A Simple Baseline with Cascade Association for Robust Multi-Objects Tracking

Jiasheng Tang, Fei Du, Weihua Chen, Hao Luo, Fan Wang, Hao Li
Alibaba Group, Hangzhou, China

Alibaba Group 阿里巴巴集团

CVPR VIRTUAL JUNE 19-25

Summary

Challenges:
- Robustly tracking on 8 benchmarks.
- Extend tracking to 80 COCO-categories, not just traditional "person" or "car".

Contributions:
- We prove the strength of our strategy on training unlimited-category object ReID models again, which firstly proposed in our solution to ECCV-TAO in 2020.
- We adopt a simple cascade association method corresponding with the appearance models.
- We reformulate the post association (PA) module as a Minimum Cost Perfect Matching Problems (MCPM) in general graphs.
- We achieve **0.5859** HOTA socre with the **provided detections**, yield the **2nd** place in the competition.

Fundamental Components: Appearance Models and Post Association

CNN-based Appearance Models:

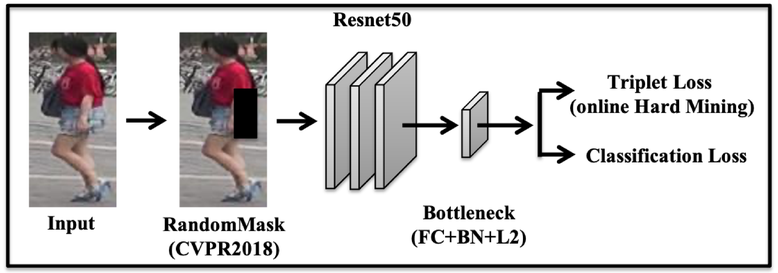

Figure 1. The framework of our CNN-based baseline BoT [1].

- Person ReID Model:
  - 23,000 IDs
  - Training Datasets: Market1501, DukeMTMC, MSMT17, MCT_NLPR, CUHK03, EndToEnd, RQEN, RPIfiled, Airport
- Car ReID model:
  - 30,000 IDs
  - Training Datasets: Veri-wild
- Anything-Appearance model:
  - 31,0000 IDs from SOT + 24,000 IDs from RobMOTS
  - Training Datasets: YouTube- BB, GOT-10k, ImageNet VID, RobMOTS

Tracklet-level Post Association:

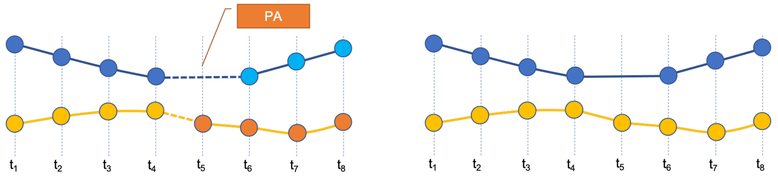

Figure 2. Why PA.

Cascade Association

Algorithm 1 Algorithm for Cascade Association

Input: Detections $\{DET^T\}$, Tracklets $\{TRK^{T-1}\}$
Output: Tracklets $\{TRK^T\}$

1: Initialize **Cost Matrix** M, **Threshold Embedding** Tr_f, **Threshold IoU** Tr_{IoU}, **Smooth Control** $alpha = 0.8$, **Life cycle** $lc = 40$
2: **for** $i = 0$ to $len(TRK^{T-1})$ **do**
3: **for** $j = 0$ to $len(DET^T)$ **do**
4: Set M_{ij} Calculate **Cosine** distance c of embedding feature between TRK_i^{T-1} and DET_j^T
5: **end for**
6: **end for**
7: Thresholding M with Tr_f.
8: Apply Hungarian Algorithm to M, get associated detection DET_a and tracklets TRK_a
9: Update TRK_a with DET_a
10: $DET^T \leftarrow DET^T \setminus DET_a$
11: $TRK^{T-1} \leftarrow TRK^{T-1} \setminus TRK_a$
12: $TRK^T \leftarrow TRK_a$
13: Clean M
14: Repeat line 2 to line 8 with IoU as distance and Tr_{IoU}, get associated detection DET_d and tracklets TRK_d
15: Update TRK_d with DET_d
16: $TRK^T \leftarrow TRK^T \cup TRK_d \cup (DET^T \setminus DET_d)$
17: Remove TRK^{T-1} with life cycle $> lc$
18: $TRK^T \leftarrow TRK^T \cup TRK^{T-1}$
19: Smooth Feature by $TRK_i^T = alpha * TRK_i^{T-1} + (1 - alpha) * DET^T$
20: **return** $\{TRK^T\}$

Comparison of Two Anything-Appearance Model

Models	HOTA	AssA	DetA
ReID-Rob-FT	60.42	64.97	57.69
ReID-SOT	60.44	65.09	57.66
Concatenation	60.53	65.21	57.68

Ablation Study

Steps/Matching Protocols	HOTA	AssA	DetA
Appearance(**APP.**) First	53.10	53.60	54.63
+ **Mask IoU**	59.16	62.47	57.70
+ **PA.**	59.48	63.10	57.70
+ Feature Smoothing(**FS.**)	60.53	65.21	57.68
Mask IoU First	55.73	56.42	57.08
+ **App.** + **PA.** + **FS.**	57.95	60.03	57.69

Overall Performance on Val and Test Board

Entires	HOTA	AssA	DetA
RobTrack	63.16	66.72	61.32
SBT	60.53	65.21	57.68
SIAtrackV1-0.5	58.73	61.71	57.64
SIAtrackV1	58.53	61.58	57.49
qd-mots	57.16	59.01	57.33
UbiTrack	57.11	58.68	57.52
OSNv1	57.08	58.79	57.35
BigColT	56.00	60.00	54.00
MeNToS	56.00	59.00	55.00
STP	55.90	56.34	57.62
OWTB	55.65	55.83	57.66
rrl2	46.80	49.95	46.24
BottomUp	45.11	52.81	40.18

Entires	HOTA	AssA	DetA
RobTrack	61.20	64.76	59.43
SBT	58.59	63.07	55.92
SIA	56.87	59.81	55.83
MeNToS	55.52	60.80	52.38
STP	54.35	55.04	55.78

Reference & Code

More Details Can be Found in:

[1] *Du Fu, Xu Bo, Tang Jiasheng, Zhang Yuqi, Wang Fan, and Li Haop*. "1st place solution to eccv-tao-2020: Detect and represent any object for tracking." arXiv preprint arXiv:2101.08040 (2021).
[2] *Jonathon Luiten, Arne Hoffhues, Blin Beqa, Paul Voigtlaender, István Sárándi, Patrick Dendorfer, Aljosa Osep, Achal Dave, Tarasha Khurana, Tobias Fischer, Xia Li, Yuchen Fan, Pavel Tokmakov, Song Bai, Linjie Yang, Federico Perazzi, Ning Xu, Alex Bewley, Jack Valmadre, Sergi Caelles, Jordi Pont-Tuset, Xinggang Wang, Andreas Geiger, Fisher Yu, Deva Ramanan, Laura Leal-Taixé, and Bastian Leibe*. RobMOTS: A Benchmark and Simple Baselines for Robust Multi-Object Tracking and Segmentation. CVPR RVSU Workshop 2021.
[3] *Luo et al*. Bag of Tricks and A Strong Baseline for Deep Person Re-identification. CVPRW 2019.