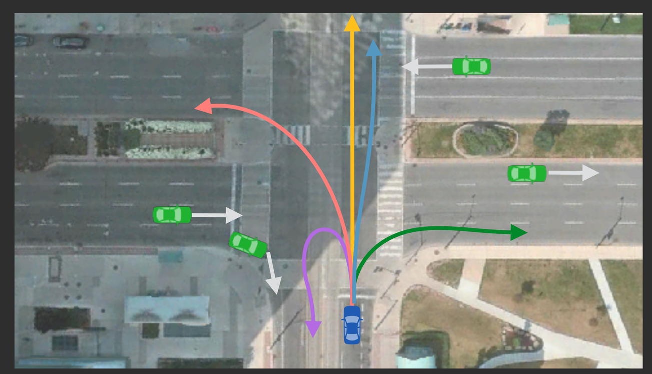
Multiple Futures Prediction

Yichuan Charlie Tang, Ruslan Salakhutdinov
Apple




Challenge

How to make multi-agent interactive and multimodal future predictions?

Our Contributions

- Multimodality: Learning semantically meaningful latent modes for *without* annotations
- Interaction: Parallel and interactive *step-wise* rollouts
- Dynamic *attention* to capture agent relationships
- Hypothetical rollouts and conditional inference
- State-of-the-art performances on several vehicle trajectory prediction datasets

Equations

$\mathbf{X}$ Past $\quad$ $\mathbf{Y}$ Future $\quad$ Z Latent modes $\quad$ $\mathcal{I}$ Scene context

Objective:

$$\log p(\mathbf{Y}|\mathbf{X},\mathcal{I}) = \log(\sum_Z p(\mathbf{Y},Z|\mathbf{X},\mathcal{I})) = \log(\sum_Z p(\mathbf{Y}|Z,\mathbf{X},\mathcal{I})p(Z|\mathbf{X},\mathcal{I}))$$

Timesteps

$$p(\mathbf{Y}|Z,\mathbf{X},\mathcal{I}) = \prod_{\delta=t+1}^{T} p(\mathbf{Y}_\delta|\mathbf{Y}_{t:\delta-1},Z,\mathbf{X},\mathcal{I})$$

All vehs

$$p(\mathbf{Y}_\delta|\mathbf{Y}_{t:\delta-1},Z,\mathbf{X},\mathcal{I}) = \prod_{n=1}^{N} p(\mathbf{y}_\delta^n|\mathbf{Y}_{t:\delta-1},z^n,\mathbf{X},\mathcal{I})$$

