TO BENCHMARK OR NOT TO BENCHMARK

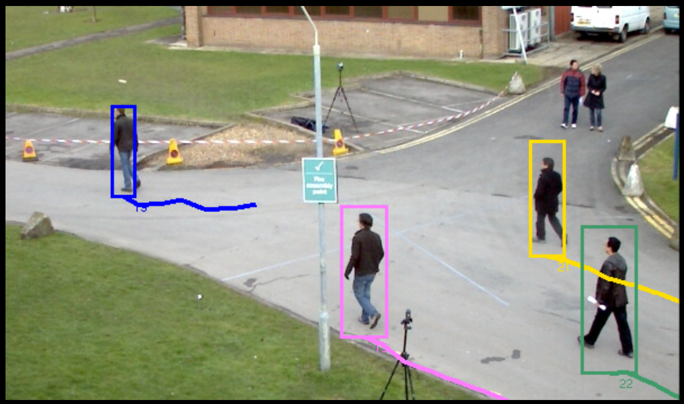
Laura Leal-Taixé

To Benchmark or Not to Benchmark: the MotChallenge case

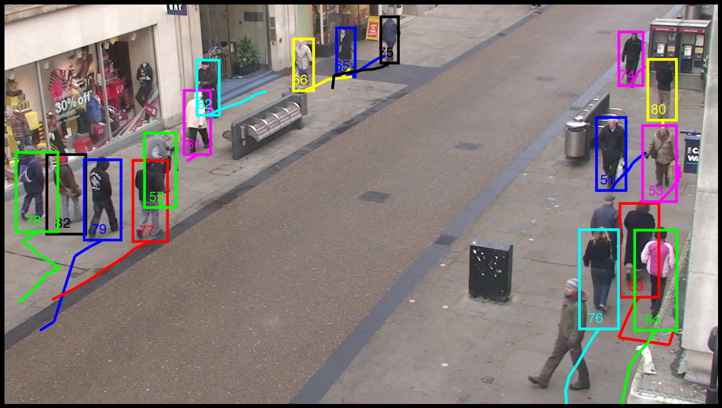
Laura Leal-Taixé

The Tracking Problem

Sparse



Semi-crowded



Crowded

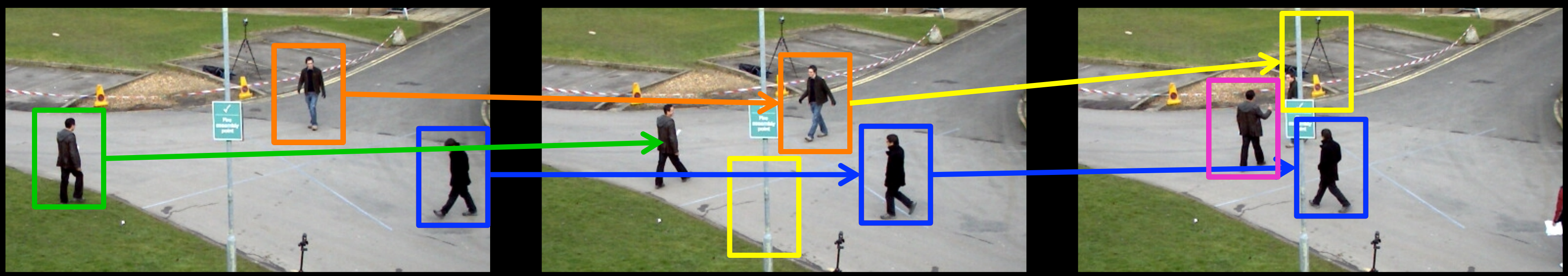
Macroscopic

Tracking-by-Detection

Video sequence

Person detection

Tracking-by-Detection



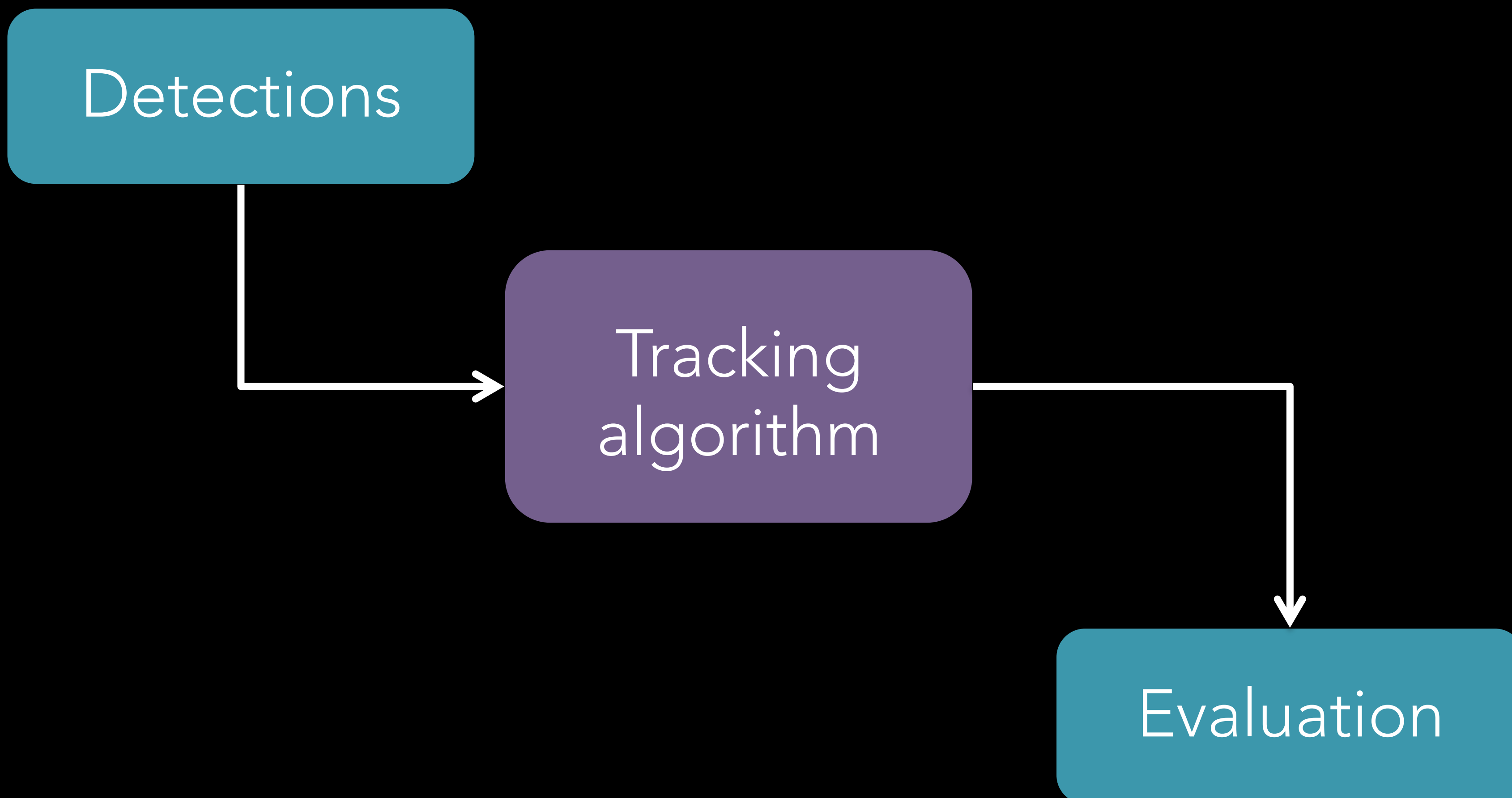
Person detection

- Missed detections
- False alarms

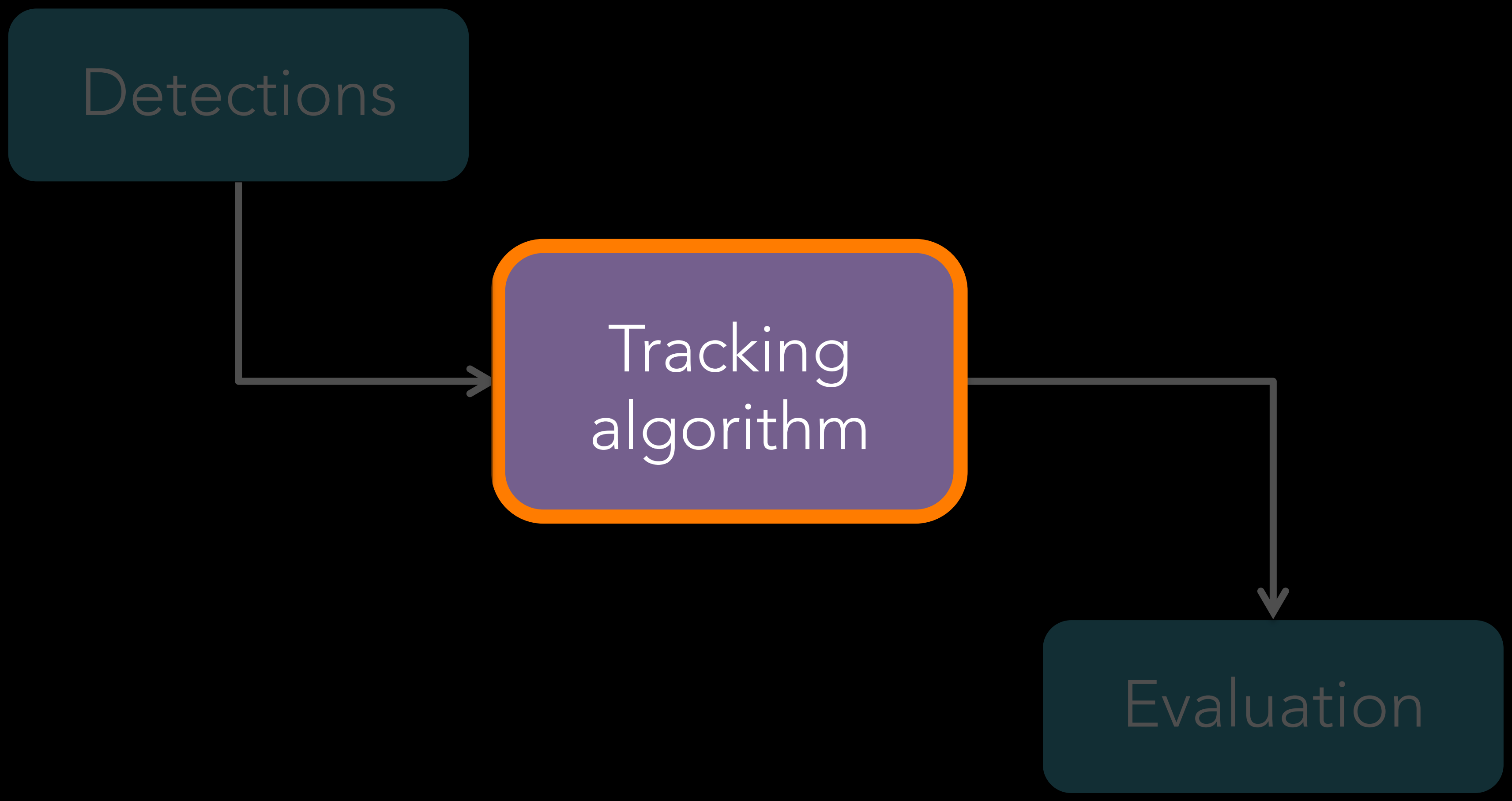
Person tracking

- Identity switches
- Gaps

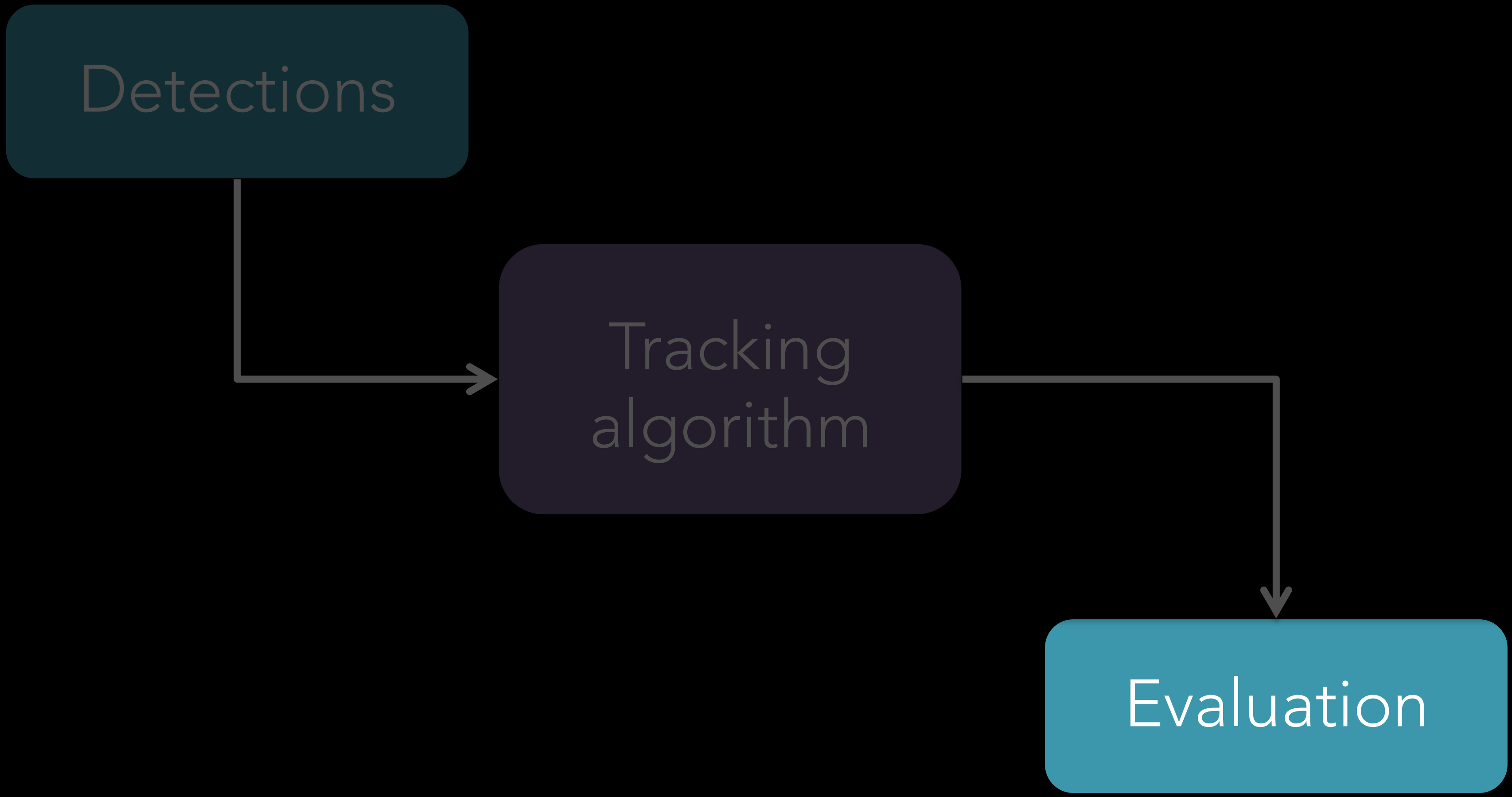
Tracking-by-Detection



