Uber ATG

Discrete Residual Flow for Probabilistic Pedestrian Behavior Prediction

Ajay Jain, Sergio Casas, Renjie Liao, Yuwen Xiong, Song Feng, Sean Segal, Raquel Urtasun

UNIVERSITY OF TORONTO

BAIR BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Introduction

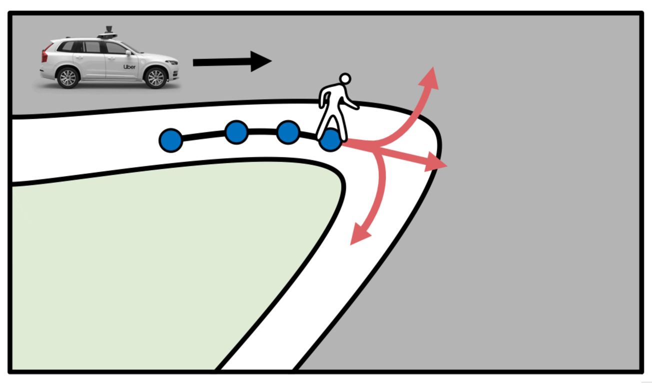
Goal: Forecast future pedestrian spatial occupancy over long horizon (10 seconds) in cities

Motivation: Safe motion planning in self-driving

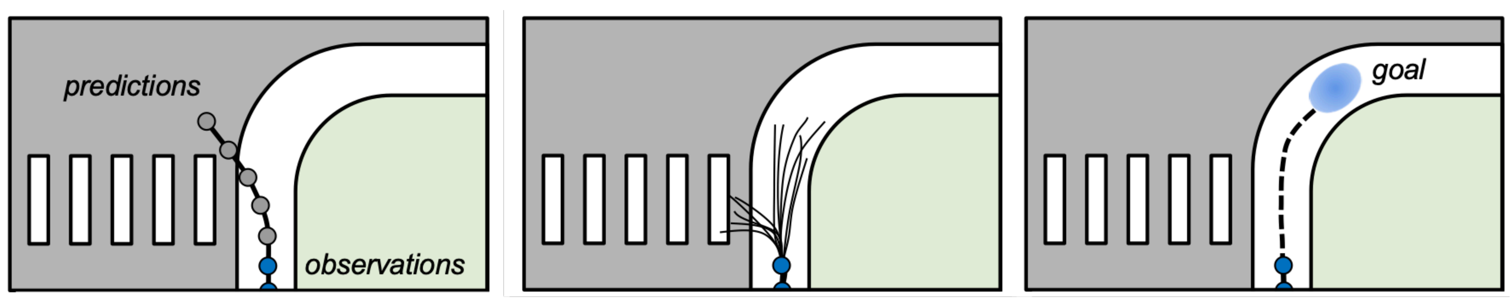
Input: Semantic map, dynamic actor tracks

Challenges:

- Multiple intentions
- Significant uncertainty
- Partial observability
- Non-gaussian posteriors
- Spatiotemporal inputs



Prior approaches



Regression [Alahi et al. 2016] [Becker et al. 2018] [Luo et al. 2018] [Casas et al. 2018]

