Datasets object tracking

Version 2

TAO: A Large-Scale Benchmark for Tracking Any Object

year: 2020

paper: https://arxiv.org/abs/2005.10356 (https://arxiv.org/abs/2005.10356)

GOT-10k: A Large High-Diversity Benchmark for Generic Object Tracking in the Wild

year: 2020

paper: https://ieeexplore.ieee.org/abstract/document/8922619 (https://ieeexplore.ieee.org/abstract/document/8922619)

LASOT (Large-scale Single Object Tracking)

Number of videos: 1,400 sequences with more 3.52 millions frames

year: 2019

page: https://cis.temple.edu/lasot/ (https://cis.temple.edu/lasot/)

paper: http://openaccess.thecvf.com/content_CVPR_2019/papers/Fan_LaSOT_A_High-Quality_Benchmark_for_Large-Scale_Single_Object_Tracking_CVPR_2019_paper.pdf (http://openaccess.thecvf.com/content_CVPR_2019/papers/Fan_LaSOT_A_High-Quality_Benchmark_for_Large-Scale_Single_Object_Tracking_CVPR_2019_paper.pdf)

Table 1. Comparison of LaSOT with the most popular dense benchmarks in the literatures.

Benchmark	Videos	Min frames	Mean frames	Median frames	Max frames	Total frames	Total duration	frame rate	Absent labels	Object classes	Class balance	Num. of attributes	Lingual feature
OTB-2013 [52]	51	71	578	392	3,872	29K	16.4 min	30 fps	✗	10	✗	11	✗
OTB-2015 [53]	100	71	590	393	3,872	59K	32.8 min	30 fps	✗	16	✗	11	✗
TC-128 [35]	128	71	429	365	3,872	55K	30.7 min	30 fps	✗	27	✗	11	✗
VOT-2014 [26]	25	164	409	307	1,210	10K	5.7 min	30 fps	✗	11	✗	n/a	✗
VOT-2017 [27]	60	41	356	293	1,500	21K	11.9 min	30 fps	✗	24	✗	n/a	✗
NUS-PRO [28]	365	146	371	300	5,040	135K	75.2 min	30 fps	✗	8	✗	n/a	✗
UAV123 [39]	123	109	915	882	3,085	113K	62.5 min	30 fps	✗	9	✗	12	✗
UAV20L [39]	20	1,717	2,934	2,626	5,527	59K	32.6 min	30 fps	✗	5	✗	12	✗
NfS [14]	100	169	3,830	2,448	20,665	383K	26.6 min	240 fps	✗	17	✗	9	✗
GOT-10k [22]	10,000	-	-	-	-	1.5M	-	10 fps	✓	563	✗	6	✗
LaSOT	1,400	1,000	2,506	2,053	11,397	3.52M	32.5 hours	30 fps	✓	70	✓	14	✓

TrackingNet

Number of videos: 30,132 (train) + 511 (test)

Number of annotations: 14,205,677 (train) + 225,589 (test)

Annotation density: high, variable, state-of-the-art trackers to fill in missing annotations, weighted average between a forward and a backward pass using the DCF tracker

sample duration: ~ 16.6s

Samples: dirived from YouTube-Bounding Boxes (YT-BB): contains a large variety of frame rates, resolutions, context and object classes. Building process: Filtered out 90% of the videos by selecting the videos that a) are longer than 15 seconds; b) include bounding boxes that cover less than 50% of the frame; c) contain a reasonable amount of motion between bounding boxes.

Year: 2018

official page: https://tracking-net.org/ (https://tracking-net.org/)

Cloud dataset: https://exrcsdrive.kaust.edu.sa/exrcsdrive/index.php/s/MAaiTPdOwiPDNlp (https://exrcsdrive.kaust.edu.sa/exrcsdrive/index.php/s/MAaiTPdOwiPDNlp)

Python devkit: https://github.com/SilvioGiancola/TrackingNet-devkit| (https://github.com/SilvioGiancola/TrackingNet-devkit|)

Papers: http://openaccess.thecvf.com/content_ECCV_2018/papers/Matthias_Muller_TrackingNet_A_Large-Scale_ECCV_2018_paper.pdf (http://openaccess.thecvf.com/content_ECCV_2018/papers/Matthias_Muller_TrackingNet_A_Large-Scale_ECCV_2018_paper.pdf)

Dataset Structure:

TrackingNet:

- Test / Train_X (with X from 0 to 11)
  - zips
  - frames
  - anno (Test: annotation only for 1st frame)

