Supplementary: Multimodal Future Locatization and Emergence Prediction for Objects in Egocentric View with a Reachability Prior

Osama Makansi[1] Özgün Çiçek[1] Kevin Buchicchio[2] Thomas Brox[1]
[1]University of Freiburg [2]IMRA-EUROPE
makansio,cicek,brox@cs.uni-freiburg.de buchicchio@imra-europe.com

1. Video

We provide a supplemental video to present our results better. Since the task inherits a temporal dependency, we refer the reader to our video where the driving scenarios are presented as they happen.

2. Egocentric Future Localization

For each dataset, we split the testing scenarios into challenging and very challenging categories based on their errors when Kalman Filter is used for future prediction (see more details in the main paper). Table 1 shows the quantitative comparison of our future localization framework against all baselines on the nuScenes [2] testing dataset for all scenarios, only the challenging ones, and only the very challenging ones. We clearly show that our framework outperforms all baselines in all difficulties. The benefit gained from our methods is even larger as the difficulty of the scenarios increases.

To show zero-shot transfer to unseen datasets, we report the same evaluation on the testing split of the Waymo Open dataset [7] in Table 2. The ranking of the methods is preserved as in the evaluation on nuScenes dataset. This shows that our framework using the reachability prior generalizes well to unseen scenarios. Note we also report the size of the testing dataset for each category where a significant drop in the number of scenarios is observed when the difficulty level increases.

To show robustness to datasets with noisy annotation, we report the same evaluation on our FIT dataset in Table 3. Similarly, our framework outperforms all baselines in all difficulties. Note that this simulates the real world applications where accurate annotations (e.g, object detection and tracking) are expensive to obtain.

3. Egocentric Emergence Prediction

We show two emergence prediction examples in Figure 1 for cars (1st row) and pedestrians (2nd row). In the first scenario, a car can emerge from the left street, from far distance, or from the occluded area by the truck. In the second scenario with a non-straight egomotion, a pedestrian can emerge from different occluded areas by the left moving car, the left parking cars, or the right truck. Note how the reachability prior helps the emergence prediction framework to cover more possible modes. Interestingly, the reachability prior prediction is different from the emergence prediction where close by objects (cars and pedestrians) are only part of the reachability prior.

4. Failure Cases

Our method is mainly based on the sampling network from Makansi et al. [4] and thus inherits its failures. The sampling network is trained with the EWTA objective which leads sometimes to generating few bad hypotheses (outliers). Figure 2 shows few examples for this phenomena. One promising direction in future work is finding strategies for better sampling to overcome this limitation.

	All ($11k$)			Challenging ($3.3k$)			Very Challenging ($1.4k$)		
	FDE ↓	IOU ↑	NLL ↓	FDE ↓	IOU ↑	NLL ↓	FDE ↓	IOU ↑	NLL ↓
Kalman [3]	45.02	0.31	–	114.50	0.03	–	179.92	0.01	–
DTP [6]	35.88	0.34	–	77.91	0.11	–	111.49	0.05	–
RNN-ED-XOE [8]	30.47	0.34	–	56.43	0.19	–	78.54	0.13	–
STED [5]	27.71	0.39	–	57.32	0.21	–	82.71	0.13	–
Baysian based on [1]	28.51	0.37	19.75	58.14	0.20	26.16	82.23	0.13	28.44
FLN w/o Reachability	15.91	0.54	19.46	32.36	0.38	24.62	47.15	0.29	26.85
FLN + Reachability	**12.82**	**0.55**	**17.90**	**24.23**	**0.40**	**22.08**	**32.68**	**0.33**	**24.17**

Table 1. Quantitative results of the future localization task on the nuScenes [2] dataset. The bottom three methods predict multimodal distribution allowing the NLL evaluation. Three categories are shown with their sizes in parentheses.

	All ($47.2k$)			Challenging ($13.9k$)			Very Challenging ($7.1k$)		
	FDE ↓	IOU ↑	NLL ↓	FDE ↓	IOU ↑	NLL ↓	FDE ↓	IOU ↑	NLL ↓
Kalman [3]	31.69	0.39	–	85.51	0.05	–	124.71	0.02	–
DTP [6]	28.31	0.38	–	62.29	0.14	–	82.64	0.10	–
RNN-ED-XOE [8]	25.23	0.36	–	47.09	0.21	–	59.23	0.18	–
STED [5]	20.73	0.42	–	44.03	0.24	–	58.14	0.20	–
Baysian based on [1]	23.75	0.38	18.80	48.66	0.21	25.06	64.67	0.17	27.54
FLN w/o Reachability	13.20	0.54	18.84	26.62	0.40	23.90	36.57	0.34	26.19
FLN + Reachability	**10.35**	**0.58**	**16.63**	**20.73**	**0.42**	**21.26**	**27.15**	**0.37**	**22.95**

Table 2. Quantitative results of the future localization on the Waymo Open dataset [7]. The bottom three methods predict multimodal distribution allowing the NLL evaluation. Three categories are shown with their sizes in parentheses.

