IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR 2017

DESIRE: DISTANT FUTURE PREDICTION IN DYNAMIC SCENES WITH INTERACTING AGENTS

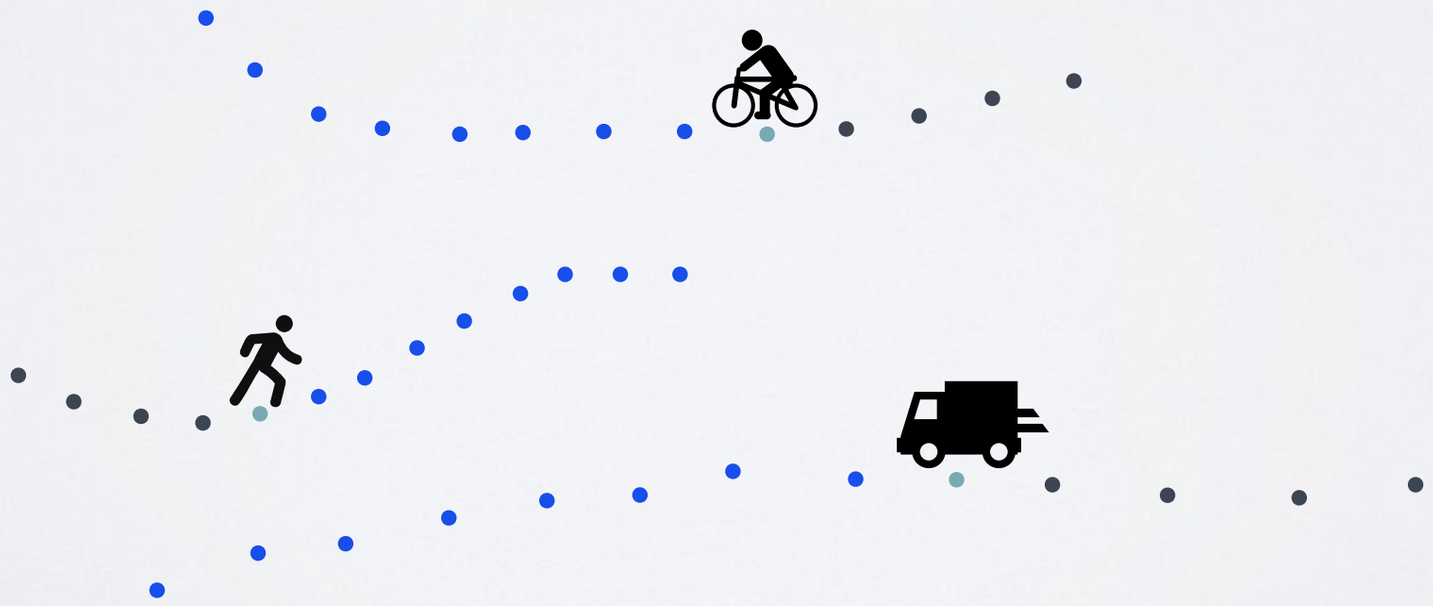
Namhoon Lee[1], Wongun Choi[2], Paul Vernaza[2], Christopher B. Choy[3],
Philip H. S. Torr[1], Manmohan Chandraker[2,4]

1: University of Oxford, 2: NEC Labs, 3: Stanford University, 4: UCSD

FUTURE PREDICTION

- We address the problem of ***future prediction*** for multiple agents in dynamic scenes.

- Future prediction is defined as ***predicting agents' future locations in terms of trajectories***.



FUTURE PREDICTION IS ***DIFFICULT***

- ***Various factors***

  A prediction entails reasoning about probable outcomes from multiple influences (e.g., ***past motion, scene context, interactions***).

