate selection, according to their weight.

- Each tracker’s weight represents the probability that the tracker’s estimation will be selected.



Comparision with High group

- I compared the proposed method with other aggregation-based trackers.
- In other trackers, performance was lower than most experts.

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
ATOM	0.67	0.87	*0.60*	*0.81*	**0.62**	*0.82*	0.58	0.69	*0.51*	*0.51*	*0.52*
DaSiamRPN	0.65	0.88	0.53	0.75	0.57	0.78	0.55	0.67	0.43	0.42	0.43
SiamMCF	0.65	0.85	0.57	0.78	0.54	0.77	0.57	0.70	0.44	0.45	0.45
SiamRPN++	*0.69*	*0.90*	0.58	0.77	0.60	0.80	0.60	*0.74*	0.49	0.51	0.50
SPM	0.67	0.87	0.58	0.79	0.59	0.77	0.57	0.67	0.47	0.48	0.48
THOR	0.64	0.85	0.52	0.72	0.57	0.77	0.57	0.68	0.40	0.41	0.47
MCCT[22]	0.64	0.83	0.53	0.72	0.58	0.76	0.57	0.69	0.42	0.44	0.40
Random	0.66	0.87	0.56	0.77	0.58	0.79	0.57	0.69	0.46	0.46	0.48
Max	0.68	0.89	0.56	0.77	0.58	0.78	*0.61*	0.74	0.46	0.46	0.46
AAA	**0.70**	**0.91**	**0.62**	**0.84**	*0.62*	**0.83**	**0.61**	**0.75**	**0.53**	**0.55**	**0.52**

[22] Wang, Ning, et al, "Multi-cue correlation filters for robust visual tracking," in Proc. CVPR, 2018.

Comparision with Low group

- Even when we aggregate experts in Low group, no other trackers can outperform the experts.

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
GradNet	0.63	0.85	*0.56*	*0.76*	0.51	0.74	0.51	0.64	0.36	0.38	0.40
MemTrack	0.63	0.82	0.54	0.74	0.49	0.70	0.50	0.61	0.34	0.35	0.39
SiamDW	**0.67**	**0.91**	0.53	0.75	0.46	0.69	0.50	0.63	0.35	0.34	0.37
SiamFC	0.59	0.78	0.52	0.70	0.51	0.74	0.51	0.60	0.35	0.36	0.33
SiamRPN	0.63	0.83	0.52	0.71	**0.58**	*0.77*	*0.56*	*0.66*	**0.45**	*0.45*	**0.48**
Staple	0.60	0.79	0.51	0.68	0.45	0.64	0.41	0.48	0.24	0.23	0.30
MCCT[22]	0.59	0.79	0.49	0.66	0.50	0.70	0.51	0.63	0.32	0.34	0.32
Random	0.62	0.83	0.53	0.72	0.50	0.71	0.50	0.60	0.35	0.35	0.38
Max	0.63	0.83	0.52	0.71	0.51	0.73	0.53	0.64	0.35	0.35	0.38
AAA	*0.66*	*0.87*	**0.59**	**0.82**	*0.56*	**0.78**	**0.58**	**0.69**	*0.45*	**0.46**	*0.45*

[22] Wang, Ning, et al, "Multi-cue correlation filters for robust visual tracking," in Proc. CVPR, 2018.

Comparision with Mix group

- Mix group included three high-performance experts and three low-performance experts.

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
ATOM	0.67	0.87	*0.60*	*0.81*	**0.62**	**0.82**	0.58	0.69	**0.51**	*0.51*	**0.52**
SiamRPN++	**0.69**	**0.90**	0.58	0.77	*0.60*	0.80	**0.60**	**0.74**	0.49	0.51	*0.50*
SPM	0.67	0.87	0.58	0.79	0.59	0.77	0.57	0.67	0.47	0.48	0.48
MemTrack	0.63	0.82	0.54	0.74	0.49	0.70	0.50	0.61	0.34	0.35	0.39
SiamFC	0.59	0.78	0.52	0.70	0.51	0.74	0.51	0.60	0.35	0.36	0.33
Staple	0.60	0.79	0.51	0.68	0.45	0.64	0.41	0.48	0.24	0.23	0.30
MCCT[22]	0.60	0.78	0.50	0.66	0.53	0.71	0.53	0.65	0.36	0.38	0.33
Random	0.64	0.84	0.55	0.75	0.54	0.74	0.53	0.63	0.40	0.41	0.42
Max	0.65	0.84	0.56	0.76	0.55	0.76	0.57	0.68	0.40	0.42	0.43
AAA	*0.68*	*0.89*	**0.62**	**0.83**	0.60	*0.81*	*0.59*	*0.71*	*0.51*	**0.53**	0.49

