Tracking without bells and whistles

ICCV 2019

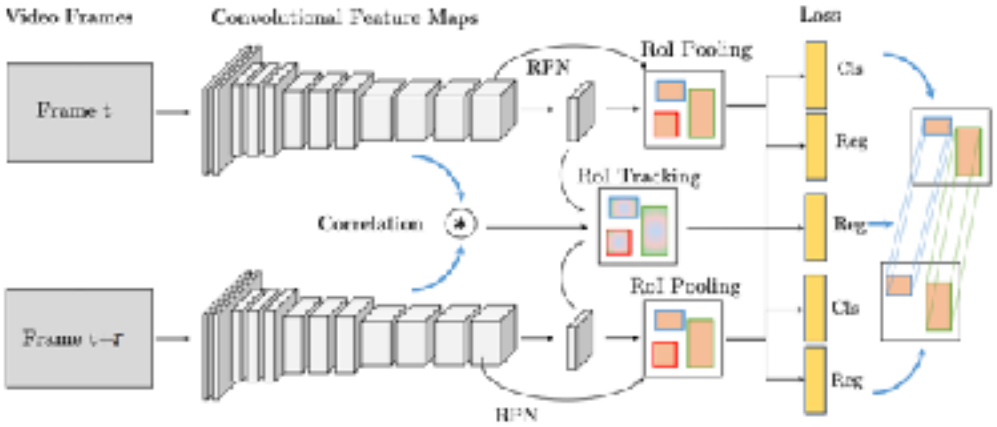
Philipp Bergmann, Tim Meinhardt, Laura Leal-Taixe

Problem Overview

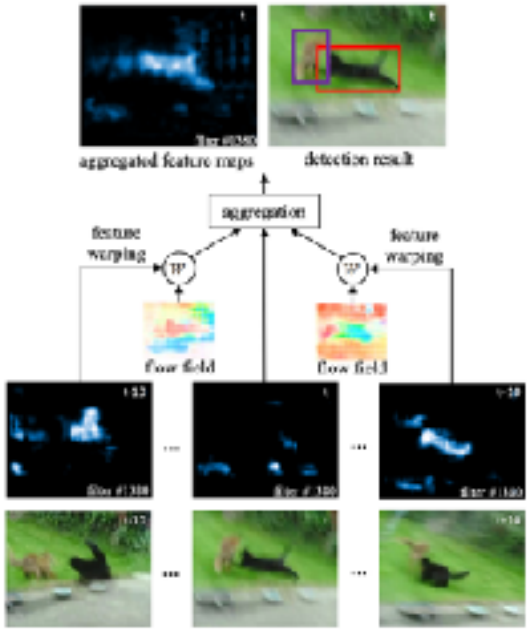
- The goal is to track multiple pedestrians in a given video sequence.

Prior Work

- Prior approaches are effective but complicated.



Detect to Track and Track to Detect, ICCV 2017



“Flow-Guided Feature Aggregation for Video Object Detection”, ICCV 2017

Motivation

- Spatial and temporal modeling require different types of models.
- Which of these are more important for the problem of object tracking?

Spatial Modeling:

- 2D Convolution Based.
- Large spatial resolution (~1000x1000).
- Static image inputs.

Temporal Modeling:

- 3D Convolution Based.
- Small spatial resolution (~200x200).
- Long video clips as inputs.

Frame-level detection

Tracking

Motivation

- Spatial and temporal modeling require different types of models.
- Which of these are more important for the problem of object tracking?

Spatial Modeling:

- 2D Convolution Based.
- Large spatial resolution (~1000x1000).
- Static image inputs.

