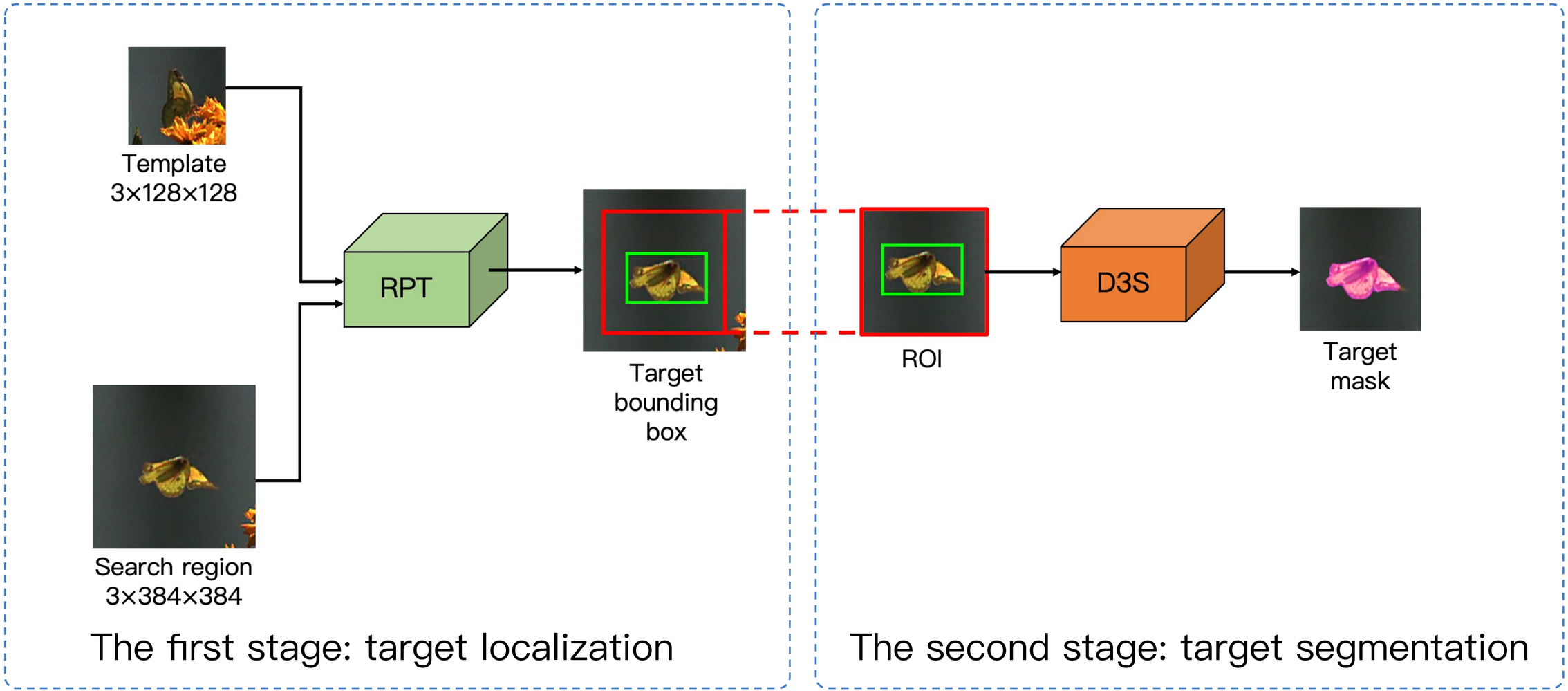
RPTMask

Winner of VOT2021 short-term challenge

Zhihong Fu[1], Liangliang Wang[2], Qili Deng[2], Kang Du[2], Min Zheng[2], Qingjie Liu[1]

[1] Beihang University, China
[2] ByteDance Inc. China

Pipeline of VOT2020-ST winner

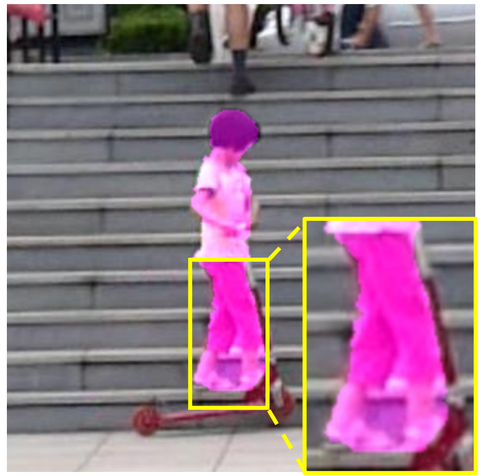


RPT: Learning Point Set Representation for Siamese Visual Tracking (ECCV 2020 Workshops)