[22] Wang, Ning, et al, "Multi-cue correlation filters for robust visual tracking," in Proc. CVPR, 2018.

Offline tracking results with Mix group

- In some cases, the performance of an offline tracker was worse than that of an expert.

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
ATOM	0.69	0.89	*0.61*	*0.81*	**0.66**	**0.87**	0.61	0.73	*0.55*	*0.55*	**0.54**
SiamRPN++	**0.70**	**0.92**	0.58	0.78	0.64	0.84	*0.64*	**0.78**	0.52	0.54	*0.52*
SPM	0.67	0.88	0.58	0.80	0.63	0.82	0.60	0.70	0.51	0.52	0.51
MemTrack	0.64	0.84	0.55	0.75	0.54	0.75	0.54	0.66	0.38	0.40	0.42
SiamFC	0.60	0.79	0.53	0.72	0.55	0.78	0.56	0.66	0.39	0.40	0.37
Staple	0.62	0.81	0.52	0.69	0.50	0.70	0.47	0.56	0.28	0.28	0.35
AAA	*0.69*	*0.90*	**0.63**	**0.84**	*0.65*	*0.87*	**0.64**	*0.77*	**0.55**	**0.58**	0.51

Average regret with Mix group

- However, the regret of the proposed method was still smaller than that of any expert, and it can be seen that the regret was minimized.

Tracker	OTB2015	TColor128	UAV123	NFS	LaSOT	VOT2018
ATOM	32.36	*24.73*	*65.72*	479.28	281.13	*41.65*
SiamRPN++	*27.21*	42.01	84.93	*279.48*	*220.56*	42.34
SPM	30.20	31.36	72.44	440.12	266.94	47.40
MemTrack	45.66	47.10	123.54	488.33	414.40	53.20
SiamFC	63.63	58.49	109.16	463.40	440.14	76.74
Staple	66.24	64.67	147.94	855.97	753.89	94.19
AAA	**-31.25**	**-23.75**	**-31.27**	**-229.27**	**-98.45**	**-16.80**

Comparision with SiamDW[10]

- we implemented experts using several parameter sets (e.g., backbone networks, the weight of the network, and hyper-parameters) in SiamDW[10]

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
SiamDW_SiamFCRes22	0.64	0.84	*0.58*	*0.79*	0.51	0.73	0.52	0.64	*0.38*	*0.39*	0.38
SiamDW_SiamFCIncep22	0.61	0.81	0.55	0.76	0.50	0.72	0.51	0.64	0.36	0.38	0.35
SiamDW_SiamFCNext22	0.62	0.82	0.57	0.76	0.49	0.71	0.51	0.63	0.37	0.38	0.32
SiamDW_SiamRPNRes22	**0.67**	**0.91**	0.53	0.75	0.46	0.69	0.50	0.63	0.35	0.34	0.37
SiamDW_SiamFCRes22_VOT	0.63	0.84	0.56	0.77	0.51	0.73	0.52	0.65	0.36	0.37	0.37
SiamDW_SiamFCIncep22_VOT	0.60	0.80	0.54	0.75	0.50	0.73	0.49	0.61	0.35	0.36	0.35
SiamDW_SiamFCNext22_VOT	0.61	0.81	0.54	0.74	0.49	0.72	0.51	0.63	0.35	0.38	0.34
SiamDW_SiamRPNRes22_VOT	0.66	*0.90*	0.53	0.74	0.46	0.69	0.51	*0.66*	0.35	0.35	**0.43**
MCCT[22]	0.63	0.83	0.54	0.74	0.50	0.71	0.51	0.65	0.34	0.36	0.34
Random	0.63	0.84	0.55	0.76	0.49	0.71	0.51	0.64	0.36	0.37	0.37
Max	0.64	0.86	0.55	0.76	*0.51*	*0.74*	*0.53*	0.66	0.36	0.36	0.37
AAA	*0.66*	0.88	**0.60**	**0.82**	**0.52**	**0.75**	**0.55**	**0.68**	**0.42**	**0.43**	*0.42*