Tracking-by-Detection



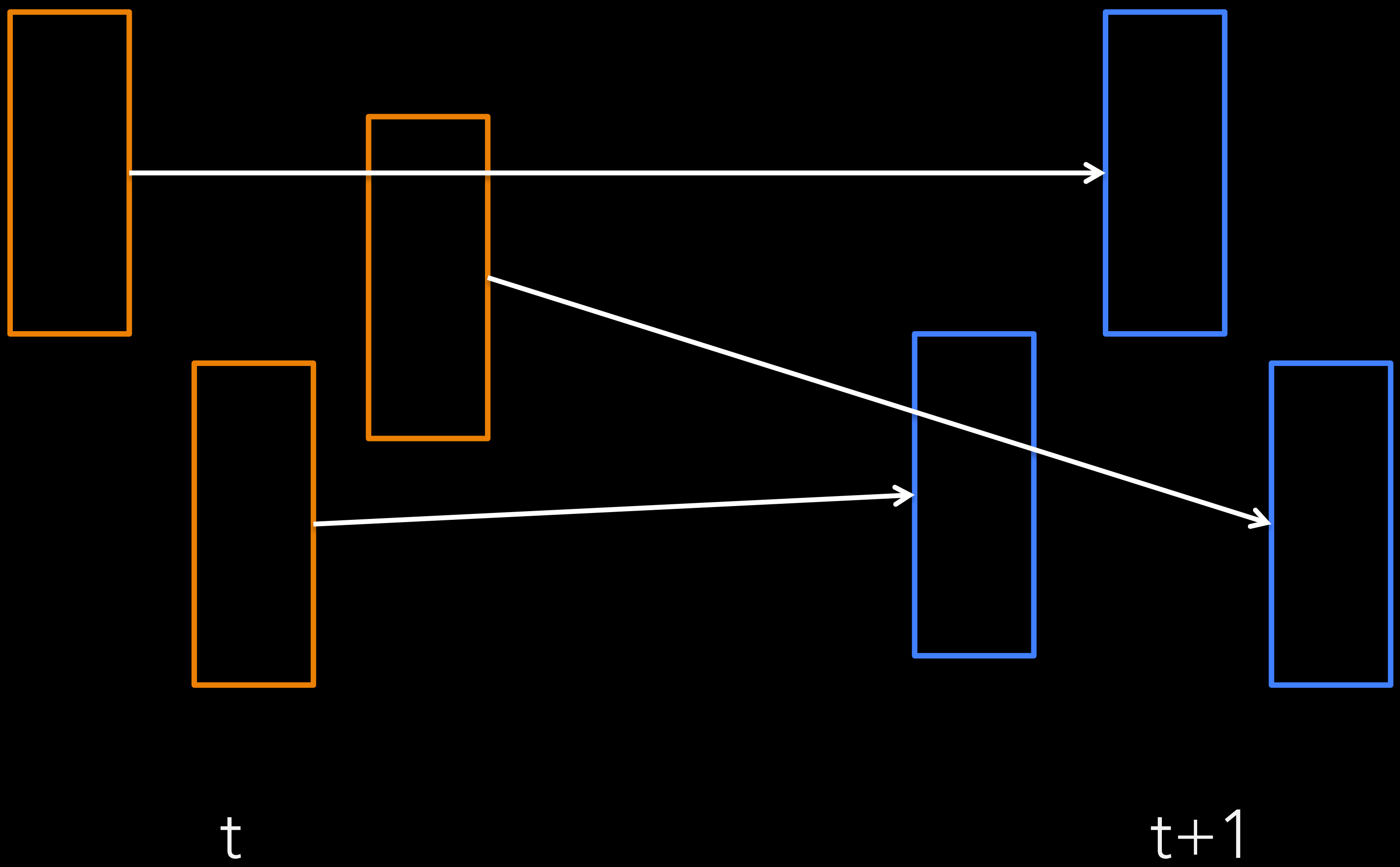
Tracking-by-Detection



Evaluation: the MOTA score

Overlap > 50%

Evaluation: the MOTA score

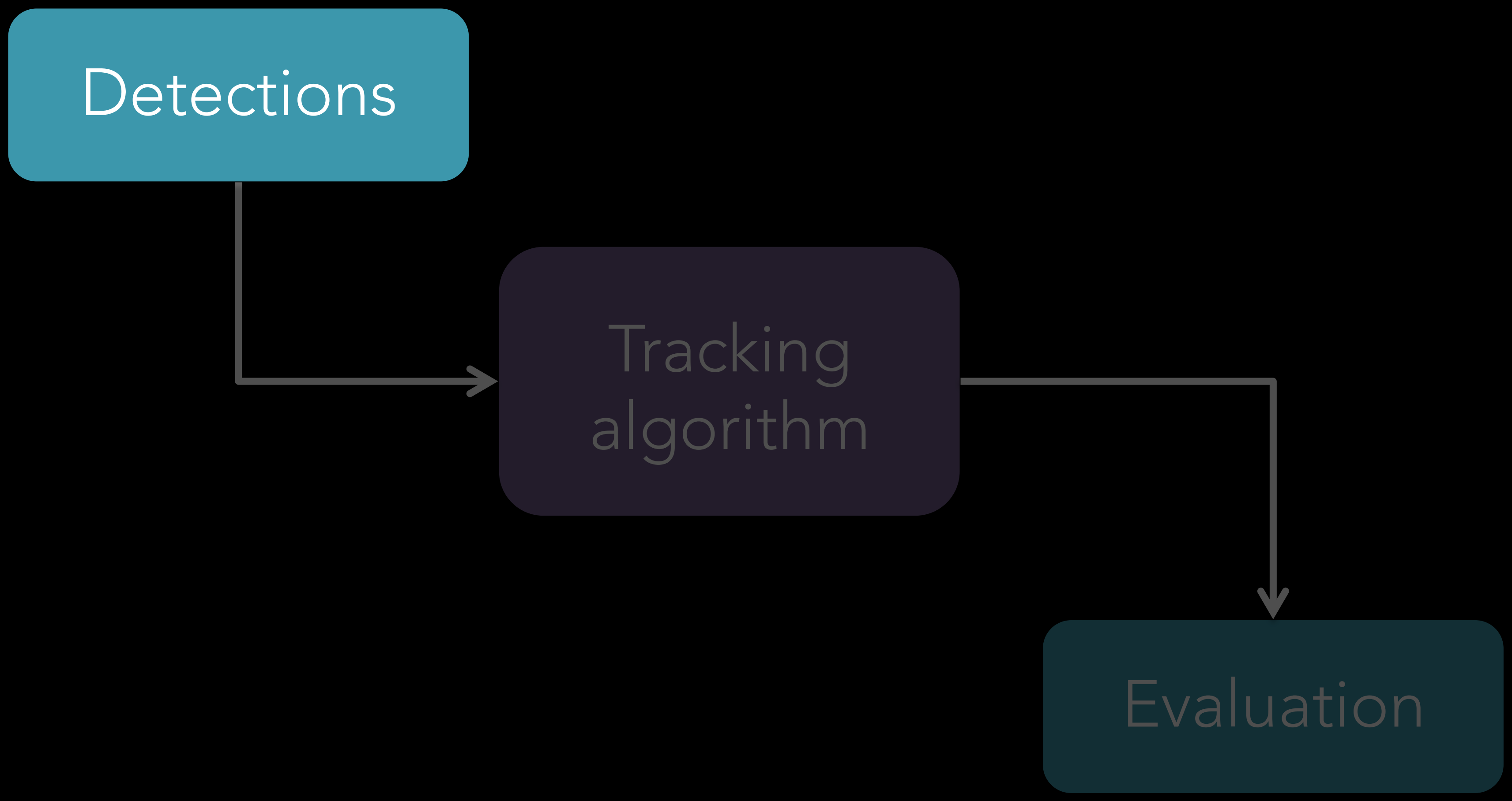


Evaluation: the MOTA score

MOTA: Multiple object tracking accuracy

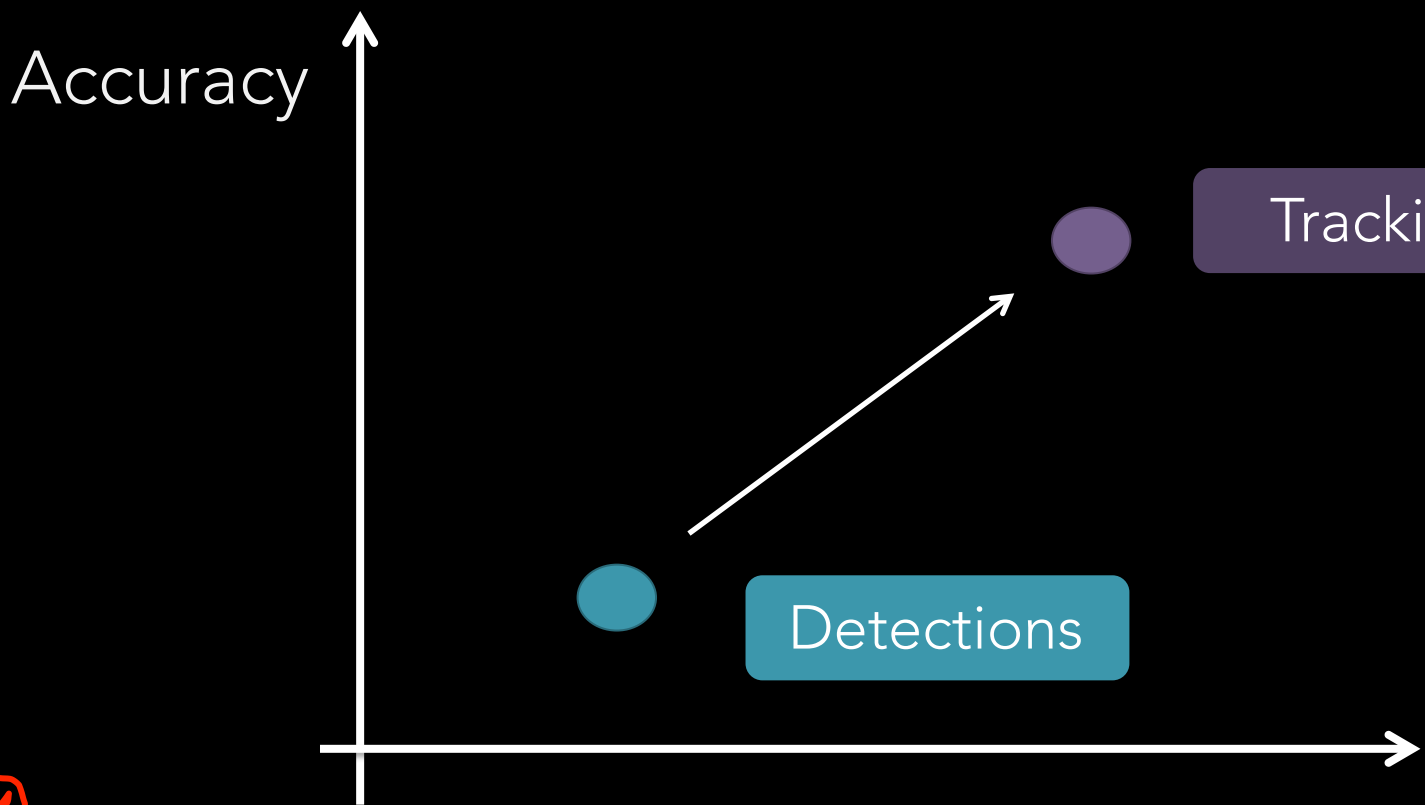
Missed objects	Detections
False alarms	Detections
Identity switches	Tracking

Tracking-by-Detection



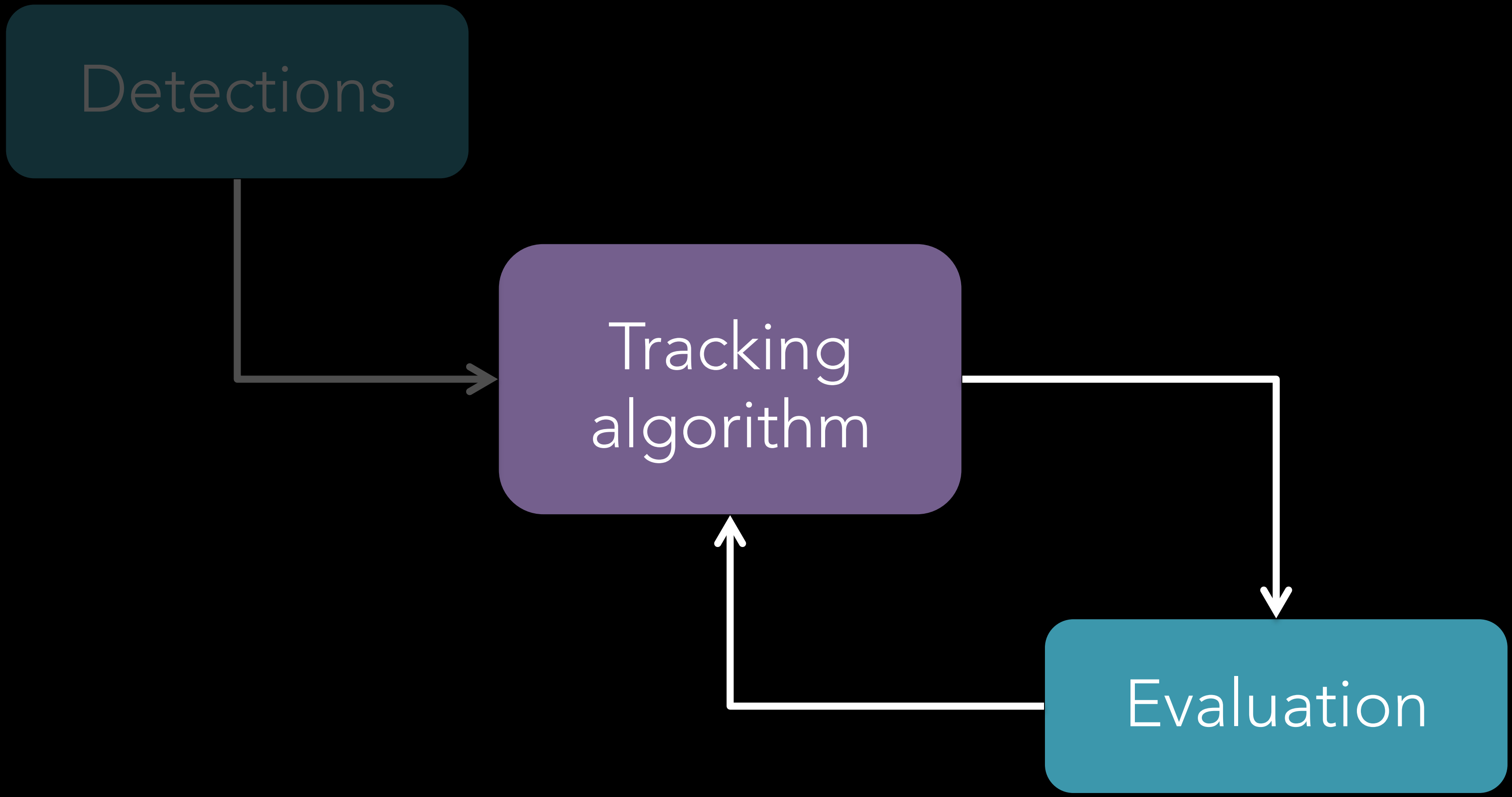
Detection Dependency

Detections are the "starting point" for the tracker



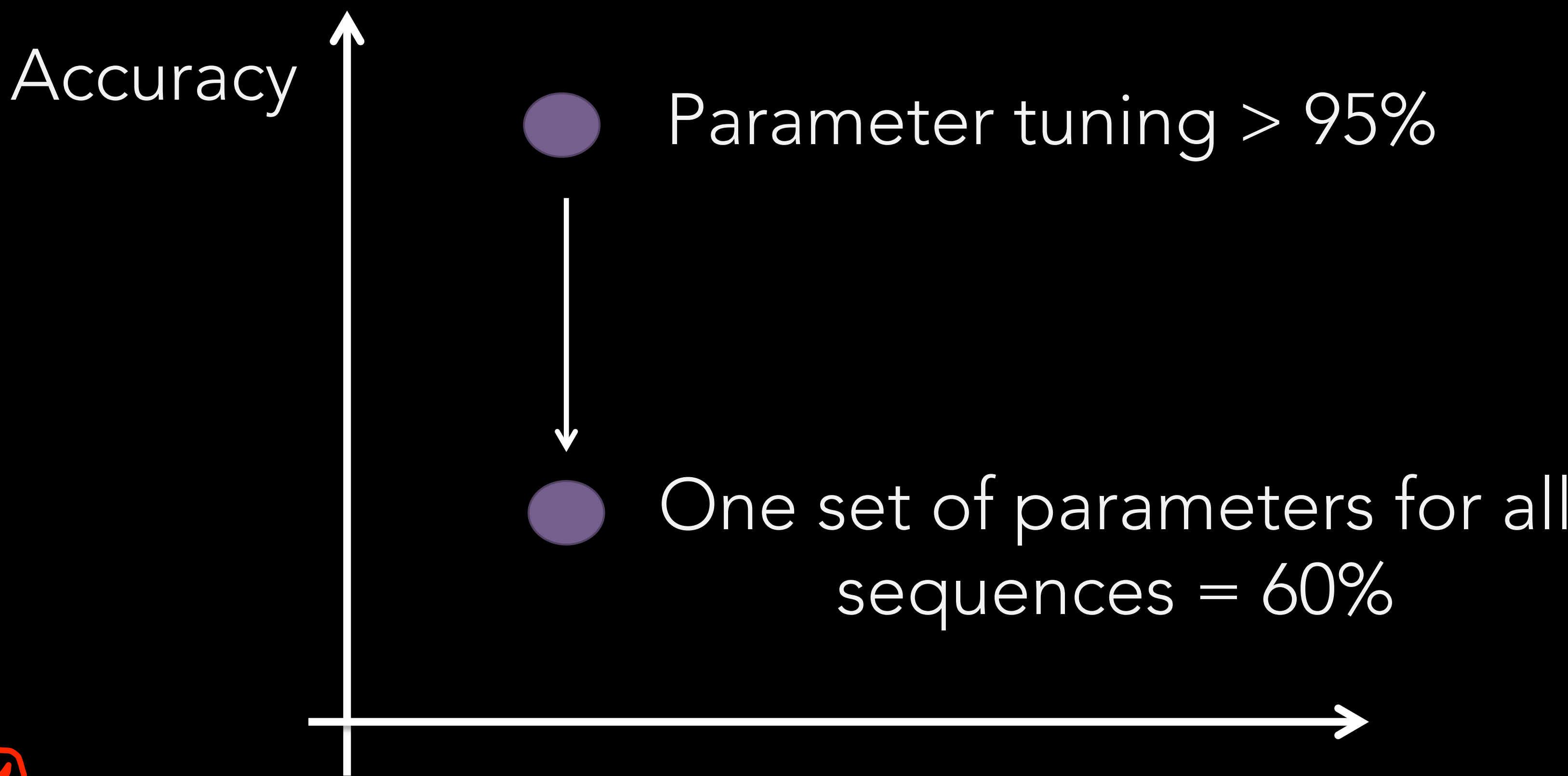
Detection accuracy is often not mentioned, results cannot be compared

