VOT-ST2020 Winners Presentation

RPT: Learning Point Set Representation for Siamese Visual Tracking

Ziang Ma, Linyuan Wang, Haitao Zhang, Wei Lu, Jun Yin

Zhejiang Dahua Technology CO., LTD.

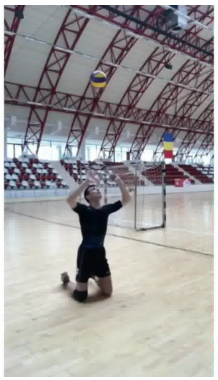
What is Visual Object Tracking?

Given the target region in the initial frame

Locate an arbitrary target of interest during a whole video sequence

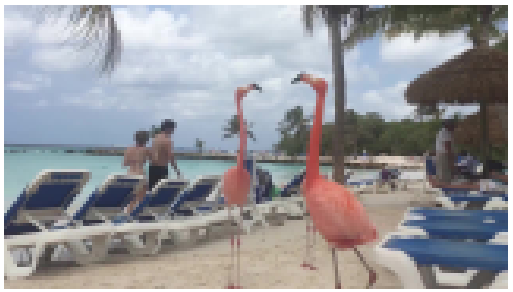
VOT2020 Challenges

Blur

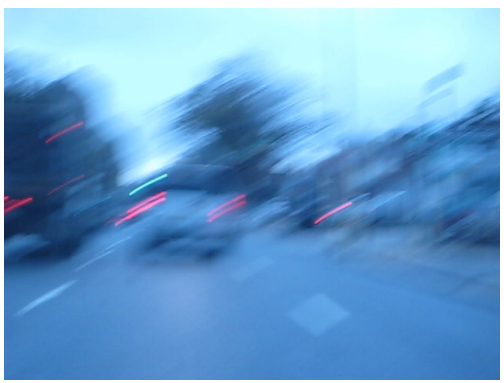


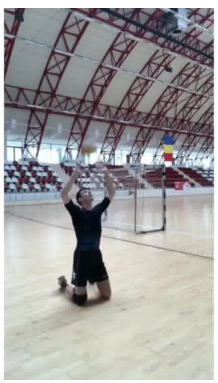
Large Motion

Posture Variations

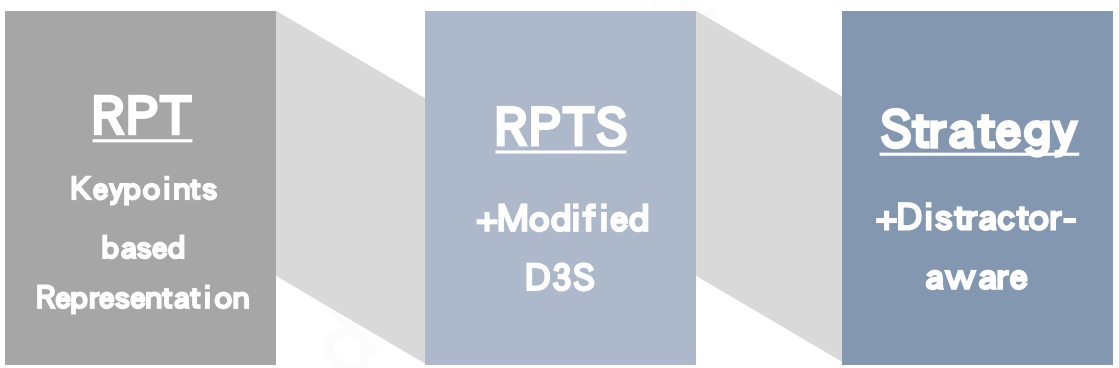


Occlusions



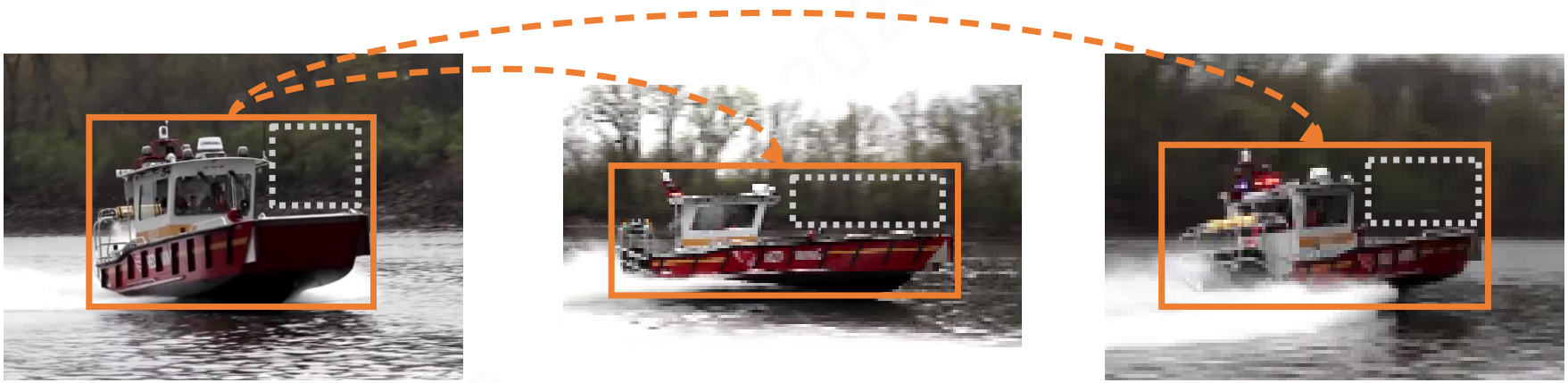


Content


RPT