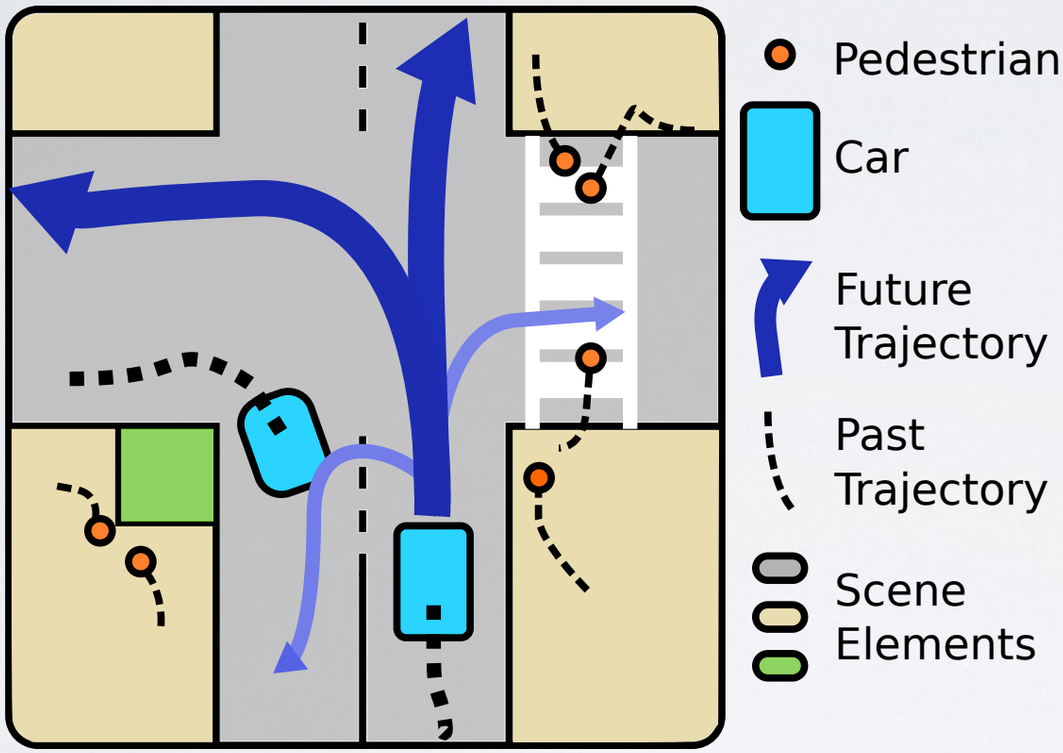
  It requires accurate ***time-profile*** for inter-influence between agents.

- ***Multi-modality***

  Future prediction is inherently riddled with ***uncertainty*** and is fundamentally different from path prediction.

  A system needs to produce a ***distribution over all probable outcomes*** (future), instead of one deterministic output (a path).

FUTURE PREDICTION IS ***DIFFICULT***



problem scenario

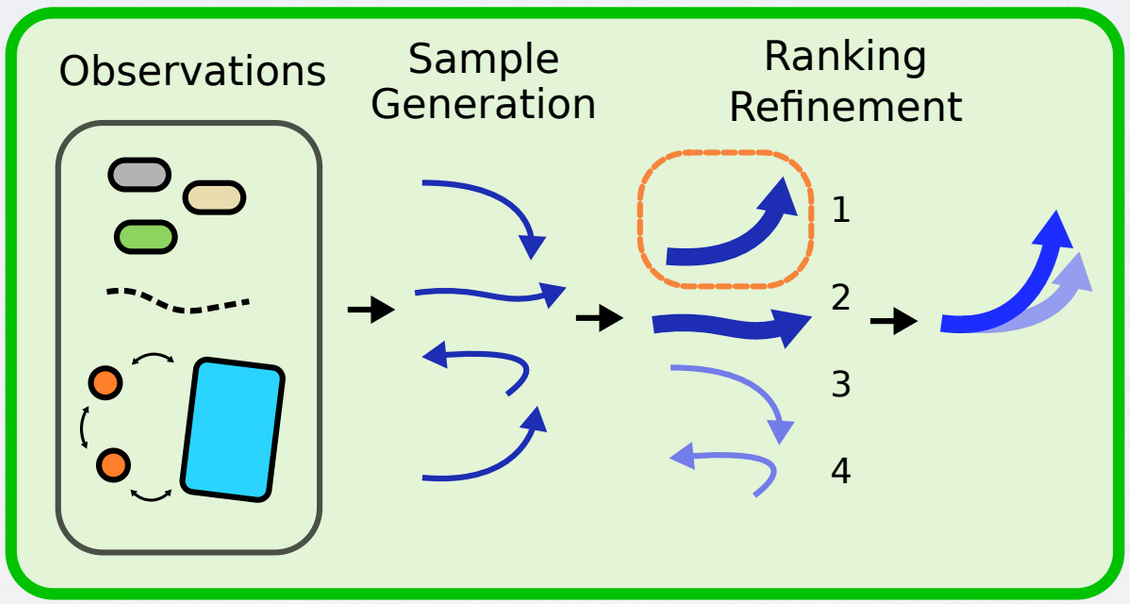
- ***Various factors***
  (past motion, scene context, interactions).