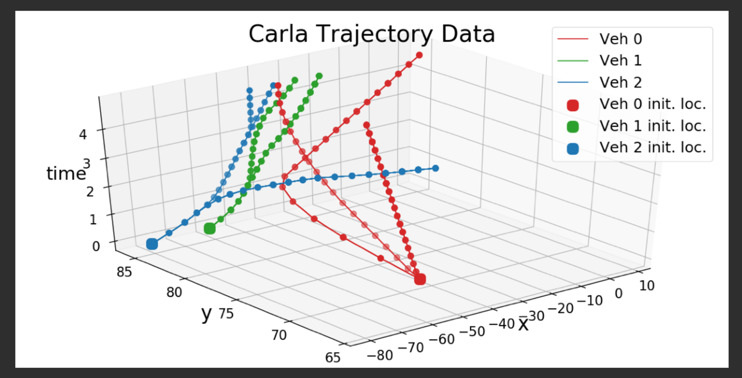
Final factorization

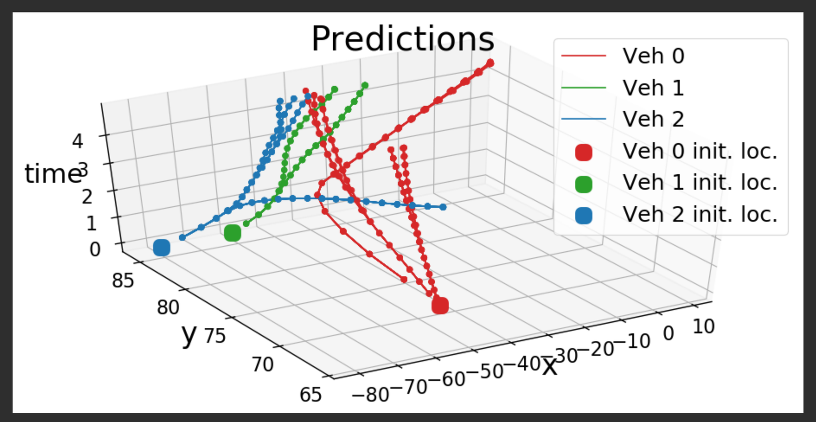
$$\log\left(\sum_Z p(\mathbf{Y}|Z,\mathbf{X},\mathcal{I})p(Z|\mathbf{X},\mathcal{I})\right) = \log\left(\sum_Z \prod_{\delta=t+1}^{T}\prod_{n=1}^{N} p(\mathbf{y}_\delta^n|\mathbf{Y}_{t:\delta-1},z^n,\mathbf{X},\mathcal{I})p(z^n|\mathbf{X},\mathcal{I})\right)$$
$$= \log\left(\sum_Z \prod_{n=1}^{N} p(z^n|\mathbf{X},\mathcal{I}) \prod_{\delta=t+1}^{T} p(\mathbf{y}_\delta^n|\mathbf{Y}_{t:\delta-1},z^n,\mathbf{X},\mathcal{I})\right)$$

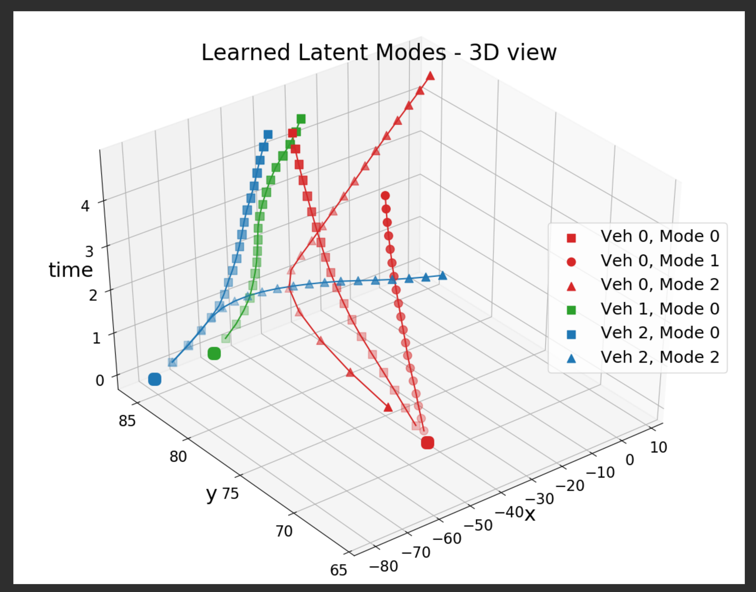
ELBO:

$$\log p(\mathbf{Y}|\mathbf{X};\theta) = \sum_Z q(Z|\mathbf{Y},\mathbf{X})\log\frac{p(\mathbf{Y},Z|\mathbf{X};\theta)}{q(Z|\mathbf{Y},\mathbf{X})} + D_{KL}(q||p) \geq \sum_Z q(Z|\mathbf{Y},\mathbf{X})\log p(\mathbf{Y},Z|\mathbf{X};\theta) + H(q)$$