The well-known Parameter Tuning



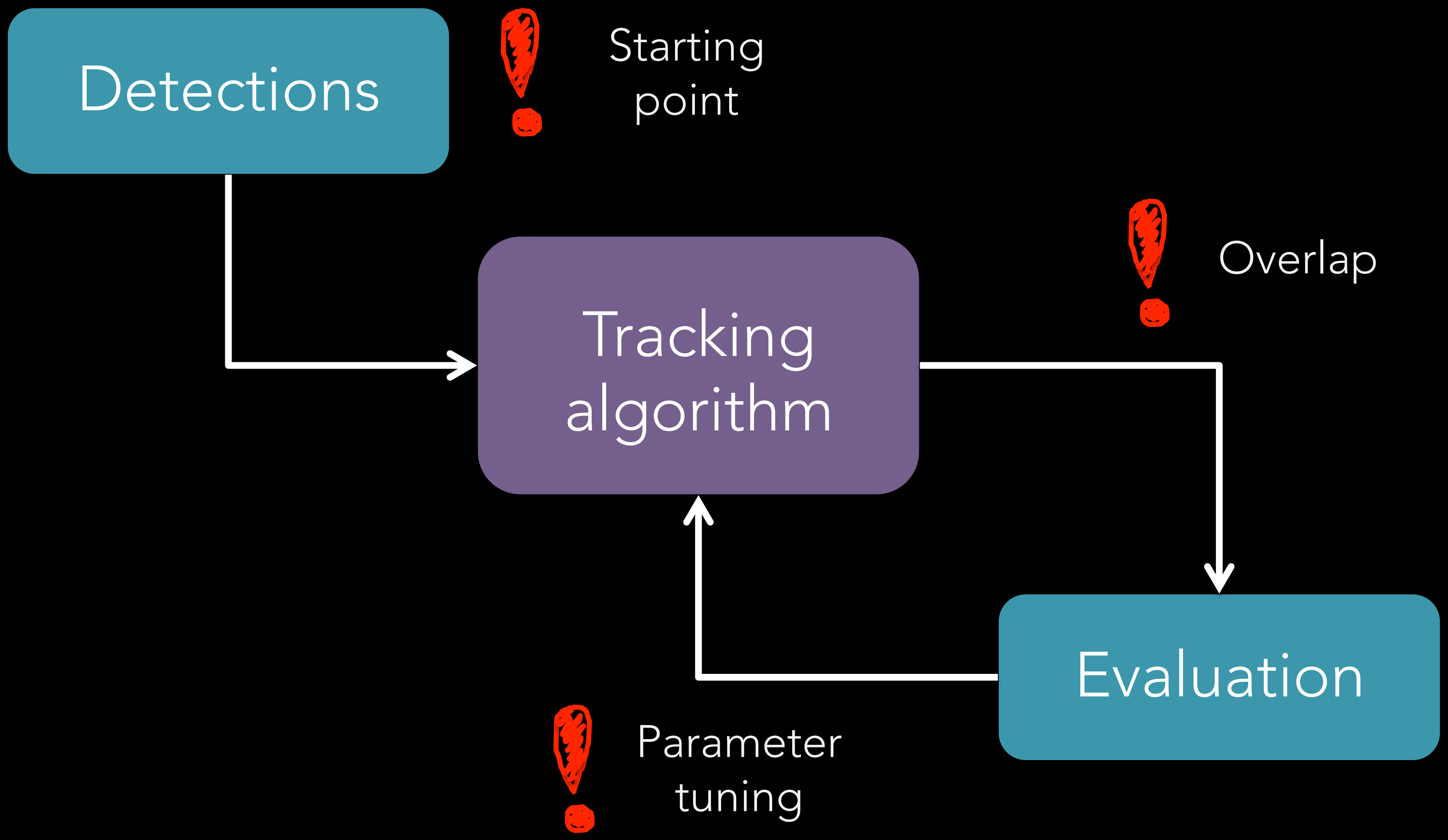
The well-known Parameter Tuning

PETS 2009 sequence: obsolete?

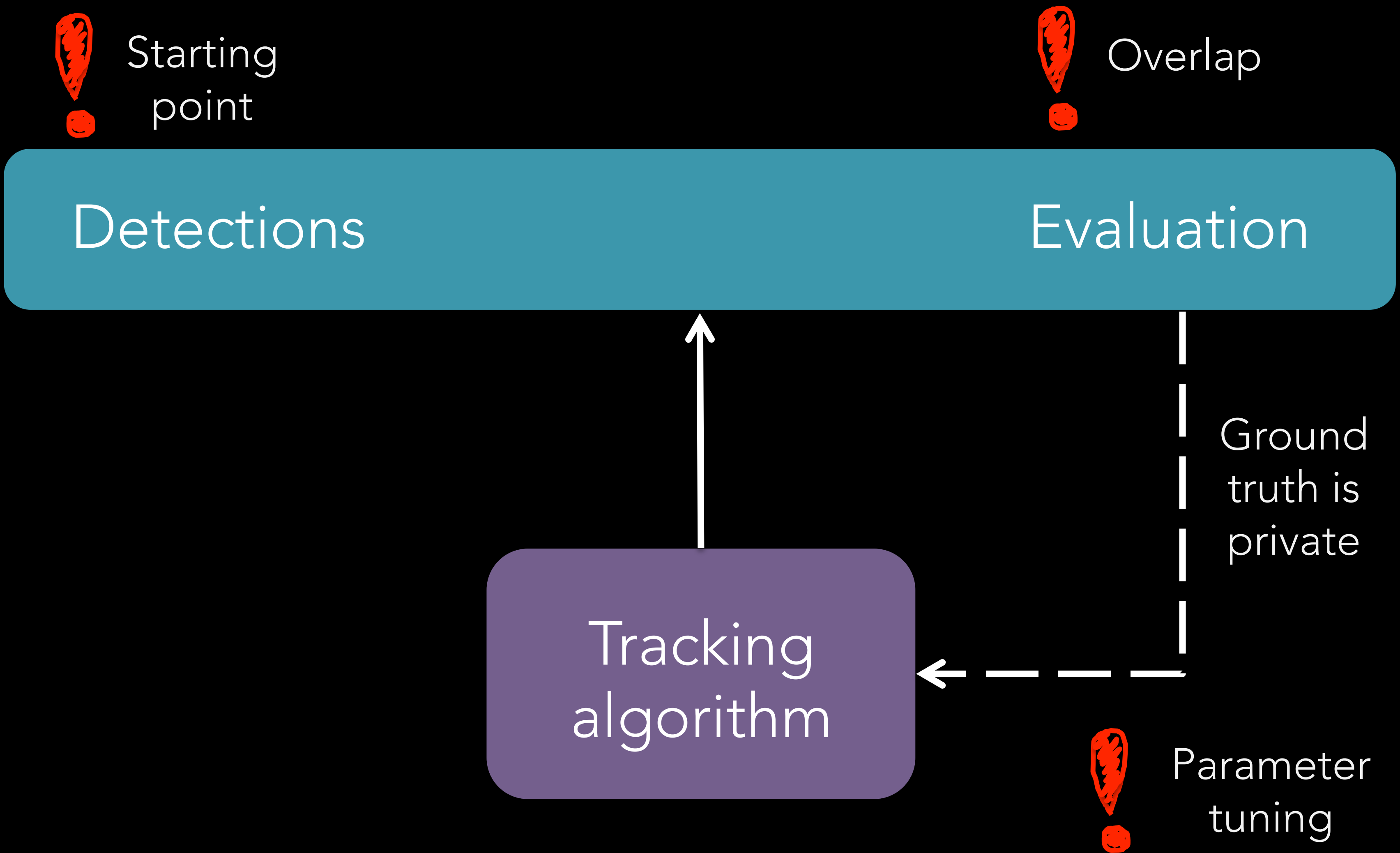


Parameter tuning doesn’t allow us to see the real challenges tracking still has to face

We need a New Evaluation Platform



MOTChallenge



Multiple Object Tracking Benchmark

home | data | results | vis | submit | people | login | sign up

Welcome to the Multiple Object Tracking Benchmark!

In the recent past, the computer vision community has relied on several centralized benchmarks for performance evaluation of numerous tasks including object detection, pedestrian detection, 3D reconstruction, optical flow, single-object short-term tracking, and stereo estimation. Despite potential pitfalls of such benchmarks, they have proved to be extremely helpful to advance the state-of-the-art in the respective research fields. Interestingly, there has been rather limited work on the standardization of multiple target tracking evaluation. One of the few exceptions is the well-known PETS dataset, targeted primarily at surveillance applications. Even for this widely used benchmark, a common technique for presenting tracking results to date involves using different subsets of the available data, inconsistent model training and varying evaluation scripts.

With this benchmark we would like to pave the way for a unified framework towards more meaningful quantification of multi-target tracking.

What do we provide?

We have created a framework for the fair evaluation of multiple people tracking algorithms. In this framework we provide:

- A large collection of datasets, some already in use and some new challenging sequences!
- Detections for all the sequences.
- A common evaluation tool providing several measures, from recall to precision to running time.
- An easy way to compare the performance of state-of-the-art tracking methods.
- Several challenges with subsets of data for specific tasks such as 3D tracking, surveillance, sports analysis (updates coming soon).

MOT 2015 EDITION

- 16 well-known datasets in the community (PETS, TUD, ETH, KITTI….)

Are they really obsolete?

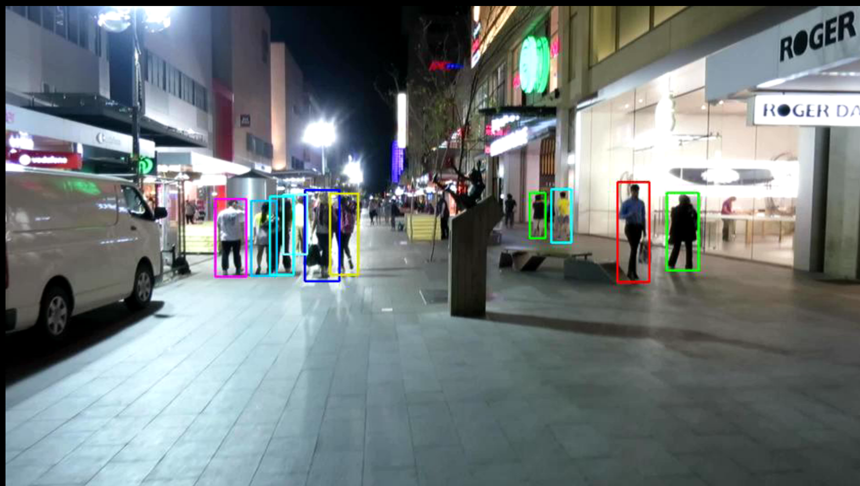
- 3 new challenging datasets

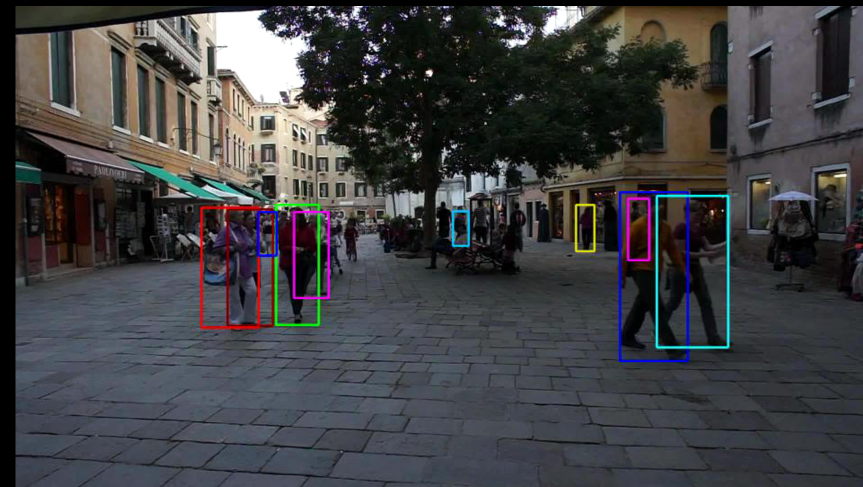
Increasing the difficulty

Old density: 7.29 New density: 12.8

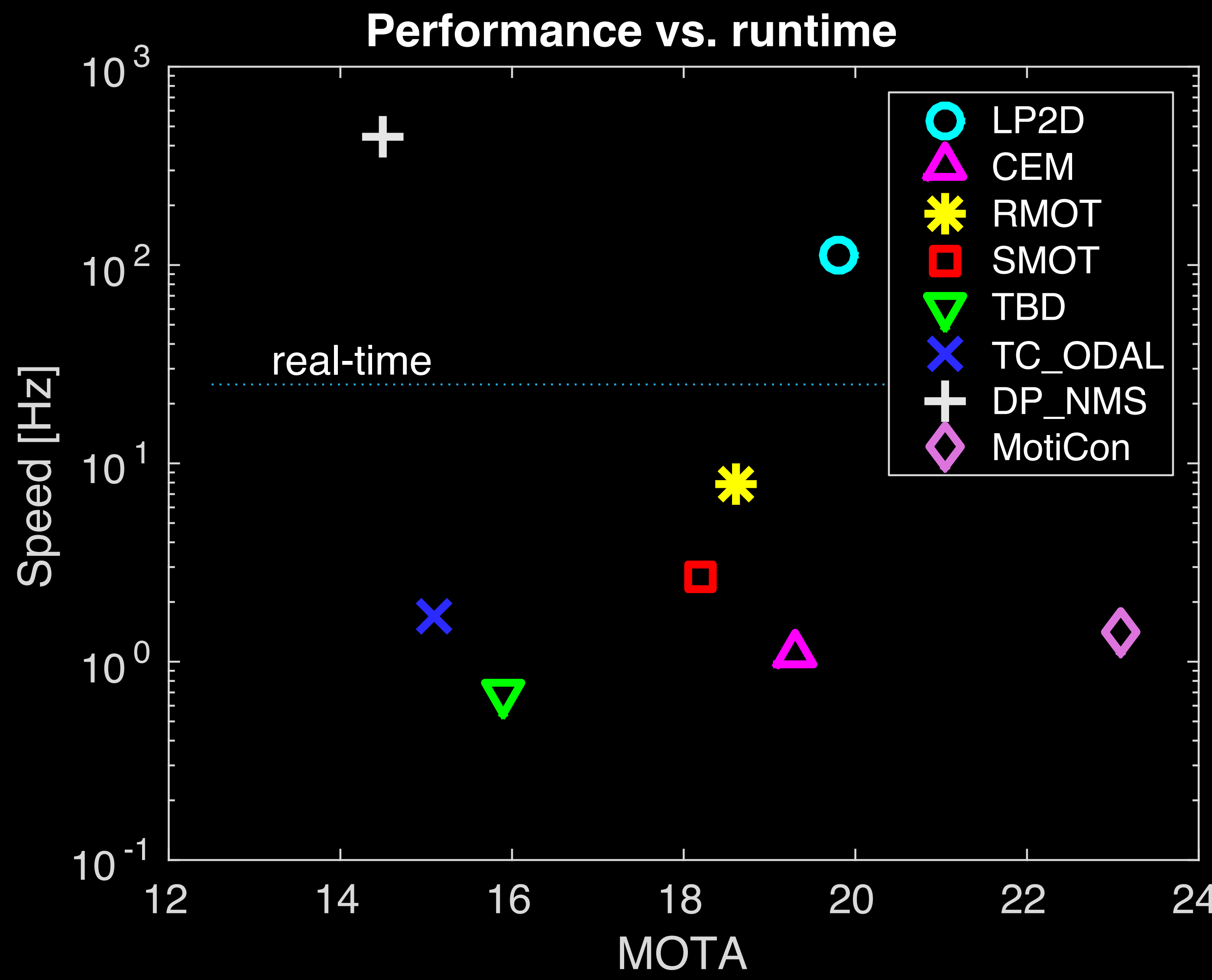
MOT 2015 EDITION

MOT 2015 EDITION





Baselines



Performance vs. runtime
10^3
10^2
10^1
10^0
10^-1
Speed [Hz]
real-time
12
14
16
18
20
22
24
MOTA
LP2D
CEM
RMOT
SMOT
TBD
TC_ODAL
DP_NMS
MotiCon