Temporal Modeling:

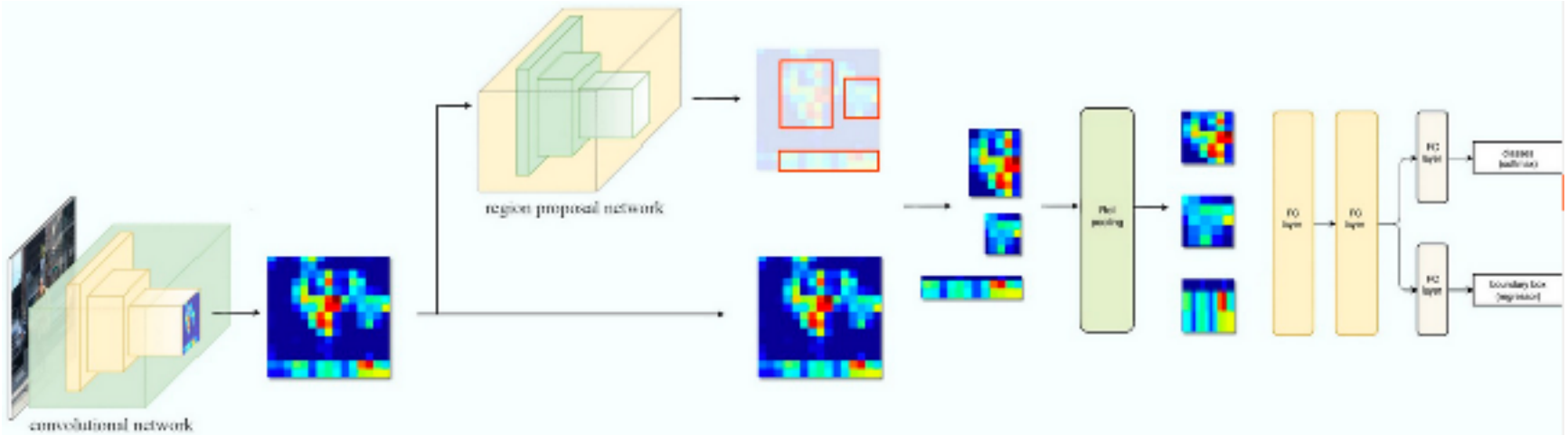
- 3D Convolution Based.
- Small spatial resolution (~200x200).
- Long video clips as inputs.

Frame-level detection

Tracking

Faster R-CNN

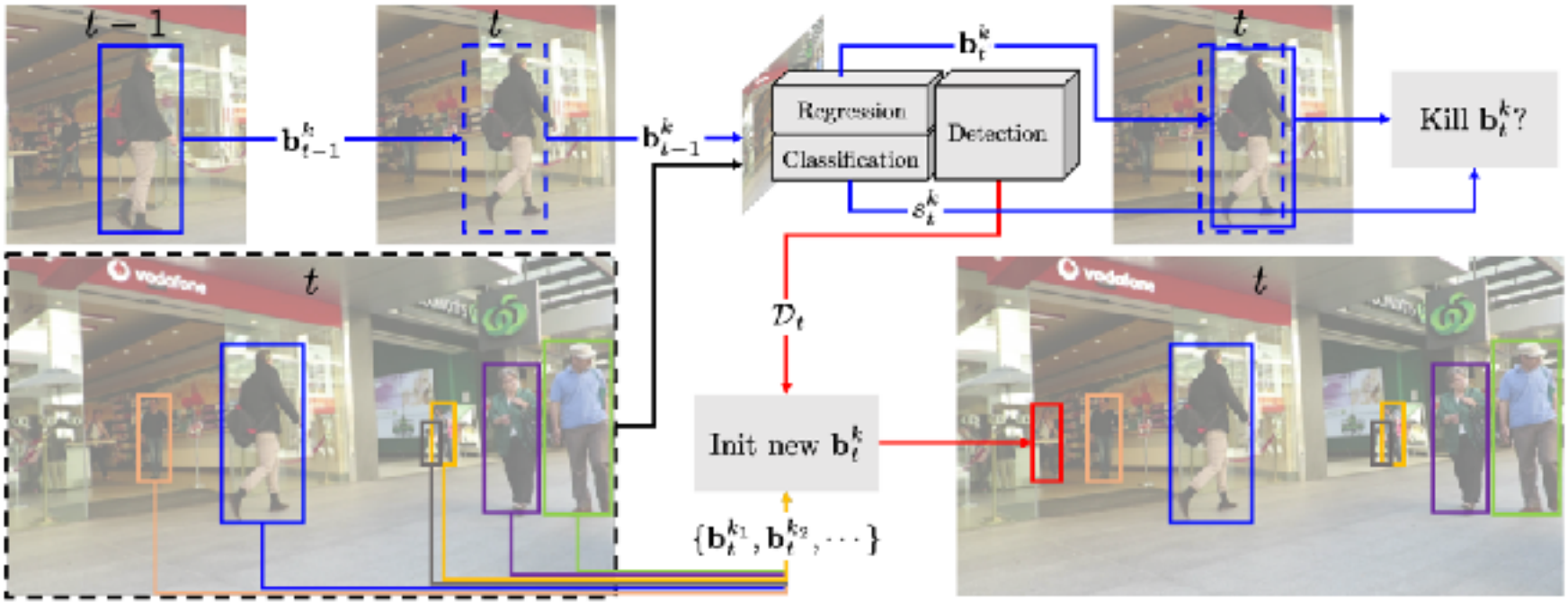
- The classification head assigns an object score to each region proposal (i.e., the likelihood of the proposal showing a pedestrian).
- The regression head refines the bounding box location tightly around an object.



Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, NIPS 2015

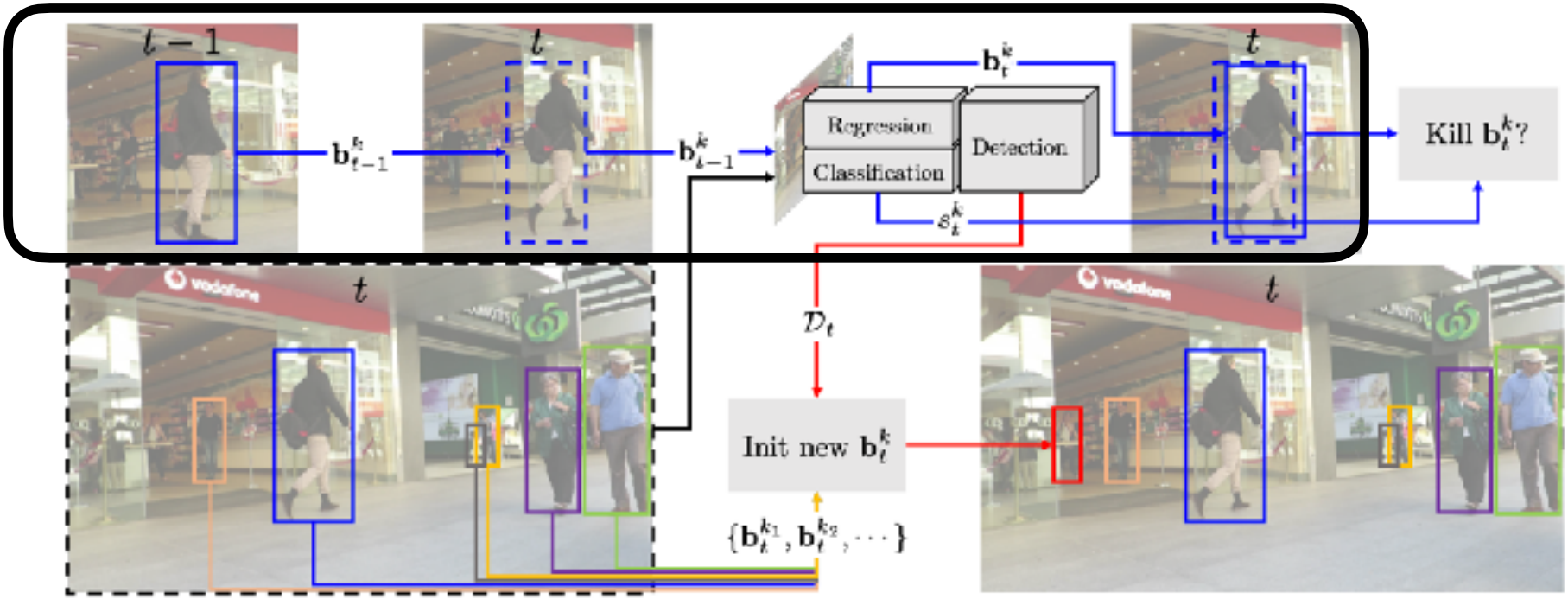
Tracktor

- The authors convert a detector into a tracker that performs multiple object tracking.
- The method does not require any tracking specific training, nor complex optimization at test time.



