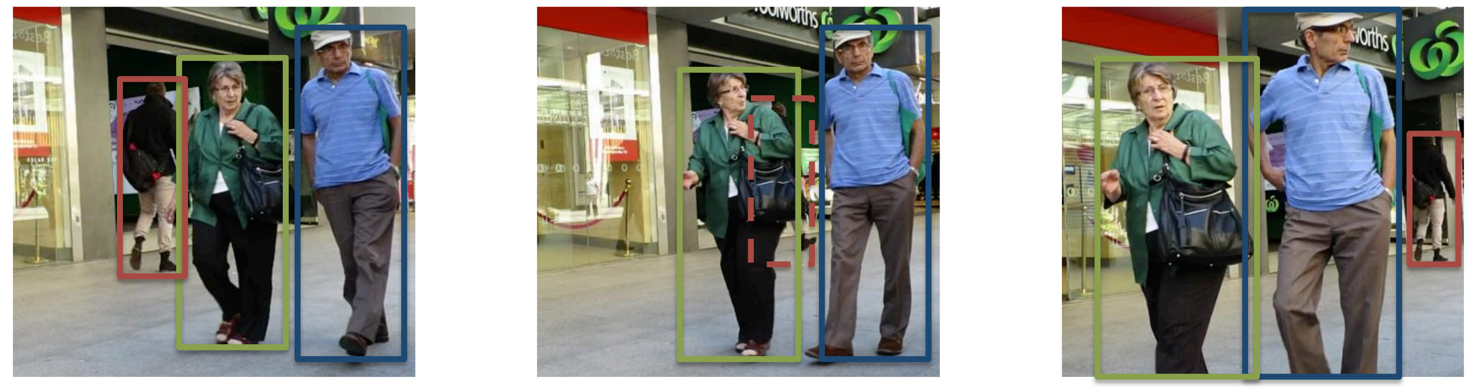
Recurrent Autoregressive Networks for Online Multi-Object Tracking

Kuan Fang, Yu Xiang, Xiaocheng Li, Silvio Savarese
Stanford University

Online Multi-Object Tracking with Neural Networks

Goal:
Reliably associate object trajectories with detections in each video frame based on their tracking history.

