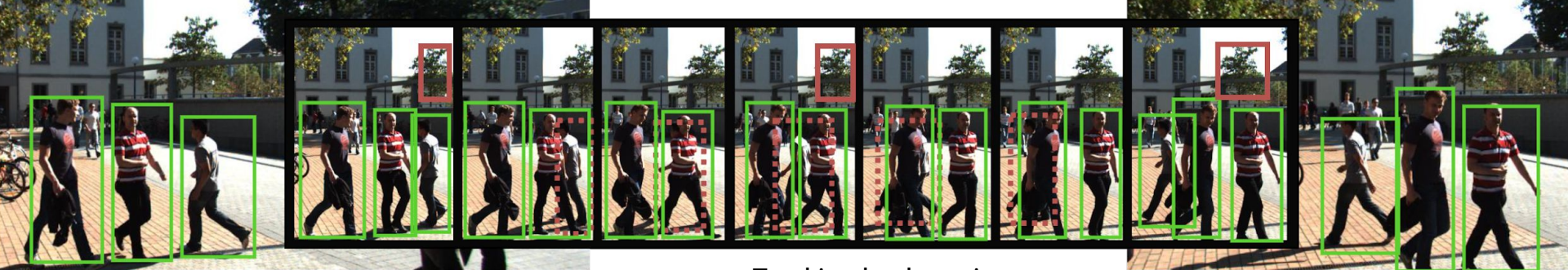


Compact and discriminative multi-object tracking with siamese CNNs

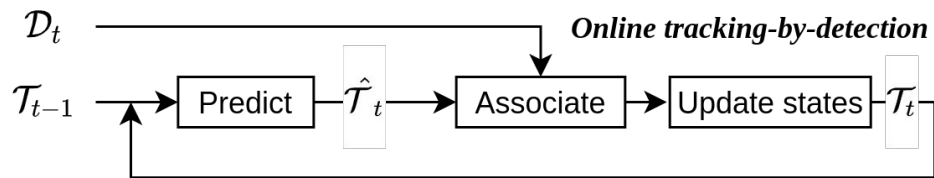
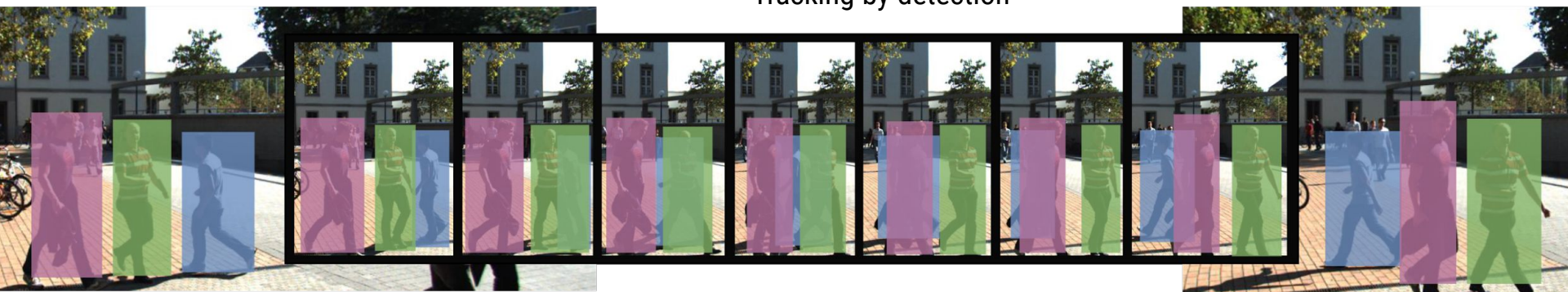
25th International Conference on Pattern Recognition - 10 | 15 January 2021

Claire Labit-Bonis, Jérôme Thomas, Frédéric Lerasle

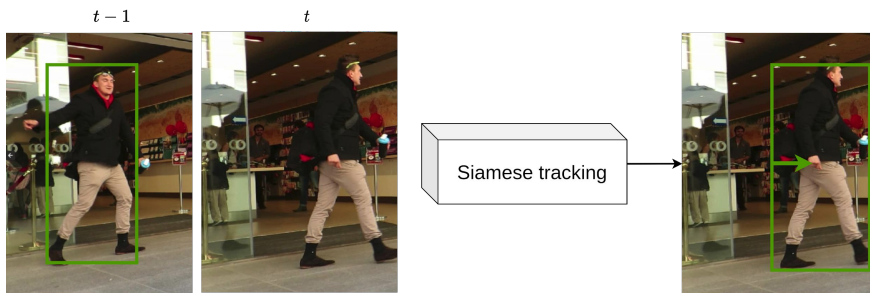




Tracking-by-detection

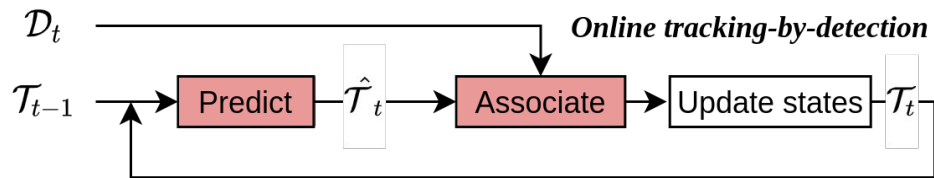
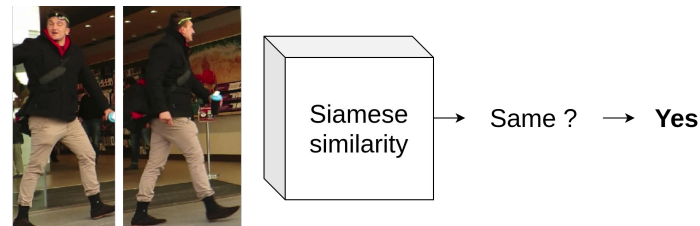
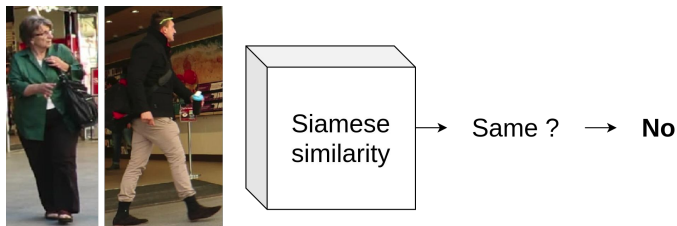


GOAL



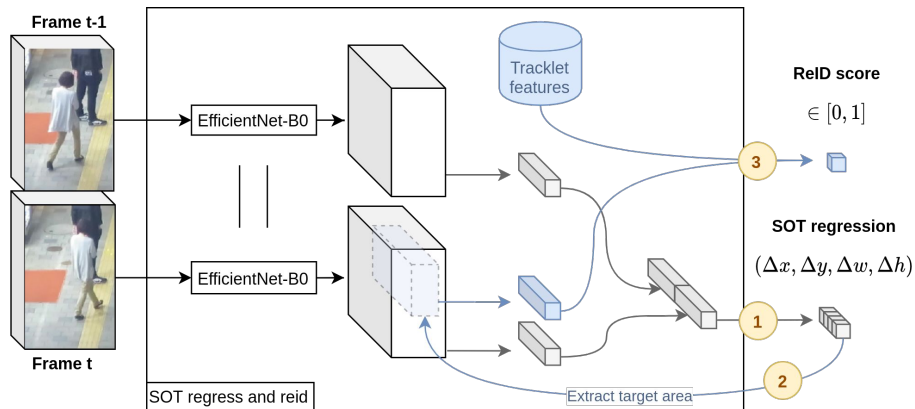
Visual object tracking “VOT”

Similarity learning, verification, reidentification...



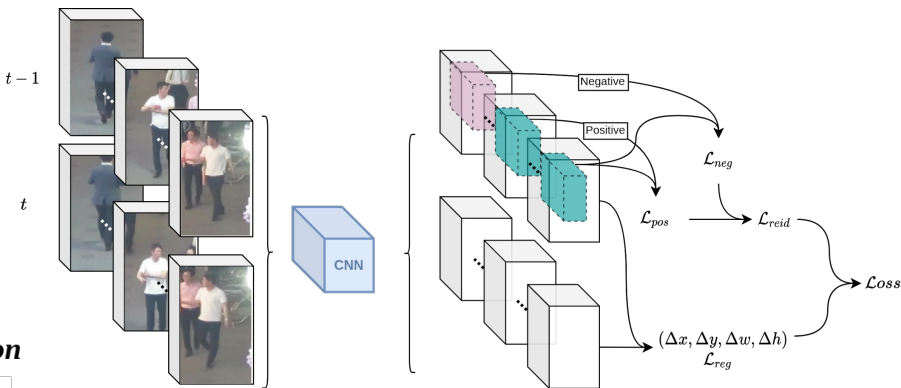
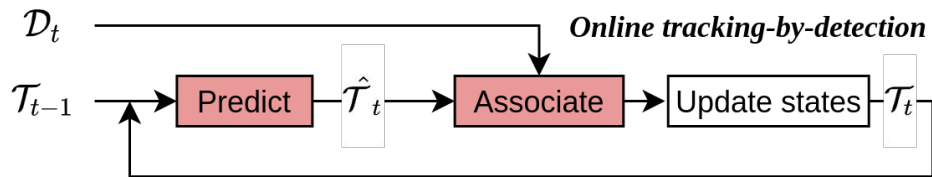
- Siamese VOT rarely used within MOT, but performed well on MOT17
- Position prediction and appearance embedding usually done sequentially, with two different networks or approaches

CONTEXT



- One-shot position prediction and identity propagation
- Light-weight architecture

- Multi-task joint learning of SOT regression and reidentification score
- Training data generation



CONTRIBUTIONS



Association cost:

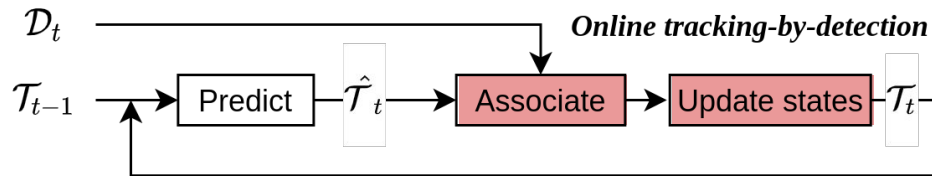
$$\mathcal{S}_{\mathcal{D}, \mathcal{T}} = \mathcal{T}_{cf} \times \text{IoU}_{\mathcal{D}, \mathcal{T}} \times \cos_{\mathcal{D}, \mathcal{T}} \times \mathcal{D}_{cf}$$

$$\mathcal{C}_{\mathcal{D}, \mathcal{T}} = 1 - \mathcal{S}_{\mathcal{D}, \mathcal{T}}$$

Tracklet confidence update inspired by LSST:

$$\mathcal{T}_{cft} = \begin{cases} \frac{(\mathcal{T}_{cft-1} + \text{IoU}_{\mathcal{D}, \mathcal{T}} \times \cos_{\mathcal{D}, \mathcal{T}} \times \mathcal{D}_{cf})}{2} & \text{if match}(\mathcal{D}, \mathcal{T}) \\ \mathcal{T}_{cft-1} \times \text{decay} \times \mathcal{T}_{reid}^k, & \text{otherwise} \end{cases}$$

- Unthresholded detections
- Multi-cue, one-pass data association
- “Keep active” tracking strategy



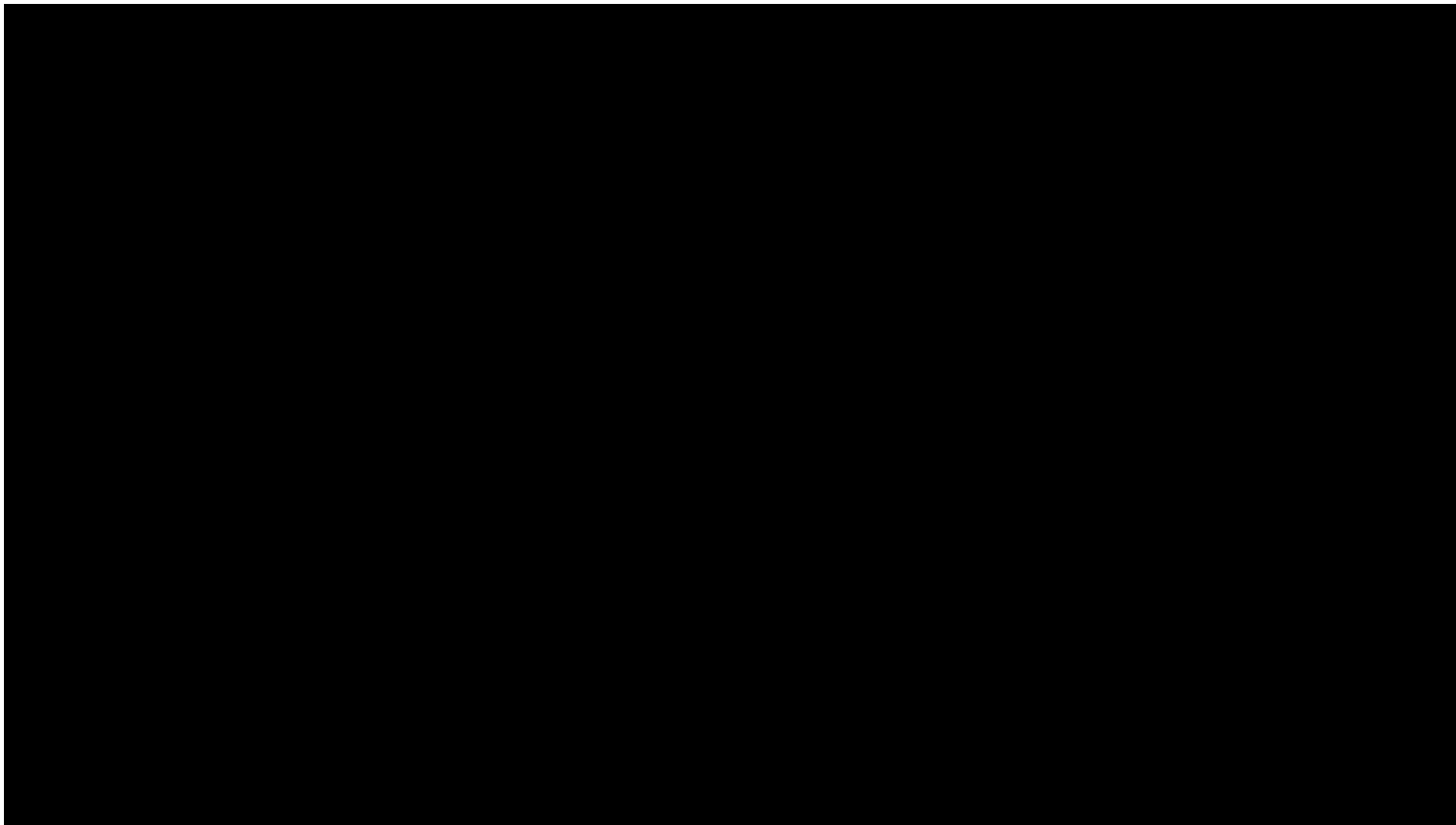
Ablation study

IoU	ReID	Classif.	Keep	Init	FP ↓	FN ↓	IDS ↓	MOTA ↑
✓					4259	34314	1309	64.49
✓				✓	-1014	+1004	-267	+0.24
✓			✓	✓	+1693	-1187	-220	-0.25
✓	✓			✓	-1152	+994	-307	+0.41
✓	✓		✓	✓	+191	-802	-263	+0.78
✓	✓	✓	✓	✓	-462	-325	-297	+0.96

CONTRIBUTIONS

MOT17 SDP test set benchmark - comparison to CNN-based approaches

Rank	Year	Conf.	Method	MOTA $\uparrow$	MT $\uparrow$	ML $\downarrow$	FP $\downarrow$	FN $\downarrow$	IDS $\downarrow$	FPS $\uparrow$	Hardware
1	2020	arXiv	UnsupTrack	61.99%	217	248	5986	64860	651	2	GTX 1080Ti
2	2019	arXiv	LSSTO [9]	61.63%	194	241	7285	64091	792	1.8	GTX 1060
3	2020	IEEEAccess	YOONKJ	61.11%	225	248	9468	62895	788	3.4	GTX 1080Ti
5	2018	AVSS	HAM_SADF	60.30%	212	251	7177	66729	759	5	Titan X
6	2019	arXiv	DEEP_TAMA	59.95%	215	246	8445	66103	779	1.5	Titan X
7	2019	ICCV	FAMNet [8]	59.16%	198	233	4822	70900	1097	0.6	Titan X
9	2020		Ours	58.48%	198	234	8585	68168	1328	10.4	Titan X
10	2018	ICME	MOTDT [12]	58.39%	177	241	6317	71005	941	18.3	GTX 1060
11	2020	IJCAI	GSM_Tracktor	58.34%	192	256	5772	72050	537	8.7	Titan X
15	2019	arXiv	GMPHD_Rd	57.15%	230	200	15661	63225	1686	20.4	GTX 1050
16	2020	AAAI	DASOT [23]	57.13%	217	213	13205	65320	2102	9.1	Titan X
17	2019	TM	MTDF	57.01%	190	223	10183	68898	1764	1.2	GTX 1060
18	2019	ICCV	STRN	56.99%	188	228	9262	70725	908	13.8	GPU (?)
19	2018	IEEEAccess	FPSN	56.52%	184	230	7682	71089	3005	10.1	Titan X
20	2019	arXiv	HISP_DAL	56.08%	174	239	7944	71012	3642	3.2	GTX 1050
21	2019	IEEEAccess	OTCD [19]	55.81%	159	295	3715	78160	1228	5.5	Titan X
22	2020	arXiv	TrctrD17	55.56%	171	271	5247	77615	715	4.9	Titan X
24	2019	ICCV	Tracktor++ [11]	55.32%	169	269	5375	77868	789	1.5	Titan X



DEMO

References

- [LSST]** W. Feng et al. "Multi-Object Tracking with Multiple Cues and Switcher-Aware Classification". 2019
- [FAMNet]** P. Chu and H. Ling. "Famnet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking". ICCV 2019
- [Tracktor]** P. Bergmann et al. "Tracking without bells and whistles". ICCV 2019
- [MOTDT]** L. Chen et al. "Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification". ICME 2018