Keypoints based Representation
RPTS
+Modified D3S
Strategy
+Distractor-aware

Bounding Box Representation

Problem: Redundant background & Geometric transformations



Point Set Representation

● Predicted Point Set ▭ Ground Truth

- ✔ Fine-grained localization
- ✔ Modeling of object appearance

Reppoints: Point set representation for object detection. In ICCV 2019

Point Set Representation

- Target Candidates

Initialization

$$R = \{(x_k, y_k)\}_{k=1}^{n}, x_k = i, y_k = j, k = 1, 2, ..., n$$

Point Set Representation

Target Candidates

Predicted Point Set

Refinement

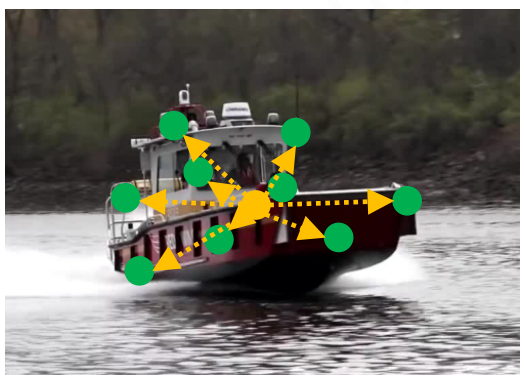
$$R_r = \{(x_k + \Delta x_k, y_k + \Delta y_k)\}_{k=1}^{n}$$

