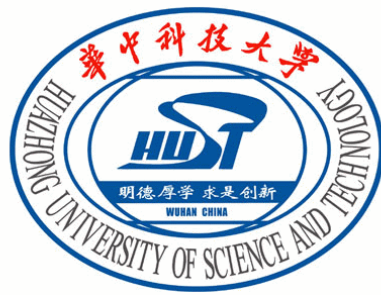


# ATP: Accurate Tracking by Progressively Refining

Bi Li<sup>1\*</sup>, Dejia Song<sup>1\*</sup>, LiangLiang Wang<sup>1\*</sup>, Xu Tang<sup>2</sup>, Chengquan Zhang<sup>2</sup>,  
Yang Liu<sup>2</sup>, Zihan Ni<sup>2</sup>, Shihu Li<sup>2</sup>, KangKang Wang<sup>2</sup>,  
Yu Zhou<sup>1</sup>, Xiang Bai<sup>1</sup>, Wenyu Liu<sup>1</sup>, Bin He<sup>2</sup>, Jingtuo Liu<sup>2</sup>

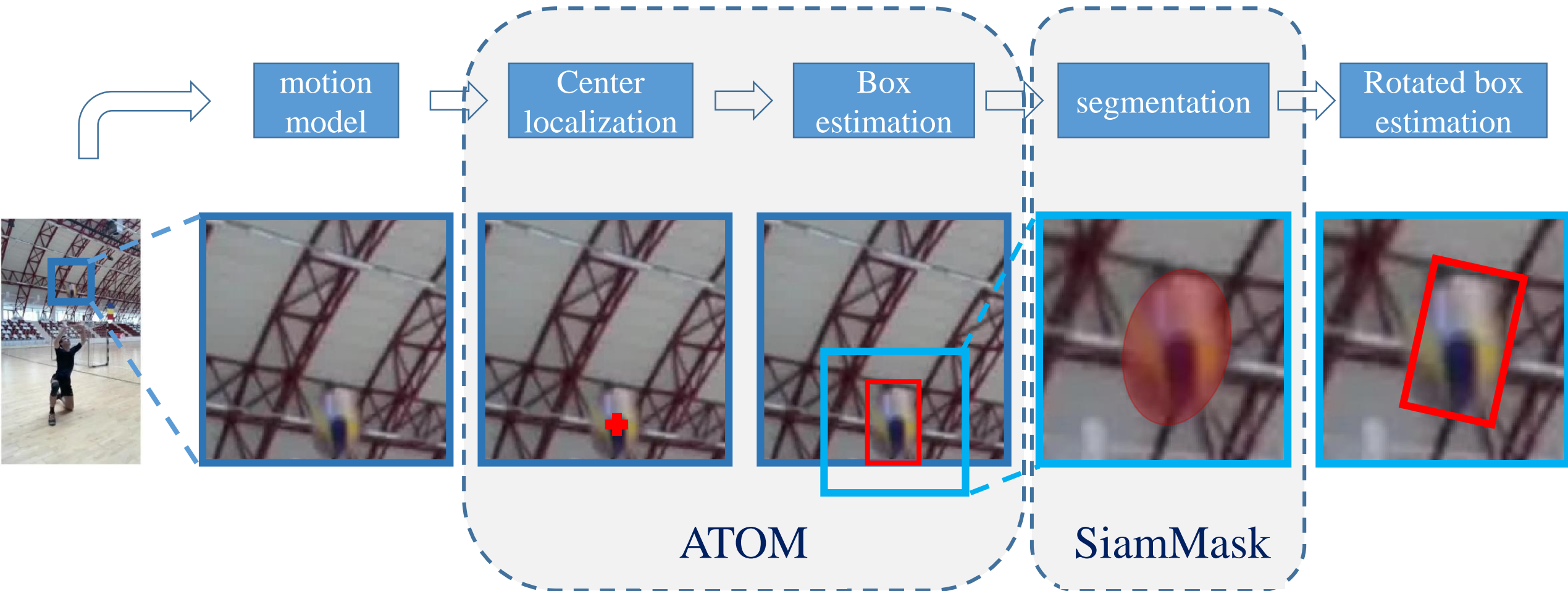
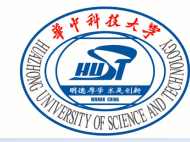
1. Huazhong University of Science and Technology
2. Baidu

\* interns of Baidu



- Tracking Pipeline
- Improvements
- Good and Bad Cases

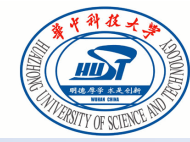
# Tracking Pipeline



Danelljan, Martin, et al. "Atom: Accurate tracking by overlap maximization." CVPR 2019.