Current State-Of-the-Art

Tracker	Avg Rank	↑MOTA		MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
JointMC	27.9	35.6	±18.9	71.9	1.8	23.2%	39.3%	10,580	28,508	457 (8.5)	969 (18.1)	0.6	Public
14.	M. Keuper, S. Tang, Y. Zhongjie, B. Andres, T. Brox, B. Schiele. A Multi-cut Formulation for Joint Segmentation and Tracking of Multiple Objects. In , 2016.												
PRMOT	25.6	35.6	±14.5	69.8	1.4	17.8%	40.5%	7,915	31,303	338 (6.9)	1,066 (21.7)	2.2	Public
15.	Anonymous submission												
mLK	22.8	35.1	±12.9	71.5	1.0	12.3%	38.3%	5,678	33,815	383 (8.5)	1,175 (26.1)	1.0	Public
17.	Yuan Zhang, Di Xie and Shiliang Pu (Hikvision Research Institute)												
HybridDAT	24.1	35.0	±15.0	72.6	1.5	11.4%	42.2%	8,455	31,140	358 (7.3)	1,267 (25.7)	4.6	Public
18.	Anonymous submission												
TSMLCDEnew	23.1	34.3	±13.1	71.7	1.4	14.0%	39.4%	7,869	31,908	618 (12.9)	959 (20.0)	6.5	Public
19.	B. Wang, G. Wang, K. L. Chan, L. Wang. Tracklet Association by Online Target-Specific Metric Learning and Coherent Dynamics Estimation. In arXiv:1511.06654, 2015.												
NOMT	21.6	33.7	±16.2	71.9	1.3	12.2%	44.0%	7,762	32,547	442 (9.4)	823 (17.5)	11.5	Public
20.	W. Choi. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor. In ICCV, 2015.												

MOTA +12%

Current State-Of-the-Art

Tracker	Avg Rank	↑MOTA	MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
NOMTwSDP	7.5	**55.5** 11.2	**76.6**	1.0	39.0%	25.8%	5,594	21,322	427 (6.5)	**701** (10.7)	6.4	Private
1.												W. Choi. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor. In ICCV, 2015.
FOMT	18.4	53.0 12.2	74.8	1.2	32.7%	14.6%	6,974	20,776	1,143 (17.3)	2,043 (30.9)	16.0	Private
2.												Anonymous submission
EAMTT	15.4	53.0 11.1	75.3	1.3	35.9%	19.6%	7,538	20,590	776 (11.7)	1,269 (19.1)	11.5	Private
3.												Anonymous submission
AMPL	13.6	51.9 11.9	75.0	1.2	26.4%	24.8%	6,963	22,225	372 (5.8)	1,130 (17.7)	2.8	Private
4.												Anonymous submission
LKDAT_CNN	17.1	49.3 11.8	74.5	1.0	20.8%	28.4%	6,009	24,550	563 (9.4)	1,155 (19.2)	1.2	Private
5.												Yuan Zhang, Di Xie and Shiliang Pu (Hikvision Research Institute)
MOT_DL	20.2	49.1 12.9	73.9	1.5	35.4%	25.0%	8,488	22,281	511 (8.0)	1,390 (21.8)	3.9	Private
6.												Anonymous submission
TSML_CDE	13.9	49.1 13.0	74.3	0.9	30.4%	26.4%	5,204	25,460	637 (10.9)	1,034 (17.7)	6.5	Private
7.												B. Wang, G. Wang, K. L. Chan, L. Wang. Tracklet Association by Online Target-Specific Metric Learning and Coherent Dynamics Estimation. In arXiv:1511.06654, 2015.
MDP_SubCNN	22.4	47.5 11.9	74.2	1.5	30.0%	18.6%	8,631	22,969	628 (10.0)	1,370 (21.9)	2.1	Private
8.												Y. Xiang, A. Alahi, S. Savarese. Learning to Track: Online Multi-Object Tracking by Decision Making. In International Conference on Computer Vision (ICCV), 2015.

MOTA +20%

Do we Need a New Dataset?

- Methods reached a MOTA plateau
- Annotations are inaccurate, especially in the sequences with moving camera
- Detections have a poor recall

MOT 2016 EDITION

- Increase the challenge!
- 14 all new sequences
- Improved annotations:
  - Increasing the accuracy of the annotations
  - Annotating occluding elements, distractors, vehicles…

Edition Comparison

	MOT 2015	MOT 2016
Pedestrian BB	101345	292733
Total BB	101345	476532
Tracks	1221	1342
Density	8.9	25.8
Annotation classes	1	11
HD sequences	27 %	86 %

Annotations

Target

Pedestrian

- Walking
- Running
- Standing

On vehicle

Ambiguous

Static person

- Sitting
- Lying
- ...