Relative Offsets $\{(\Delta x_k, \Delta y_k)\}_{k=1}^{n}$

Point Set Representation

Target Candidates

Predicted Point Set



Pseudo Box

Supervision

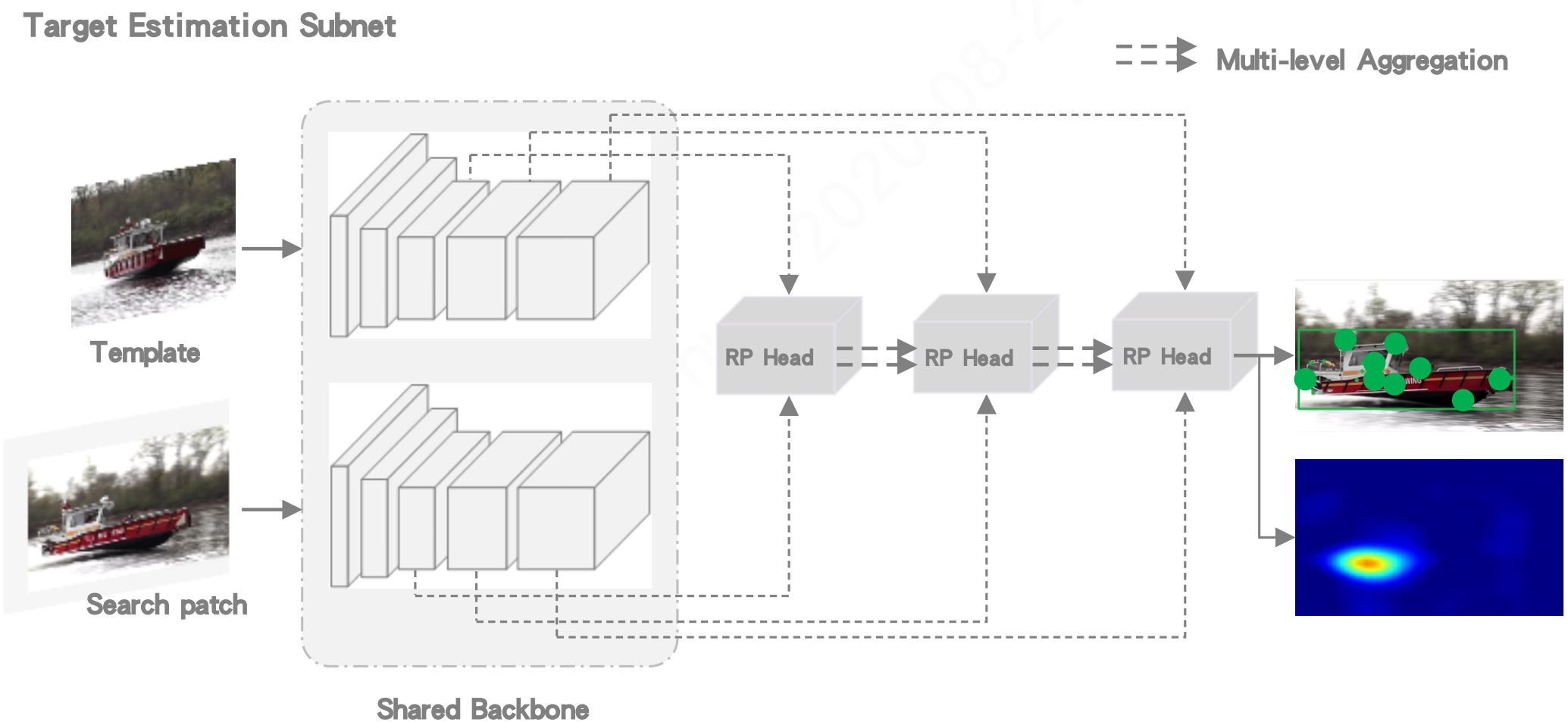
✔ *Pseudo* Box & IOU Loss

$$R_p = (\min\{x_k + \Delta x_k\}, \min\{y_k + \Delta y_k\}, \max\{x_k + \Delta x_k\}, \max\{y_k + \Delta y_k\})$$

