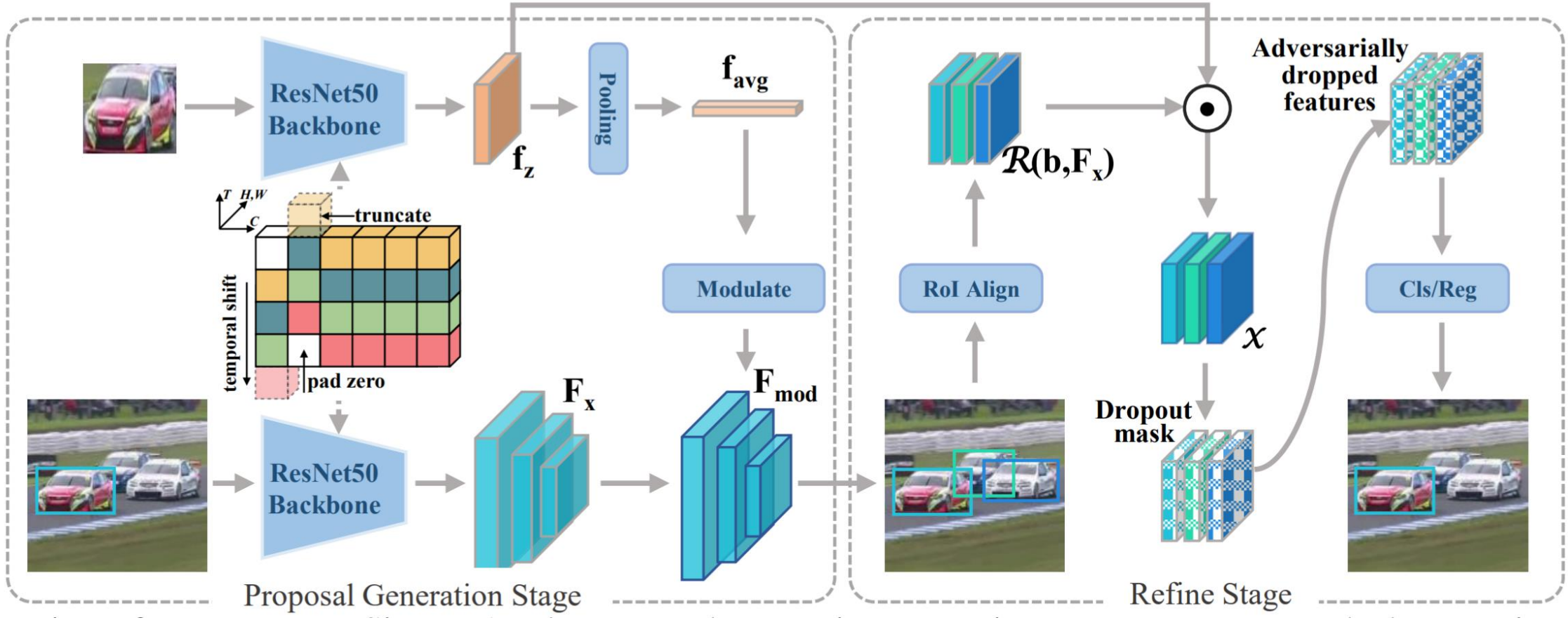
END-TO-END TEMPORAL FEATURE AGGREGATION FOR SIAMESE TRACKERS

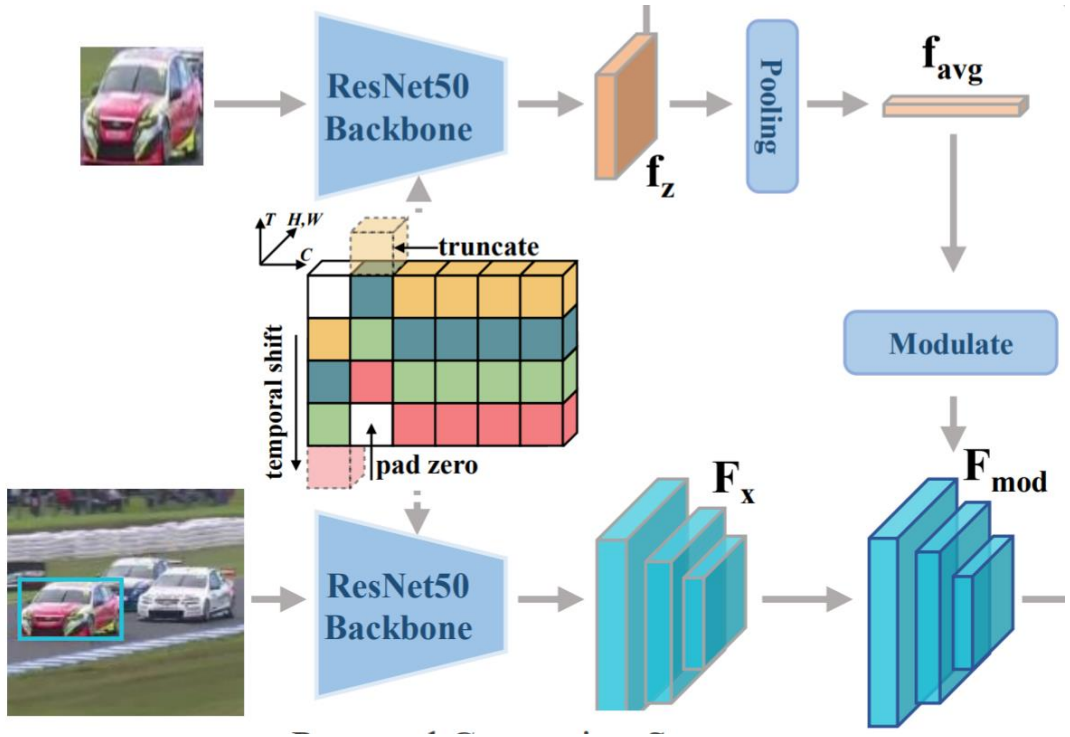
Zhenbang Li, Qiang Wang, Jin Gao, Bing Li, Weiming Hu
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, University of Chinese Academy of Sciences

