Online Multi-Target Tracking Using Recurrent Neural Networks

Anton Milan   S. Hamid Rezatofighi   Anthony Dick   Ian Reid   Konrad Schindler

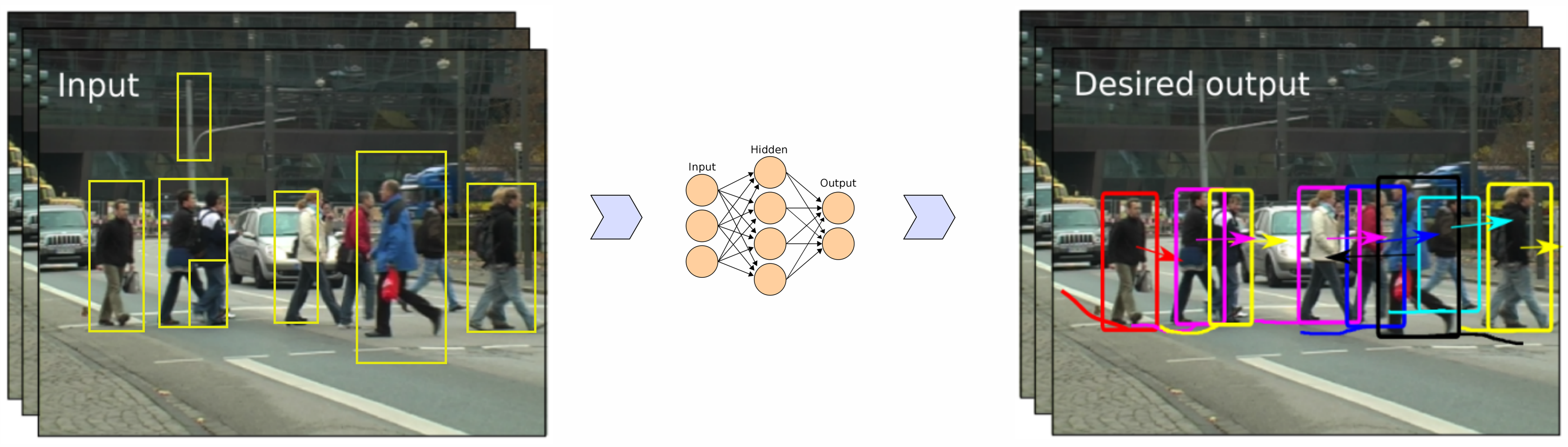
School of Computer Science, The University of Adelaide   Photogrammetry and Remote Sensing, ETH Zürich

Abstract

We present a novel approach to online multi-target tracking based on recurrent neural networks (RNNs). Tracking multiple objects in real-world scenes involves many challenges, including a) an a-priori unknown and time-varying number of targets, b) a continuous state estimation of all present targets, and c) a discrete combinatorial problem of data association. Our solution addresses all of the above points in a principled way.

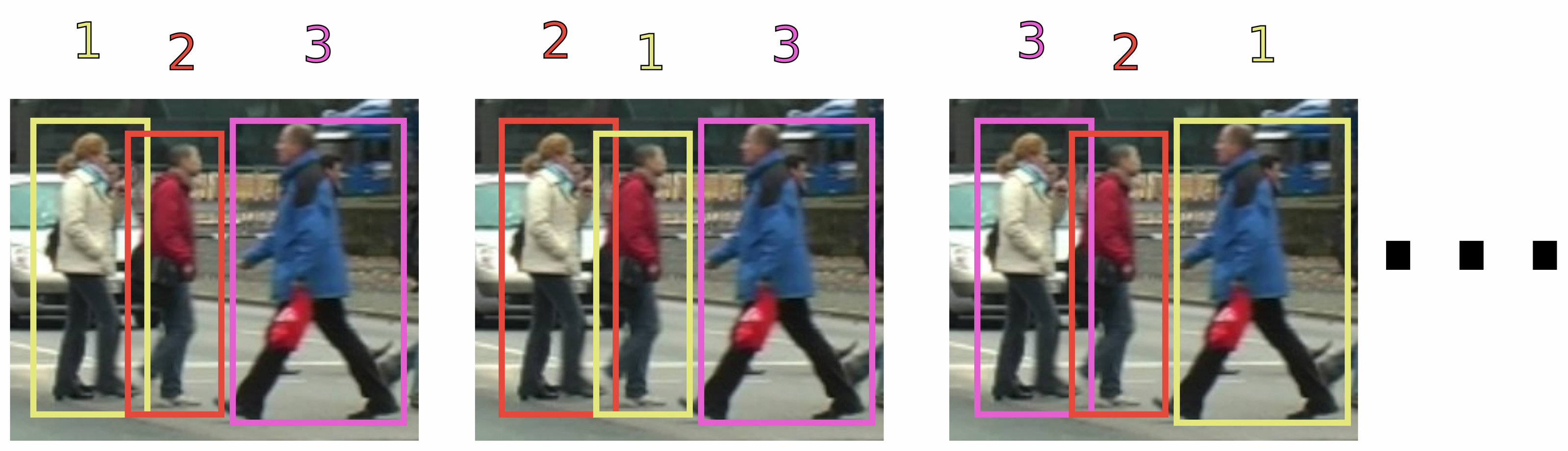
Motivation

- Exploit power of deep learning for multi-target tracking
- Data-driven approach, first step towards end-to-end learning
- Efficient inference (up to 300Hz on a CPU)



