GLOBALLY SPATIAL-TEMPORAL PERCEPTION: A LONG-TERM TRACKING SYSTEM

Zhenbang Li, Qiang Wang, Jin Gao, Bing Li, Weiming Hu
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, University of Chinese Academy of Sciences

Local search mechanism

- Pros
  - Works well if the target only has a small displacement between two adjacent frames.
  - Avoids interference from the distractors in the background.
- Cons
  - Causes irreversible cumulative errors if the predictions of the target positions in the previous frames drift away due to challenging illumination change, motion blur, etc.
  - Difficult for the local mechanism-based trackers to meet the needs of long-term tracking.

Globally Spatial-Temporal Perception tracking system

- Global perception mechanism
  - Is always able to perceive the target over the entire image
  - Even if the tracker makes a mistake due to the challenging target appearance variations, the target can still be retrieved in time once its appearance returns to normal
- Temporal motion model
  - Mitigates the interference of distractors
  - Predicts the target current position distribution using its historical trajectory information and current target appearance information

Globally Spatial-Temporal Perception tracking system

- We use an entire image instead of a small image patch as the input to the tracker to provide the global spatial information for it.

- In order to better perceive the global spatial information, we propose a two-stage tracking component, which is able to detect candidate targets that are visually similar to the ground truth target.

- To perceive the temporal information, we propose a motion model, which is able to exclude the distractors by predicting the location distribution to obtain the final tracking result.

Tracking component

- Feature extractor

$$\begin{aligned} f_z &= \phi_1(z) \\ f_x &= \phi_1(x) \end{aligned}$$

- Object feature

$$f_{obj} = \mathcal{R}(b_{obj}, f_z)$$

- Fusion feature

$$f_{corr} = f_{obj} * f_x$$

- RoI feature

$$f^i_{roi} = \mathcal{R}(b^i_{roi}, f_{corr})$$

Motion model

- Trajectory tensor

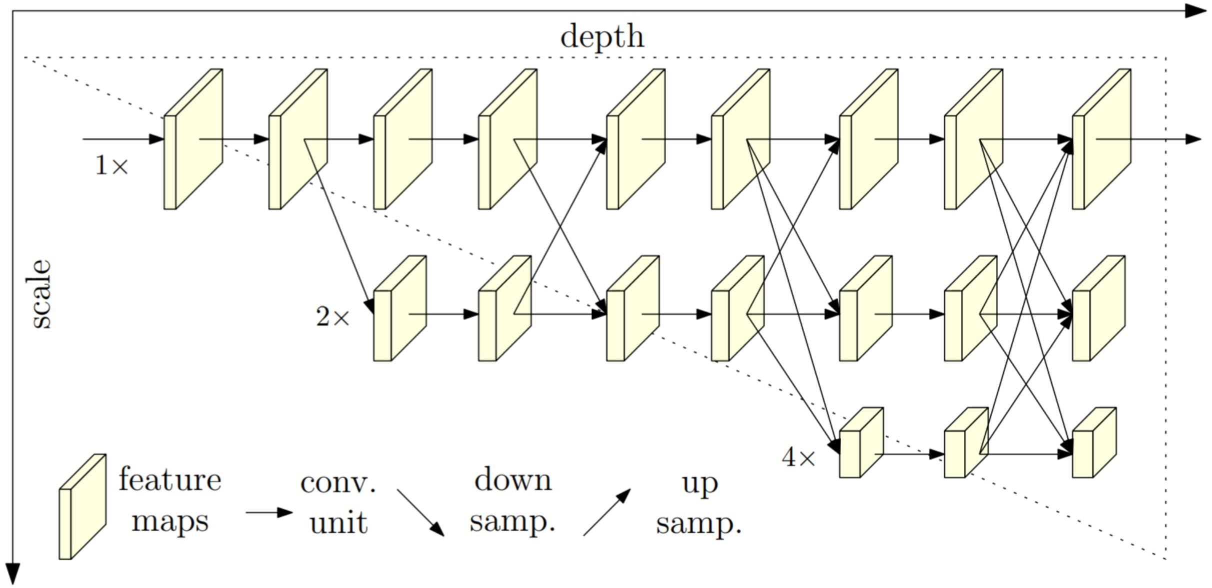
$$\mathcal{M}_t^k = \mathcal{C}(m_{t-k}, m_{t-k+1}, \ldots, m_{t-1})$$