Introduction



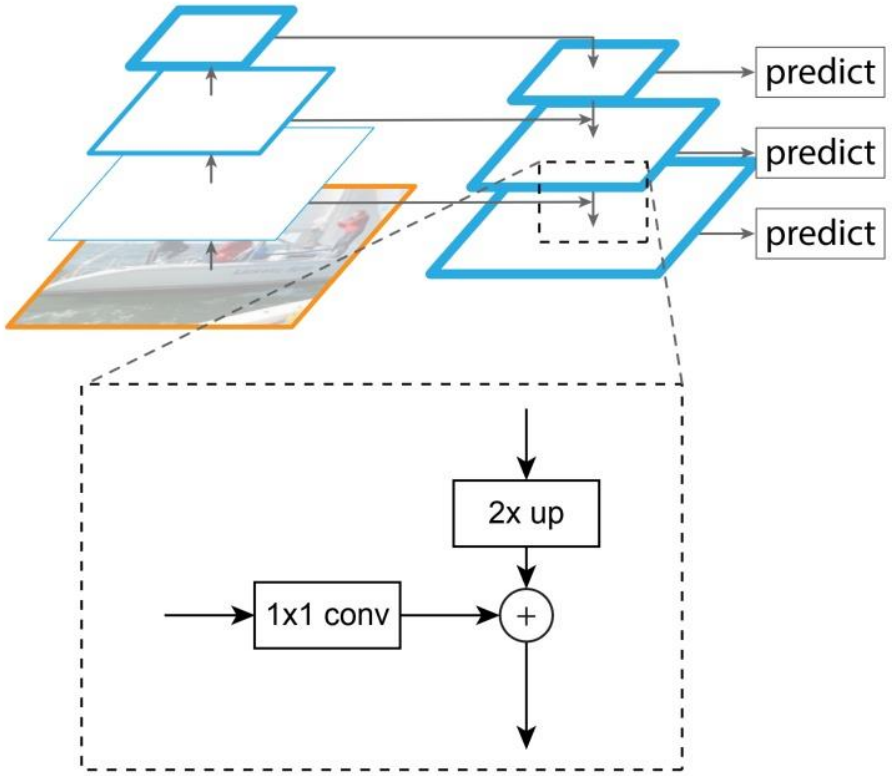
Proposal Generation Stage

- The proposal generation stage consists of 3 components:
  - (1) feature extractor
  - (2) temporal aggregation module
  - (3) feature modulation module