Challenges

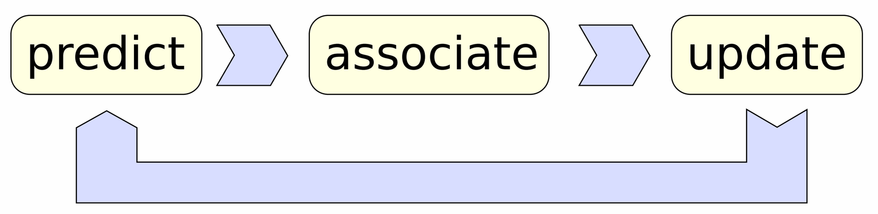
- Unknown and time-varying number of targets
- Missing, false and noisy detections
- Class (ID) assignment is arbitrary

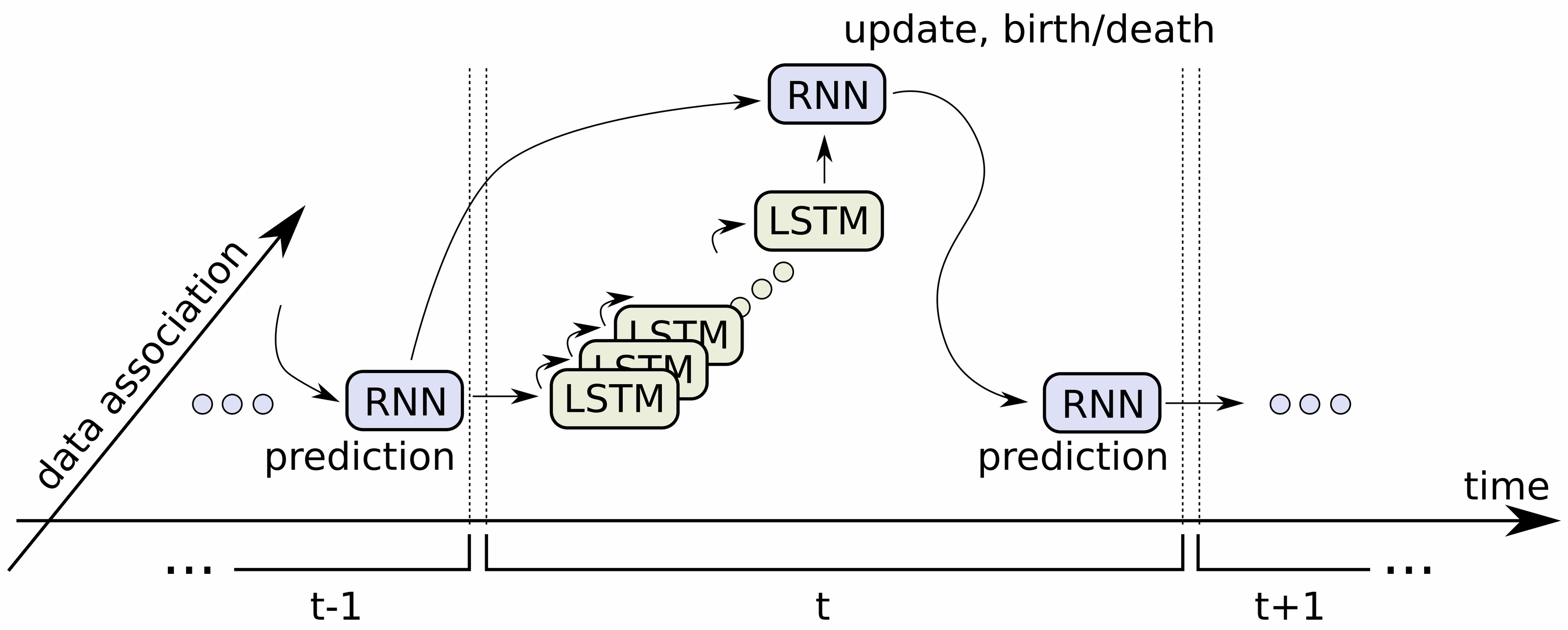


