Uncertainty-Aware Spatiotemporal Perception for Autonomous Vehicles

Thesis Defense

Masha Itkina

February 7th, 2022

SISL
Stanford Intelligent
Systems Laboratory

Robots are here!

Waymo
TechCrunch, 2020.

Boston Dynamics
The Verge, 2020.

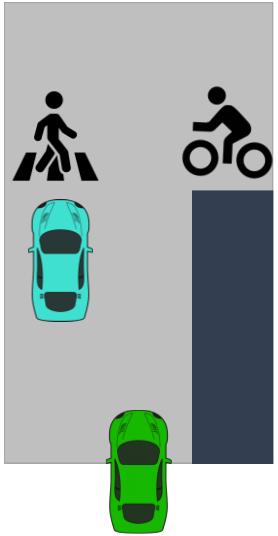
Toyota Research Institute [3]
The Robot Report, 2020.

But navigating human environments is still difficult!

[Credit: https://www.youtube.com/watch?v=gkbmUI0MsdM]

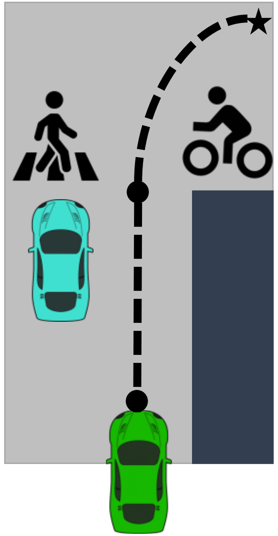
Robot Autonomy Stack for Autonomous Vehicles (AVs)

Perception



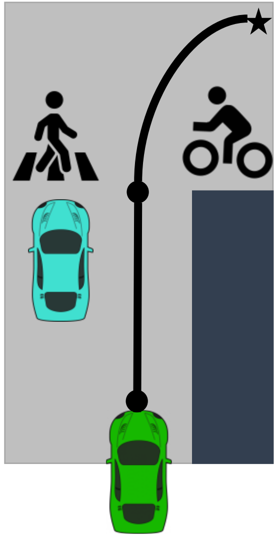
Robot Autonomy Stack for Autonomous Vehicles (AVs)

Perception ⇨ **Planning**



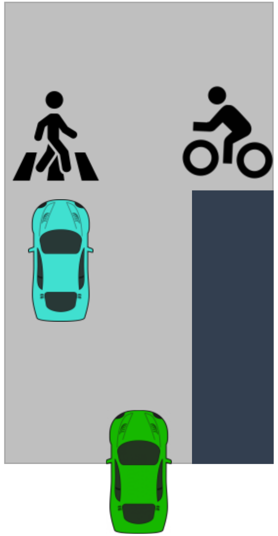
Robot Autonomy Stack for Autonomous Vehicles (AVs)

Perception → Planning → Control



Robot Autonomy Stack for Autonomous Vehicles (AVs)

Perception ⟹ Planning ⟹ Control



Key Insight

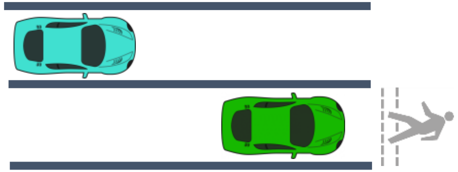
We can extend AV perception capabilities ***beyond on-board sensors*** *by making* ***uncertainty-aware inferences in time and space*** *from observed* ***human behaviors****.*

Perception Challenges

Temporal predictions in cluttered spaces

Perception Challenges

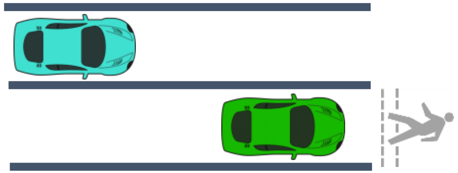
Temporal predictions in cluttered spaces