Point Set Representation

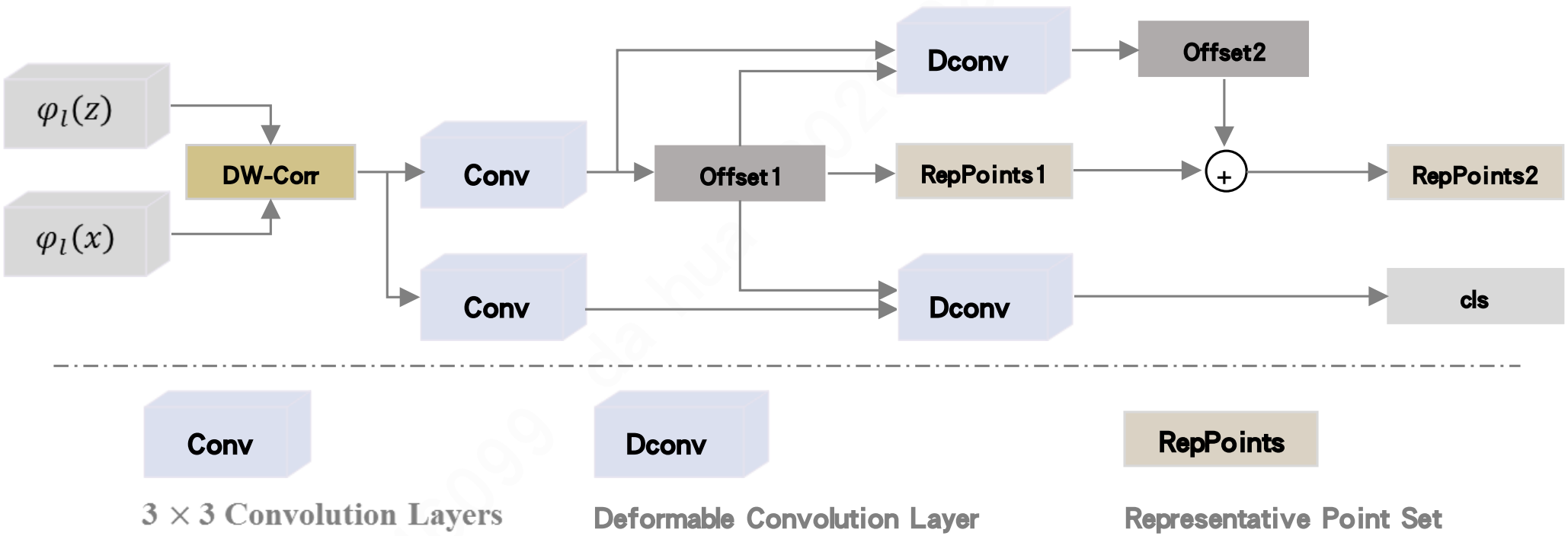
RPT Framework

Target Estimation Subnet



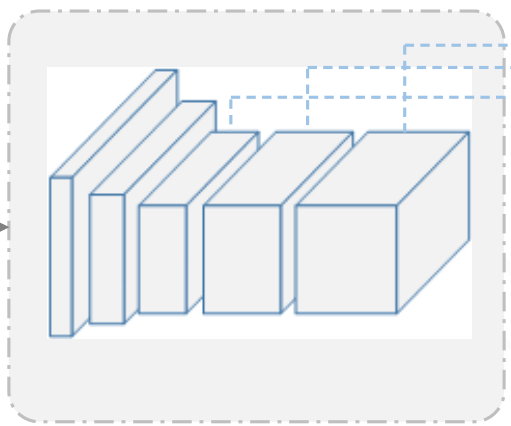
Target Estimation Subnet

RP Head



Online Classification Subnet

Training set



Shared Backbone

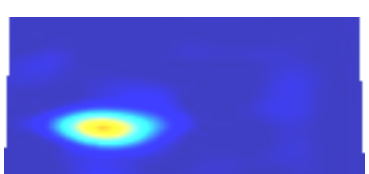
Classifier

Classifier

Classifier

Classification Head Output

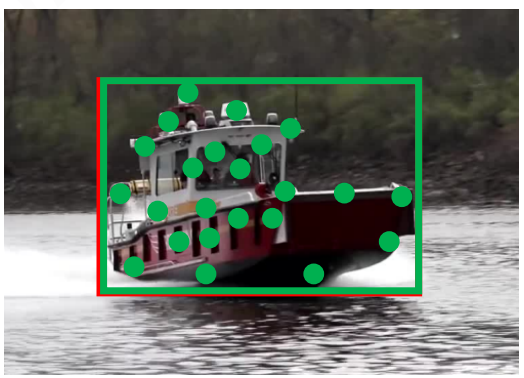
$$f_l = \alpha f^l_{online} + (1-\alpha) f^l_{offline}$$