- ***Multi-modality***
  distribution over all probable outcomes

DESIRE:
DEep ***S***tochastic ***I***OC ***R***NN ***E***ncoder-decoder

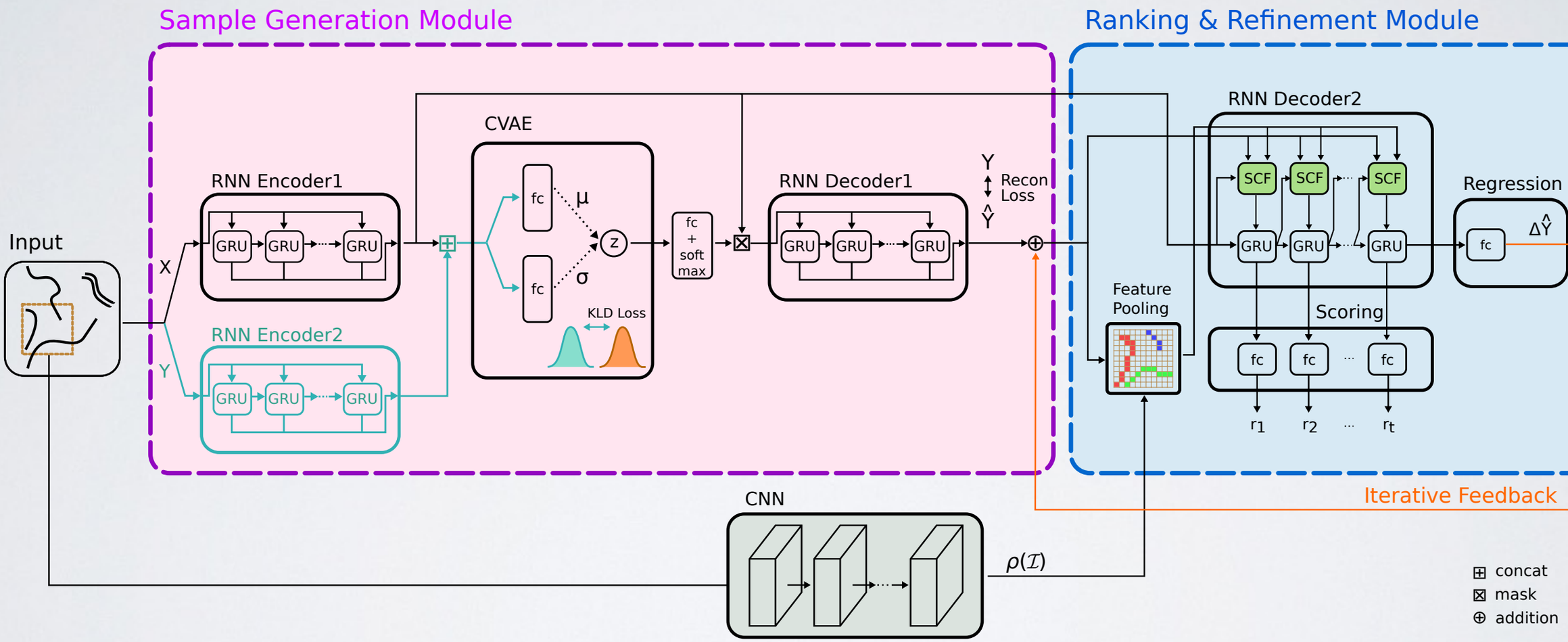
- DESIRE is a framework for distant future prediction of multiple interacting agents in dynamic scenes.

- We generate multiple prediction hypothesis using ***Variational Auto-Encoder*** and rank-and-refine them within ***Inverse Optimal Control*** framework.