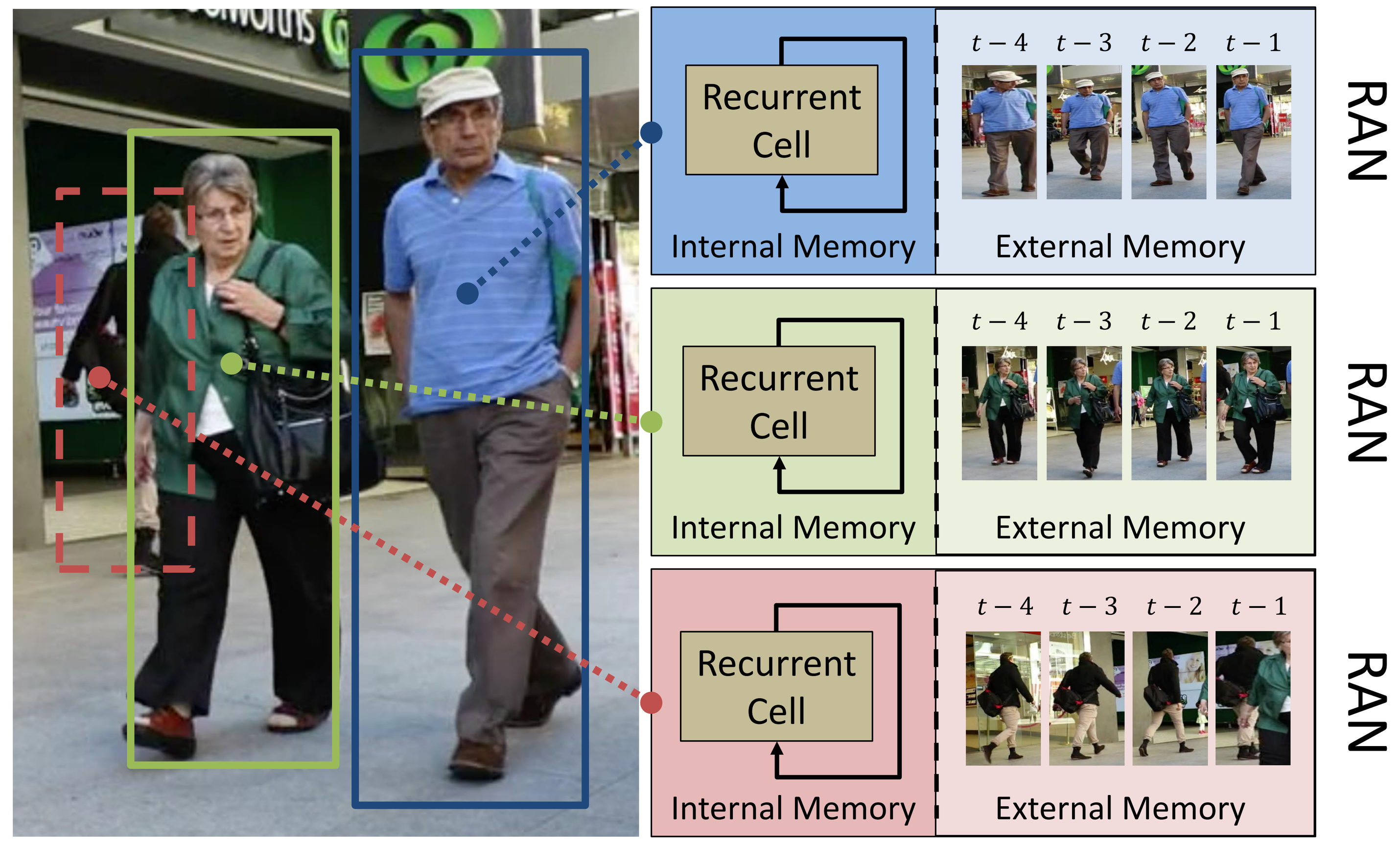


Challenges:
- Extract feature representations that can handle false alarms and occlusions.
- Data-efficient models trained on limited labeled videos.

Internal Memory v.s. External Memory

Internal memory: Hidden layer of recurrent cells.

External memory: Templates directly storing the previous input features.