Experimental Results - Carla Interactions

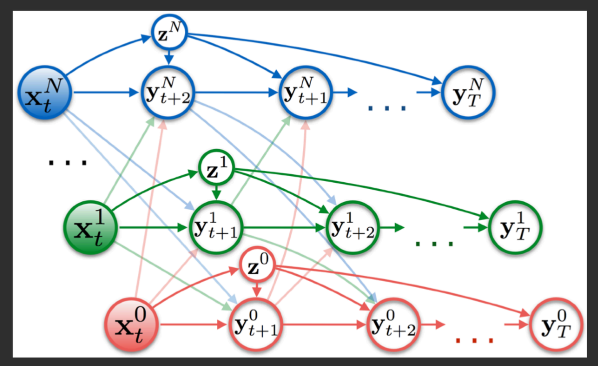





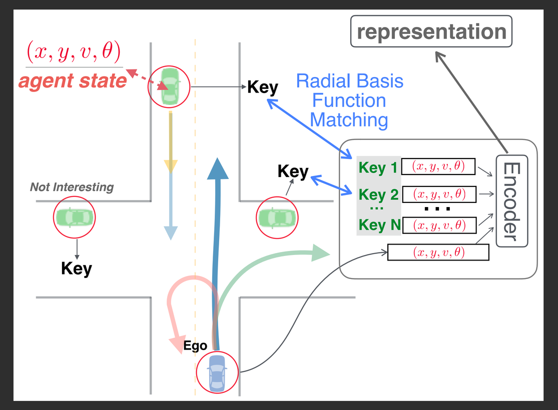


NLL	C.V.	RNN	MFP1	MFP2	MFP3	MFP4	MFP5
nats	11.46	5.64	5.23	3.37	1.72	**1.39**	**1.39**

Graphical Model

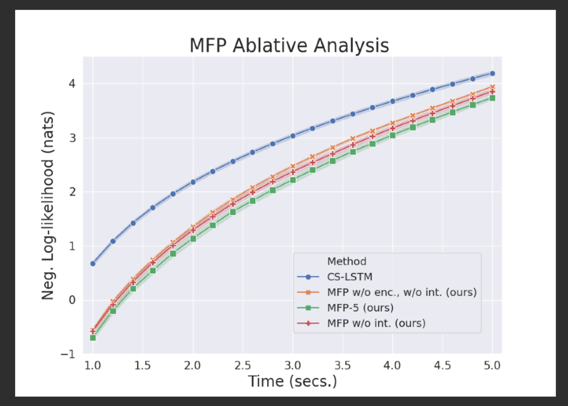


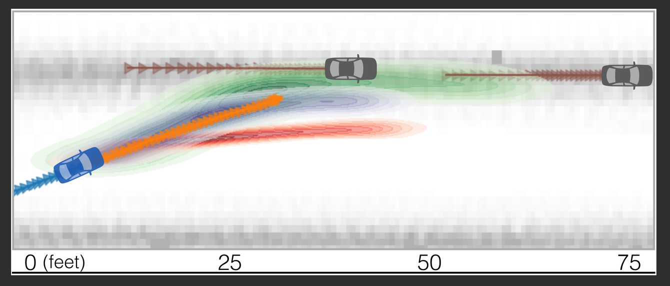
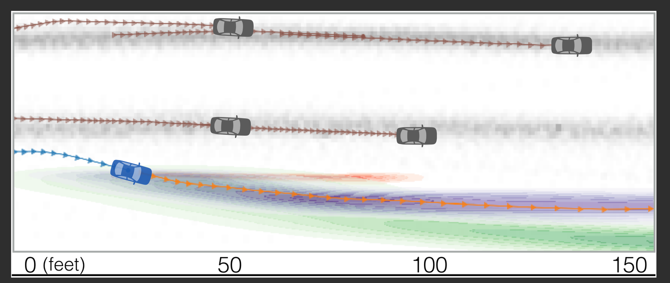
Dynamic Attention