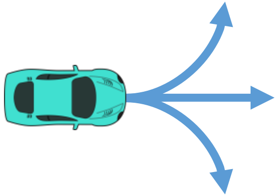
Spatial occlusions

Perception Challenges

Temporal predictions in cluttered spaces



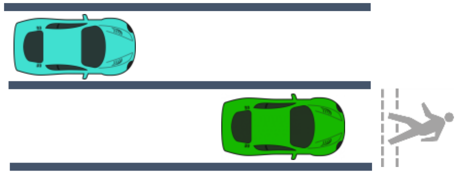
Spatial occlusions



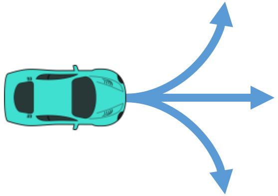
Multimodality in trajectory predictions

Perception Challenges

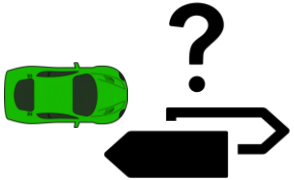
Temporal predictions in cluttered spaces



Spatial occlusions



Multimodality in trajectory predictions

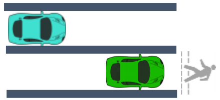


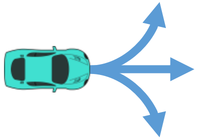
Previously unseen scenarios

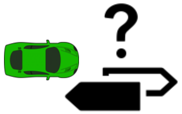
Contributions

Predict the spatiotemporal environment end-to-end from a map-centric (i.e., occupancy grid) representation of the world using recurrent representation learning.

[ITSC 2019*, IROS 2021, ICRA 2021]





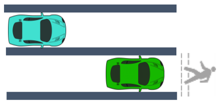


* M. Itkina is the first author. Otherwise second author.

Contributions

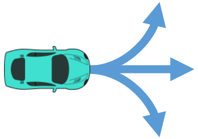
Predict the spatiotemporal environment end-to-end from a map-centric (i.e., occupancy grid) representation of the world using recurrent representation learning.

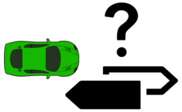
[ITSC 2019*, IROS 2021, ICRA 2021]



Infer occluded occupancy in multi-agent settings by modeling observed human behaviors as sensor measurements for the environment occupancy.

[ICRA 2022*]





* M. Itkina is the first author. Otherwise second author.

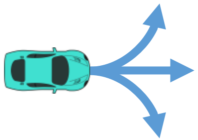
Contributions

Predict the spatiotemporal environment end-to-end from a map-centric (i.e., occupancy grid) representation of the world using recurrent representation learning.

[ITSC 2019*, IROS 2021, ICRA 2021]

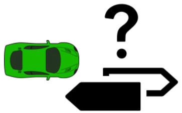
Infer occluded occupancy in multi-agent settings by modeling observed human behaviors as sensor measurements for the environment occupancy.

[ICRA 2022*]



Sparsify the latent sample space in deep generative models for trajectory prediction, while maintaining multimodality, using evidential theory.

[NeurIPS 2020*, NeurIPS 2021]



* M. Itkina is the first author. Otherwise second author.

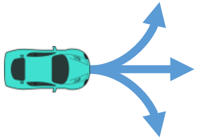
Contributions

Predict the spatiotemporal environment end-to-end from a map-centric (i.e., occupancy grid) representation of the world using recurrent representation learning.

[ITSC 2019*, IROS 2021, ICRA 2021]

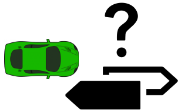
Infer occluded occupancy in multi-agent settings by modeling observed human behaviors as sensor measurements for the environment occupancy.

[ICRA 2022*]



Sparsify the latent sample space in deep generative models for trajectory prediction, while maintaining multimodality, using evidential theory.

[NeurIPS 2020*, NeurIPS 2021]



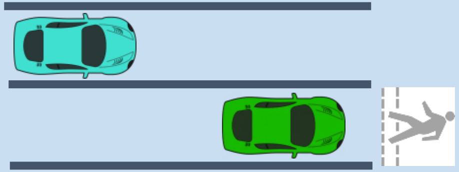
Model epistemic uncertainty for trajectory prediction using deep generative models and second-order distributions over low-dimensional, interpretable latent variables.

[In Progress*, ITSC 2021]

* M. Itkina is the first author. Otherwise second author.