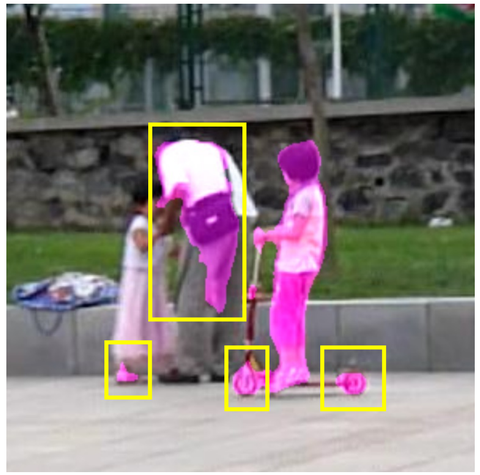
Motivation

baseline's results

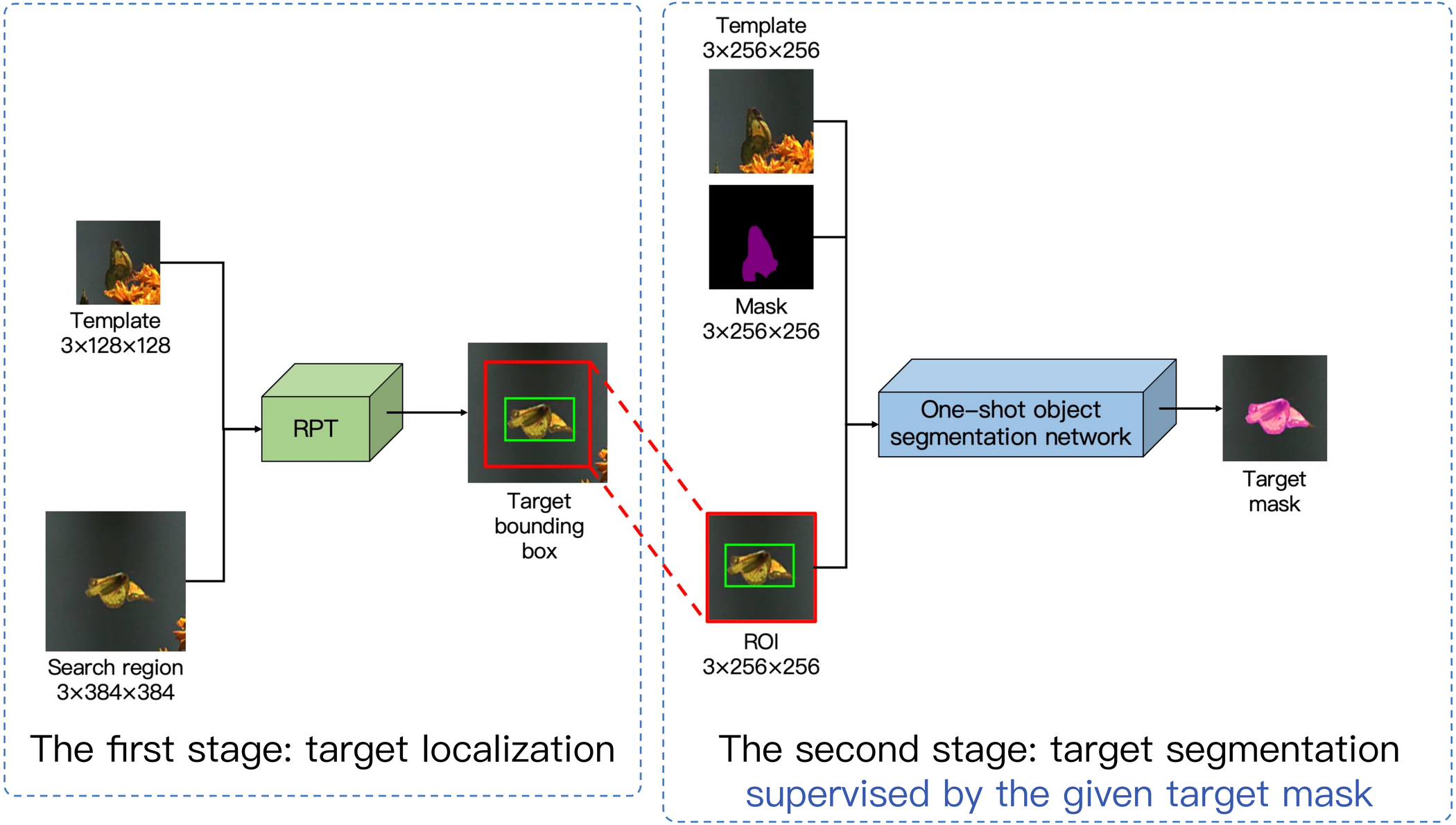
Poor detail segmentation ability