Feature extractor

- Res50-FPN
- Template feature: f_z
- Search feature pyramid: $F_x = \{f_x^i\}_{i=1:5}$



Temporal aggregation module

$$f^t \in \mathbb{R}^{C\times H\times W}$$

split

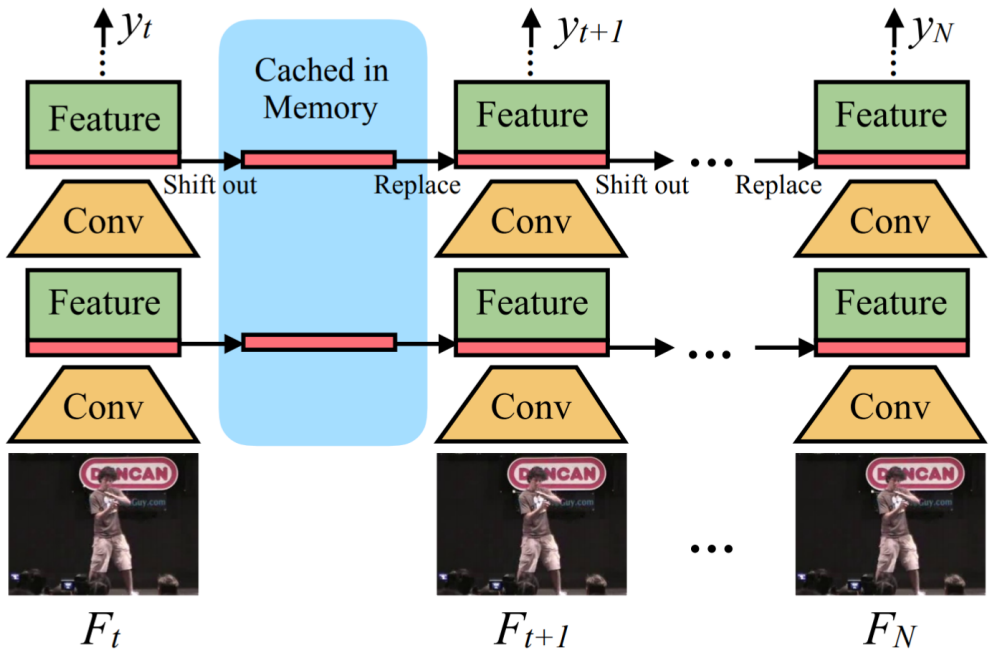
$$f^t_{1:K} \in \mathbb{R}^{K\times H\times W}$$