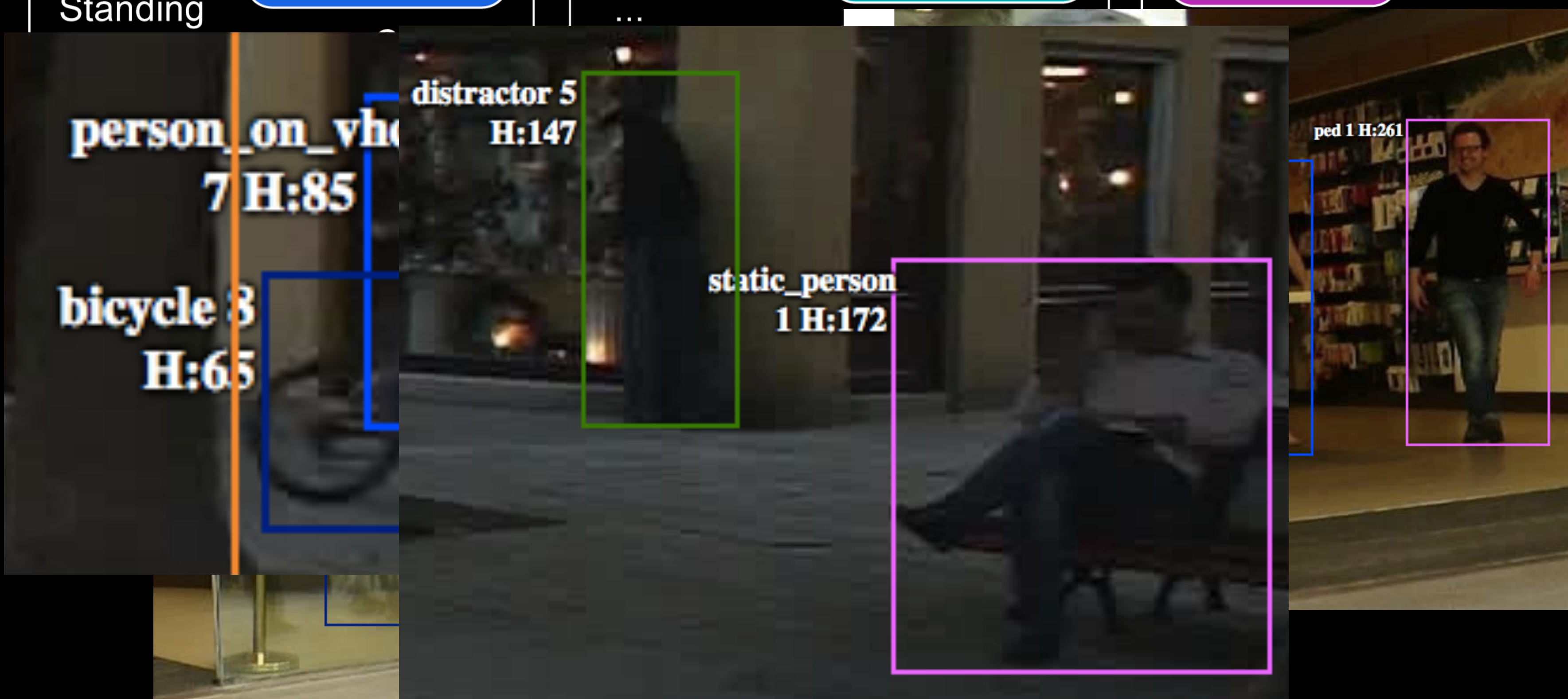
Distractor

Other

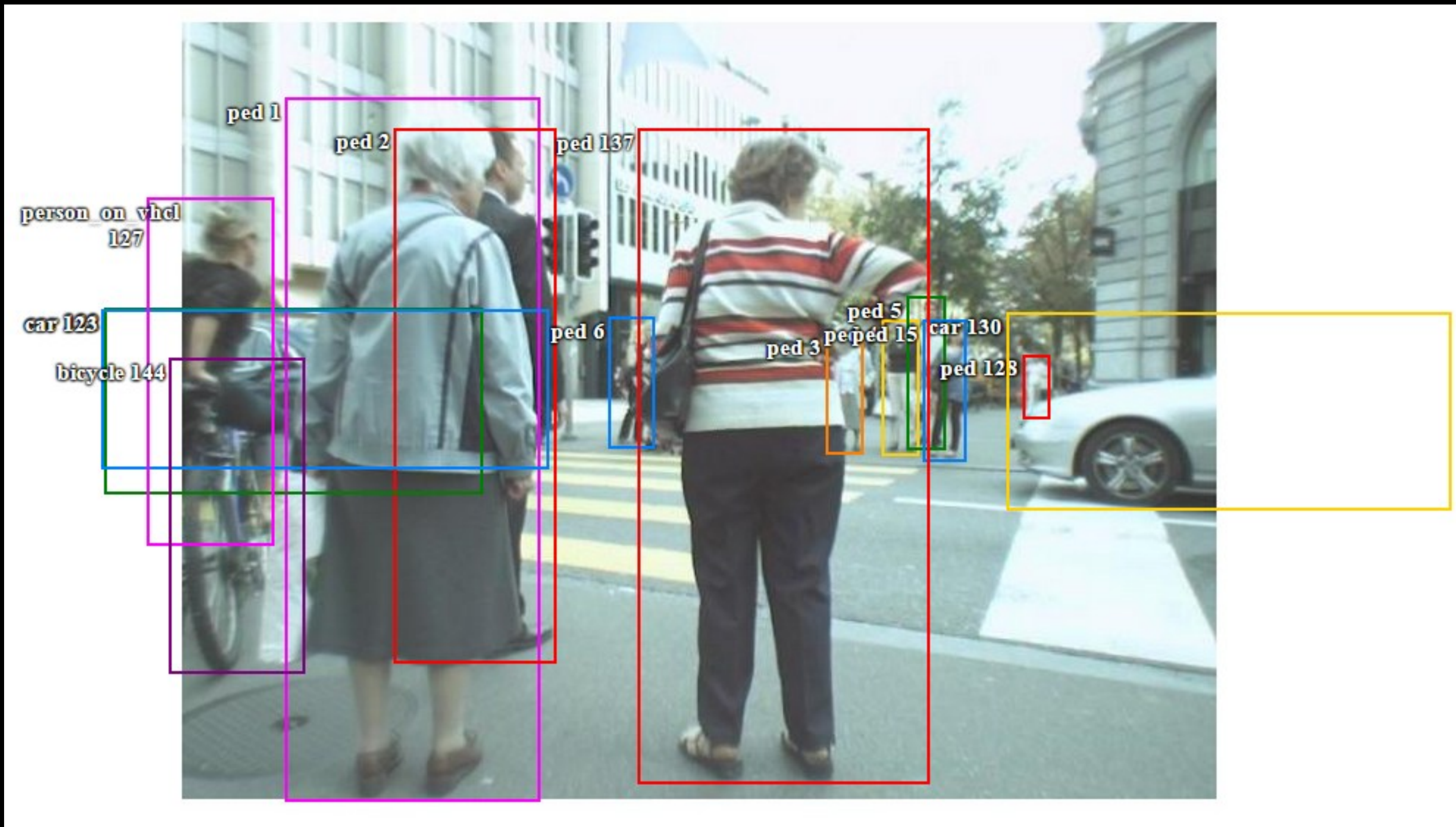
Car

Motorbike

Occluder

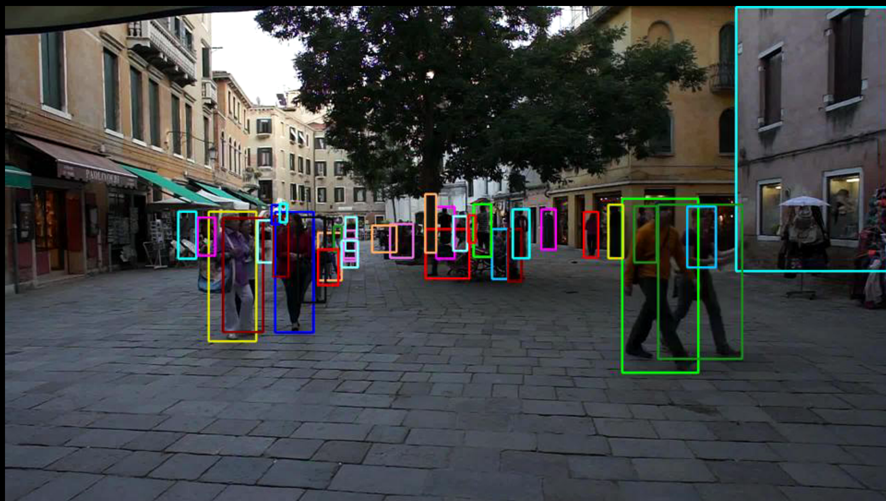


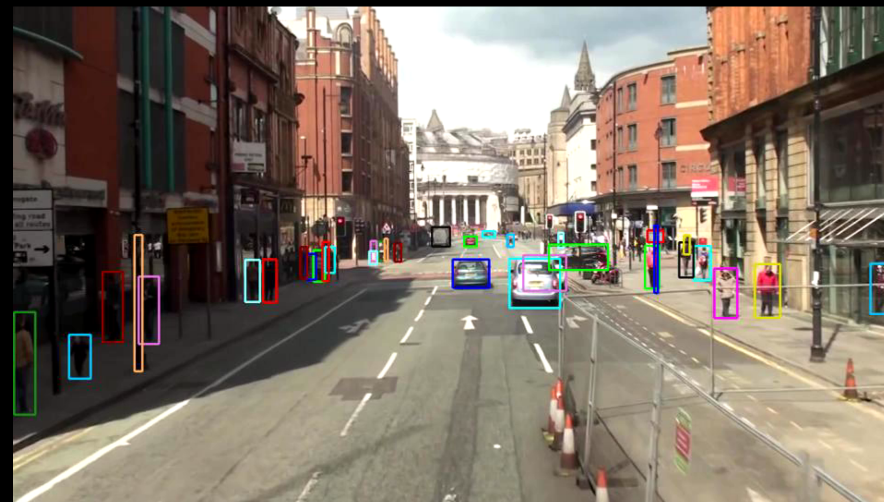
ANNOTATIONS

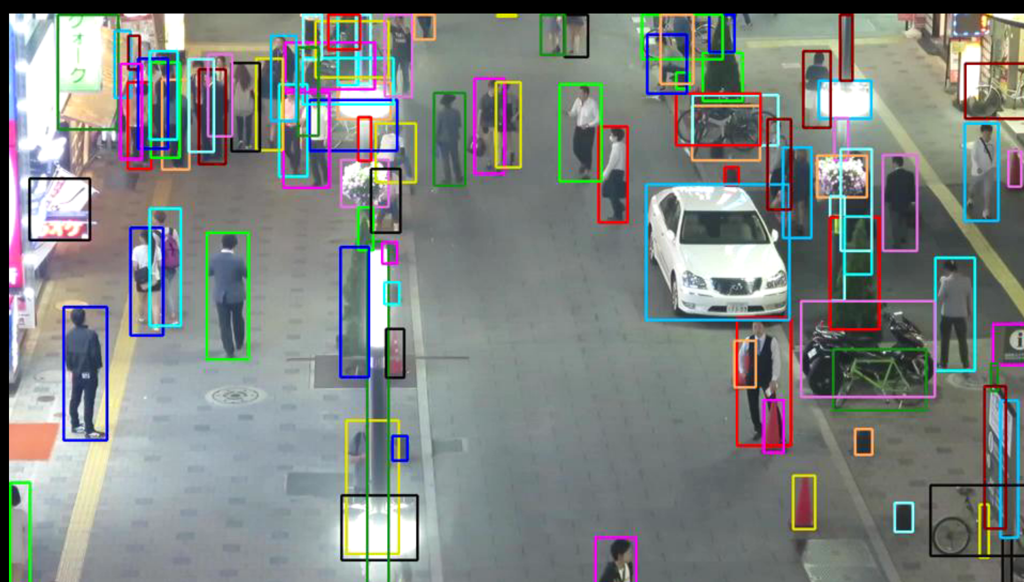


MOT 2016 EDITION

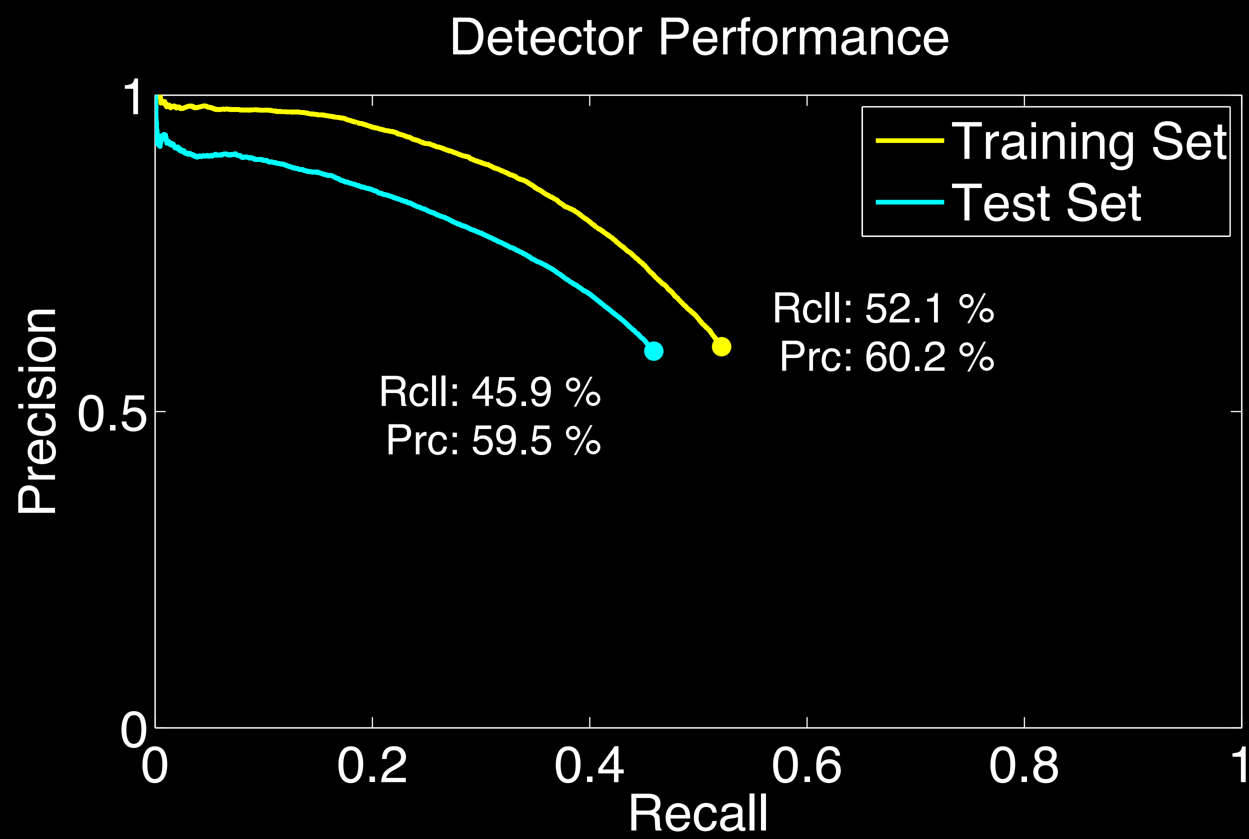
MOT 2016 EDITION





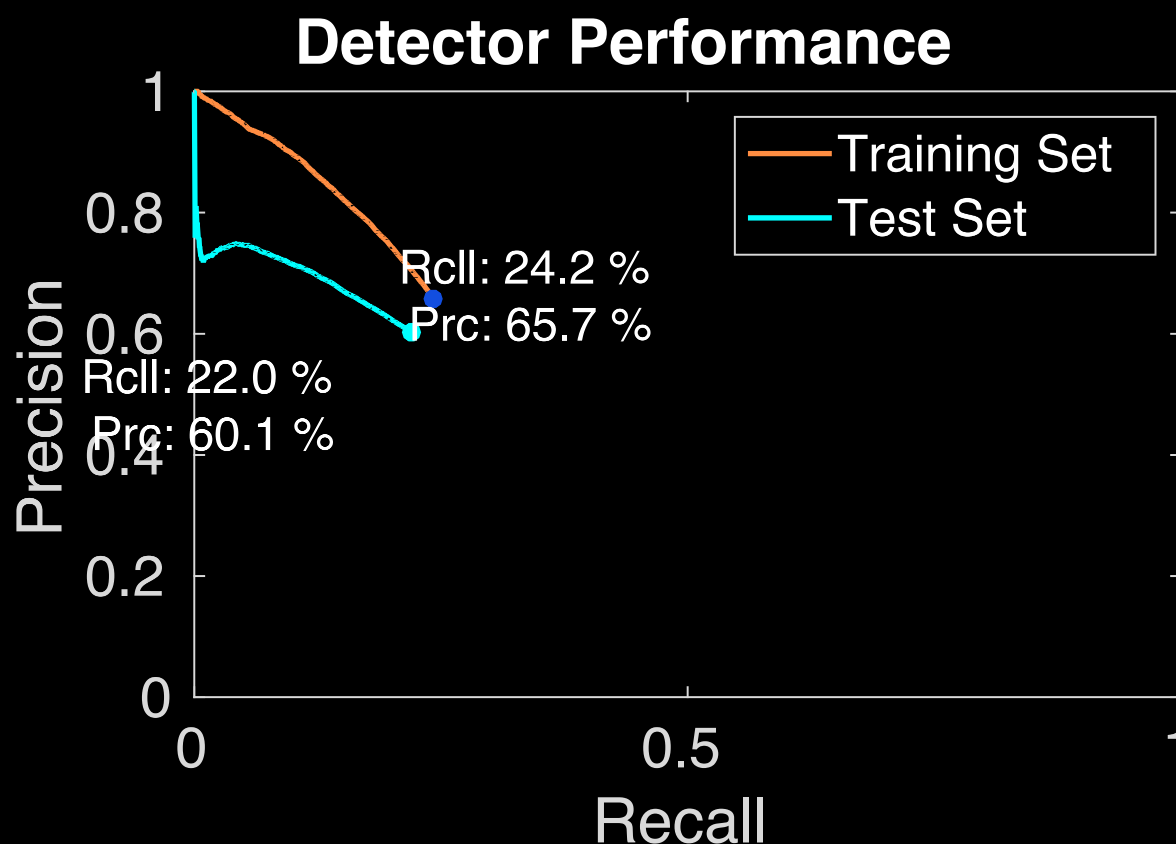


Detections



P. Dollár, R. Appel, S. Belongie, and P. Perona. *Fast feature pyramids for object detection*. PAMI, 2014.

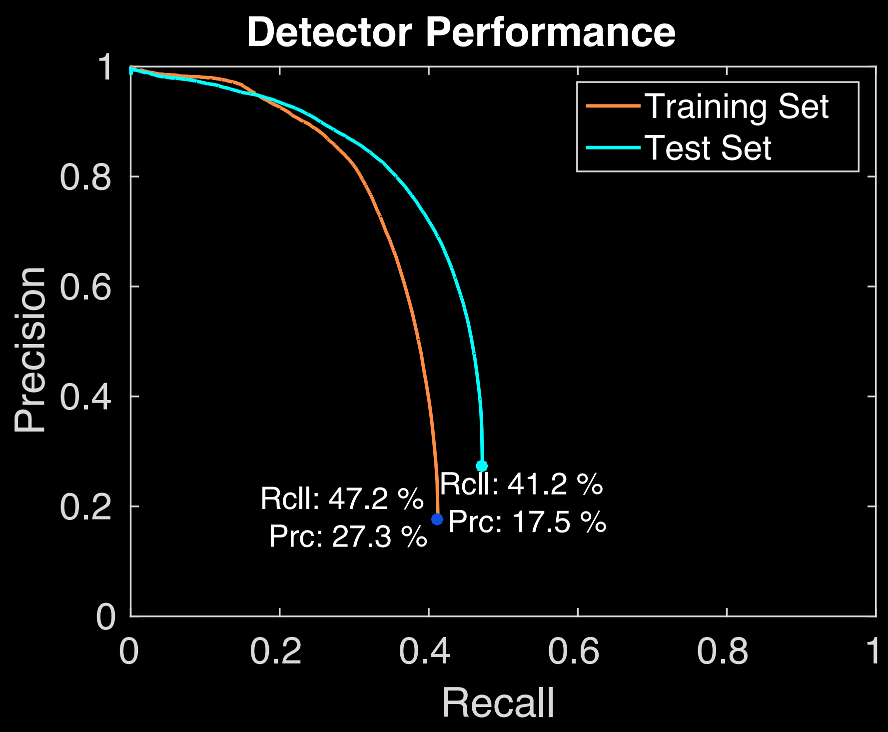
DETECTIONS



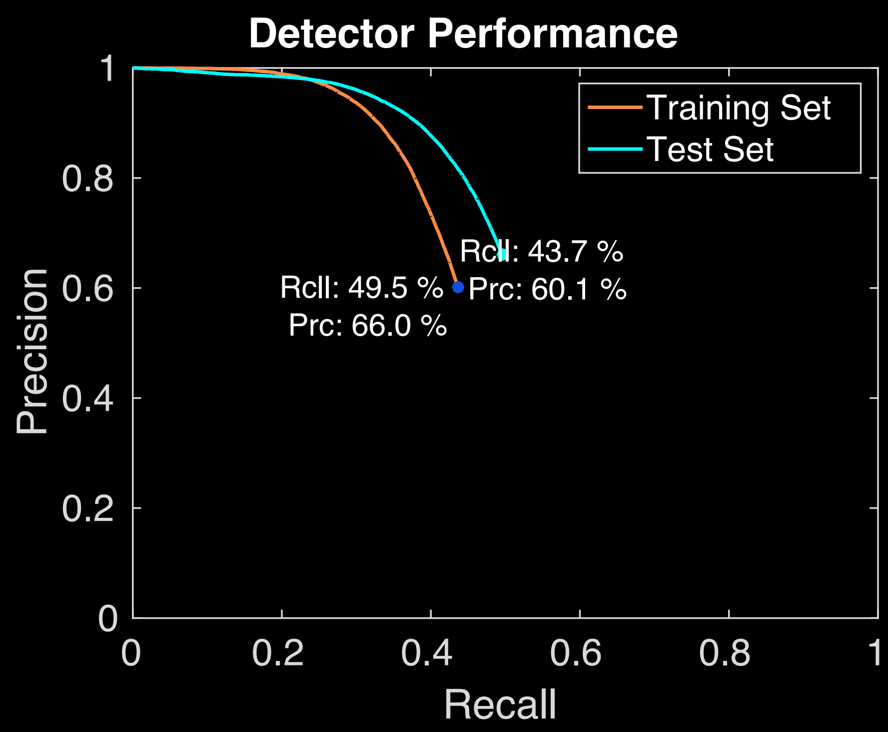
P. Dollár, R. Appel, S. Belongie, and P. Perona. *Fast feature pyramids for object detection*. PAMI, 2014.

DETECTIONS

[1]



[2]



[1] R. Girshick. Fast R-CNN. ICCV 2015

[2] P.F.Felzenszwalb,R.B.Girshick,D.McAllester,andD.Ramanan. *Object detection with discriminatively trained part based models.* PAMI, 2010.

Baselines

Method	MOTA	MOTP	FAR	MT(%)	ML(%)
TBD	33.3 ± 9.6	76.5	1.0	6.6	58.2
CEM	32.6 ± 8.6	75.9	1.3	7.0	59.4
DP_NMS	31.9 ± 9.9	76.4	0.2	4.8	65.2
SMOT	29.2 ± 7.9	75.2	3.0	4.9	53.3
JPDA_m	25.9 ± 6.4	76.4	0.7	3.7	70.4

MOTA +10% vs MOT15

Current State-Of-the-Art

Tracker	Avg Rank	↑MOTA		MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
NOMT	9.5	46.4	±9.9	76.6	1.6	18.3%	41.4%	9,753	87,565	359 (6.9)	**504** (9.7)	2.6	Public
8.	W. Choi. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor. In ICCV, 2015.												
JMC	11.6	46.3	±9.0	75.7	1.1	15.5%	39.7%	6,373	90,914	657 (13.1)	1,114 (22.2)	0.8	Public
9.	S. Tang, B. Andres, M. Andriluka, B. Schiele. Subgraph Decomposition for Multi-Object Tracking. In Computer Vision and Pattern Recognition (CVPR), 2015.												
YGT	15.4	45.2	±9.0	73.9	2.2	18.6%	41.9%	13,130	86,090	676 (12.8)	1,008 (19.1)	0.6	Public
10.	Anonymous submission												
SMMUML	13.9	43.3	±13.6	74.3	1.4	11.9%	42.8%	8,463	93,892	985 (20.3)	1,509 (31.1)	182.7	Public
11.	Anonymous submission												
oICF	15.1	43.2	±10.2	74.3	1.1	11.3%	48.5%	6,651	96,515	381 (8.1)	1,404 (29.8)	0.4	Public
12.	H. Kieritz, S. Becker, W. Hübner, M. Arens. Online Multi-Person Tracking using Integral Channel Features. In IEEE Advanced Video and Signal-based Surveillance (AVSS) 2016, 2016.												
MHT_DAM	11.6	42.9	±8.9	76.6	1.0	13.6%	46.9%	5,668	97,919	499 (10.8)	659 (14.2)	0.8	Public
13.	C. Kim, F. Li, A. Ciptadi, J. Rehg. Multiple Hypothesis Tracking Revisited. In ICCV, 2015.												
LINF1	15.0	41.0	±9.5	74.8	1.3	11.6%	51.3%	7,896	99,224	430 (9.4)	963 (21.1)	1.1	Public
14.	L. Fagot-Bouquet, R. Audigier, Y. Dhome, F. Lerasle. Improving Multi-Frame Data Association with Sparse Representations for Robust Near-Online Multi-Object Tracking. In ECCV, (to appear) 2016.												