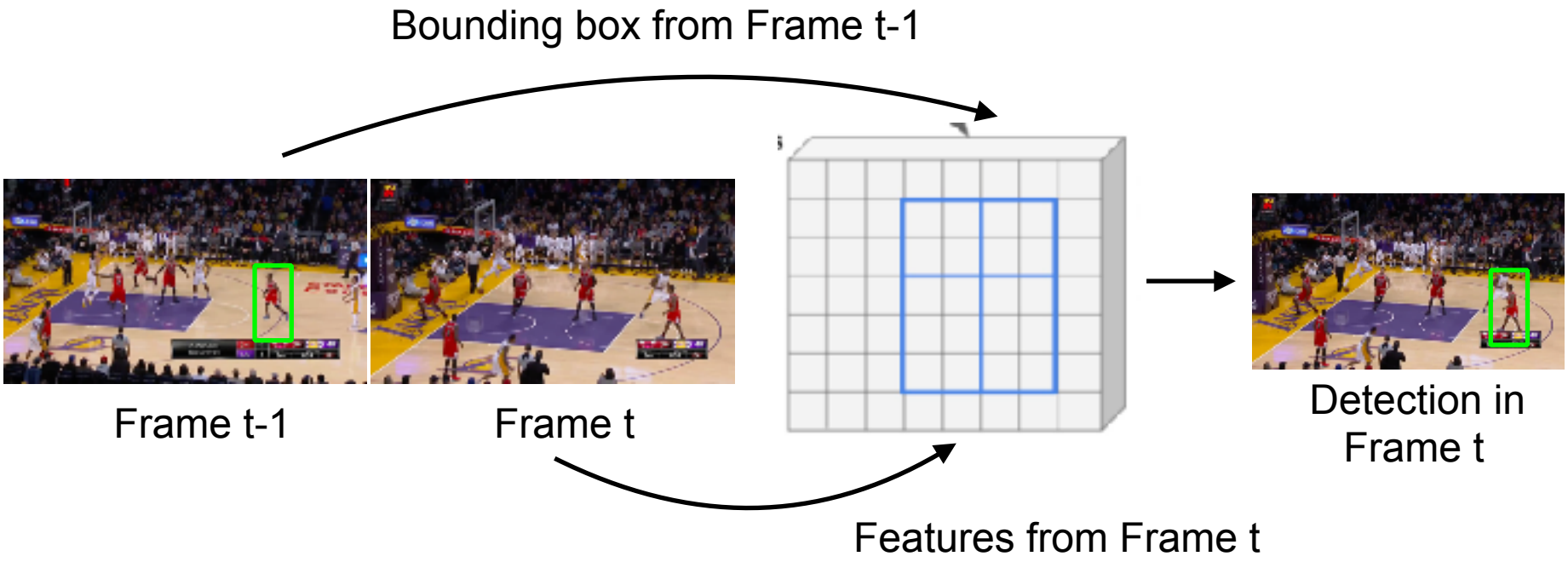
Tracktor

- The authors convert a detector into a tracker that performs multiple object tracking.
- The method does not require any tracking specific training, nor complex optimization at test time.