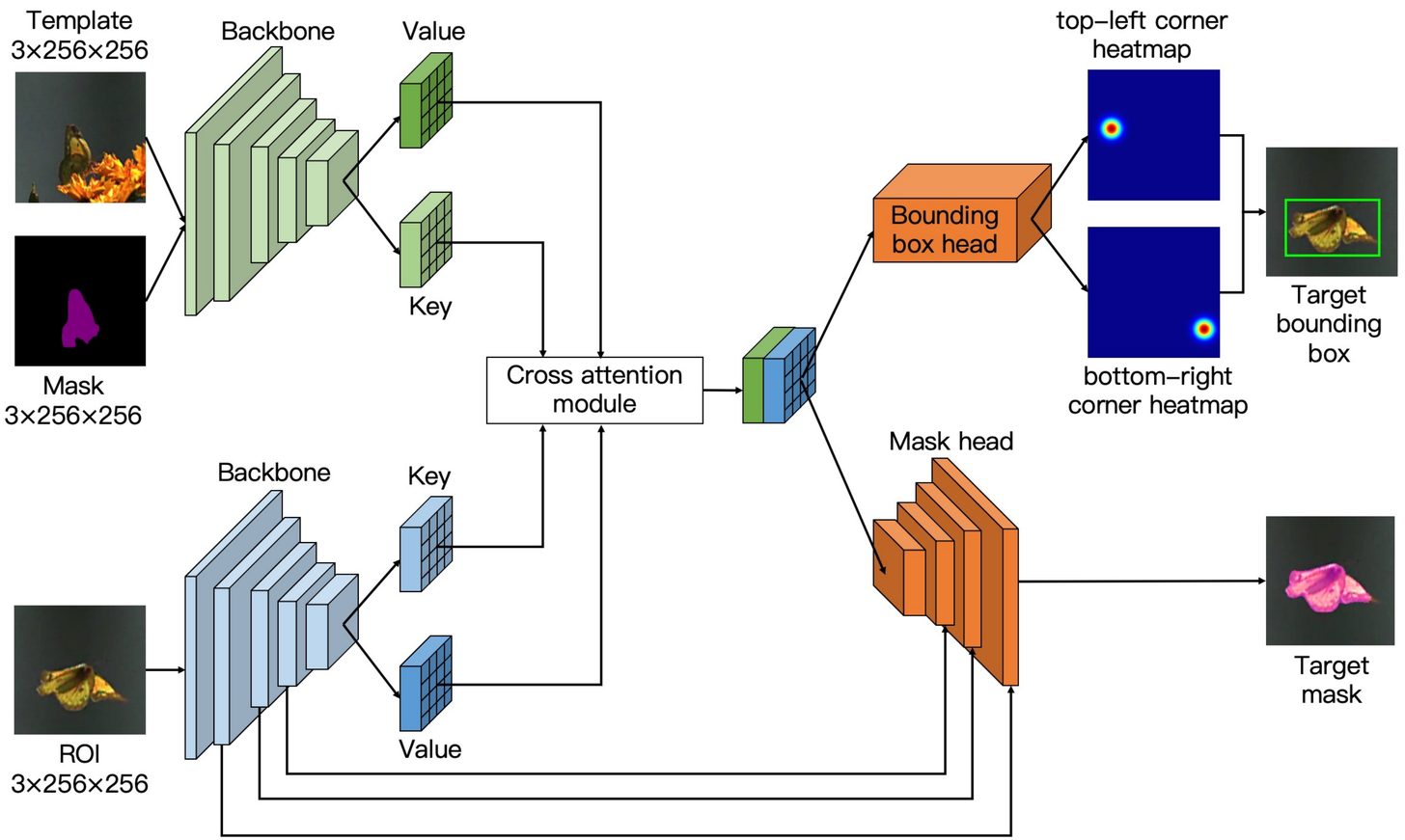
Poor discriminative ability



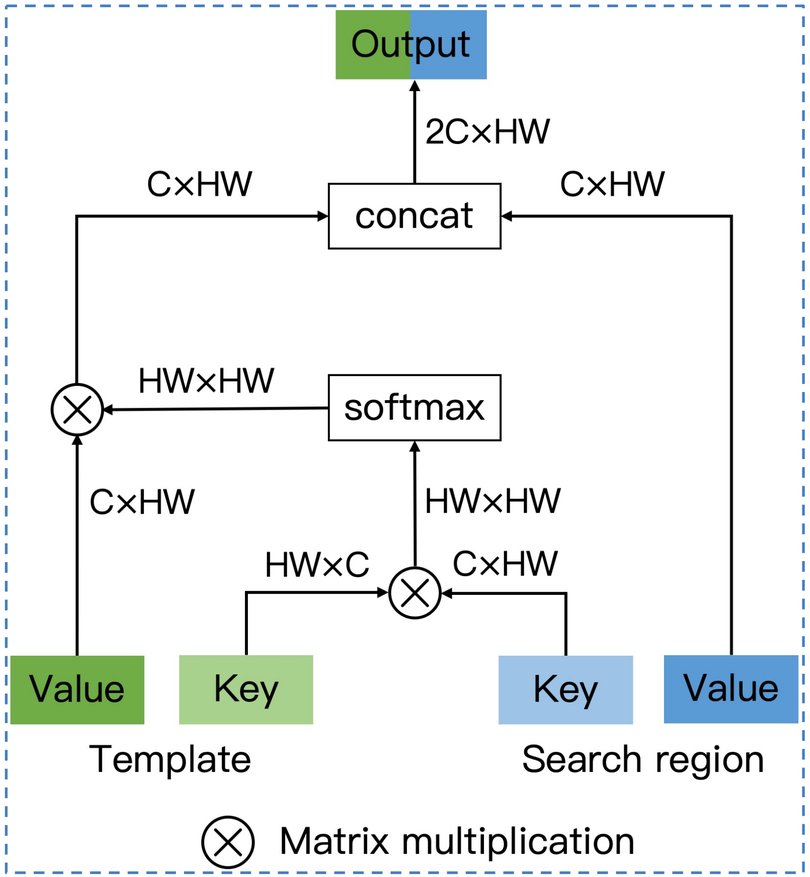

Pipeline of RPTMask (Ours)