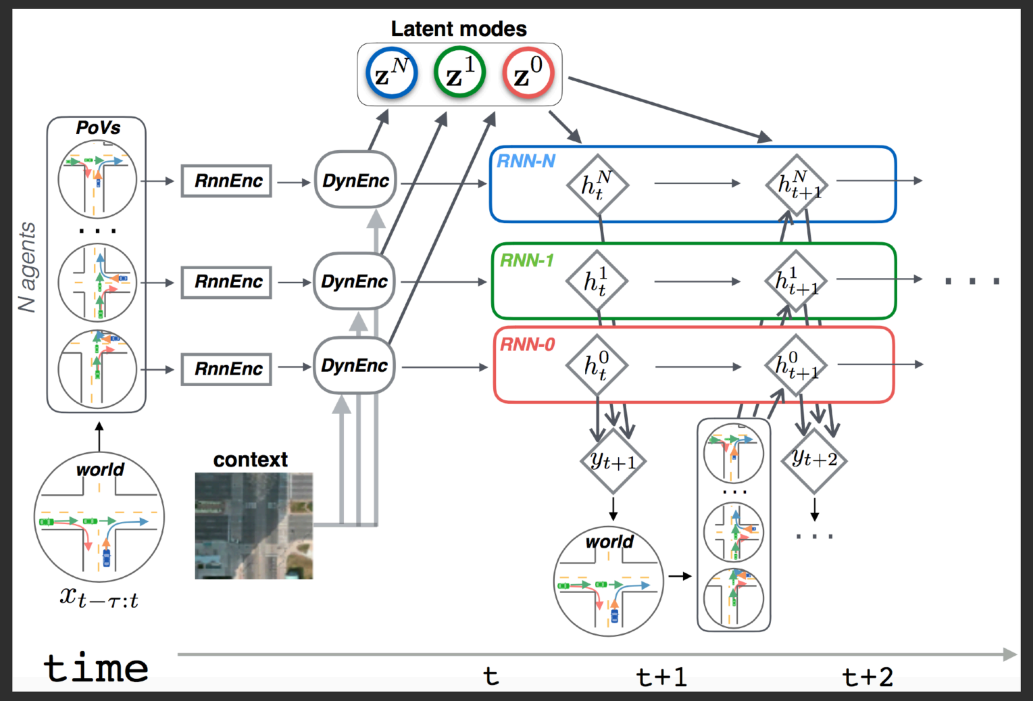
Experimental Results - NGSIM

time (secs) / NLL (nats)	C.V.	CVGMM	LSTM	Social LSTM	Conv SLSTM	CLSTM Multi	MFP1	MFP2	MFP3	MFP4	MFP5
1	3.72	2.02	1.17	1.01	0.89	0.58	0.73	-0.32	-0.58	**-0.65**	-0.45
2	5.37	3.63	2.85	2.49	2.43	2.14	2.33	1.43	1.26	**1.19**	1.36
3	6.40	4.62	3.80	3.36	3.30	3.03	3.17	2.45	2.32	**2.28**	2.42
4	7.16	5.35	4.48	4.01	3.97	3.68	3.77	3.21	3.07	**3.06**	3.17
5	7.76	5.93	4.99	4.54	4.51	4.22	4.26	3.81	**3.69**	**3.69**	3.76



MFP Architecture

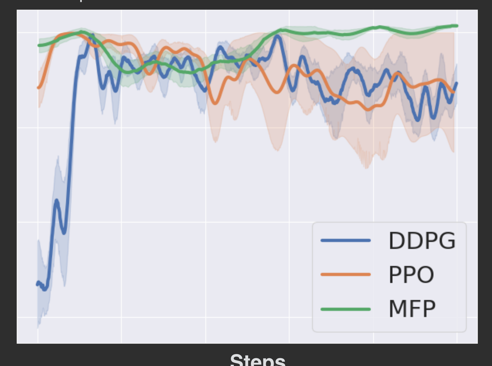


Planning & Decision Making

- MFP can be used for planning (trajectory optimization)
- MFP model estimations used for collision and safety checking
- Better sample complexity and performance to model-free RL methods

Parameters	Values
Episode Length	20 sec
Velocity	[5, 25] m/s
Reward: Success	+100
Reward: Collision	-500

Unprotected Left Turn Environment



Params	DDPG	PPO	MFP
+0 m/s	3%	4%	0%
+5 m/s	8%	4%	0%
+10 m/s	6%	15%	0%
+1 m/s^2	3%	1%	0%

Experimental Results - Carla

minMSD K=12	DESIRE	sGAN	R2P2-MA	ESP	ESP+ Lidar	MultiPath	MFP7
Town01 Test	2.599	1.464	0.843	1.213	0.716	0.68	**0.584**
Town02 Test	2.422	1.141	0.770	1.102	0.675	0.69	**0.512**

minMSD (meters)	MFP3	MFP5 no lidar	MFP5	MFP7
Town02 Test	0.632 (0.010)	0.609 (0.007)	0.595 (0.010)	**0.512 (0.004)**

Experimental Results - Argoverse

minADE K=6	C.V.	NN+Map	LSTM	LSTM+Map	MFP3(v1.0)	MFP3(v1.1)
Town01 Test	3.55	2.28	2.27	2.25	**1.411**	**1.399**