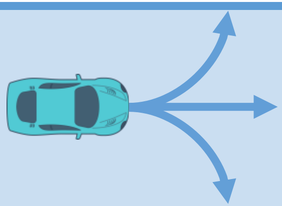
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]

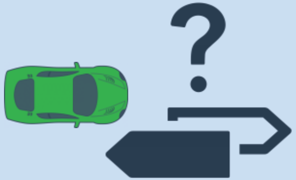


Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]



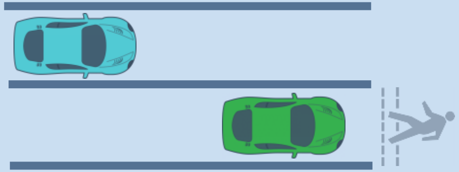
Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

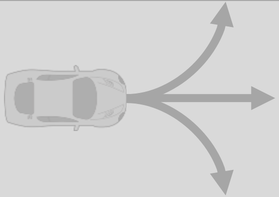
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]

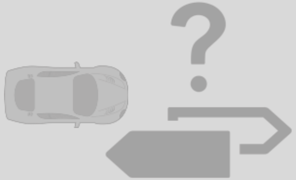


Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]



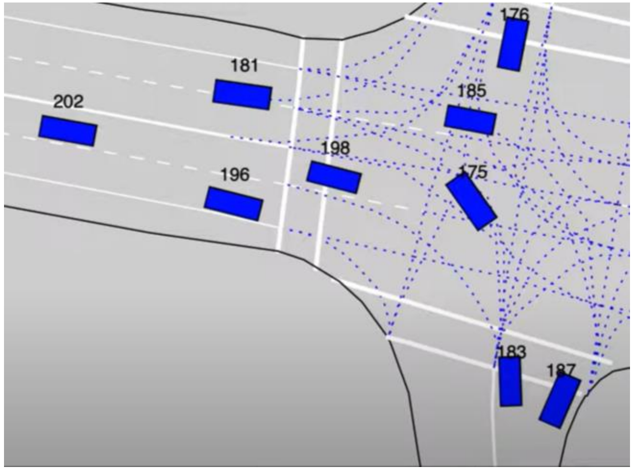
Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

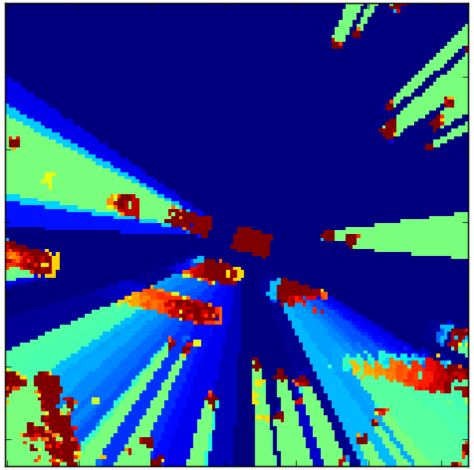
Environment Representations in Robotics

Object-centric (continuous):
e.g., Landmark approaches



Data from the INTERACTION Dataset.

Map-centric (discrete):
e.g., Occupancy grid maps (OGMs)



Data from the KITTI Dataset.

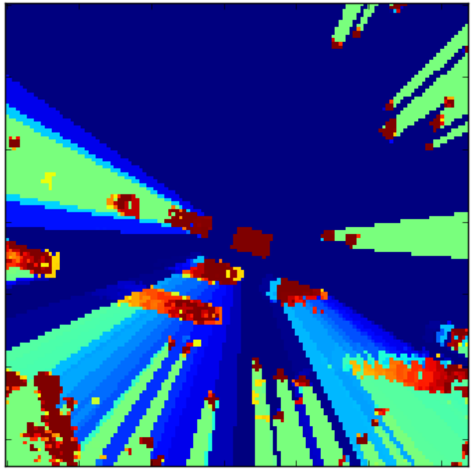
W. Zhan et al. Arxiv, 2019.
A. Geiger et al. In CVPR, 2012.

Environment Representations in Robotics

Object-centric (continuous):
e.g., Landmark approaches

Data from the INTERACTION Dataset.