- Enhanced trajectory tensor

$$\mathcal{N}_t^k = \mathcal{C}(\mathcal{I}_t, \mathcal{M}_t^k)$$

- Position distribution

$$\mathcal{O}_t^k = \phi_2(\mathcal{N}_t^k)$$



Ablation Studies

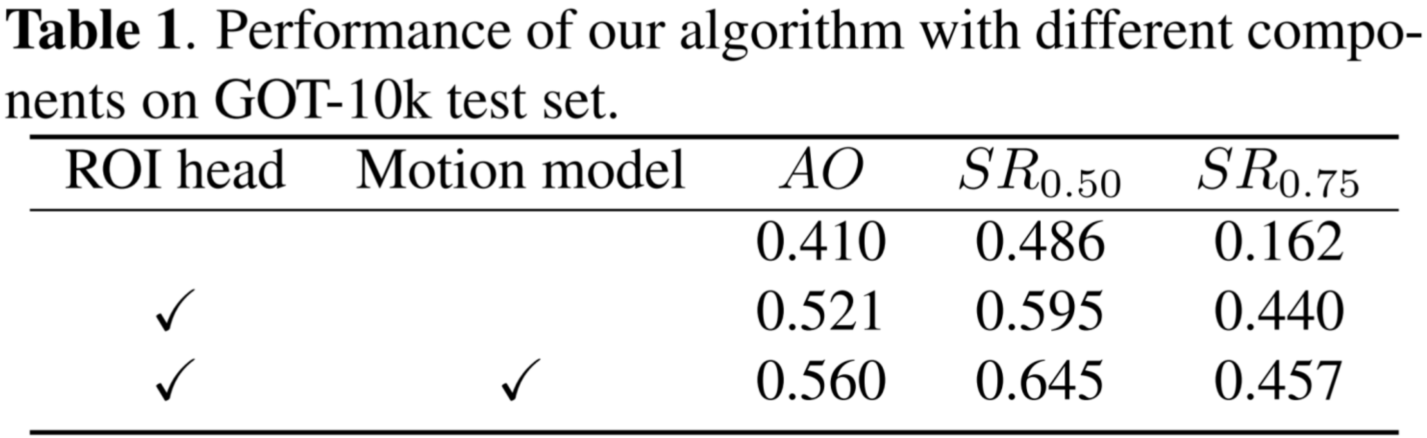


Table 1. Performance of our algorithm with different components on GOT-10k test set.

ROI head	Motion model	AO	$SR_{0.50}$	$SR_{0.75}$
		0.410	0.486	0.162
✓		0.521	0.595	0.440
✓	✓	0.560	0.645	0.457