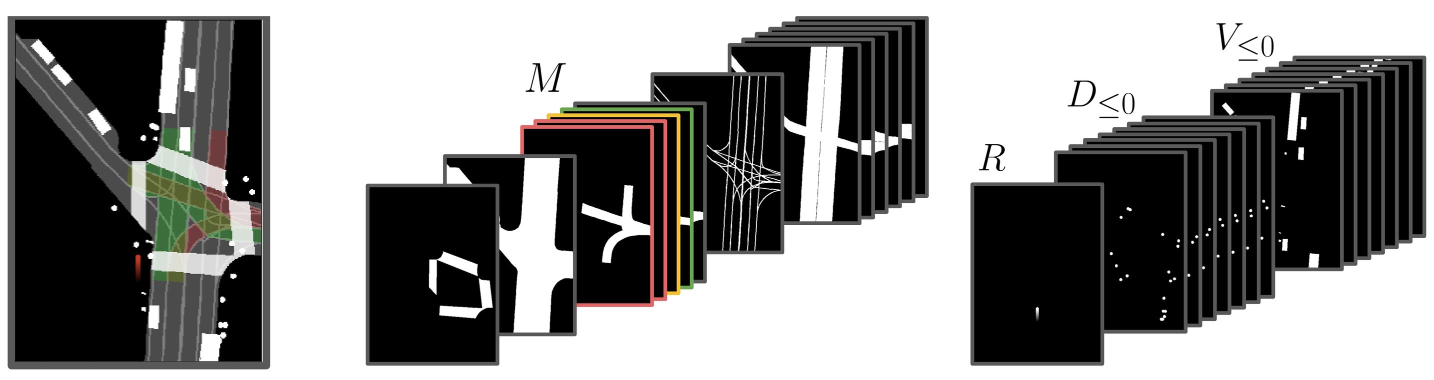
Sampling [Ma et al. 2017] [Gupta et al. 2018] [Rhinehart et al. 2018] [Sadeghian et al. 2019]

Goal based [Ziebart et al. 2009] [Wu et al. 2018] [Rehder et al. 2015, 2018] [Fisac et al. 2018]

Our approach

Multiscale scene embedding

- Spatiotemporal feature extraction from BEV scene raster with feature pyramid network



(a) Aggregated rasterization Dynamic objects shown at t=0

(b) Semantic map channels Crossings, road mask, lights, lanes