Multi-level Aggregation

Predicted Point Set and *Pseudo* Box Ground Truth

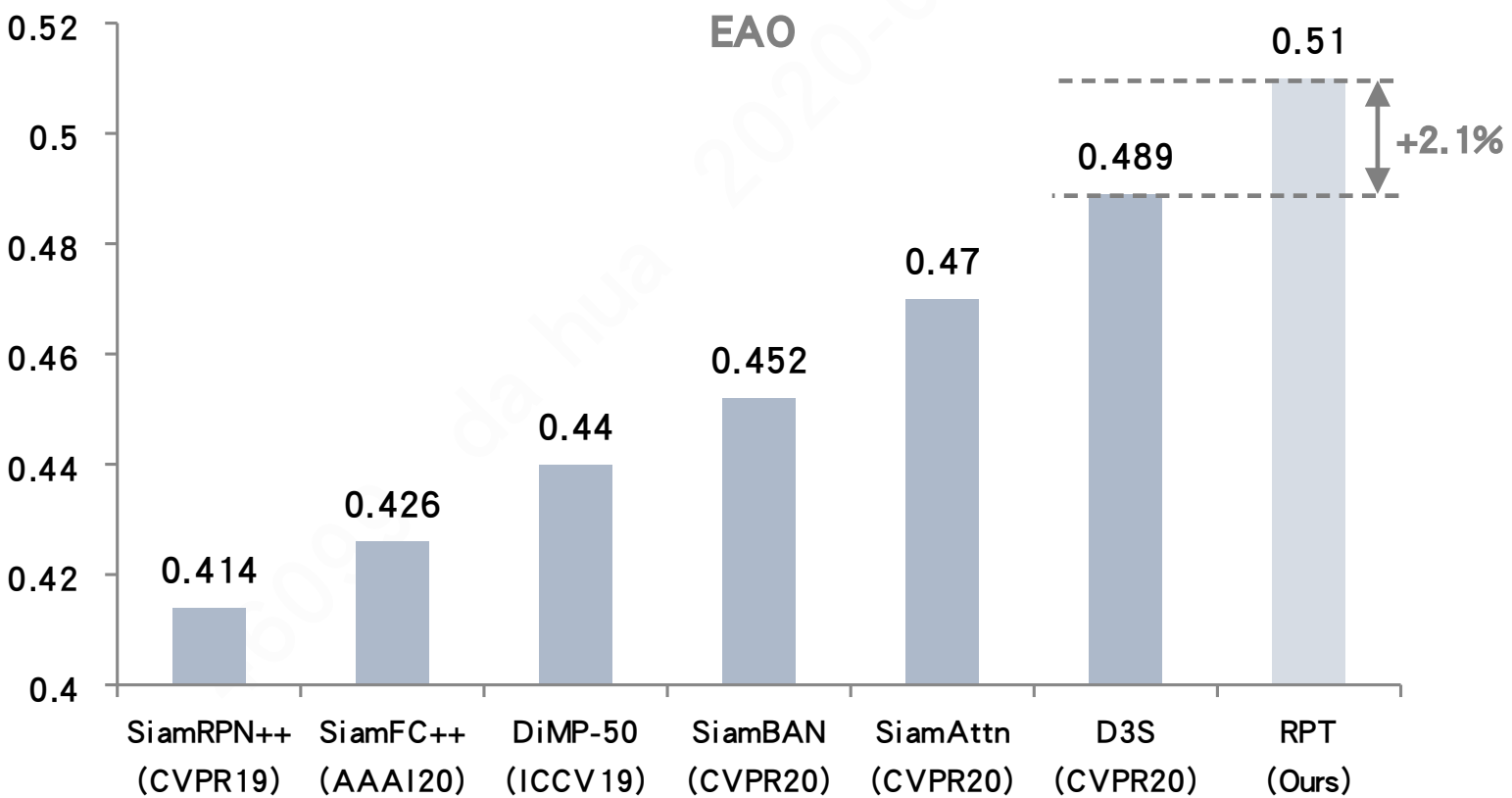
Sparse Point Set



Dense Point Set

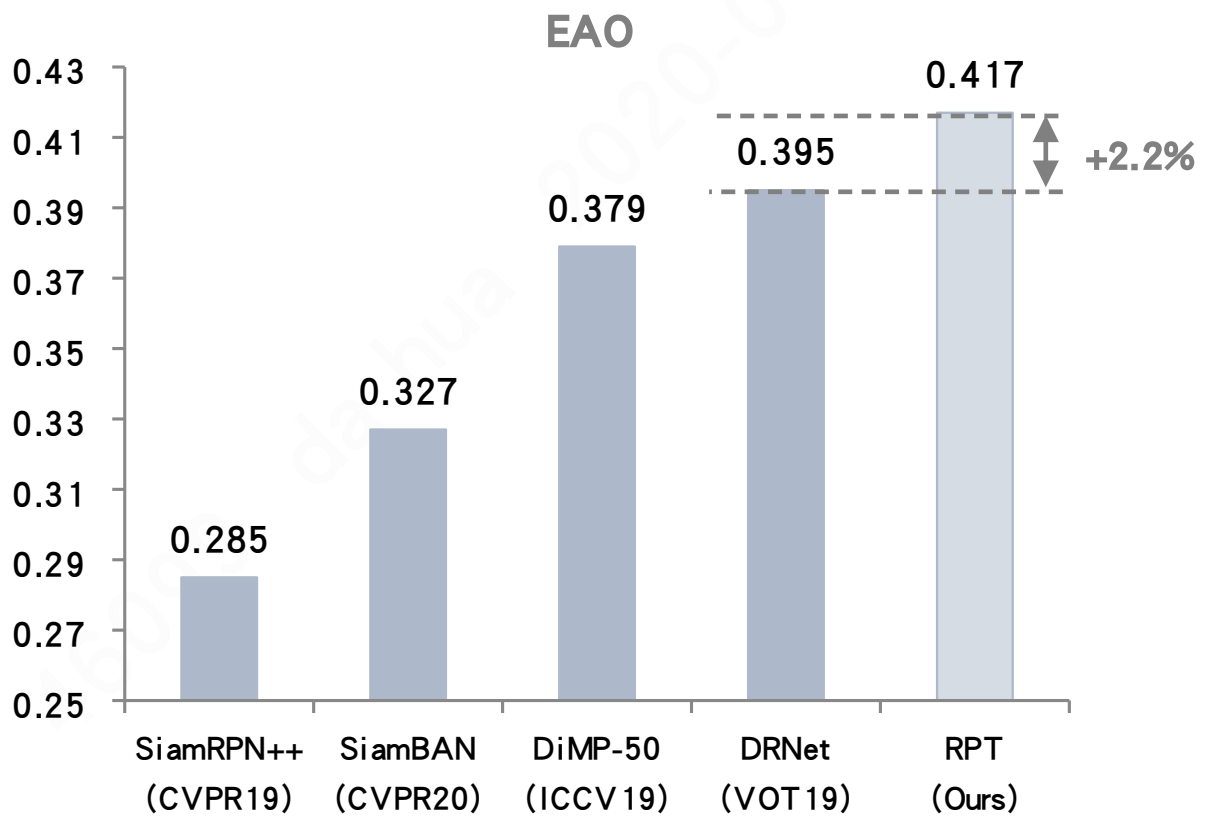