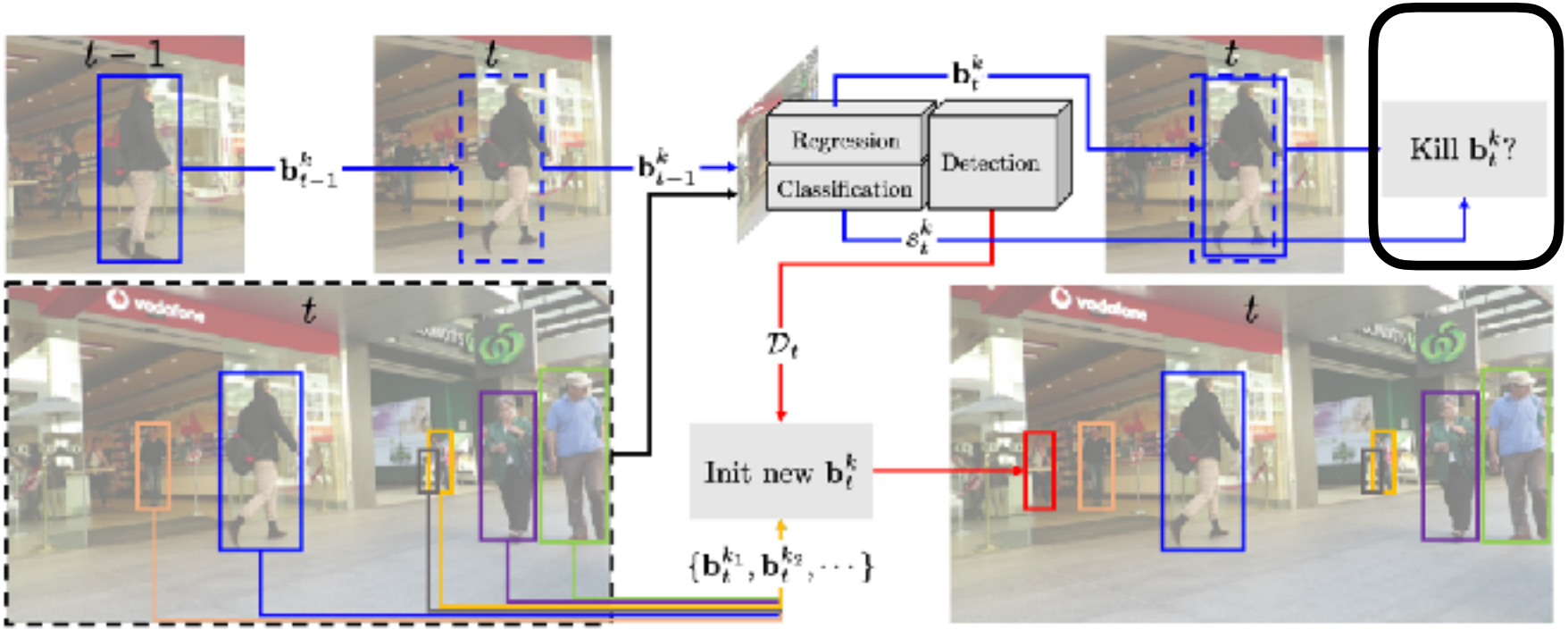
RoI Pooling

- RoI pooling is applied on the features of the current frame (Frame t) but with the previous bounding box coordinates (Frame t-1).