One-shot object segmentation network



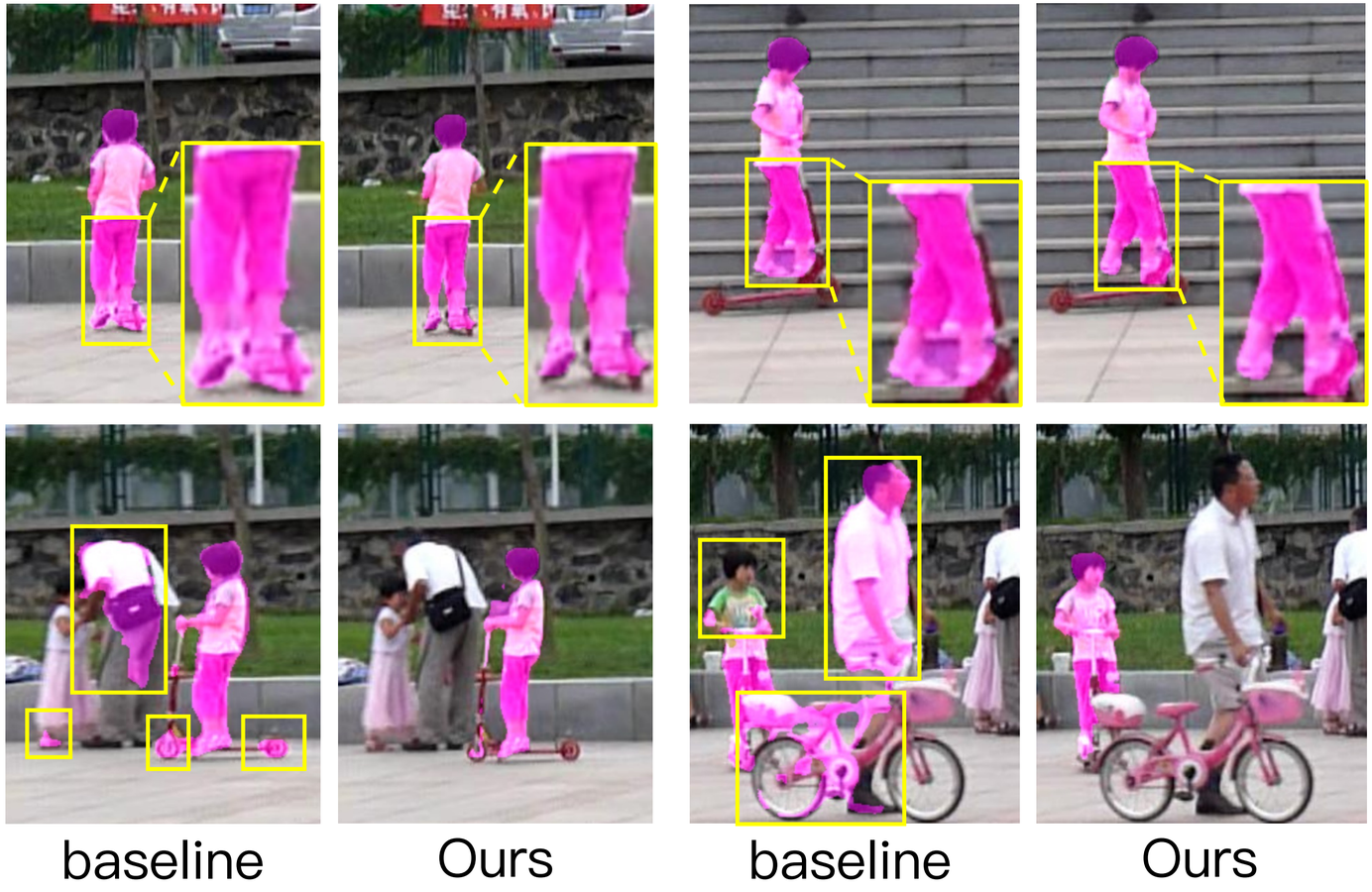
Architecture

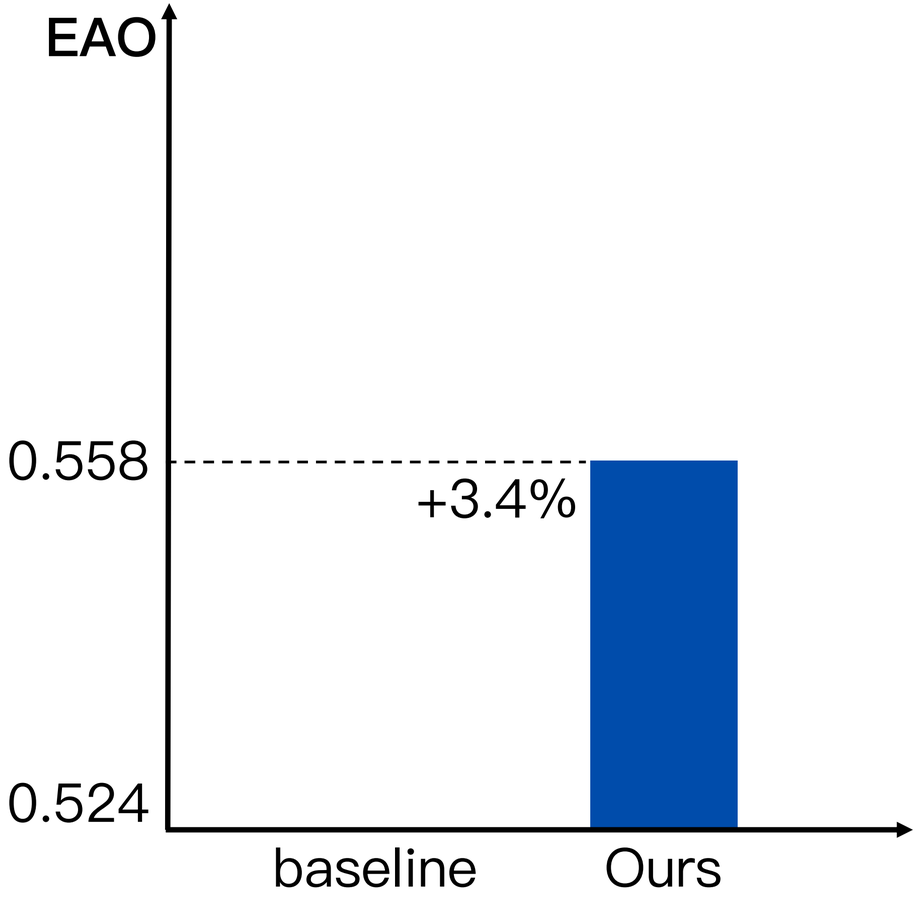


Cross attention module

Video Object Segmentation using Space-Time Memory Networks (ICCV 2019)