Comparison with SOTA

VOT2018



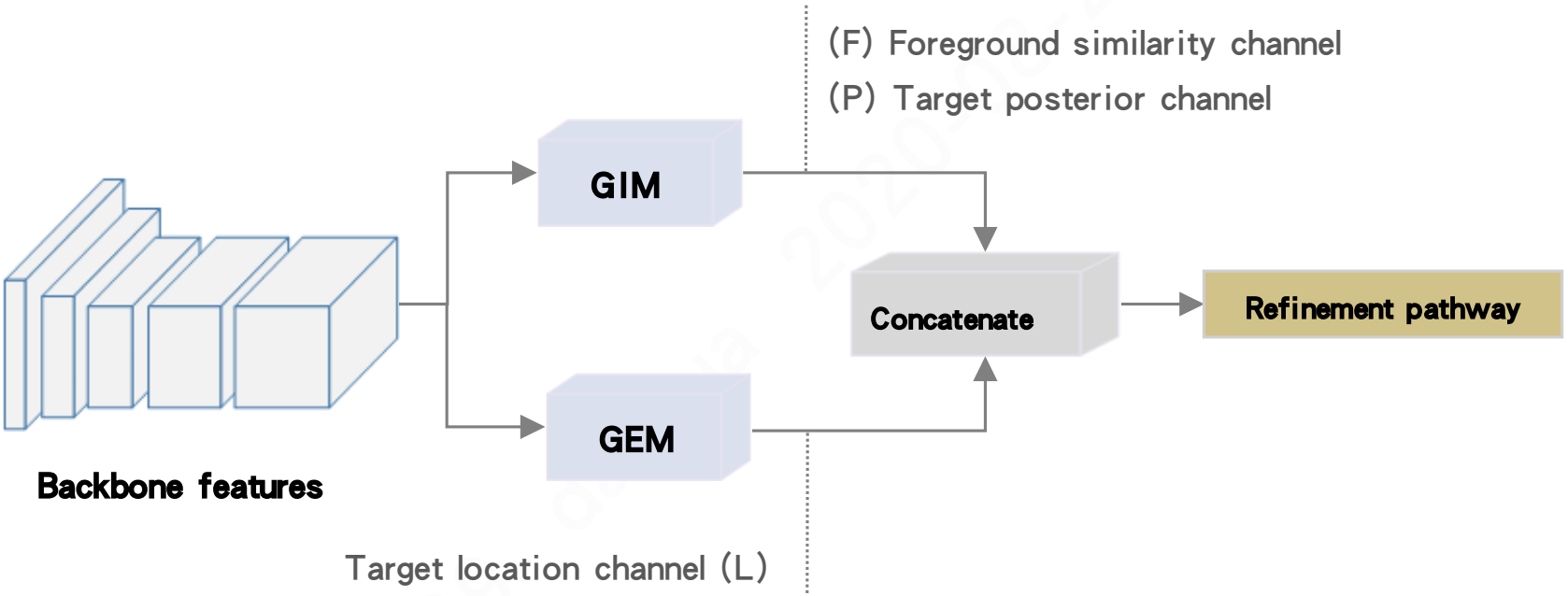
Comparison with SOTA

VOT2019



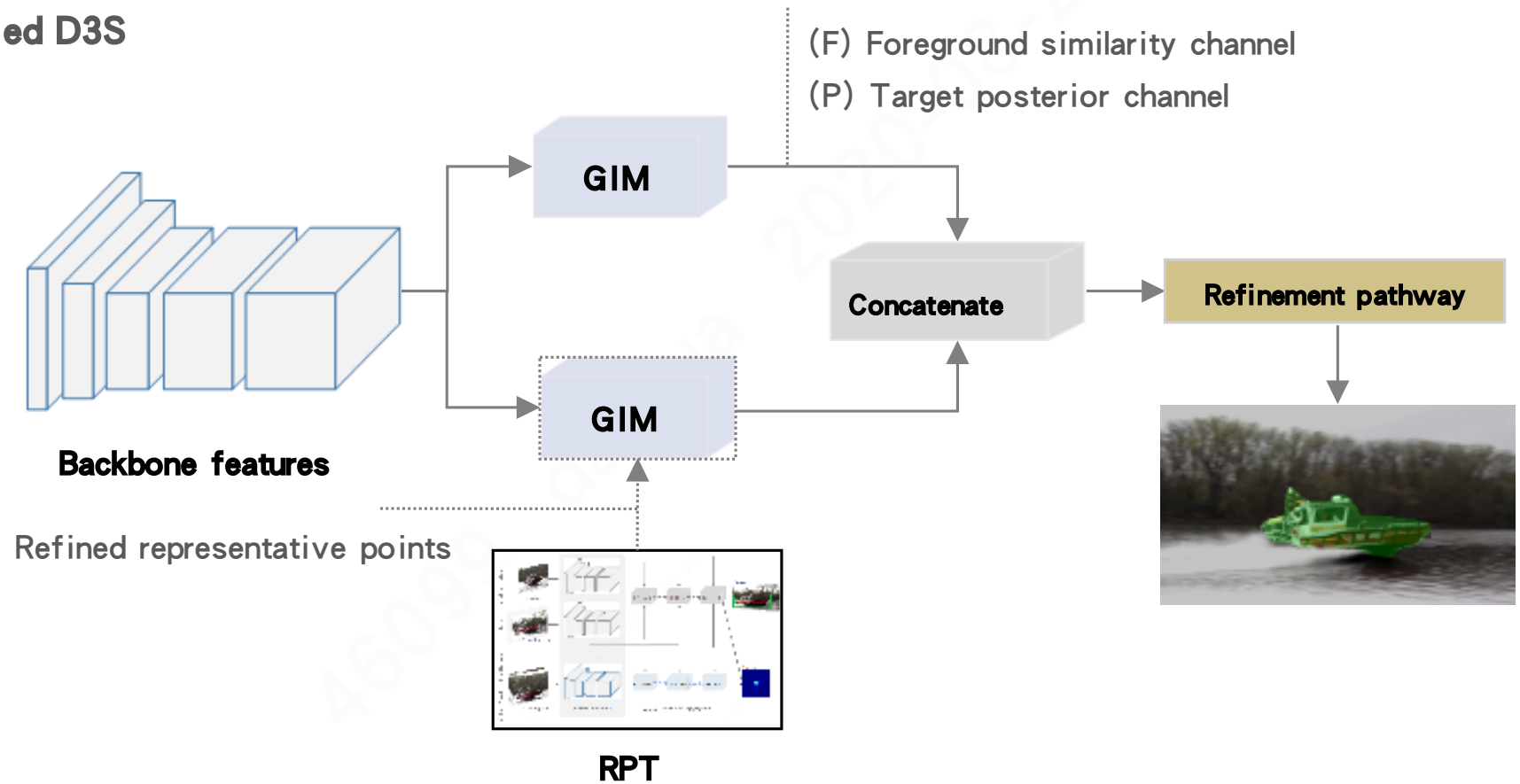
RPTS Framework

D3S