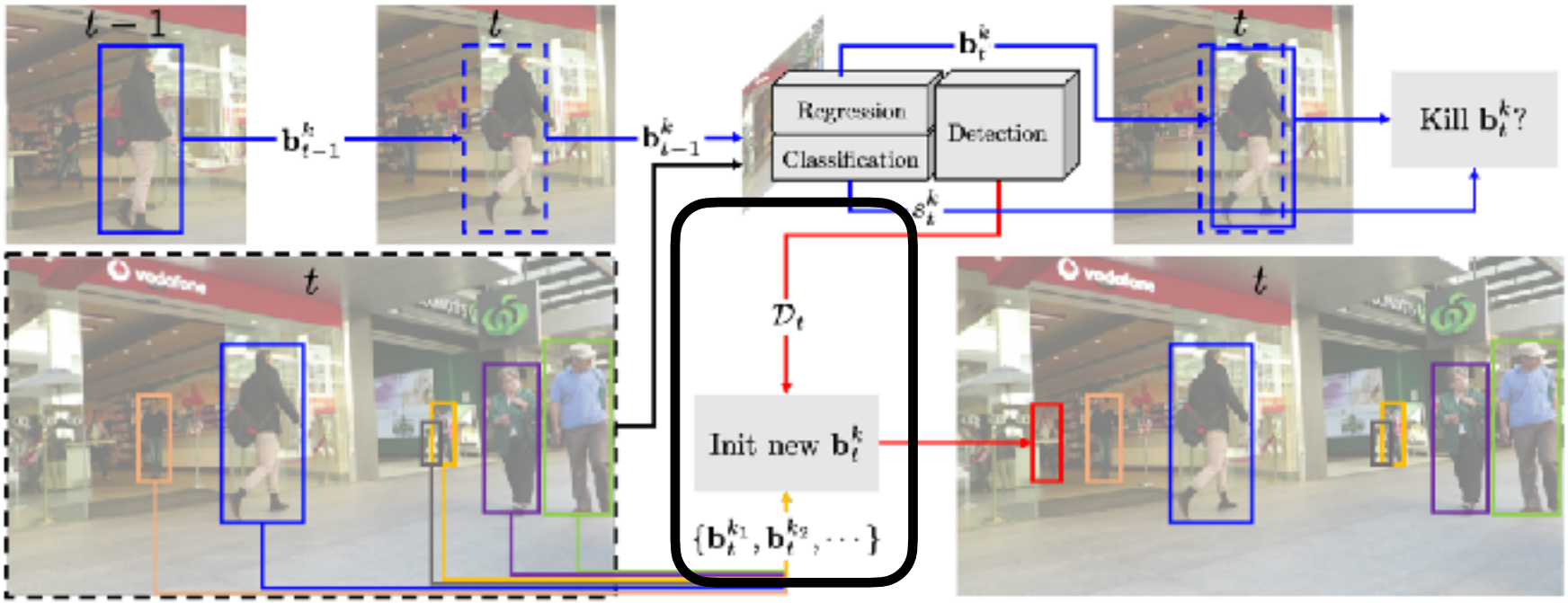
Tracktor

- The authors convert a detector into a tracker that performs multiple object tracking.
- The method does not require any tracking specific training, nor complex optimization at test time.



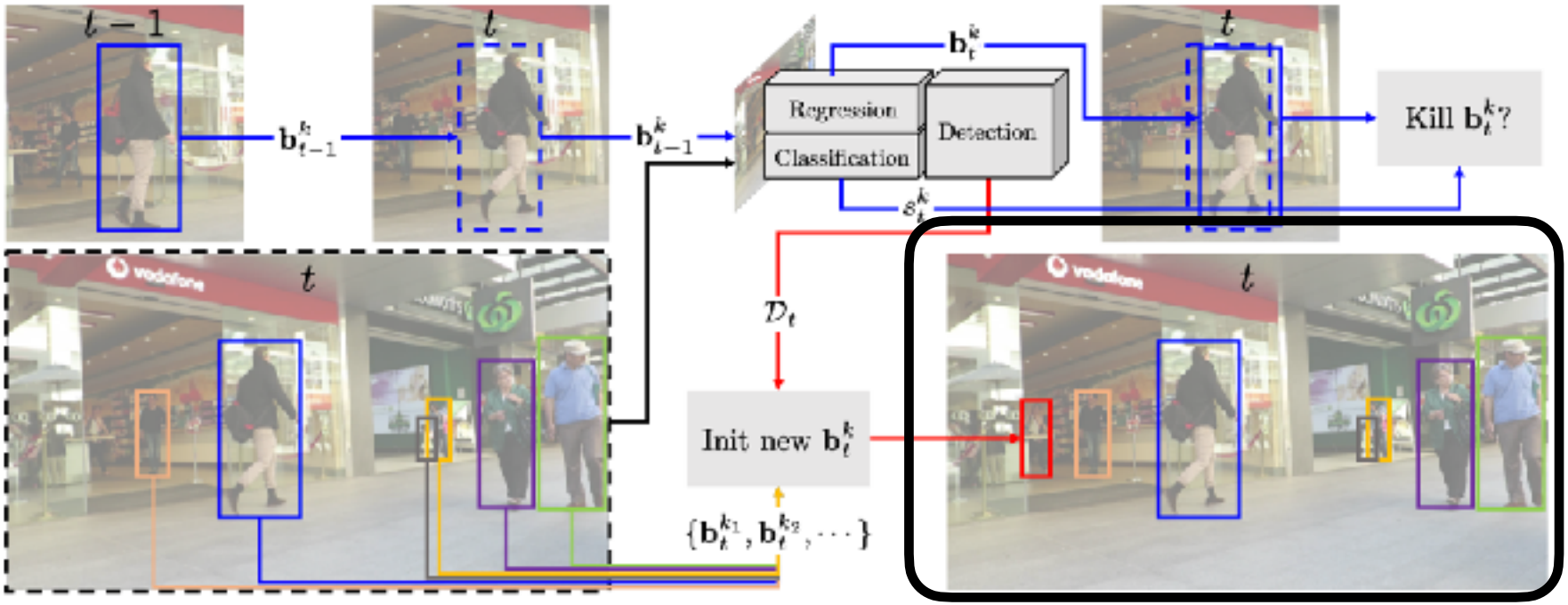
Tracktor

- The authors convert a detector into a tracker that performs multiple object tracking.
- The method does not require any tracking specific training, nor complex optimization at test time.