Our Proposed Approach

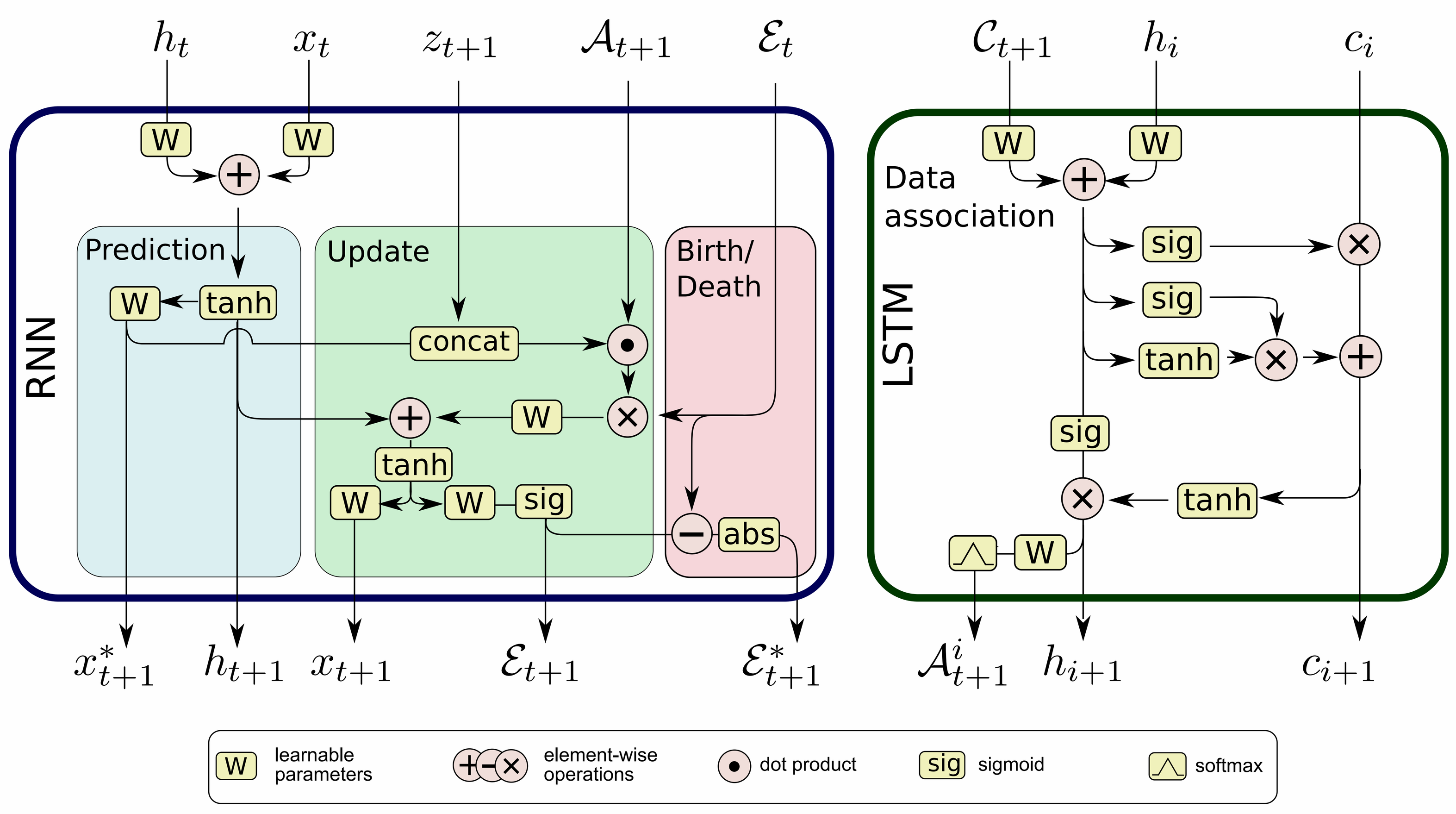
Based on **Bayesian filtering**

$$p(x_t|z_{1:t}) \propto p(z_t|x_t) \int p(x_t|x_{t-1}) p(x_{t-1}|z_{1:t-1}) dx_{t-1}$$