Map-centric (discrete):
e.g., Occupancy grid maps (OGMs)

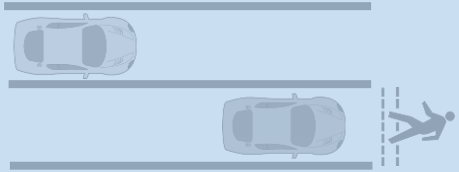


Data from the KITTI Dataset.

W. Zhan et al. Arxiv, 2019.
A. Geiger et al. In CVPR, 2012.

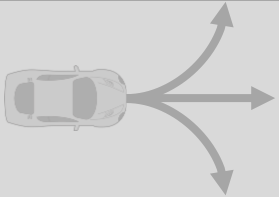
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]

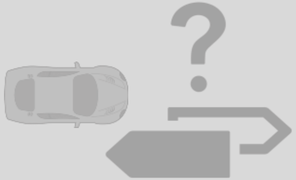


Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]



Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models

* M. Itkina is the first author. Otherwise second author.

End-to-End Spatiotemporal Environment Prediction in Cluttered Urban Scenes

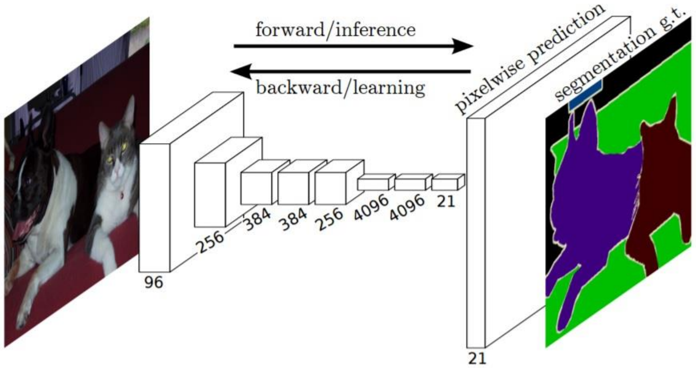
Research question:

*Can we use OGMs to generate **environment state predictions** end-to-end without additional dynamic information?*

M. Itkina, K. Driggs-Campbell, and M. J. Kochenderfer, "Dynamic environment prediction in urban scenes using recurrent representation learning". In ITSC, 2019.