Evaluation on GOT-10k Dataset

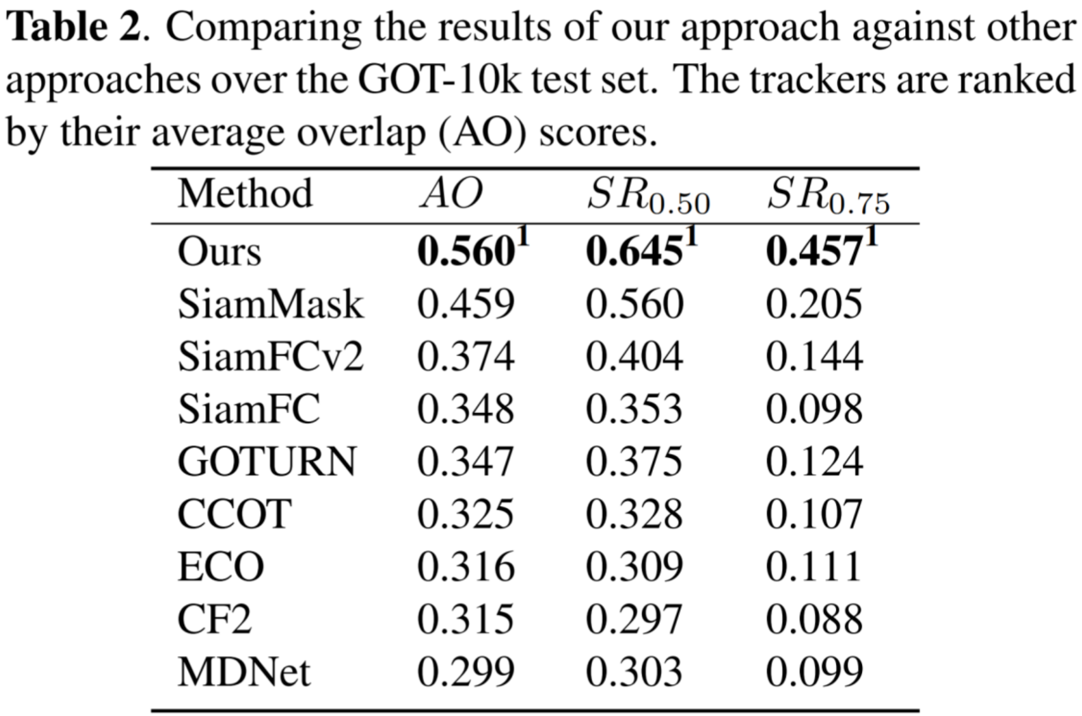
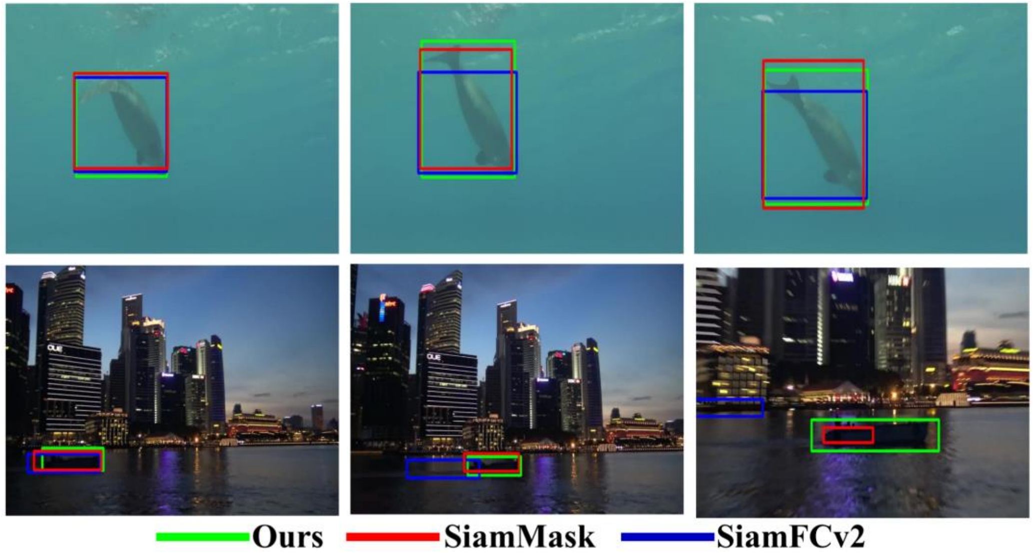


Table 2. Comparing the results of our approach against other approaches over the GOT-10k test set. The trackers are ranked by their average overlap (AO) scores.

Method	AO	$SR_{0.50}$	$SR_{0.75}$
Ours	**0.560**^{1}	**0.645**^{1}	**0.457**^{1}
SiamMask	0.459	0.560	0.205
SiamFCv2	0.374	0.404	0.144
SiamFC	0.348	0.353	0.098
GOTURN	0.347	0.375	0.124
CCOT	0.325	0.328	0.107
ECO	0.316	0.309	0.111
CF2	0.315	0.297	0.088
MDNet	0.299	0.303	0.099