Model



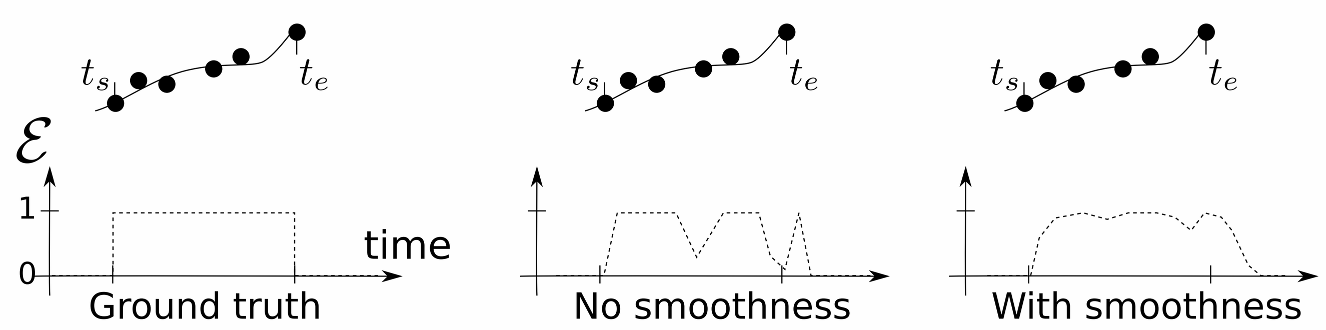
Loss

$$\mathcal{L} = \underbrace{\lambda \sum \|x^* - \tilde{x}\|^2}_{\text{prediction}} + \underbrace{\kappa \|x - \tilde{x}\|^2}_{\text{update}} + \underbrace{\nu \mathcal{L}_{\mathcal{E}} + \xi \mathcal{E}^*}_{\text{birth/death + reg.}}$$

$$\mathcal{L}_{\mathcal{E}} = \tilde{\mathcal{E}} \log \mathcal{E} + (1 - \tilde{\mathcal{E}}) \log(1 - \mathcal{E})$$

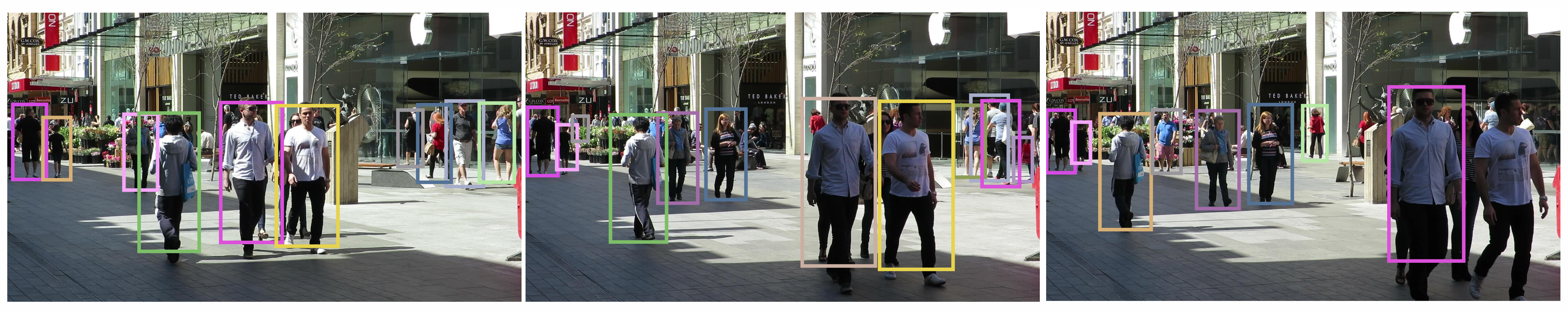
$$\mathcal{L}_{\mathcal{A}}(\mathcal{A}^i, \tilde{a}) = -\log(\mathcal{A}_{i\tilde{a}})$$

Existence smoothness $\mathcal{E}$



Experiments

MOTChallenge 2015 [4]



Baseline comparison

Method	MOTA	Recall	Precision	ID Sw.
Kalman+HA (O)	19.2	28.5	79.0	685
Kalman+HA+Post	22.4	28.3	**83.4**	**105**
RNN+HA (O)	**24.0**	**37.8**	75.2	518
RNN+LSTM (O)	22.3	37.1	73.5	572

(O) = Online method

Benchmark result

Method	MOTA	FN	FP	ID Sw.	FPS
MDP [1]	**30.3**	**32,422**	9,717	680	1.1
JPDAm [2]	23.8	40,084	**6,373**	**365**	32.6
TC_ODAL [3]	15.1	38,538	12,970	637	1.7
RNN+LSTM	19.0	38,706	11,578	1,490	**165.2**

bitbucket.org/amilan/rnntracking

Acknowledgements

This work was supported by SNSF International Short Visit 162004, ARC Linkage Project LP130100154, ARC Laureate Fellowship FL130100102 and the ARC Centre of Excellence for Robotic Vision CE140100016.

References

[1] Xiang *et al.* 2015. Learning to track: Online multi-object tracking by decision making. In ICCV
[2] Rezatofighi *et al.* 2015. Joint probabilistic data association revis ited. In ICCV.
[3] Bae, S.-H., and Yoon, K.-J. 2014. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In CVPR.
[4] Leal-Taixé, L.; Milan, A.; Reid, I.; Roth, S.; and Schindler, K. 2015. MOTChallenge 2015: Towards a benchmark for multi-target tracking. arXiv:1504.01942 [cs].