$$f^t_{(K+1):2K} \in \mathbb{R}^{K\times H\times W}$$

$$f^t_{(2K+1):C} \in \mathbb{R}^{(C-2K)\times H\times W}$$

aggregate

$$f^t_{agg} = \mathcal{C}(f^{t-1}_{1:K}, f^{t+1}_{(K+1):2K}, f^t_{(2K+1):C})$$



Refine Stage

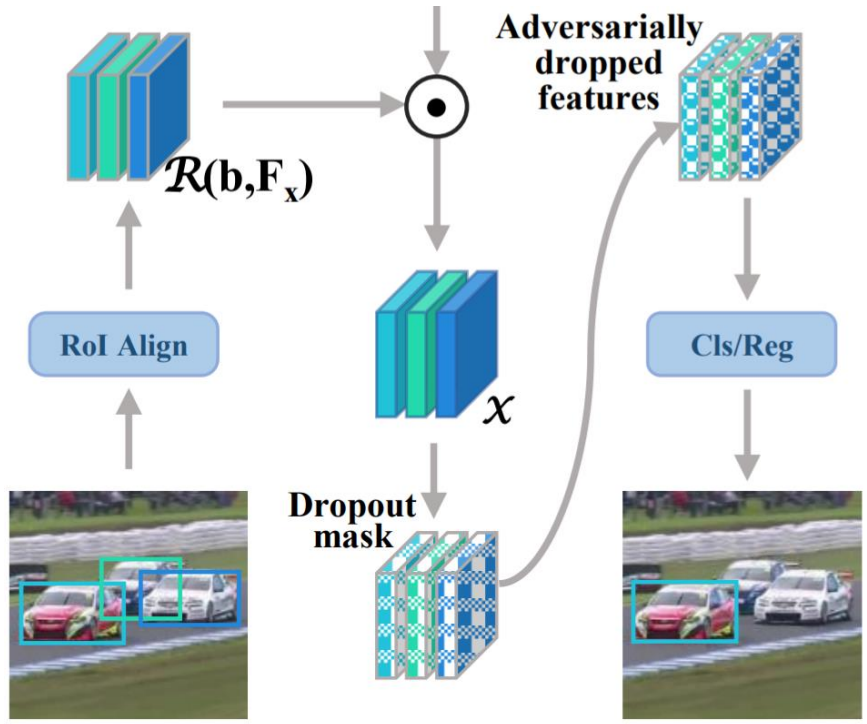
Generate the adversarial dropout mask:

$$\mathbf{m}^{adv} = \arg\max_{\mathbf{m}} D[h^{cls}(\mathcal{X} \odot \mathbf{m}^s), h^{cls}(\mathcal{X} \odot \mathbf{m})]$$

$$where\ ||\mathbf{m}^s - \mathbf{m}|| \leq \delta_e L,$$

Minimize the divergence between two predicted distribution:

$$\mathcal{L}_{adv} = \mathbb{E}[D_{KL}[h^{cls}(\mathcal{X} \odot \mathbf{m}^s)||h^{cls}(\mathcal{X} \odot \mathbf{m}^{adv}))]]$$



Ablation Studies

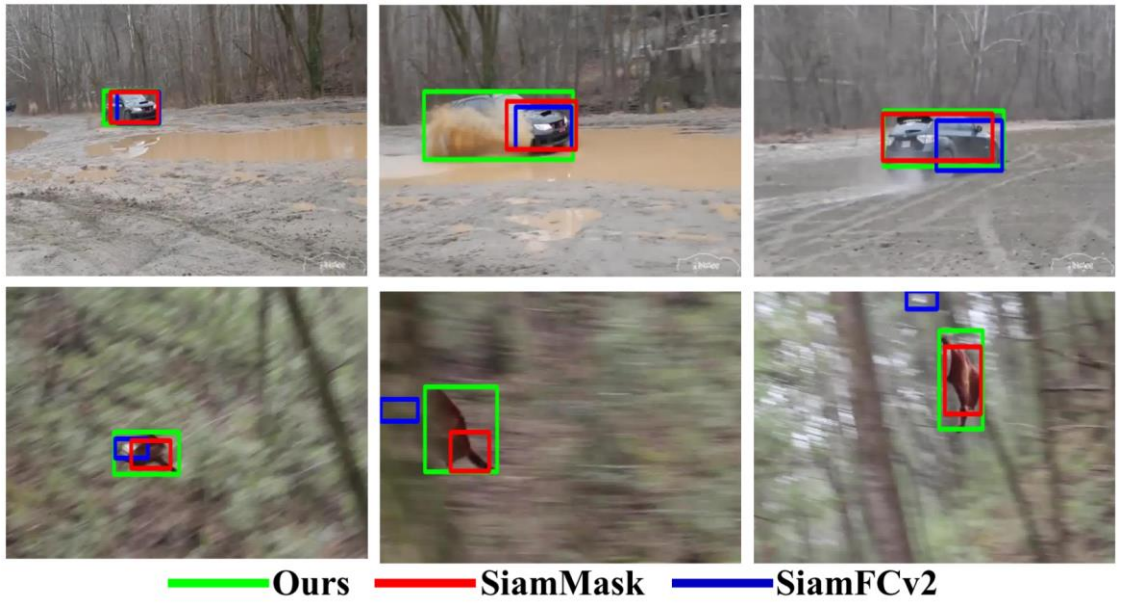
Table 1. Performance of our algorithm with different components on GOT-10k test set.