References

[1] A. Bhattacharyya, M. Fritz, and B. Schiele. Long-term on-board prediction of people in traffic scenes under uncertainty. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018.

[2] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[3] R. E. Kalman. A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering*, 1960.

[4] Osama Makansi, Eddy Ilg, Ozgun Cicek, and Thomas Brox. Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[5] Olly Styles, Tanaya Guha, and Victor Sanchez. Multiple object forecasting: Predicting future object locations in diverse environments, 2019.

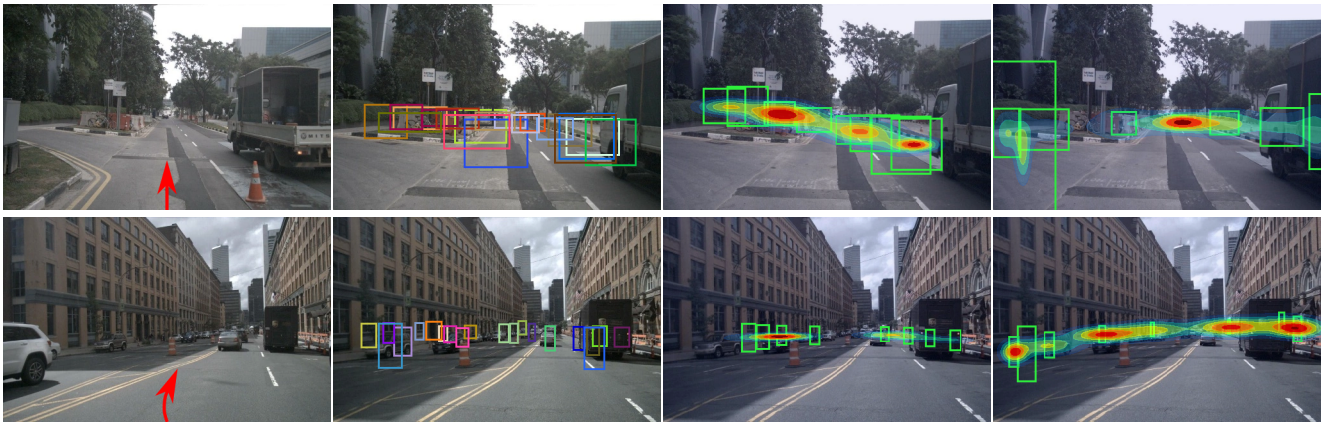
[6] O. Styles, A. Ross, and V. Sanchez. Forecasting pedestrian trajectory with machine-annotated training data. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, June 2019.

[7] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset, 2019.

[8] Y. Yao, M. Xu, C. Choi, D. J. Crandall, E. M. Atkins, and B. Dariush. Egocentric vision-based future vehicle localization for intelligent driving assistance systems. In *2019 International Conference on Robotics and Automation (ICRA)*, May 2019.

	All (1442)			Challenging (404)			Very Challenging (223)		
	FDE ↓	IOU ↑	NLL ↓	FDE ↓	IOU ↑	NLL ↓	FDE ↓	IOU ↑	NLL ↓
Kalman [3]	38.33	0.36	–	105.82	0.08	–	146.50	0.03	–
DTP [6]	34.99	0.37	–	86.13	0.14	–	118.36	0.09	–
RNN-ED-XOE [8]	35.74	0.36	–	69.30	0.21	–	88.58	0.17	–
STED [5]	31.80	0.35	–	67.00	0.20	–	86.58	0.16	–
Baysian based on [1]	32.64	0.38	20.56	67.40	0.20	26.77	87.63	0.16	28.83
FLN w/o Reachability	18.12	0.53	20.38	37.55	0.37	25.98	47.92	0.33	27.88
FLN + Reachability	**15.41**	**0.54**	**19.08**	**26.99**	**0.42**	**23.42**	**32.14**	**0.39**	**24.73**

Table 3. Quantitative results of the future localization on our FIT dataset. The bottom three methods predict multimodal distribution allowing the NLL evaluation. Three categories are shown with their sizes in parentheses.



(a) Input t (b) Reachability $t + \Delta t$ (d) EPN w/o RPN $t + \Delta t$ (e) EPN w RPN $t + \Delta t$

Figure 1. Emergence Prediction qualitative results on nuScenes [2]. For each row (scenario), we show (a) the observed image and the planned ego-motion (red arrow) to the future, (b) the reachability prior resulted from our RTN in the future, (c-d) both variants of our emergence prediction framework.

(a) Input t (b) EPN w RPN $t + \Delta t$

Figure 2. Two examples from Waymo [7] dataset illustrating the outlier hypotheses generated by our method. In both examples, a pedestrian is expected to jump into the middle of the street by changing his/her behavior. Note that our method assign almost zero likelihood for those unlikely modes.