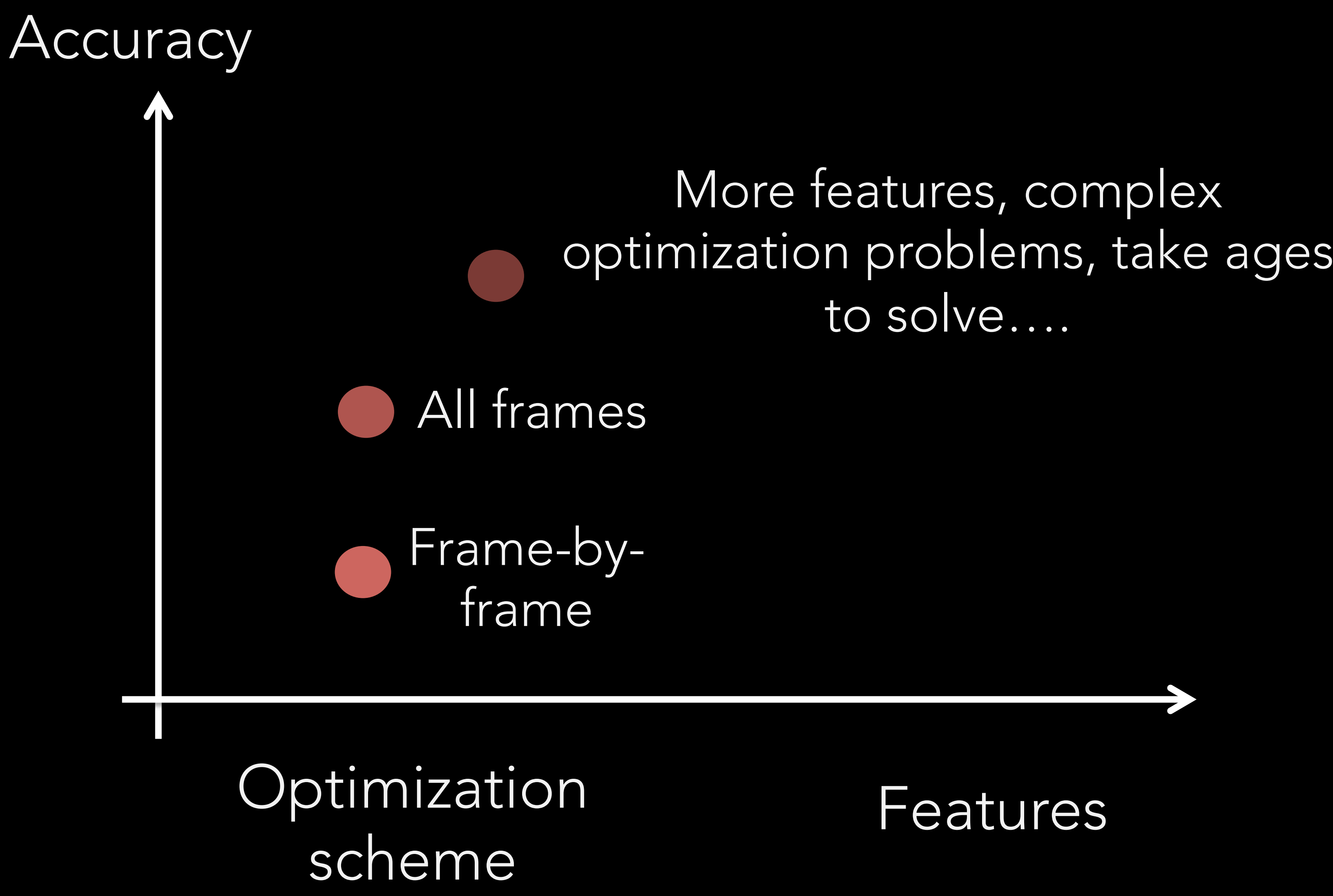
MOTA +13%

Current State-Of-the-Art

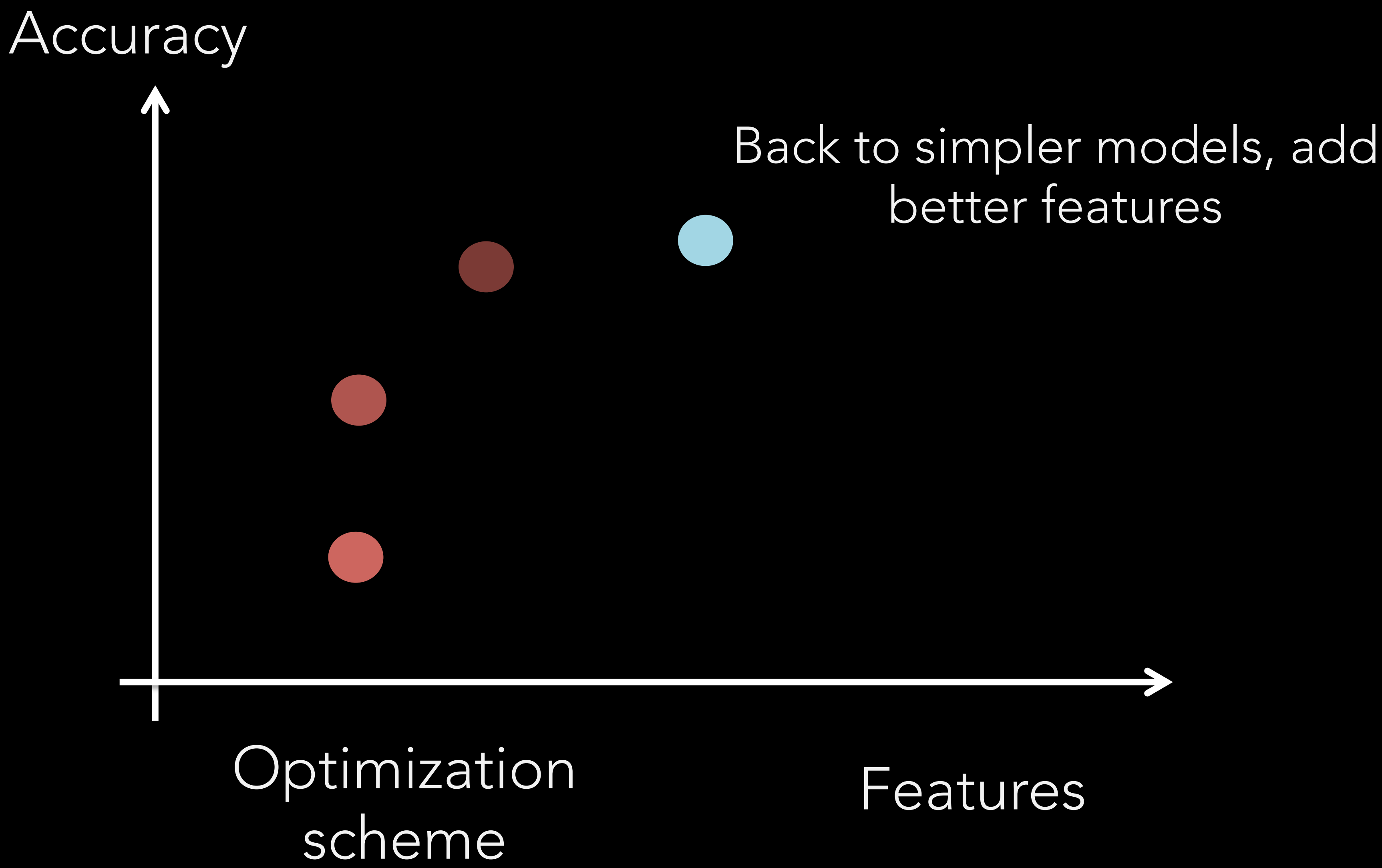
Tracker	Avg Rank	MOTA		MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
KDNT	9.6	**68.2**	±12.9	79.4	1.9	**41.0%**	**19.0%**	11,479	**45,605**	933 (12.4)	1,093 (14.6)	0.7	Private
1.													F. Yu, W. Li, Q. Li, Y. Liu, X. Shi, J. Yan. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature. In BMTT, 2016, SenseTime Group Limited.
POI	7.1	66.1	±13.3	79.5	0.9	34.0%	20.8%	5,061	55,914	805 (11.6)	3,093 (44.6)	9.9	Private
2.													F. Yu, W. Li, Q. Li, Y. Liu, X. Shi, J. Yan. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature. In BMTT, 2016, SenseTime Group Limited.
MCMOT_HDM	9.1	62.4	±10.6	78.3	1.7	31.5%	24.2%	9,855	57,257	1,394 (20.3)	1,318 (19.2)	34.9	Private
3.													B. Lee, E. Erdenee, S. Jin, M. Nam, Y. Jung, P. Rhee. Multi-Class Multi-Object Tracking using Changing Point Detection. In BMTT, 2016.
NOMTwSDP16	5.3	62.2	±11.0	**79.6**	0.9	32.5%	31.1%	5,119	63,352	406 (6.2)	642 (9.8)	3.1	Private
4.													W. Choi. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor. In ICCV, 2015.
KFILDAwSDP	13.6	57.3	±15.9	77.5	2.6	24.6%	25.3%	15,682	60,252	1,873 (28.0)	2,664 (39.8)	2.2	Private
5.													Anonymous submission
EAMTT	7.2	52.5	±11.4	78.8	0.7	19.0%	34.9%	4,407	81,223	910 (16.4)	1,321 (23.8)	12.2	Private
6.													Anonymous submission
AMPL	6.5	50.9	±7.1	77.0	0.5	16.7%	40.8%	3,229	86,123	**196** (3.7)	639 (12.1)	1.5	Private
7.													Anonymous submission

MOTA +22%

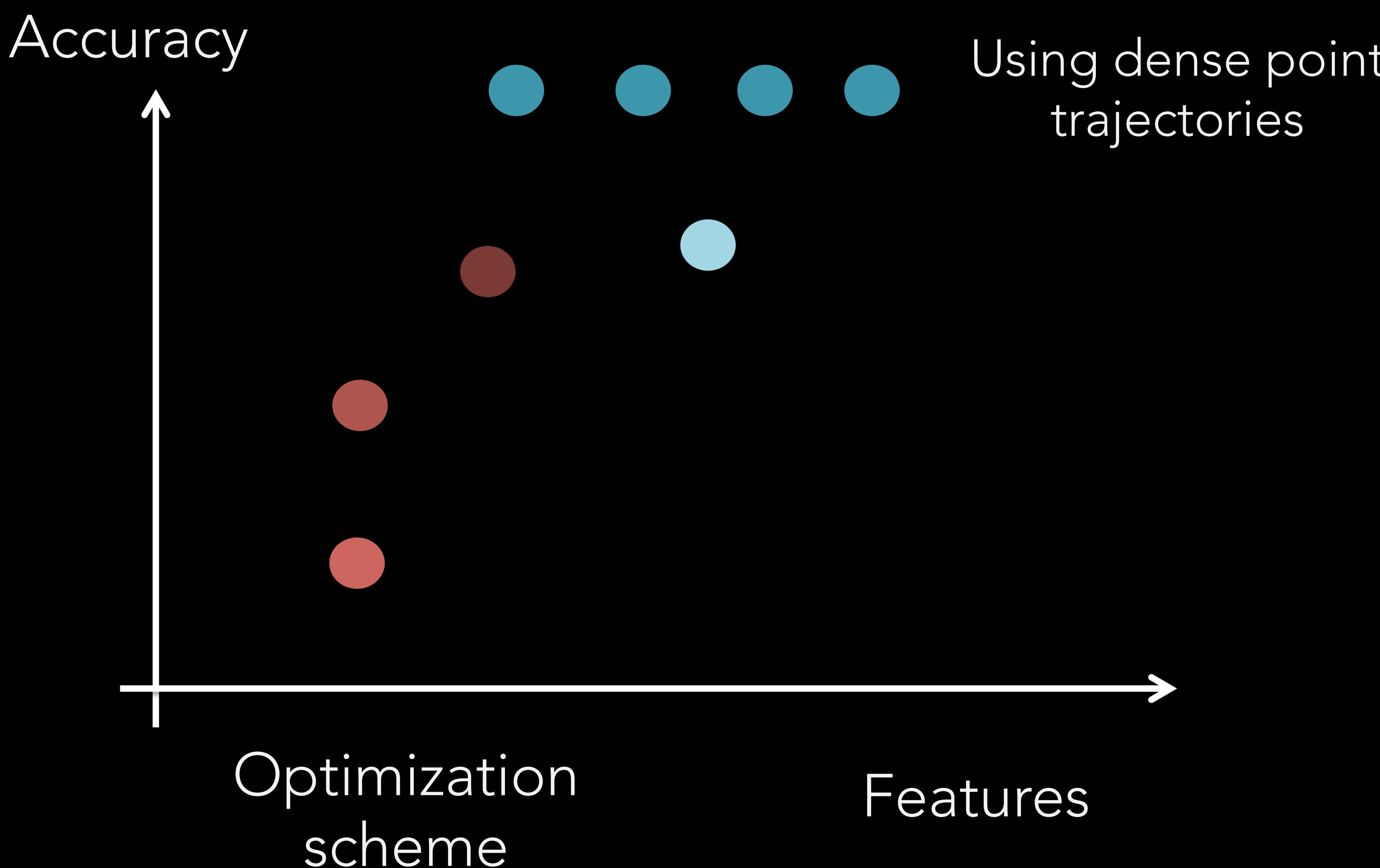
What Works?


Accuracy
More features, complex optimization problems, take ages to solve….
All frames
Frame-by-frame
Optimization scheme
Features

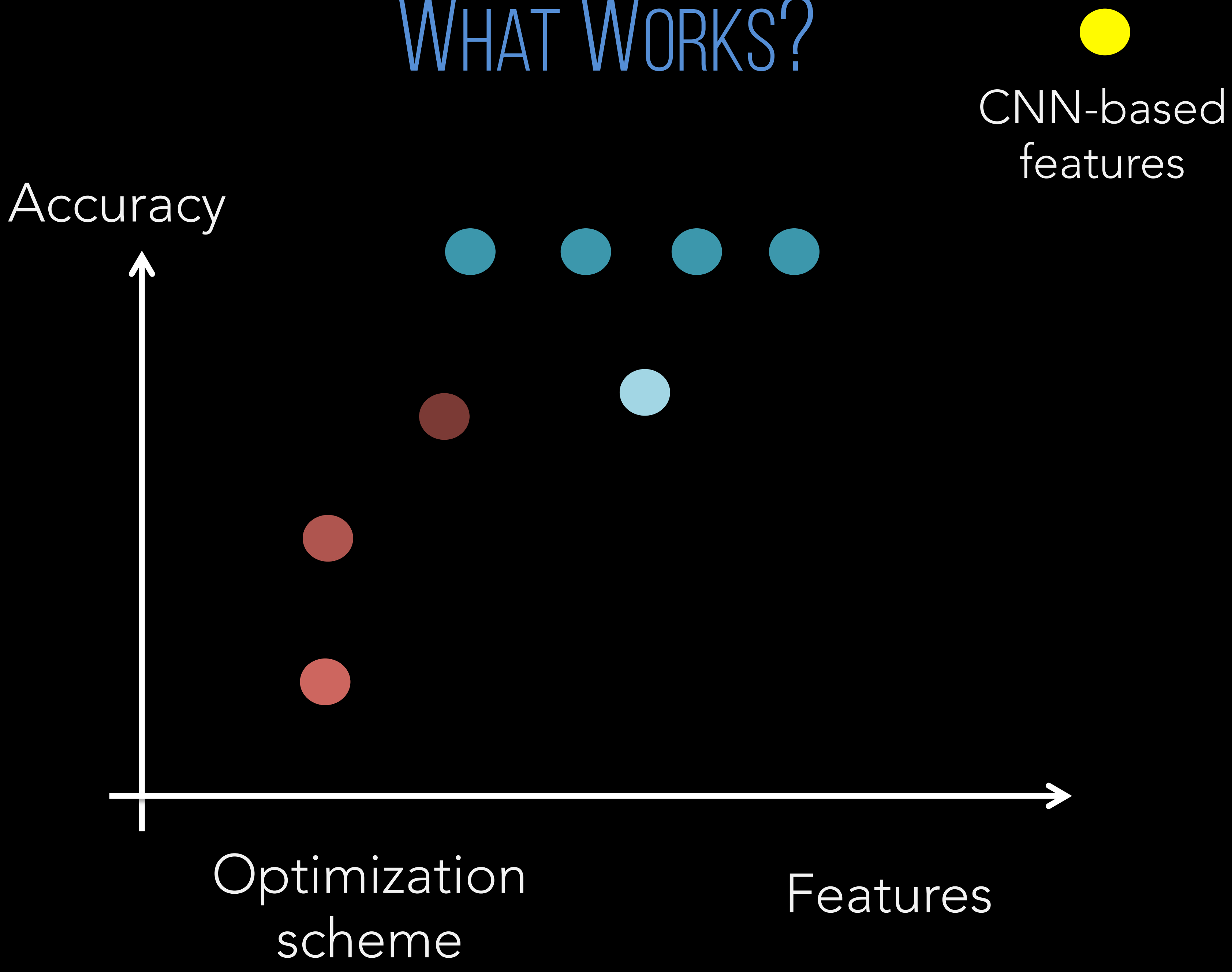
What Works?


Accuracy
Back to simpler models, add better features
Optimization scheme
Features

What Works?


Accuracy
Using dense point trajectories
Optimization scheme
Features

What Works?