Tracktor

- The authors convert a detector into a tracker that performs multiple object tracking.
- The method does not require any tracking specific training, nor complex optimization at test time.



Experiments

- The experiments are done on the multi-object tracking benchmark MOTChallenge, which contains 7 training and testing video sequences.
- Only pedestrians are annotated and evaluated.

Video Index	Resolution	FPS	Length (frames)	Boxes	Tracks	Density
02	1920 × 1080	30	600	18581	62	31.0
04	1920 × 1080	30	1050	47557	83	45.3
05	640 × 480	14	837	6917	133	8.3
09	1920 × 1080	30	525	5325	26	10.1
10	1920 × 1080	30	654	12839	57	19.6
11	1920 × 1080	30	900	9436	75	15.5
13	1920 × 1080	25	750	11642	110	8.3

Evaluation Metrics

- When objects are successfully detected, but not tracked, they are identified as an identity switch (IDSW).

Evaluation Metrics

- When objects are successfully detected, but not tracked, they are identified as an identity switch (IDSW).
- A target is mostly tracked (MT) if it is successfully tracked for at least 80% of its life span.

Evaluation Metrics

- When objects are successfully detected, but not tracked, they are identified as an identity switch (IDSW).
- A target is mostly tracked (MT) if it is successfully tracked for at least 80% of its life span.
- If a track is only recovered for less than 20% of its total length, it is said to be mostly lost (ML).

Evaluation Metrics

- When objects are successfully detected, but not tracked, they are identified as an identity switch (IDSW).
- A target is mostly tracked (MT) if it is successfully tracked for at least 80% of its life span.
- If a track is only recovered for less than 20% of its total length, it is said to be mostly lost (ML).
- The Identity F1 Score (IDF1) measures the identity preservation of a method.

Evaluation Metrics

- When objects are successfully detected, but not tracked, they are identified as an identity switch (IDSW).
- A target is mostly tracked (MT) if it is successfully tracked for at least 80% of its life span.
- If a track is only recovered for less than 20% of its total length, it is said to be mostly lost (ML).
- The Identity F1 Score (IDF1) measures the identity preservation of a method.
- Multiple Object Tracking Accuracy (MOTA) metric focuses on object coverage.

$$1 - \frac{\sum_t (FP_t + FN_t + IDSW_t)}{\sum_t GT_t}$$

Results

- This ablation study illustrates multiple aspects on the performance of Tracktor.

Method	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
D&T [18]	50.1	24.9	23.1	27.1	3561	52481	2715
Tracktor-no-FPN	57.4	58.7	30.2	22.5	2821	45042	1981
Tracktor	61.5	61.1	33.5	**20.7**	367	42903	1747

Results

- This ablation study illustrates multiple aspects on the performance of Tracktor.

Method	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
D&T [18]	50.1	24.9	23.1	27.1	3561	52481	2715
Tracktor-no-FPN	57.4	58.7	30.2	22.5	2821	45042	1981
Tracktor	61.5	61.1	33.5	**20.7**	367	42903	1747

Improving the quality of object detection leads to large improvements in tracking as well.

Results

- Comparison of Tracktor with other modern tracking methods.

	Method	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
MOT17	Tracktor++	**53.5**	52.3	19.5	36.6	**12201**	248047	2072
	eHAF [58]	51.8	**54.7**	**23.4**	37.9	33212	**236772**	1834
	FWT [23]	51.3	47.6	21.4	35.2	24101	247921	2648
	jCC [30]	51.2	54.5	20.9	37.0	25937	247822	**1802**
	MOTDT17 [9]	50.9	52.7	17.5	35.7	24069	250768	2474
	MHT_DAM [32]	50.7	47.2	20.8	36.9	22875	252889	2314
MOT16	Tracktor++	**54.4**	**52.5**	19.0	**36.9**	**3280**	**79149**	682
	HCC [44]	49.3	50.7	17.8	39.9	5333	86795	**391**
	LMP [59]	48.8	51.3	18.2	40.1	6654	86245	481
	GCRA [43]	48.2	48.6	12.9	41.1	5104	88586	821
	FWT [23]	47.8	44.3	**19.1**	38.2	8886	85487	852
	MOTDT [9]	47.6	50.9	15.2	38.3	9253	85431	792
2D MOT 2015	Tracktor++	**44.1**	46.7	18.0	**26.2**	6477	**26577**	1318
	AP_HWDPL_p [8]	38.5	**47.1**	8.7	37.4	**4005**	33203	586
	AMIR15 [56]	37.6	46.0	15.8	26.8	7933	29397	1026
	JointMC [30]	35.6	45.1	23.2	39.3	10580	28508	457
	RAR15pub [17]	35.1	45.4	13.0	42.3	6771	32717	**381**

Results

- Comparison of Tracktor with other modern tracking methods.

	Method	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
MOT17	Tracktor++	**53.5**	52.3	19.5	36.6	**12201**	248047	2072
	eHAF [58]	51.8	**54.7**	**23.4**	37.9	33212	**236772**	1834
	FWT [23]	51.3	47.6	21.4	35.2	24101	247921	2648
	jCC [30]	51.2	54.5	20.9	37.0	25937	247822	**1802**
	MOTDT17 [9]	50.9	52.7	17.5	35.7	24069	250768	2474
	MHT_DAM [32]	50.7	47.2	20.8	36.9	22875	252889	2314
MOT16	Tracktor++	**54.4**	**52.5**	19.0	**36.9**	**3280**	**79149**	682
	HCC [44]	49.3	50.7	17.8	39.9	5333	86795	**391**
	LMP [59]	48.8	51.3	18.2	40.1	6654	86245	481
	GCRA [43]	48.2	48.6	12.9	41.1	5104	88586	821
	FWT [23]	47.8	44.3	**19.1**	38.2	8886	85487	852
	MOTDT [9]	47.6	50.9	15.2	38.3	9253	85431	792
2D MOT 2015	Tracktor++	**44.1**	46.7	18.0	**26.2**	6477	**26577**	1318
	AP_HWDPL_p [8]	38.5	**47.1**	8.7	37.4	**4005**	33203	586
	AMIR15 [56]	37.6	46.0	15.8	26.8	7933	29397	1026
	JointMC [30]	35.6	45.1	23.2	39.3	10580	28508	457
	RAR15pub [17]	35.1	45.4	13.0	42.3	6771	32717	**381**

Results

- A summary of the fundamental characteristics of Tracktor and other state-of-the-art trackers.

Method	Online	Graph	reID	Appearance model	Motion model	Other
Tracktor	×					
Tracktor++	×		×		Camera	
FWT [23]		Dense				Face detection
jCC [30]		Dense				Point trajectories
MOTDT17 [9]	×		×	×	Kalman	
MHT_DAM [32]		Sparse		×	Kalman	

Results

- A summary of the fundamental characteristics of Tracktor and other state-of-the-art trackers.

Method	Online	Graph	reID	Appearance model	Motion model	Other
Tracktor	×					
Tracktor++	×		×		Camera	
FWT [23]		Dense				Face detection
jCC [30]		Dense				Point trajectories
MOTDT17 [9]	×		×	×	Kalman	
MHT_DAM [32]		Sparse		×	Kalman	

Compared to prior approaches, the proposed method is much simpler.

Contributions

- A simple method that demonstrates how to convert a standard detector into a tracker.
- Despite the simplicity, the proposed method achieves state-of-the-art performance on multiple pedestrian tracking benchmarks.
- Provides a simple yet powerful baseline for subsequent research to build on.

Discussion Questions

- What are the main limitations / assumptions made by the proposed approach?

Discussion Questions

- What are the main limitations / assumptions made by the proposed approach?

- Is detection all you need for tracking?