Results on the VOT2021 public dataset





Motivation #2

Template

ROI

Output

Template

ROI

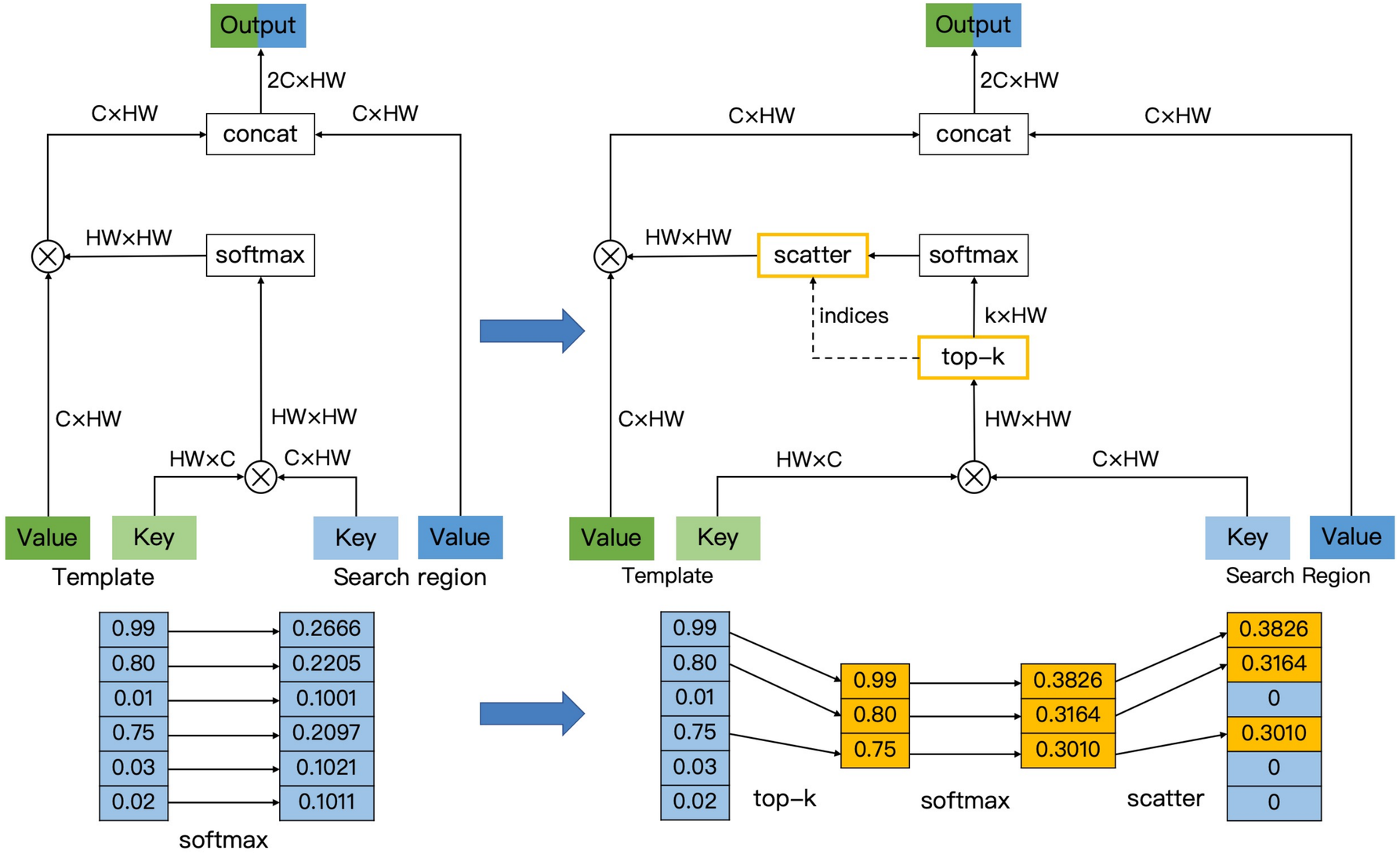
Output

Good matching😊

Bad matching☹️

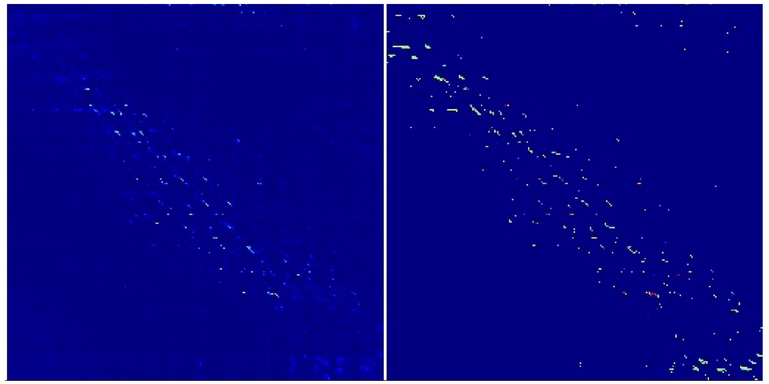
Undesired response☹️

Improved cross-attention module



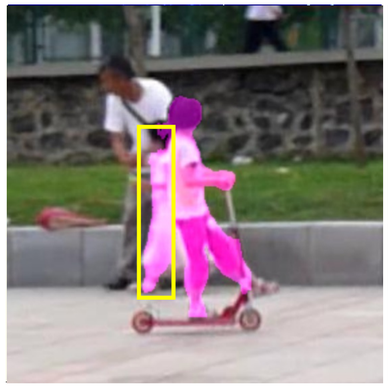
Modular Interactive Video Object Segmentation: Interaction-to-Mask, Propagation and Difference-Aware Fusion (CVPR 2021)