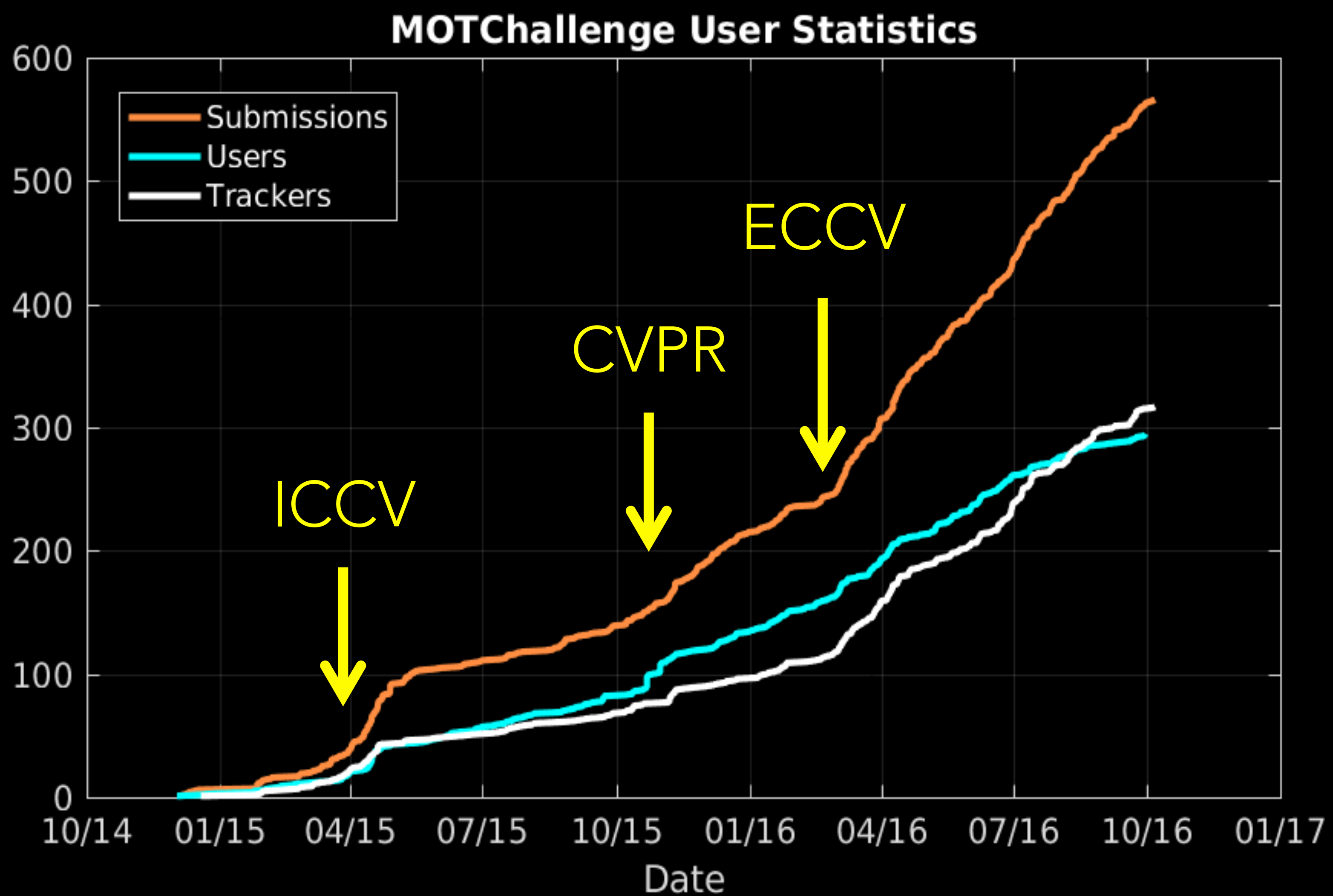
CNN-based
features
Accuracy
Optimization
scheme
Features

Current State-Of-the-Art

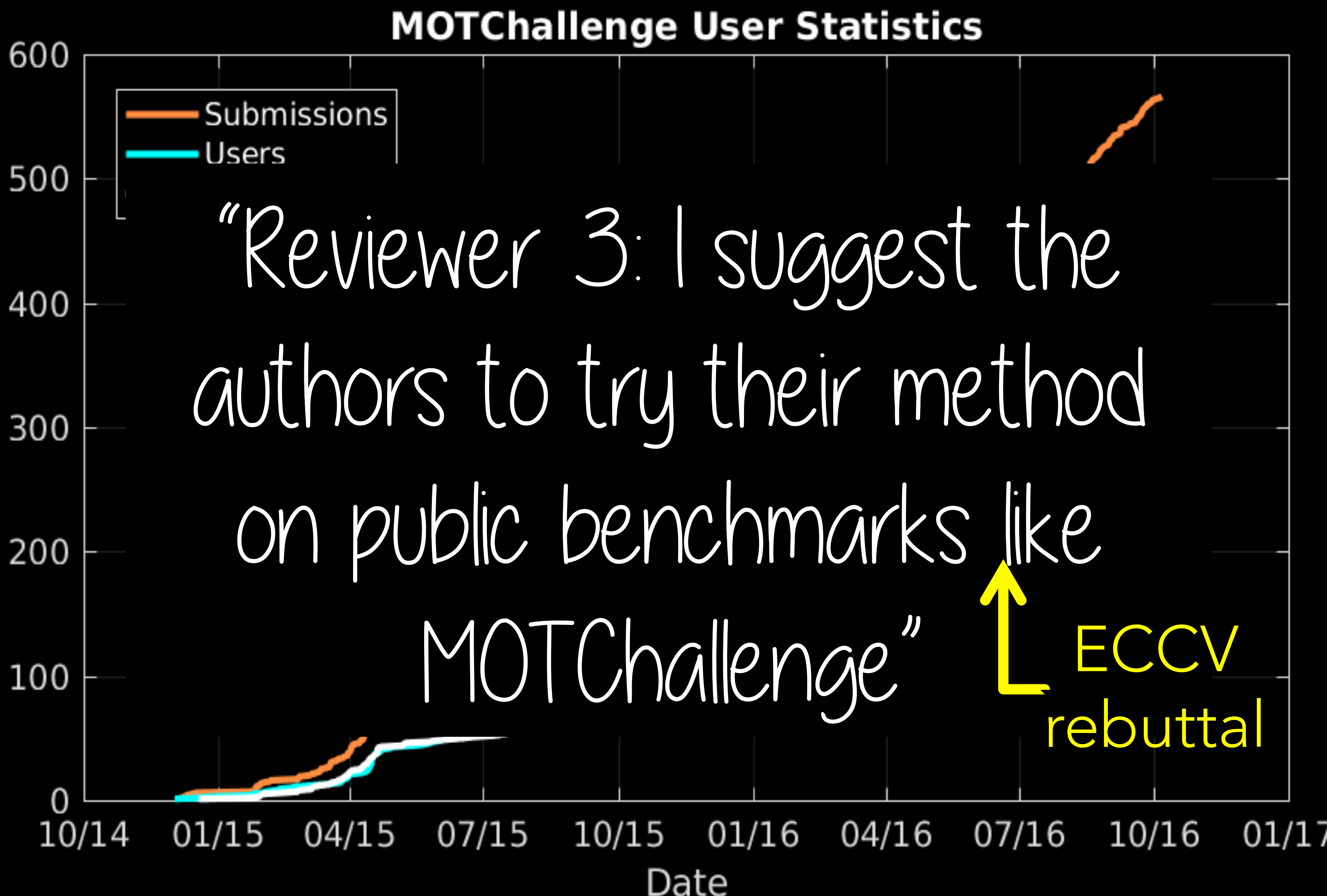
Tracker	Avg Rank	↑MOTA		MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
KDNT 1.	9.6	**68.2**	±12.9	79.4	1.9	**41.0%**	**19.0%**	11,479	**45,605**	933 (12.4)	1,093 (14.6)	0.7	Private
													F. Yu, W. Li, Q. Li, Y. Liu, X. Shi, J. Yan. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature. In BMTT, 2016, SenseTime Group Limited.
POI 2.	7.1	66.1	±13.3	79.5	0.9	34.0%	20.8%	5,061	55,914	805 (11.6)	3,093 (44.6)	9.9	Private
													F. Yu, W. Li, Q. Li, Y. Liu, X. Shi, J. Yan. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature. In BMTT, 2016, SenseTime Group Limited.
MCMOT_HDM 3.	9.1	62.4	±10.6	78.3	1.7	31.5%	24.2%	9,855	57,257	1,394 (20.3)	1,318 (19.2)	34.9	Private
													B. Lee, E. Erdenee, S. Jin, M. Nam, Y. Jung, P. Rhee. Multi-Class Multi-Object Tracking using Changing Point Detection. In BMTT, 2016.
NOMTwSDP16 4.	5.3	62.2	±11.0	**79.6**	0.9	32.5%	31.1%	5,119	63,352	406 (6.2)	642 (9.8)	3.1	Private
													W. Choi. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor. In ICCV, 2015.
KFILDAwSDP 5.	13.6	57.3	±15.9	77.5	2.6	24.6%	25.3%	15,682	60,252	1,873 (28.0)	2,664 (39.8)	2.2	Private
													Anonymous submission
EAMTT 6.	7.2	52.5	±11.4	78.8	0.7	19.0%	34.9%	4,407	81,223	910 (16.4)	1,321 (23.8)	12.2	Private
													Anonymous submission
AMPL 7.	6.5	50.9	±7.1	77.0	0.5	16.7%	40.8%	3,229	86,123	**196** (3.7)	639 (12.1)	1.5	Private
													Anonymous submission

CNN-based appearance features

Evolution of the Benchmark



Evolution of the Benchmark



"Reviewer 3: I suggest the authors to try their method on public benchmarks like MOTChallenge"

ECCV rebuttal

What Challenges Lie Ahead?

Best performing method for MOT2015

↓

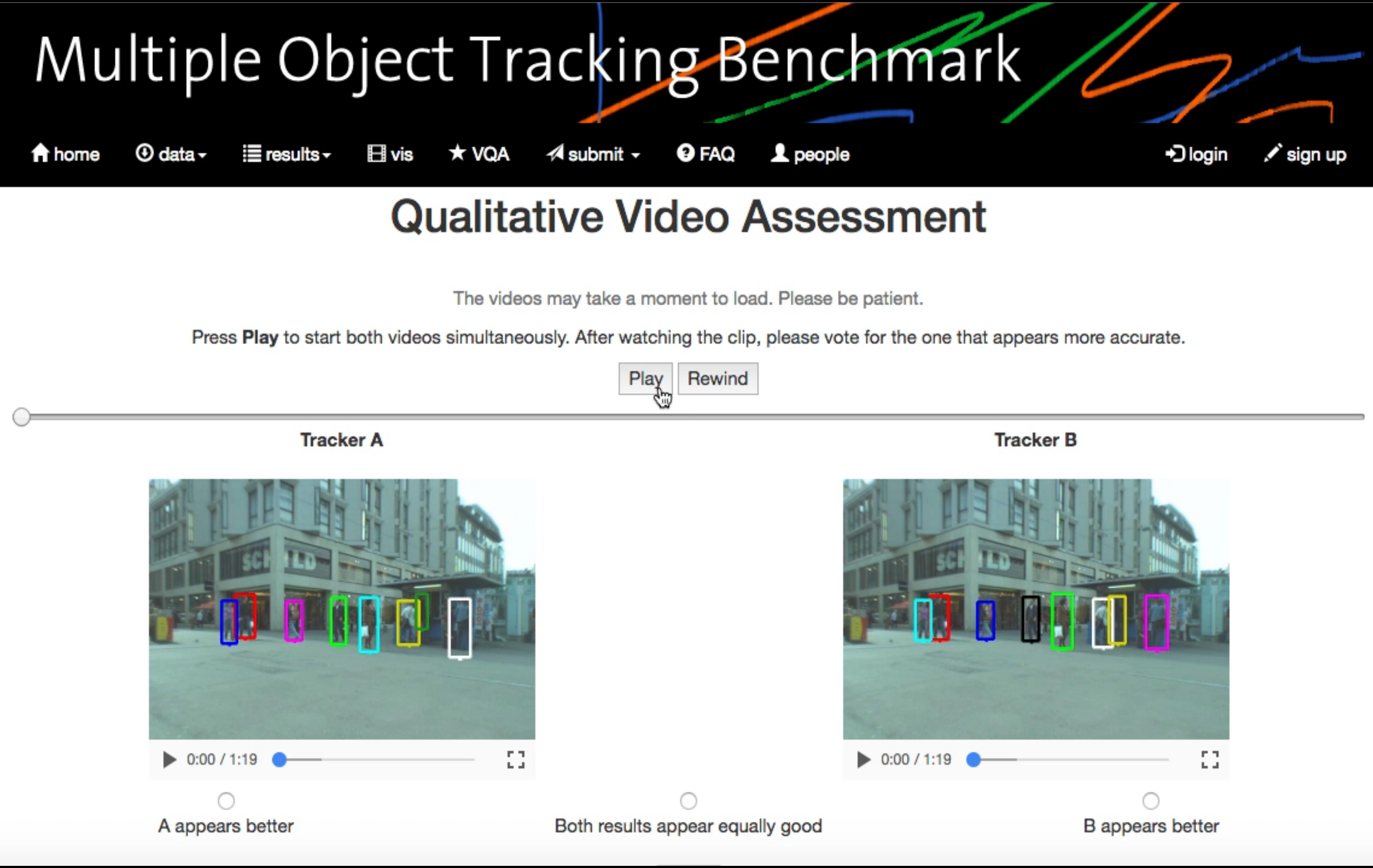
Well-known "solved" sequences reach on average MOTA of 60%

GENERALITY

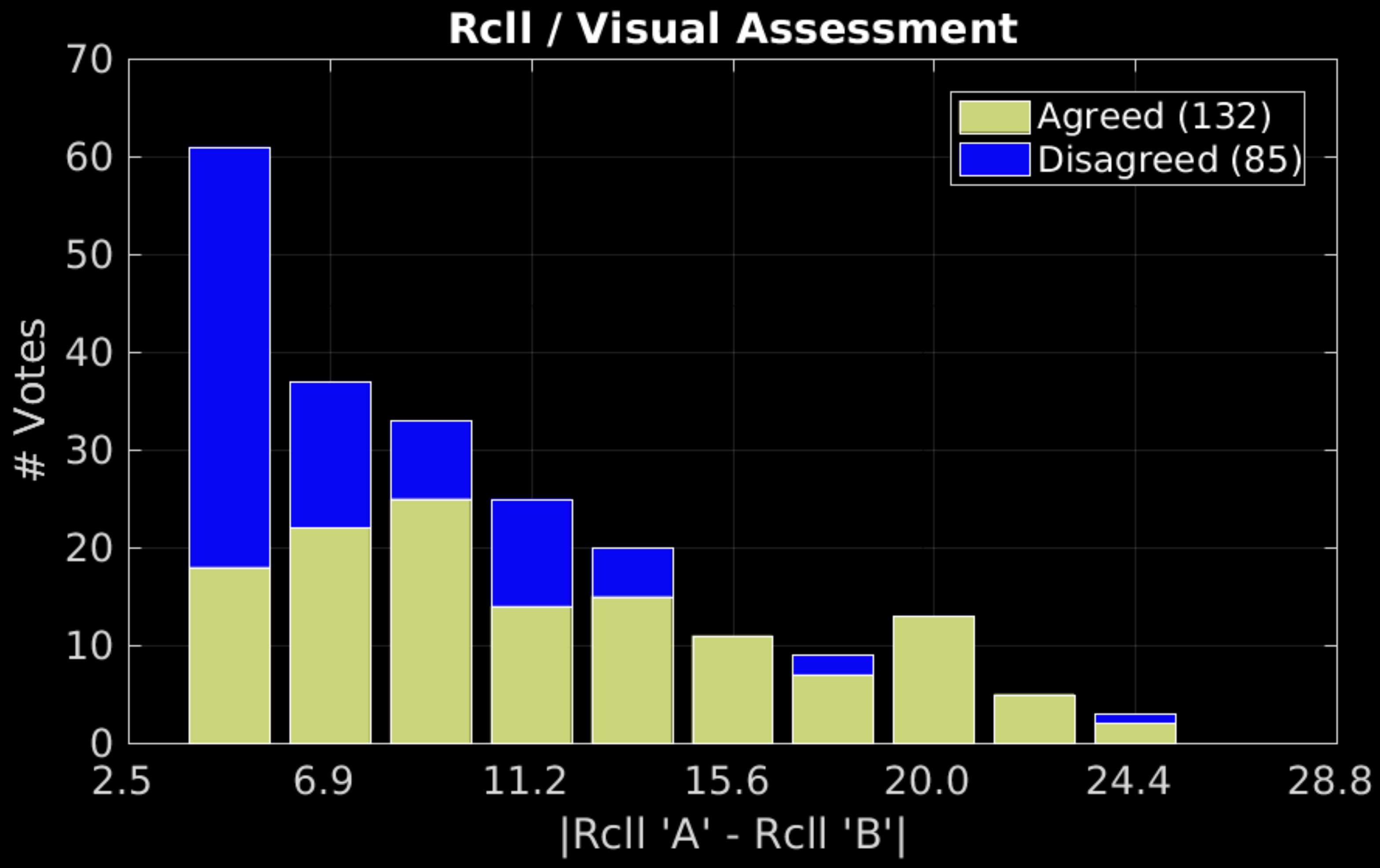
WHAT ABOUT MOTA?