Performance Analysis by Attributes

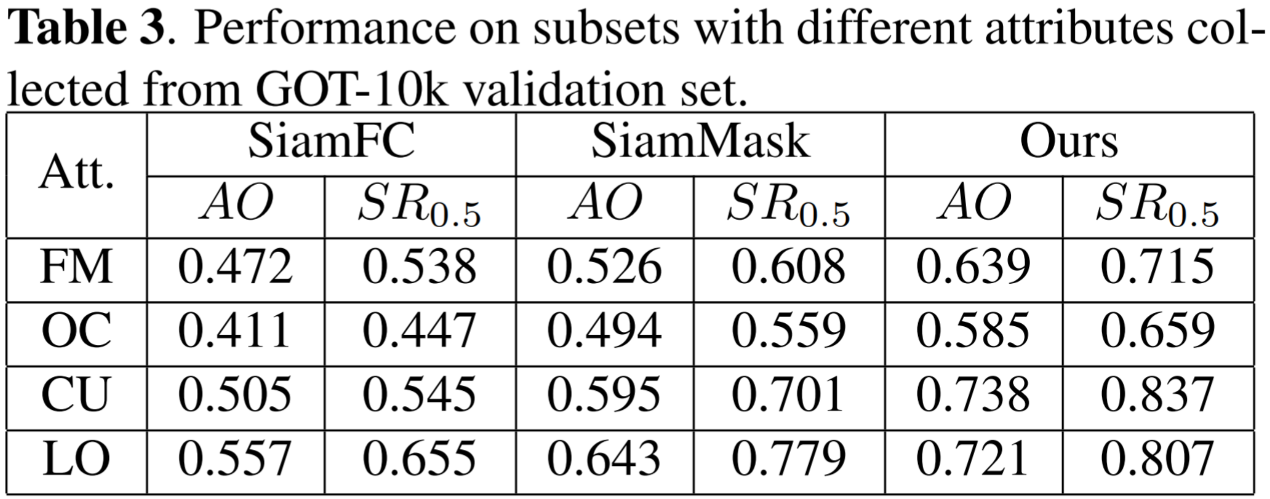


Table 3. Performance on subsets with different attributes collected from GOT-10k validation set.

Att.	SiamFC		SiamMask		Ours	
	AO	$SR_{0.5}$	AO	$SR_{0.5}$	AO	$SR_{0.5}$
FM	0.472	0.538	0.526	0.608	0.639	0.715
OC	0.411	0.447	0.494	0.559	0.585	0.659
CU	0.505	0.545	0.595	0.701	0.738	0.837
LO	0.557	0.655	0.643	0.779	0.721	0.807

Evaluation on UAV20L Dataset

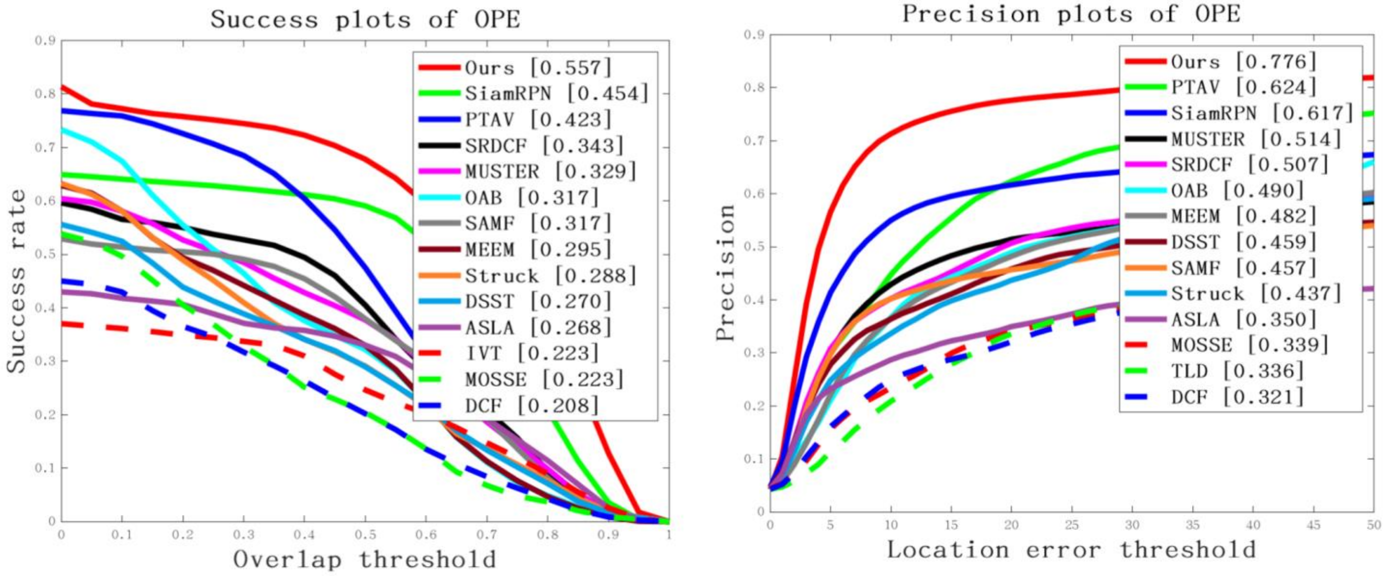


Fig. 2. Success and precision plots on UAV20L dataset.