Temporal aggregation	Adversarial dropout	AO	$SR_{0.50}$	$SR_{0.75}$
		0.542	0.607	0.456
✓		0.561	0.645	0.480
✓	✓	0.577	0.662	0.509

Evaluation on GOT-10k Dataset

Table 2. Comparing the results of our approach against other approaches over the GOT-10k test set.

Method	AO	$SR_{0.50}$	$SR_{0.75}$
Ours	**0.577**[1]	**0.662**[1]	**0.509**[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



Evaluation on UAV20L Dataset

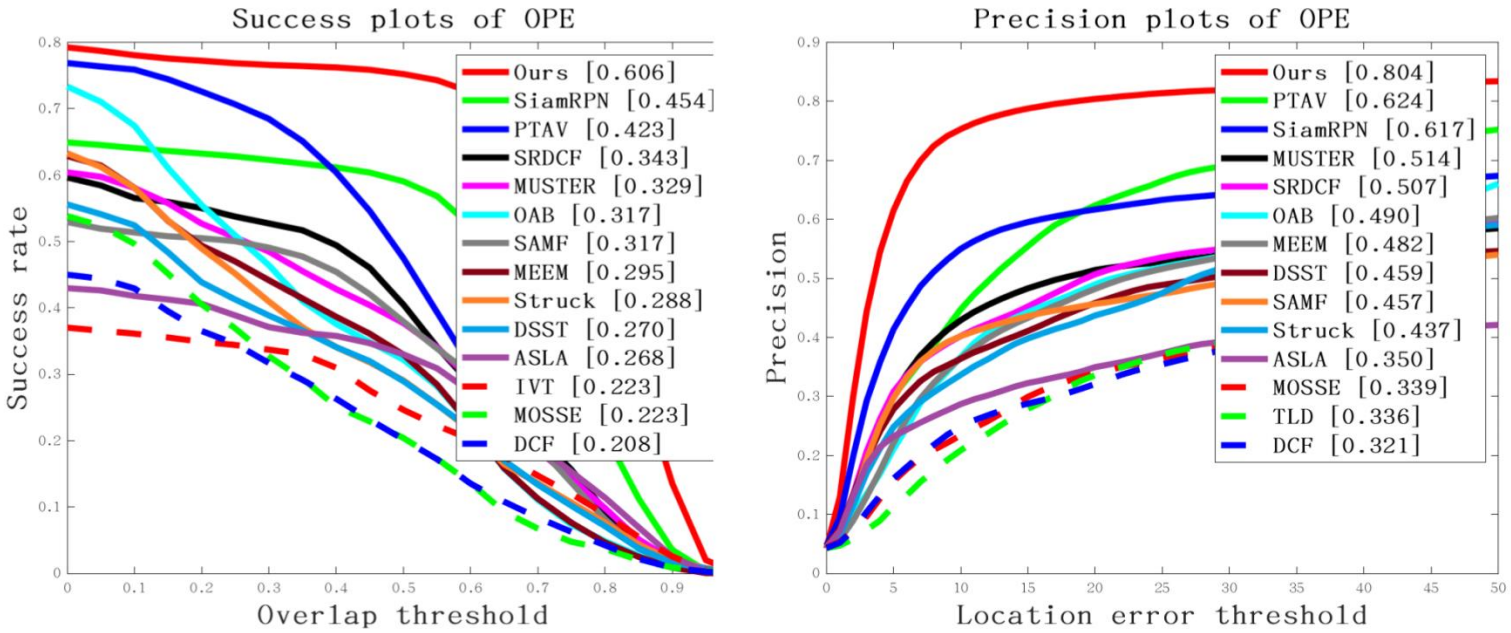


Fig. 3. Success and precision plots on UAV20L dataset.