Wang, Qiang, et al. "Fast online object tracking and segmentation: A unifying approach." CVPR 2019

# Improvements 1– backbone and training dataset

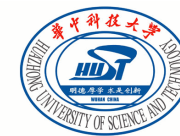


Name	Backbone	Training dataset	EAO on VOT2018
ATOM from Github	Resnet 18	VID	0.361
+ add LaSOT	Resnet 18	VID + LaSOT	0.375
+ add COCO	Resnet 18	VID + LaSOT + COCO	0.392
ATOM+Res50	Resnet 50	VID + LaSOT + COCO	0.432

## Remarks

- Using deeper backbones such as Resnet101, SE-Resnet50, InceptionV4, EfficientNet didn't bring gain.
- Adding more datasets such as GOT10K, TrackingNet, YoutubeBB didn't bring gain.

# Improvements 2– high and low level feature fusion



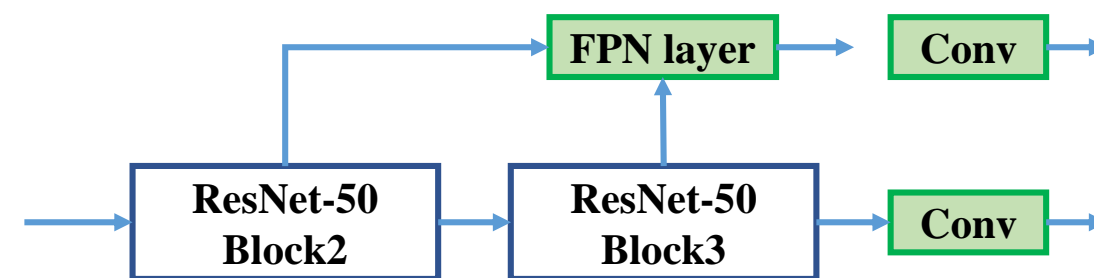
Name	Feature fusion	EAO on VOT2018
ATOM + Res50	NO	0.432
ATOM + Res50 + FPN	YES	0.443
DIMP (SOTA at the time)		0.440
SiamRPN++		0.415
SiamMask		0.423

Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." CVPR 2017.

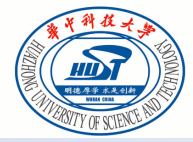
Bhat, Goutam, et al. "Learning Discriminative Model Prediction for Tracking." ICCV 2019

Li, Bo, et al. "Siamrpn++: Evolution of siamese visual tracking with very deep networks." CVPR2019

Wang, Qiang, et al. "Fast online object tracking and segmentation: A unifying approach." CVPR2019



# Improvements 3– handling fast moving objects

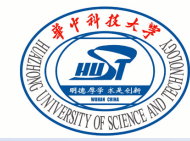


Name	EAO on VOT2019
ATOM+FPN	0.307
SiamRPN++	0.287
SiamMask	0.283



1. Due to fast motion, the target is on the edge of search region.
2. Static motion model penalizes large displacement from the center
3. The target is lost

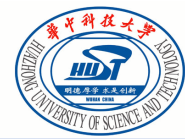
# Improvements 3– handling fast moving objects



Name	Fast motion strategy	EAO on VOT2019
ATOM+Res50+FPN	NONE	0.307
ATOM+Res50+FPN+Area	+ enlarge search image based on target size	0.331

```
ratio = (init_target_height * init_target_width) / (image_height * image_width)
if ratio < threshold:
    search_image_size = 'large_size'
else:
    search_image_size = 'default_size'
```