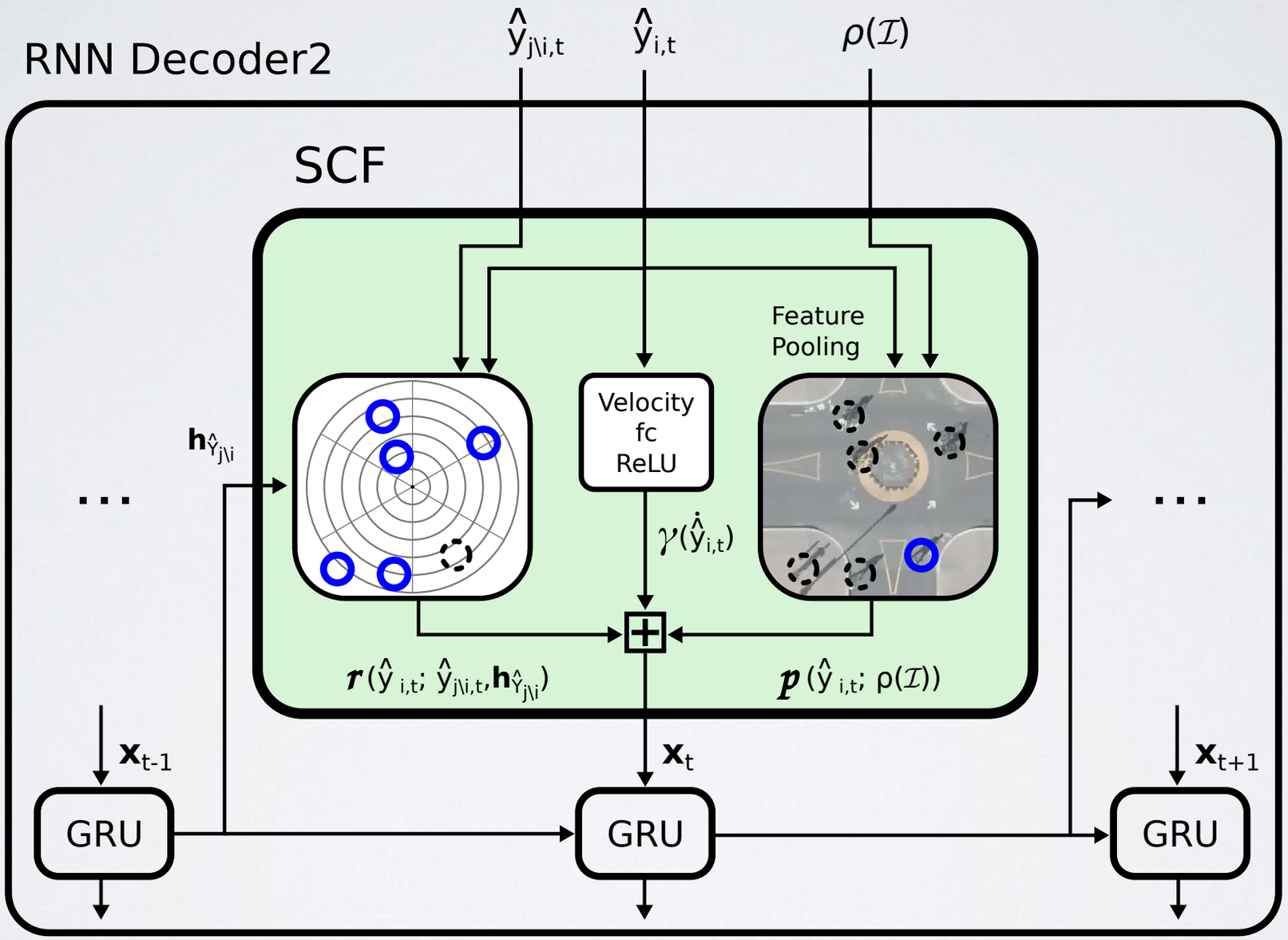


workflow

DESIRE:
***DE**ep **S**tochastic **I**OC **R**NN **E**ncoder-decoder*

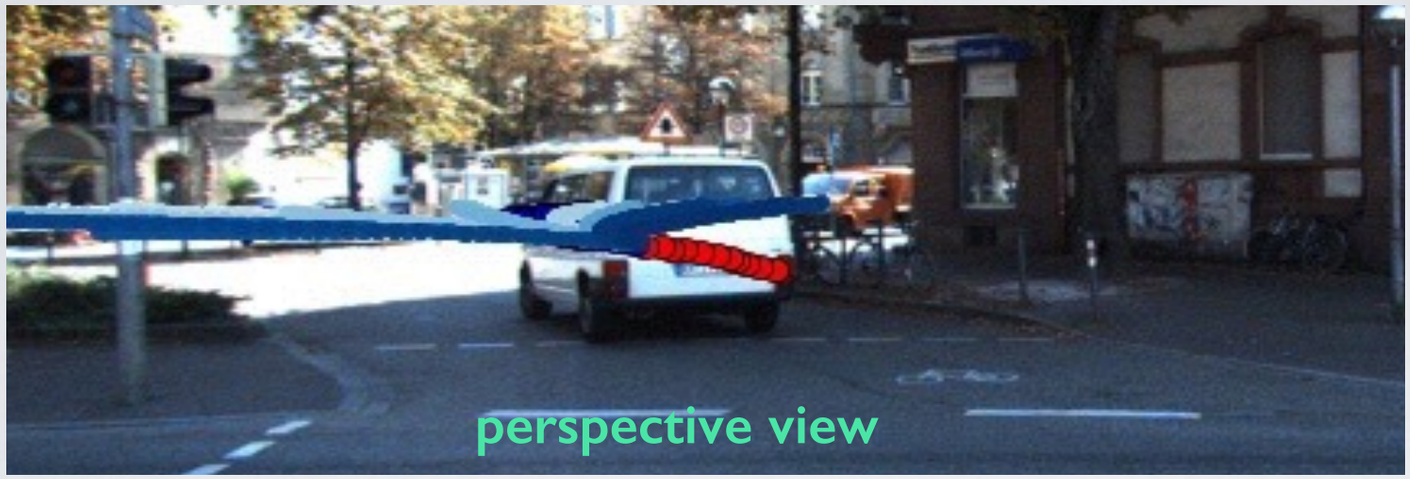


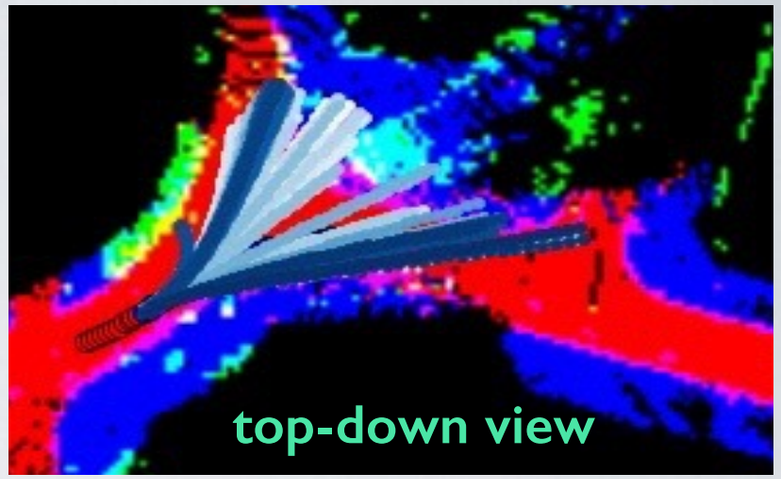
SCENE CONTEXT FUSION (SCF) UNIT



$$\mathbf{x}_t = \left[\gamma(\hat{v}_{i,t}), p(\hat{y}_{i,t}; \rho(\mathcal{I})), r(\hat{y}_{i,t}; \hat{y}_{j\backslash i,t}, \mathbf{h}_{\hat{Y}_{j\backslash i}})\right]$$

Prediction example





Iterative feedback

Prediction errors

(10% acc. for CVAE and DESIRE)

	KITTI (error in meters / miss-rate with 1m threshold)				SDD (pixel error at 1/5 resolution)			
Method	1s	2s	3s	4s	1s	2s	3s	4s
Linear	0.89 / 0.31	2.07 / 0.49	3.67 / 0.59	5.62 / 0.64	2.58	5.37	8.74	12.54
RNN ED-SI	0.56 / 0.16	1.40 / 0.44	2.65 / 0.58	4.29 / 0.65	1.51	3.56	6.04	8.80
CVAE	0.35 / 0.06	0.93 / 0.30	1.81 / 0.49	3.07 / 0.59	1.84	3.93	6.47	9.65
DESIRE-S-IT0	0.32 / 0.05	0.84 / 0.26	1.67 / 0.43	2.82 / 0.54	1.59	3.31	5.27	7.75
DESIRE-SI-IT4	**0.28 / 0.04**	**0.67 / 0.17**	**1.22 / 0.29**	**2.06 / 0.41**	**1.29**	**2.35**	**3.47**	**5.33**

DESIRE CHARACTERISTICS

- ***Scalability***:
  The use of deep learning allows for end-to-end training and easy incorporation of multiple cues.

- ***Diversity***:
  CVAE is combined with RNN encodings to generate stochastic prediction hypotheses to hallucinate multi-modalities.

- ***Accuracy***:
  The IOC-based framework accumulates long-term future rewards and the refinement module learns to estimate a deformation of the trajectory, enabling more accurate predictions.

THANK YOU