(c) Pol + dynamic object history Subset of timesteps shown

Probabilistic motion forecasting

- Predict **marginal occupancy distributions**
- **Categorical predictions** are flexible, multimodal

Bayesian approach: Learn conditional distributions and marginalize → $O(K^2)$ cost per timestep for K bins

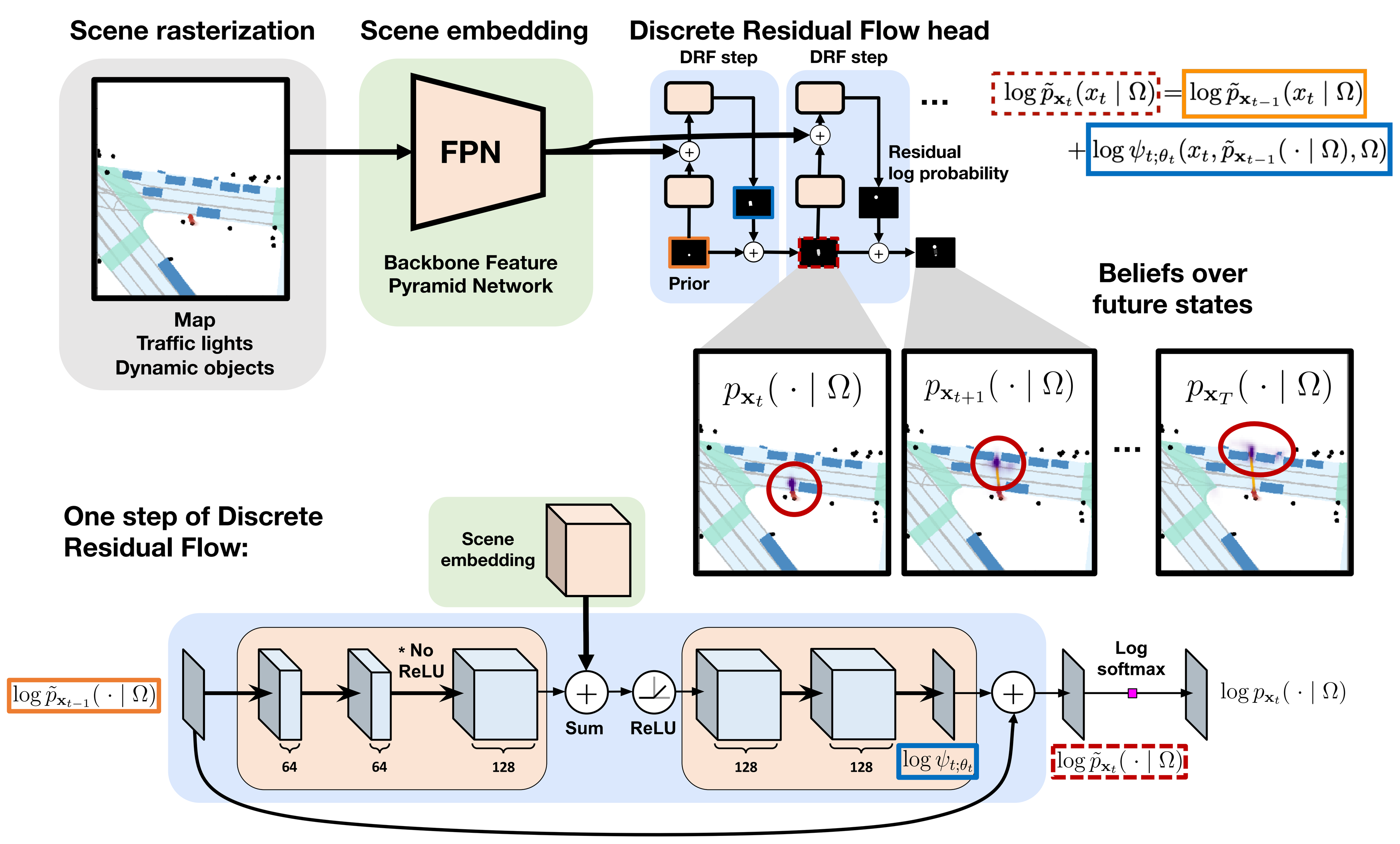
$$p_{\mathbf{x}_t}(x_t \mid \Omega) = \sum_{x_{t-1}} p_{\mathbf{x}_t|\mathbf{x}_{t-1}}(x_t \mid x_{t-1}, \Omega)\, p_{\mathbf{x}_{t-1}}(x_{t-1} \mid \Omega)$$