# Improvements 3– handling fast moving objects

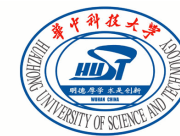


Name	Fast motion strategy	EAO on VOT2019
ATOM+FPN	NONE	0.307
	+ enlarge search image based on target size	0.331
ATOM+Res50+FPN+Area+Speed	+ enlarge search image based on target speed	0.357

```
speed = target_velocity_at_last_frame
if speed > threshold:
    search_image_size = 'large_size'
else:
    search_image_size = 'default_size'
```



# Improvements 4— increase feature dimension of classifier



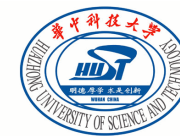
Name	Feature dim for center localization	EAO on VOT2019
ATOM+Res50+FPN+Area	64	0.331
ATOM+Res50+FPN+Area+dim256	256	0.343
ATOM+Res50+FPN+Area+dim512	512	0.348
<b>ATOM+Res50+FPN+Area+dim768</b>	<b>768</b>	<b>0.350</b>
ATOM+Res50+FPN+Area+dim1024	1024	0.345

ATOM reduces the feature dimension for center localization to 64.

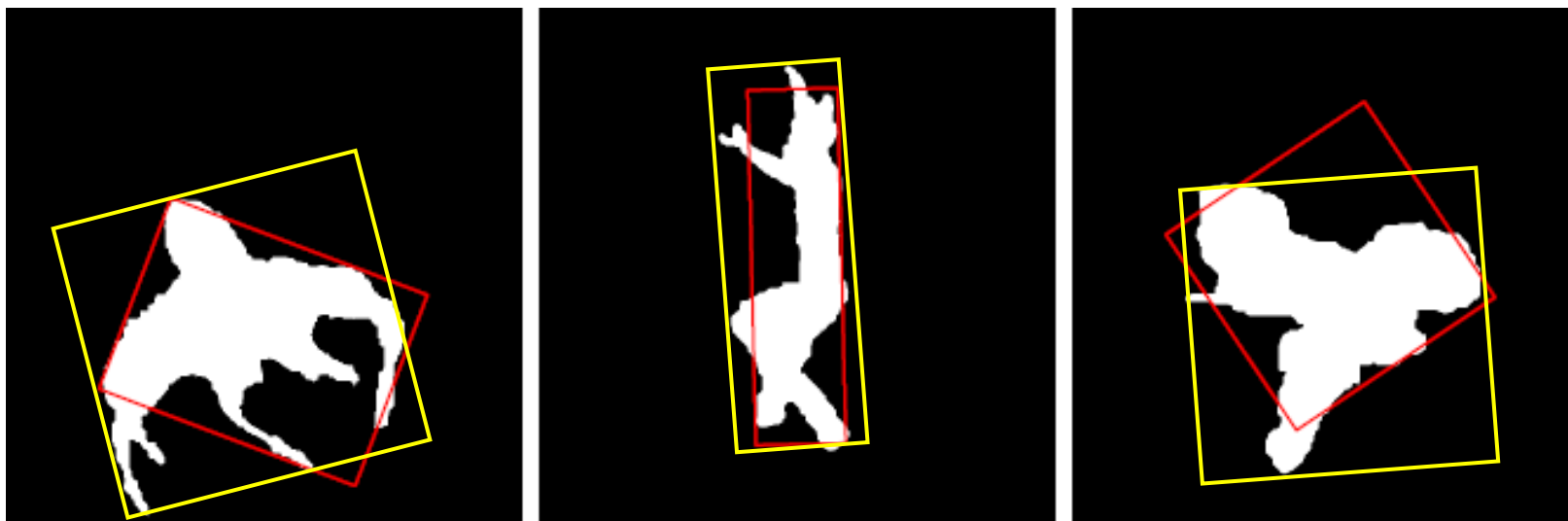
The representation power is hampered since the original dimension is 1024 for ResNet 50.

Therefore, we increase the feature dimension.