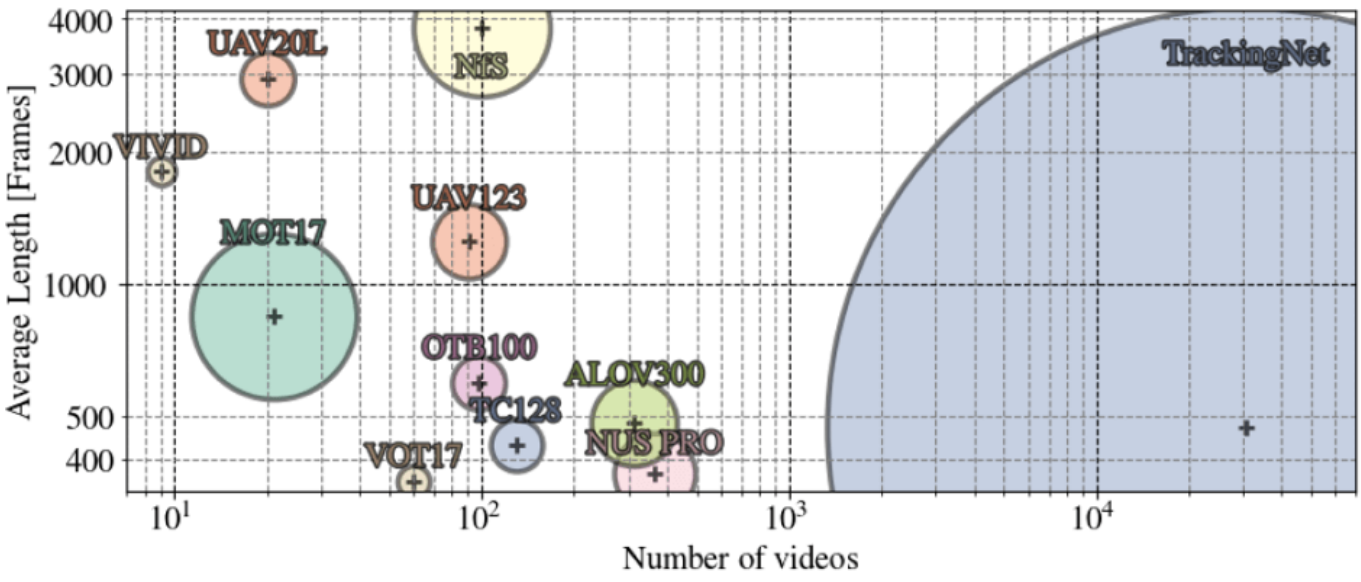

Figure - Comparison of tracking datasets distributed across the number of videos and the average length of the videos. The size of circles is proportional to the number of annotated bounding boxes. Our dataset has the largest amount of videos and frames and the video length is still reasonable for short video tracking.

Table 1. Comparison of current datasets for object tracking.

Datasets	Nb Videos	Nb Annot.	Frame per Video	Nb Classes
VIVID [5]	9	16274	1808.2	-
TC128 [33]	129	55652	431.4	-
OTB50 [48]	51	29491	578.3	-
OTB100 [49]	98	58610	598.1	-
VOT16 [22]	60	21455	357.6	-
VOT17 [23]	60	21356	355.9	-
UAV20L [36]	20	58670	2933.5	-
UAV123 [36]	91	113476	1247.0	-
NUS PRO [29]	365	135305	370.7	-
ALOV300 [43]	314	151657	483.0	-
NfS [13]	100	383000	3830.0	-
MOT16 [35]	7	182326	845.6	-
MOT17 [35]	21	564228	845.6	-
TrackingNet (Train)	**30132**	**14205677**	**471.4**	**27**
TrackingNet (Test)	**511**	**225589**	**441.5**	**27**

In [ ]:

```
#TrackingNet-devkit
#1) download TrackingNet-devkit and create appropriate python enviroment: conda
 create -n TrackingNet python=3 requests pandas tqdm numpy
#2) import scripts
import os
import TrackingNet_devkit_master as tn
import TrackingNet_devkit_master.download_TrackingNet
import TrackingNet_devkit_master.metrics
import TrackingNet_devkit_master.extract_frame
masterdir=os.path.abspath(tn.__file__)
masterdir=os.path.dirname(masterdir)
masterdir
```

In [ ]:

```
#Download the dataset
trackingnet_dir=os.path.join(masterdir,"TrackingNet")
csv_dir=os.path.join(masterdir, 'csv_link')
tn.download_TrackingNet.main(trackingnet_dir=trackingnet_dir,
            csv_dir=csv_dir,
            overwrite=False,
            chunks=["TEST"],
            data=["ANNO","ZIPS"])
```

In [ ]:

```
#Unzip the frames
trackingnet_dir=os.path.join(masterdir,"TrackingNet")
tn.extract_frame.main(trackingnet_dir="TrackingNet",
     overwrite_frames=False,
     chunks=["TEST"])
```