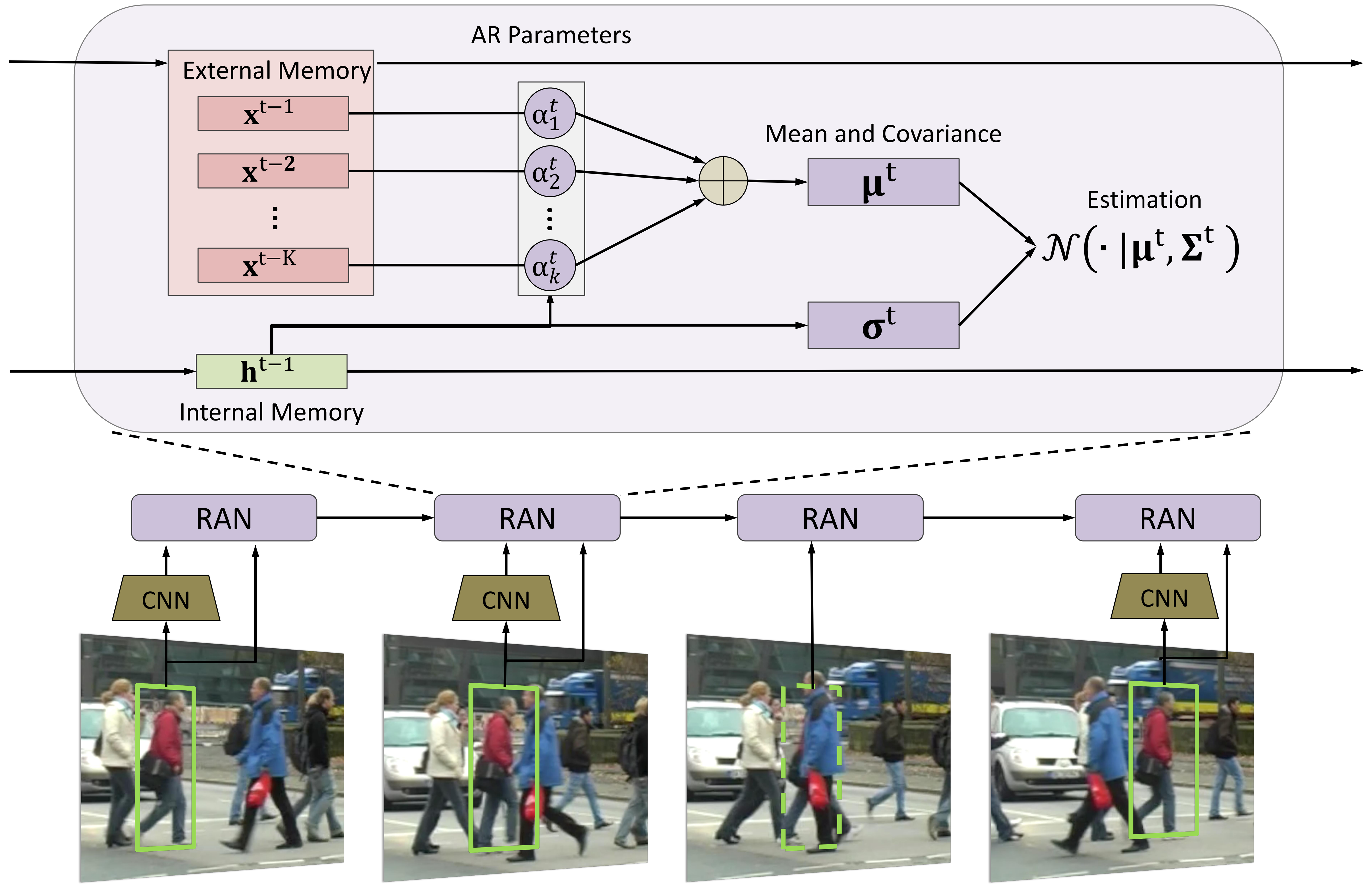
Recurrent Autoregressive Networks

Estimate the feature of the next time step by autoregressive (AR) model.

$$\Pr(\mathbf{x}^t|\mathbf{x}^{1:t-1}) = \mathcal{N}(\mathbf{x}^t|\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$$

The parameters of the AR model is predicted by Gate Recurrent Units.

$$\boldsymbol{\mu}^t = \sum_{k=1}^{K} \alpha_k^t \mathbf{x}^{t-k} \qquad \boldsymbol{\Sigma}^t = \mathrm{diag}((\boldsymbol{\sigma}^t)^2) = \mathrm{diag}\left((\sigma_1^t)^2, \ldots, (\sigma_N^t)^2\right)$$



Benefits:
- More robust to occlusions and sudden changes of the targets by maintaining an external memory.
- Use fewer neural network parameters compared to directly estimating high-dimensional features.

Results

Results on MOT Benchmark 2015

Method	Mode	MOTA(↑)	MOTP(↑)	IDS(↓)
CNNTCM	Batch	29.6	71.8	712
MHT_DAM	Batch	32.4	71.8	**435**
NOMT	Batch	**33.7**	**71.9**	442
SCEA	Online	29.1	71.1	604
MDP	Online	30.3	71.3	680
AMIR15	Online	**37.6**	**71.7**	1,026
Our Model (RAN)	Online	35.1	70.9	**381**

Results on MOT Benchmark 2016

Method	Mode	MOTA(↑)	MOTP(↑)	IDS(↓)
JMC	Batch	46.3	75.7	657
NOMT	Batch	46.4	76.6	**359**
NLLMPa	Batch	**47.6**	**78.5**	629
EAMTT	Online	38.8	**75.1**	965
oICF	Online	43.2	74.3	**381**
Our Model (RAN)	Online	**45.9**	74.8	648

Ablation Study

Model	MOTA(↑)	IDS(↓)
A-GRU	43.3	107
A-AVE	67.8	138
A-TIV	68.9	108
A-RAN	**69.9**	**80**
M-GRU	56.7	**108**
M-AVE	68.5	158
M-TIV	68.6	149
M-RAN	**68.9**	118
(A+M)-GRU	57.7	85
(A+M)-AVE	68.6	142
(A+M)-TIV	69.3	109
(A+M)-RAN	**70.7**	**77**

Analysis of the predicted AR parameters