# Improvements 5– add rotated box from segmentation



Name	Generate rotated bbox	EAO on VOT2019	EAO on VOT2018
ATOM+Res50+FPN+Area+Speed+dim768	NONE	0.366	
ATOM+Res50+FPN+Area+Speed+dim768+MinAreaRect	cv2.minAreaRect	0.382	
<b>ATOM+Res50+FPN+Area+Speed+dim768+OptimalRect</b>	<b>Follow the GT generation procedures</b>	<b>0.394</b>	<b>0.491</b>

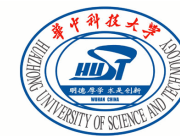


— optimal  
— min area

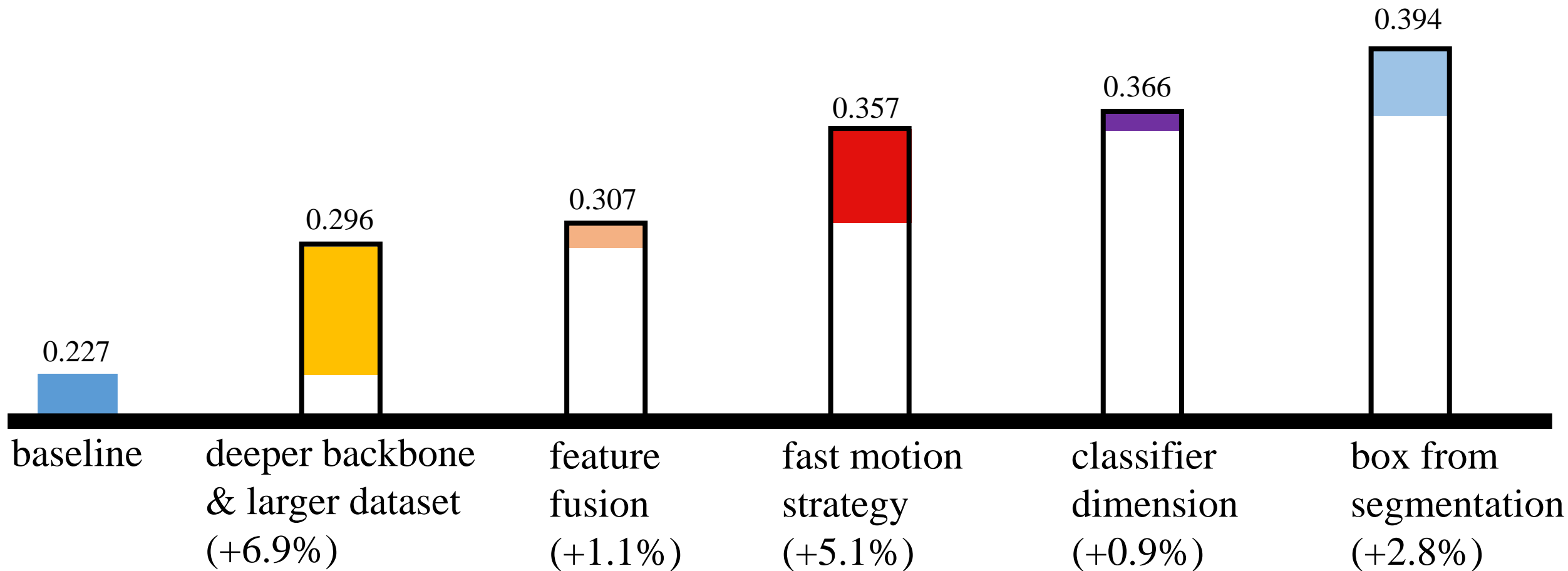
Kristan, Matej, et al. "The visual object tracking vot2017 challenge results." ICCV 2017.

[https://github.com/vojirt/bbox\\_from\\_segmentation](https://github.com/vojirt/bbox_from_segmentation)