Results on the VOT2021 public dataset



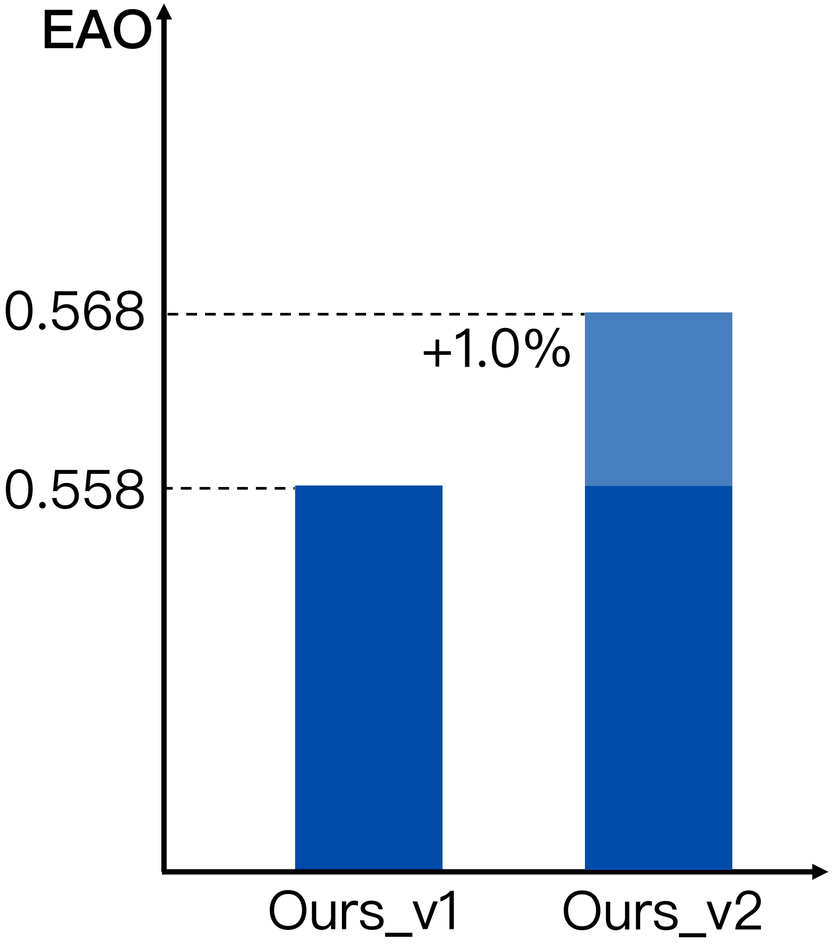
Similarity matrix with naive softmax

Similarity matrix with Top-K softmax

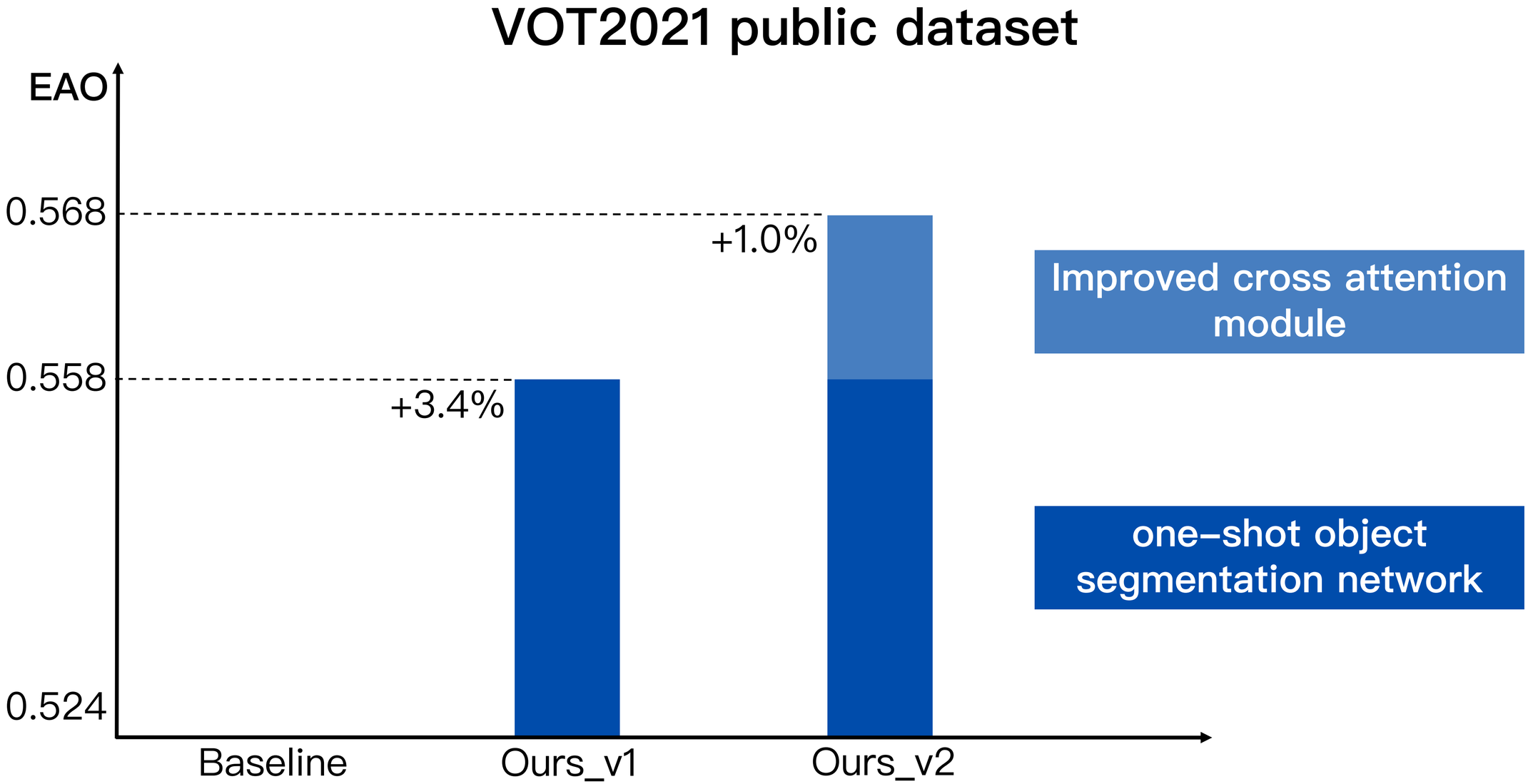


Ours_v1

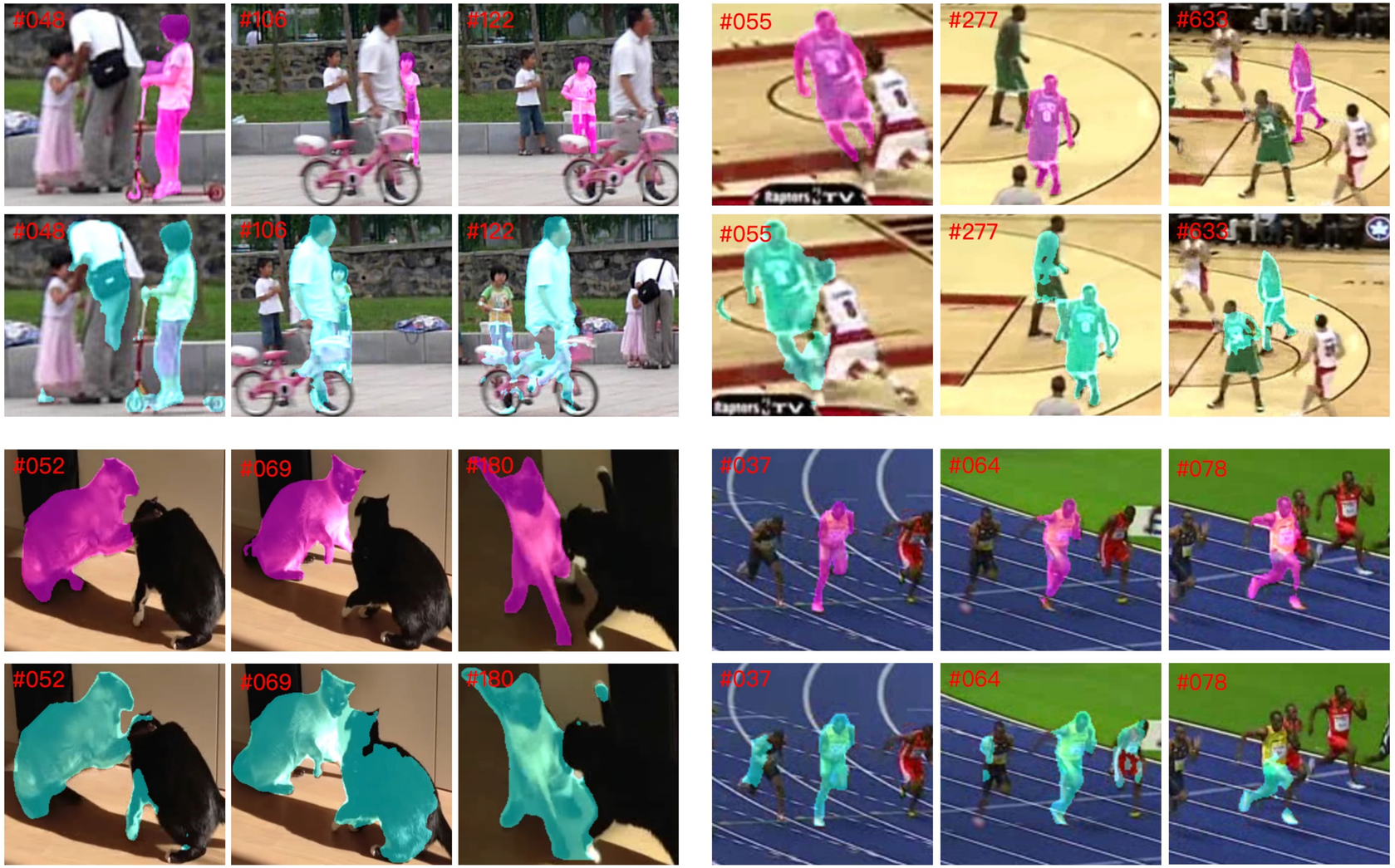
Ours_v2



Quantitative results



Qualitative results



Pink: RPTMask's results

green: baseline's results

Training dataset

- **RPT (we directly use the trained model provided by the VOT2020-ST winner)**
  - Bounding box
    - YouTube-BB
    - COCO
    - ImageNet VID
    - ImageNet DET

- **One-shot object segmentation network**
  - Mask
    - YouTube-VOS
    - Saliency
  - Bounding box
    - GOT-10k
    - LaSOT
    - COCO
    - ImageNet VID
    - ImageNet DET

Thanks