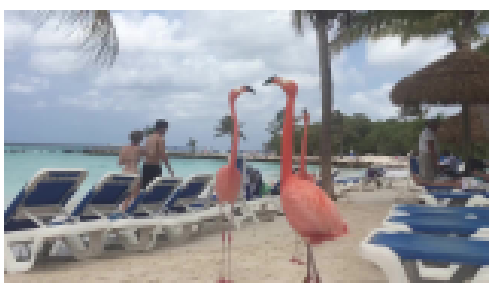


RPTS Framework

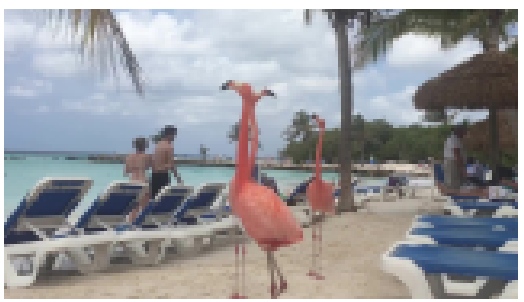
modified D3S



Distractor-aware Strategy








Improvement Details

VOT2020_Public Datasets

	Accuracy	Robustness	EAO
RPT	0.45632	0.79691	0.30817
RPTS (w/ modified D3S)	0.66204	0.86056	0.52556
RPTS (w/ tricks)	0.67949	0.86868	0.53946

「让社会更安全 让生活更智能」

ENABLING A SAFER SOCIETY AND SMARTER LIVING