DRF-NET (ours): Approximate intractable marginalization using function approximator, amortizing cost → $O(K)$ probability flow that predicts a residual update to previous timestep marginal

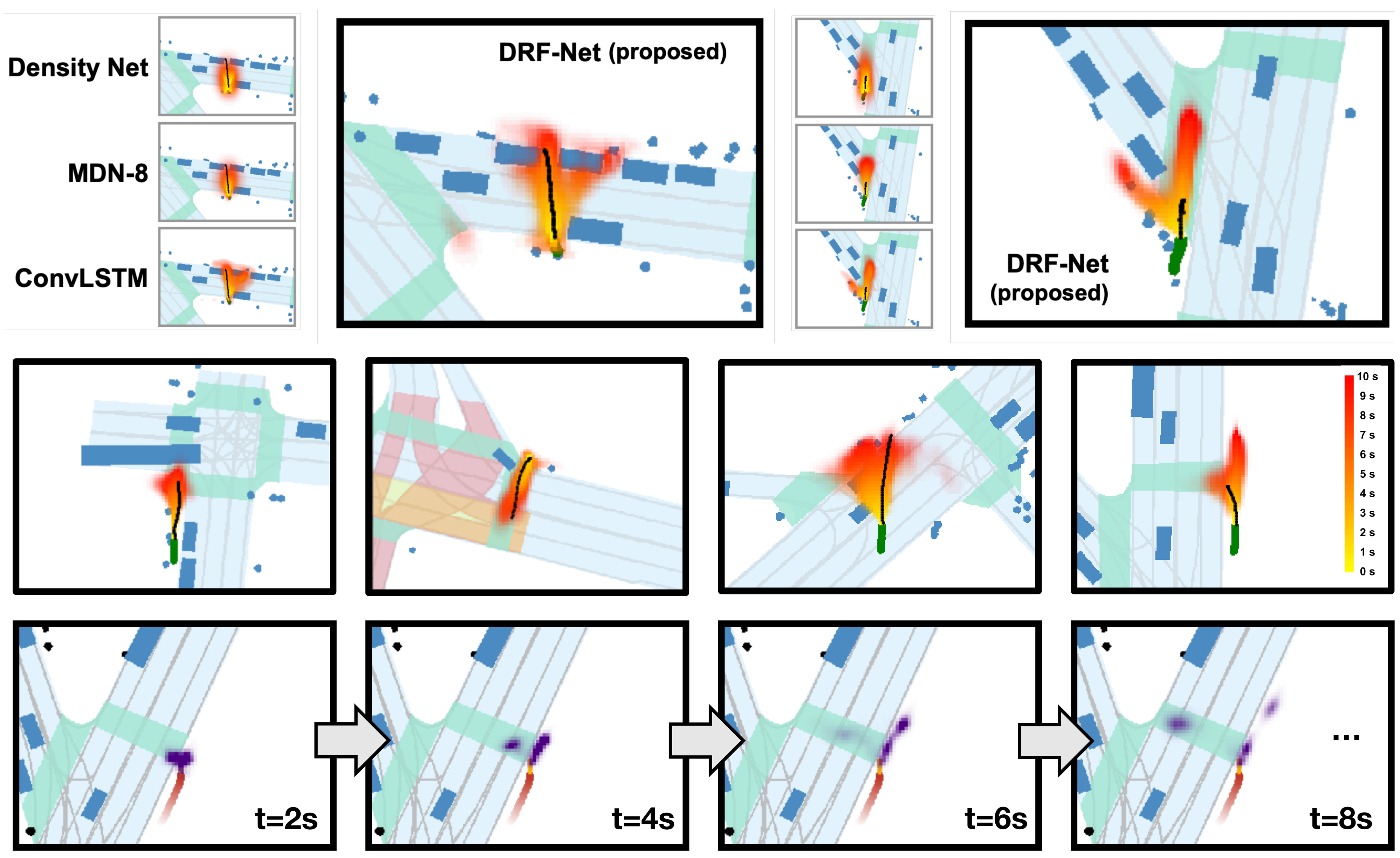
$$p_{\mathbf{x}_t}(x_t \mid \Omega) = \left[\sum_{x_{t-1}} \frac{p_{\mathbf{x}_t|\mathbf{x}_{t-1}}(x_t \mid x_{t-1}, \Omega)\, p_{\mathbf{x}_{t-1}}(x_{t-1} \mid \Omega)}{p_{\mathbf{x}_{t-1}}(x_t \mid \Omega)} \right] p_{\mathbf{x}_{t-1}}(x_t \mid \Omega)$$

$$\approx \frac{1}{Z_t} \underbrace{\psi_{t;\theta_t}\left(x_t, p_{\mathbf{x}_{t-1}}(\,\cdot \mid \Omega), \Omega\right)}_{\text{Exponentiated residual}} p_{\mathbf{x}_{t-1}}(x_t \mid \Omega)$$

Network architecture



Qualitative results



Evaluation

Model	Negative log likelihood (NLL) Mean	@ 1 s	@ 3 s	@ 10 s	ADE (m) 0.2-10s	FDE (m) @ 1 s	@ 3 s	@ 10 s	Mass Ratio (%) Acc.	Recall	Real detection data (NLL) Mean	@ 1 s	@ 3 s	@ 10 s
Density Net	5.39	2.87	3.96	6.74	3.49	0.93	1.72	7.66	77.99	81.33	5.64	1.88	4.12	7.91
MDN-4	3.01	1.64	2.00	4.33	1.47	0.38	0.69	3.38	87.85	84.12	3.21	1.52	2.54	4.71
MDN-8	3.43	1.60	2.77	4.79	1.78	0.60	0.88	3.91	85.56	84.19	3.21	1.53	2.55	4.73
ConvLSTM	2.51	0.89	1.86	4.07	1.58	0.47	1.06	3.20	88.02	85.02	3.14	1.54	2.51	4.64
DRF-NET	**2.37**	**0.76**	**1.74**	**3.83**	**1.23**	**0.35**	**0.62**	**2.71**	**89.78**	**85.41**	**2.98**	**1.47**	**2.39**	**4.36**