- Is MOTA the right measure?
- Does it correspond to the human perception of a good tracker?
- Can we use it as a single measure to classify trackers?

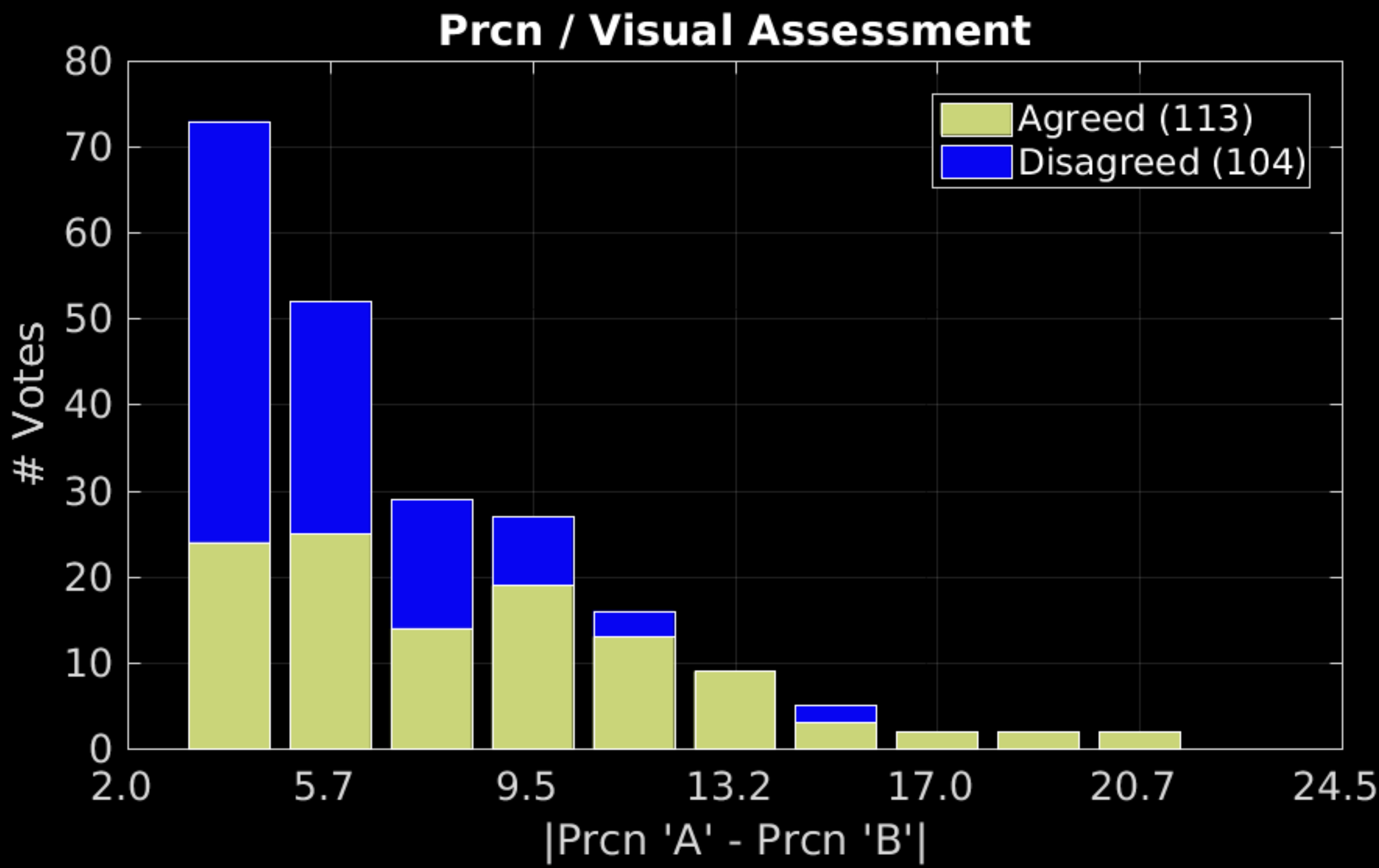
Hot-or-Not Tracker



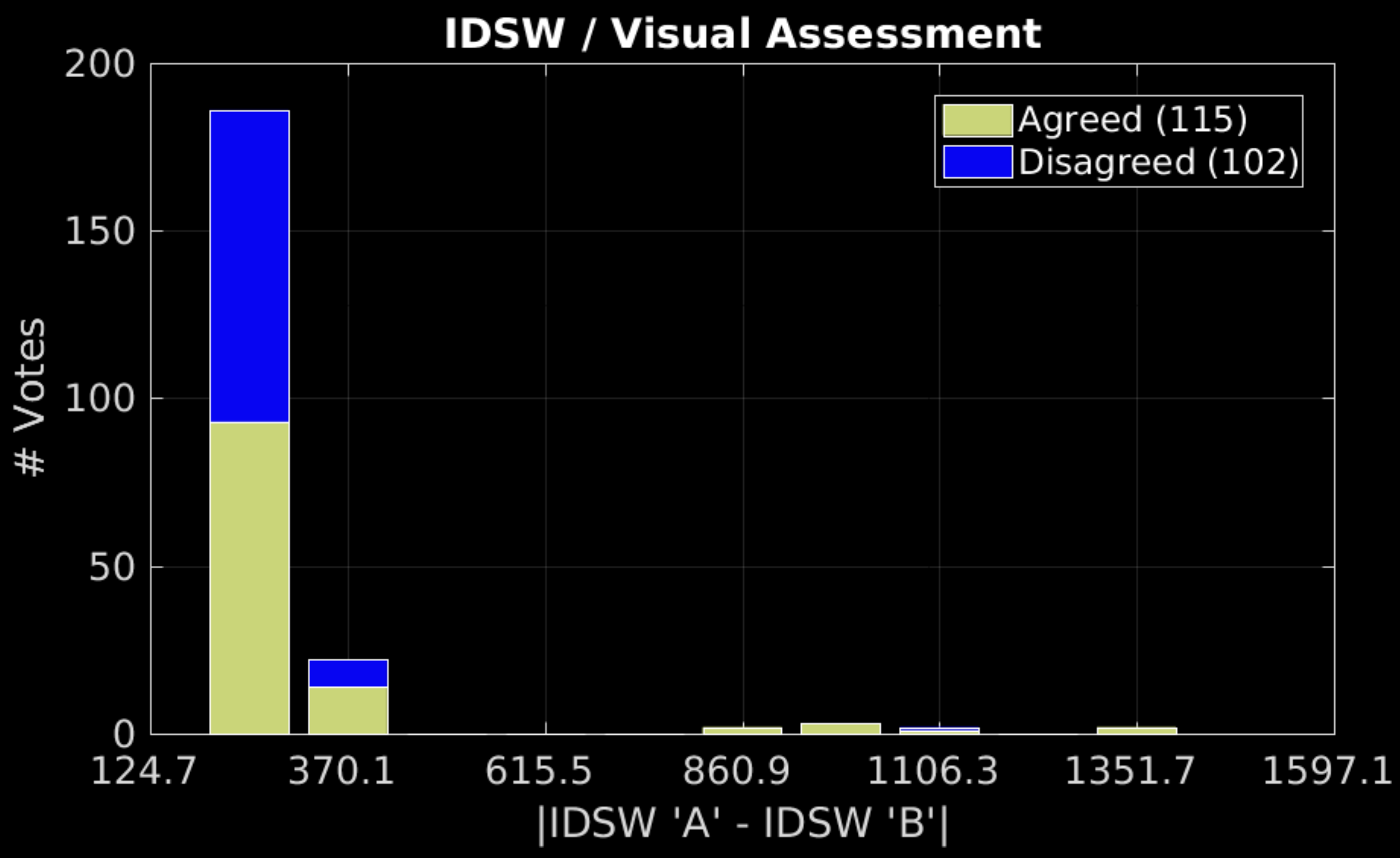
Hot-or-Not Tracker: Results



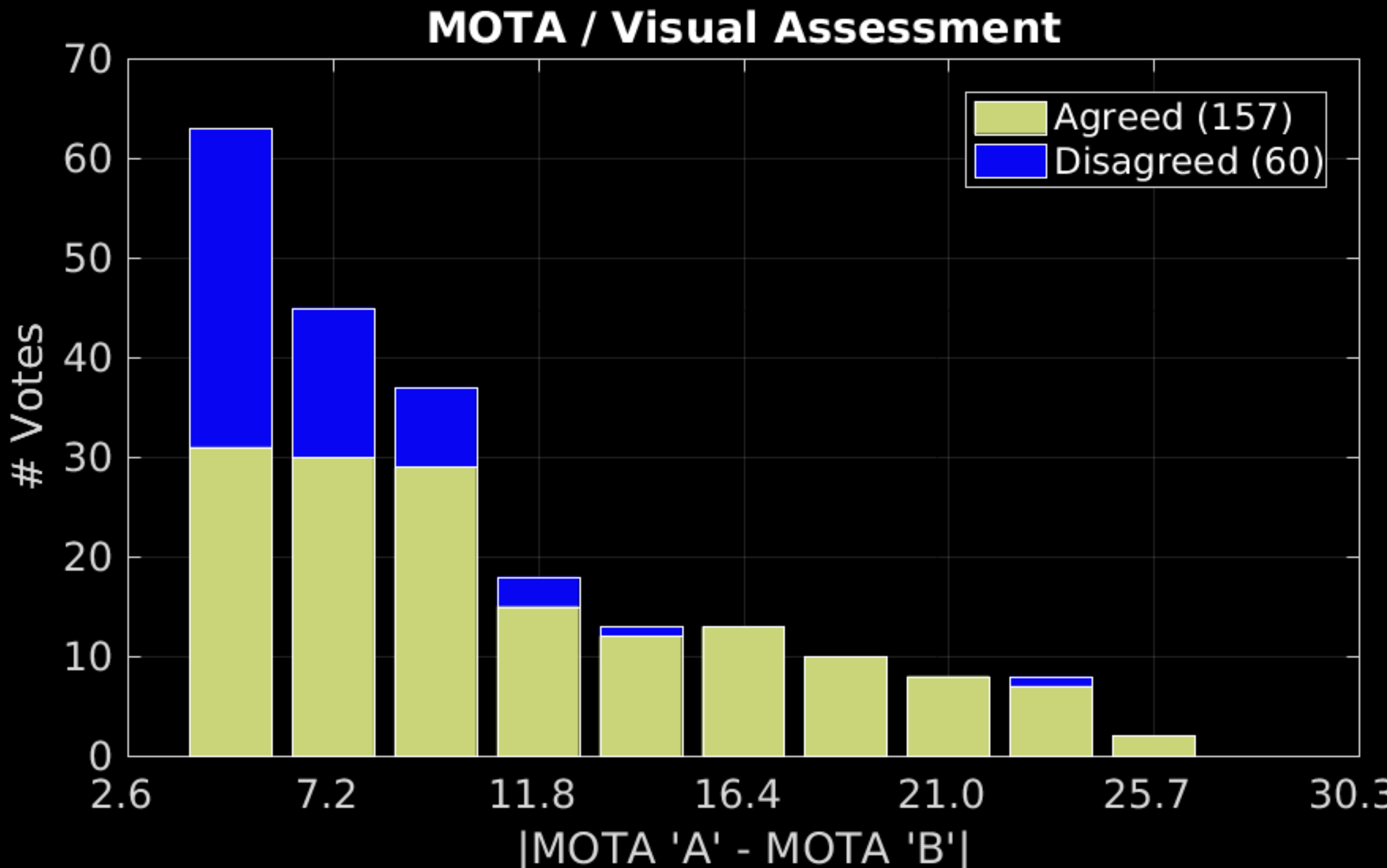
Hot-or-Not Tracker: Results



Hot-or-Not Tracker: Results

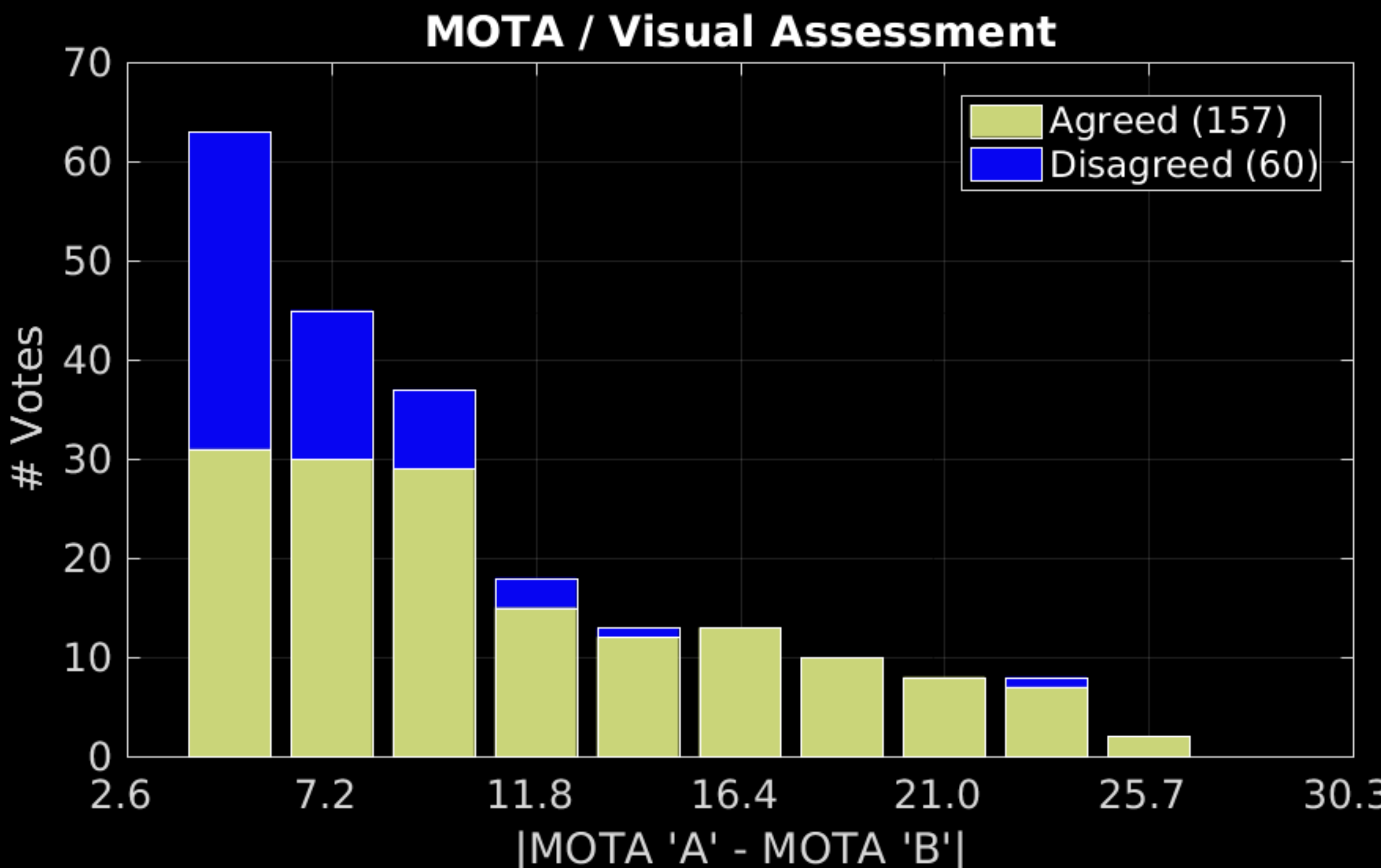


Hot-or-Not Tracker: Results



75% of the votes agree with MOTA

Hot-or-Not Tracker: Results



Disagreements come from MOTA differences < 10%

THE FUTURE OF MOTCHALLENGE

biological data

sports data

Expanding

First-person videos

SHARE YOUR TRACKING DATA

The Future of MOTChallenge

Open questions

Is this data enough for deep learning?

What would you like to see added?

Exciting Workshops and Challenges!

- 1st Workshop on Benchmarking Multi-Target Tracking organized together with WACV, 2015.

- 2nd Workshop TOMORROW!

www.motchallenge.net/workshops/bmtt2016

THE TEAM

Laura Leal-Taixé

Anton Milan

Konrad Schindler

Daniel Cremers

Stefan Roth

Ian Reid

THE UNIVERSITY of ADELAIDE

ETH

TECHNISCHE UNIVERSITÄT DARMSTADT

DAIMLER