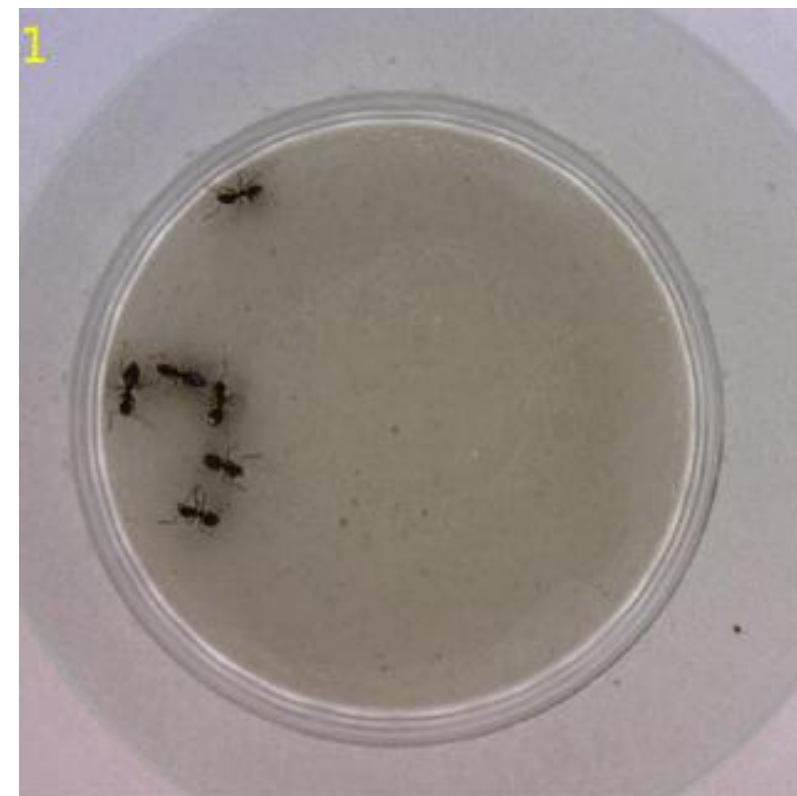
# Improvements – Summary



## VOT 2019



# Good cases



**Ours**



**ATOM**



# Bad cases – fast motion and occlusion is still challenging



fail 6 times



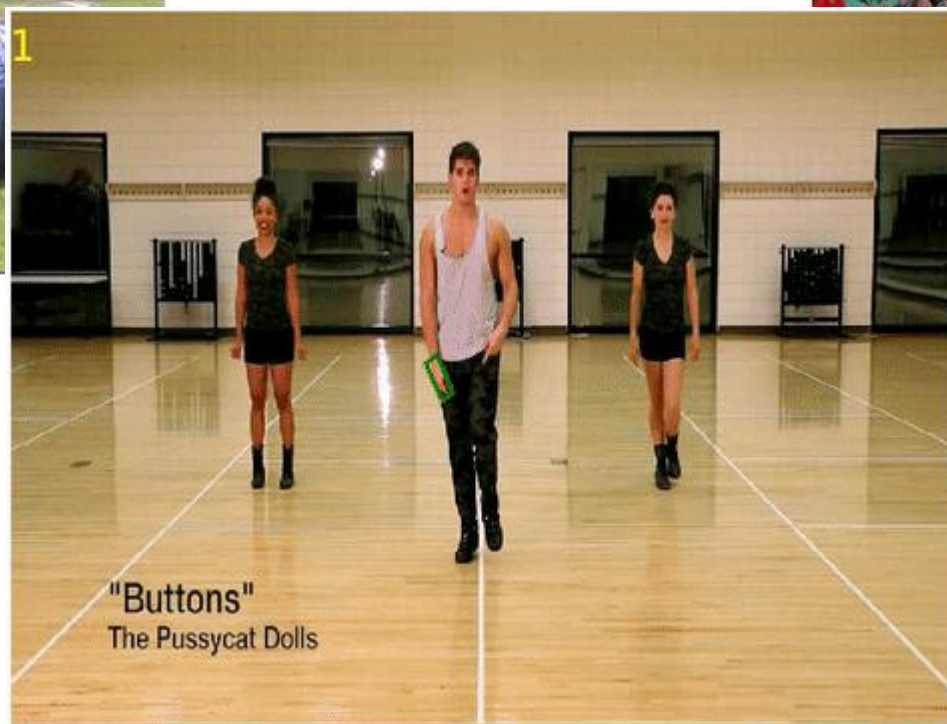
fail 8 times



Ours



GT



fail 8 times

## Code release

We are about to release a new tracking codebase with paddlepaddle including most state-of-the-art trackers:

- ATOM
- DiMP
- SiamMask
- SiamRPN
- SiamFC
- ...

## Recruitment

If you are skilled at face detection/tracking and looking for a job, please email us:

[tangxu02@baidu.com](mailto:tangxu02@baidu.com)

[hebin04@baidu.com](mailto:hebin04@baidu.com)

Questions?