[10] Zhang, Zhipeng, and Houwen Peng, "Deeper and wider siamese networks for real-time visual tracking," in Proc. CVPR, 2019.
[22] Wang, Ning, et al, "Multi-cue correlation filters for robust visual tracking," in Proc. CVPR, 2018.

Comparision with SiamRPN++[5]

- we implemented experts using several parameter sets (e.g., backbone networks, the weight of the network, and hyper-parameters) in SiamRPN++[5]

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
SiamRPN++_AlexNet	0.66	0.87	0.57	0.77	0.58	0.77	0.54	0.65	0.45	0.45	0.47
SiamRPN++_AlexNet_OTB	0.66	0.86	0.55	0.75	0.58	0.78	0.54	0.66	0.43	0.43	0.45
SiamRPN++_ResNet-50	0.65	0.86	0.56	0.75	0.61	0.80	0.58	0.71	0.50	0.51	0.51
SiamRPN++_ResNet-50_OTB	**0.69**	**0.90**	*0.58*	0.77	0.60	0.80	0.60	*0.74*	0.49	0.51	0.50
SiamRPN++_ResNet-50_LT	0.63	0.84	0.58	*0.79*	0.61	*0.81*	0.56	0.68	*0.52*	*0.54*	*0.51*
SiamRPN++_MobileNetV2	0.65	0.86	0.56	0.76	0.60	0.79	0.57	0.70	0.45	0.46	0.50
SiamRPN++_SiamMask	0.65	0.85	0.54	0.73	0.60	0.80	0.58	0.72	0.47	0.48	0.48
MCCT [22]	0.64	0.84	0.55	0.75	*0.61*	0.81	0.59	0.72	0.48	0.50	0.45
Random	0.66	0.86	0.56	0.76	0.60	0.79	0.57	0.69	0.47	0.48	0.49
Max	0.66	0.87	0.56	0.76	0.60	0.80	*0.60*	0.73	0.47	0.48	0.50
AAA	*0.68*	*0.89*	**0.61**	**0.83**	**0.64**	**0.85**	**0.61**	**0.74**	**0.54**	**0.56**	**0.52**

[5] Li, Bo, et al, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," in Proc. CVPR, 2019.
[22] Wang, Ning, et al, "Multi-cue correlation filters for robust visual tracking," in Proc. CVPR, 2018.

Accuracy in anchor frames

- We also evaluate whether the proposed method can actually detect the target location with high confidence in anchor frames

Tracker	OTB2015		TColor128		UAV123		NFS		LaSOT		VOT2018
	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AUC	DP	AO
ATOM	0.72	0.93	*0.66*	0.88	*0.71*	*0.91*	0.66	0.80	*0.66*	*0.70*	*0.61*
DaSiamRPN	0.70	0.92	0.61	0.84	0.68	0.88	0.64	0.79	0.59	0.62	0.57
SiamMCF	0.71	0.91	0.66	0.88	0.66	0.88	0.67	0.83	0.60	0.64	0.57
SiamRPN++	*0.73*	*0.94*	0.64	0.85	0.70	0.89	*0.68*	*0.84*	0.64	0.69	0.60
SPM	0.72	0.92	0.65	*0.89*	0.69	0.87	0.65	0.78	0.63	0.67	0.60
THOR	0.68	0.90	0.59	0.79	0.67	0.87	0.64	0.79	0.55	0.58	0.59
AAA	**0.74**	**0.94**	**0.70**	**0.92**	**0.72**	**0.93**	**0.70**	**0.87**	**0.70**	**0.76**	**0.64**