In [ ]:

```
#Evaluate the results of a tracker with a given ground truth
test_annotation_file=os.path.join(masterdir, 'dummy_GT.zip')
user_submission_file=os.path.join(masterdir, 'dummy_subm.zip')
Coverage, Success, Precision, Normalized_Precision = tn.metrics.evaluate(test_an
notation_file, user_submission_file)


print("Coverage", Coverage)
print("Precision", Precision)
print("Normalized Precision", Normalized_Precision)
print("Success", Success)
```

MOT challenge (Multiple Object Tracking)

page: https://motchallenge.net/ (https://motchallenge.net/)

MOT20

year:2020

page: https://motchallenge.net/data/MOT20/ (https://motchallenge.net/data/MOT20/)

paper: https://arxiv.org/abs/2003.09003 (https://arxiv.org/abs/2003.09003)

MOT17

Number of videos: 21 (train) + 21 (test)

Number of annotations: 564,228

Annotation density: ~15fps

Year: 2017

Page: https://motchallenge.net/data/MOT17/ (https://motchallenge.net/data/MOT17/)

NfS

Number of videos: 100

Number of annotations: 383,000

video fps: 240 fps

Year: 2017

page: http://ci2cv.net/nfs/index.html (http://ci2cv.net/nfs/index.html)

YouTube-Bounding Boxes (YT-BB)

Number of videos:300K video segments

Number of annotations: annotated every second with upright bounding boxes

Annotation density: 1fps

paper: https://arxiv.org/abs/1702.00824 (https://arxiv.org/abs/1702.00824)

Additional info

https://neurohive.io/en/datasets/new-datasets-for-object-tracking/ (https://neurohive.io/en/datasets/new-datasets-for-object-tracking/)

paper: http://openaccess.thecvf.com/content_ECCV_2018/papers/Matthias_Muller_TrackingNet_A_Large-Scale_ECCV_2018_paper.pdf (http://openaccess.thecvf.com/content_ECCV_2018/papers/Matthias_Muller_TrackingNet_A_Large-Scale_ECCV_2018_paper.pdf)

" Object Tracking Datasets. Numerous datasets are available for object tracking, the most common ones being OTB [49], VOT [25], ALOV300 [43] and TC128 [33] for single-object tracking and MOT [28,35] for multi-object tracking. VIVID [5] is an early attempt to build a tracking dataset for surveillance purposes. OTB50 [48] and OTB100 [49] provide 51 and 98 video sequences annotated with 11 different attributes and upright bounding boxes for each frame. TC128 [33] comprises 129 videos, based on similar attributes and upright bounding boxes. ALOV300 [43] comprises 314 videos sequences labelled with 14 attributes. VOT [25] proposes several challenges with up to 60 video sequences. It introduced rotated bounding boxes as well as extensive studies on object tracking annotations. VOT-TIR is a specific dataset from VOT focusing on Thermal InfraRed videos. NUS PRO [29] gathers an application-specific collection of 365 videos for people and rigid object tracking. UAV123 and UAV20L [36] gather another application-specific collection of 123 videos and 20 long videos captured from a UAV or generated from a flight simulator. NfS [11] provides a set of 100 videos with high framerate, in an attempt to focus on fast motion. Table 1 provides a detailed overview of the most popular tracking datasets. "

paper: http://openaccess.thecvf.com/content_CVPR_2019/papers/Fan_LaSOT_A_High-Quality_Benchmark_for_Large-Scale_Single_Object_Tracking_CVPR_2019_paper.pdf (http://openaccess.thecvf.com/content_CVPR_2019/papers/Fan_LaSOT_A_High-Quality_Benchmark_for_Large-Scale_Single_Object_Tracking_CVPR_2019_paper.pdf)

" In addition to the dense tracking benchmarks above, there exist other benchmarks which may not provide high-quality annotations for each frame. Instead, these benchmarks are either annotated sparsely (e.g., every 30 frames) or labeled (semi-)automatically by tracking algorithms. Representatives of this type of benchmarks include ALOV [47], TrackingNet [41] and OxUvA [51]. ALOV [47] consists of 314 sequences labeled in 14 attributes. Instead of densely annotating each frame, ALOV provides annotations every 5 frames. TrackingNet [41] is a subset of the video object detection benchmark YT-BB [43] by selecting 30K videos, each of which is annotated by a tracker. Though the tracker used for annotation is proven to be reliable in a short period (i.e., 1 second) on OTB 2015 [53], it is difficult to guarantee the same performance on a harder benchmark. Besides, the average sequence length of TrackingNet does not exceed 500 frames, which may not demonstrate the performance of a tracker in long-term scenarios. OxUvA [51] also comes from YT-BB [43]. Unlike TrackingNet, OxUvA is focused on long-term tracking. It contains 366 videos with an average length of around 4,200 frames. However, a problem with OxUvA is that it does not provide dense annotations in consecutive frames. Each video in OxUvA is annotated every 30 frames, ignoring rich temporal context between consecutive frames when developing a tracking algorithm

In [ ]: