Compact and discriminative multi-object tracking with siamese CNNs

Claire Labit-Bonis[1,2], Jérôme Thomas[2], Frédéric Lerasle[1]
[1]LAAS-CNRS, [2]ACTIA Automotive - contact: claire.labit-bonis@laas.fr

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

GOALS AND CONTRIBUTIONS

Context: Visually track and maintain identities of multiple objects in a videosteam, as long as they remain in the frame. It relies on the multi-object tracking-by-detection strategy : (i) detect objects frame-by-frame, (ii) link detections over time and maintain consistent trajectories.

Constraints: Aim at further embedded applications (= CPU cost constraint).

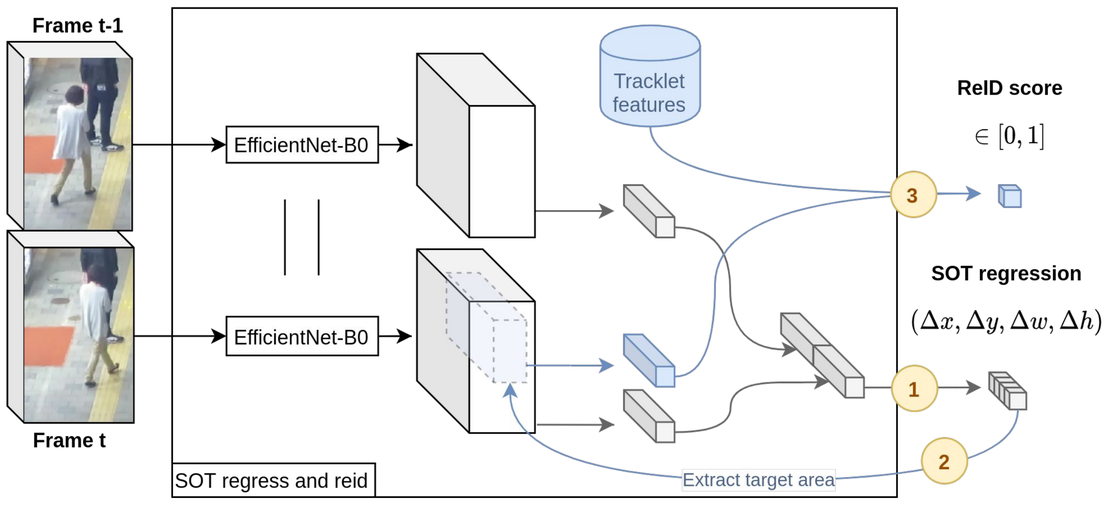
Scientific contributions:

- all-in-one compact siamese CNN architecture to address both single object tracking and frame-by-frame reidentification within MOT context
- tracklet management strategy based on data association between current targets' tracklet pool and input detections without any extra pre-processing cost
- comparative quantitative and qualitative state-of-the-art evaluation on MOT17 challenge with CPU cost reduction

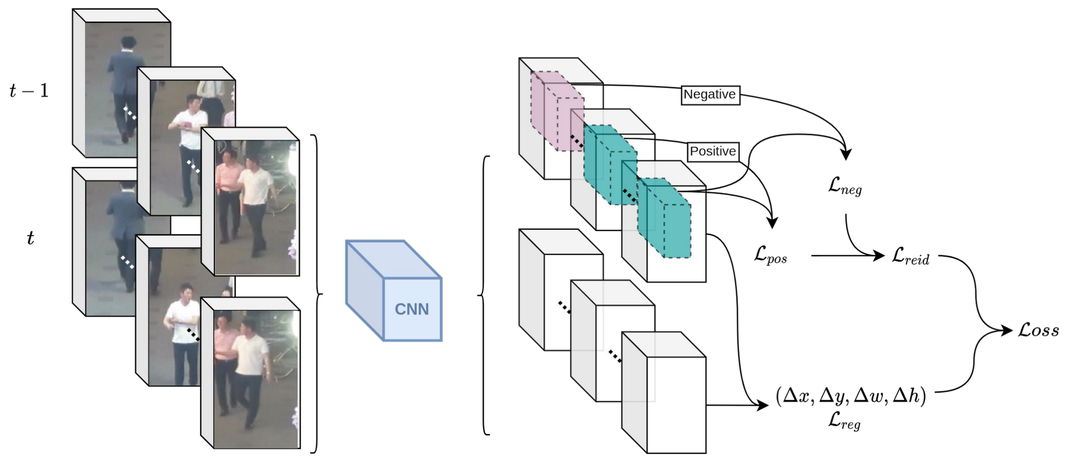
METHODOLOGY

SINGLE OBJECT TRACKING AND REIDENTIFICATION FOR MOT

- **Two branch** siamese CNN dealing with position regression and reidentification at the same time
- **Light-weight** EfficientNet-B0 based feature extractor



MULTI-TASK LEARNING AND ALL-IN-ONE TRAINING PAIRS GENERATION

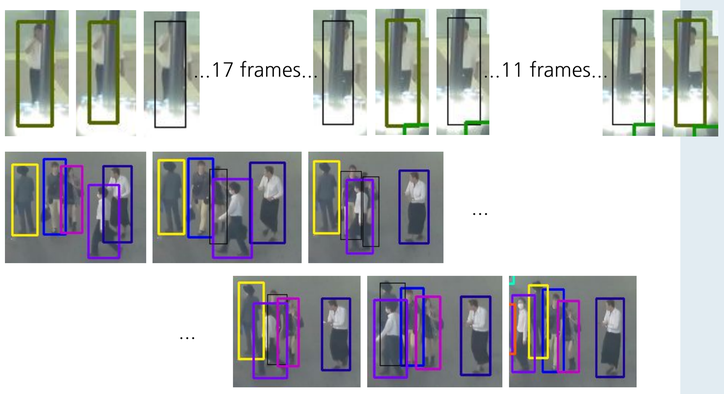


TRACKLET SET MANAGEMENT STRATEGY

- **No thresholding** done on the input detections
- **Multi-cue score** taken into account for data association (detector & tracker confidence, targets' positions and appearances) and tracklet confidence update
- Targets **appearance model update**
- **"Keep active"** strategy based on tracklet confidence to compensate for missing detections
- Straight forward with **one-stage management of active/inactive tracks**

QUALITATIVE RESULTS

Robust scene occlusions



Total target-target occlusions still challenging...

QUANTITATIVE RESULTS

Ablation study showing the impact of the tracklets multi-cue management strategy on the MOT17 SDP training set.

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

- *Complementary reidentification and keep active strategies*
- *Benefiting from non-thresholded detections*

EXPERIMENTS & RESULTS

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

Comparison to the litterature on the MOT17 test set with SDP detector inputs. We achieve state-of-the-art. Demo at https://bit.ly/2VNl9ks

CONCLUSION

- Online tracking-by-detection in a multi-object tracking context with an original architecture integrating a jointly trained all-in-one single light-weight siamese CNN for both tracklet position prediction and reidentification
- Comparison to the literature on the MOT17 benchmark, presenting state-of-the-art performance and high frame rates
- **Work in progress / Perspectives:** benefit from recent single object tracking techniques using intercorrelation to improve speed and performance ; find an in-between strategy within appearance feature management.

ACKNOWLEDGEMENTS

This work is partially supported by the french **National Association for Research and Technology (ANRT)** within a CIFRE PhD agreement.