Computer Vision Advancements

Spatial: Semantic Segmentation



J. Long et al. In CVPR, 2015.

Spatiotemporal: Video Frame Prediction

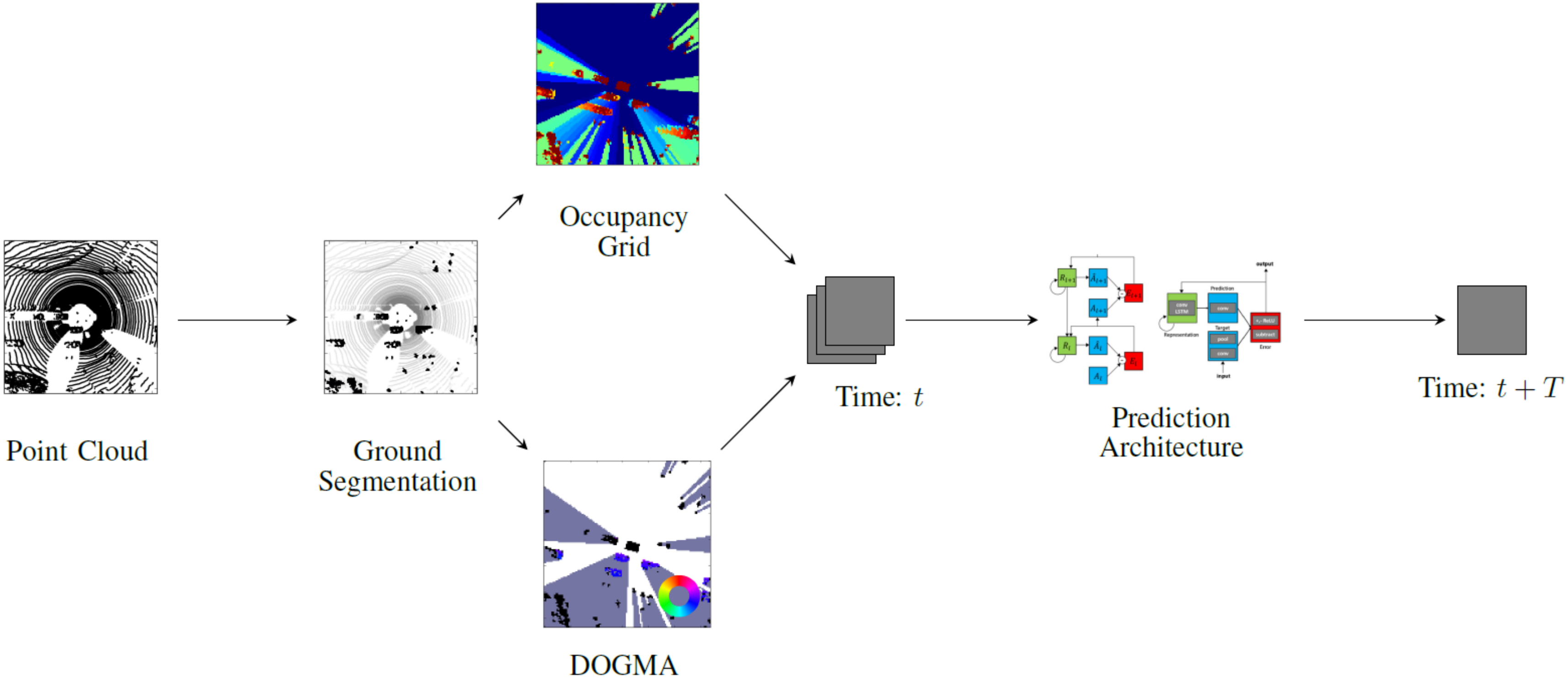
W. Lotter et al. In ICLR, 2017.

Spatiotemporal Environment Prediction Contributions

1. Formulate **environment prediction** in cluttered urban settings as a video frame prediction task.

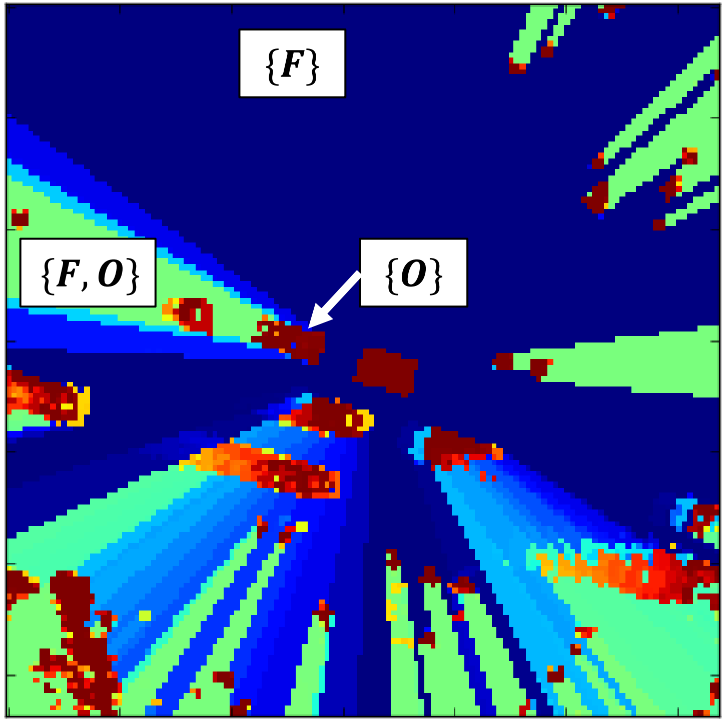
2. Demonstrate capacity of a ConvLSTM to learn an internal dynamic representation of the environment from OGM data **without additional dynamic input**.

Environment Prediction Approach



Evidential OGMs

- Frame of discernment: $\Omega = \{F, O\}$
- A belief mass m is assigned to each set in 2^{Ω}
- $\sum_{A \in 2^{\Omega}} m(A) = 1$
- Belief masses can be combined according to the DST fusion rule
- Probabilities are estimated from belief masses using pignistic probability

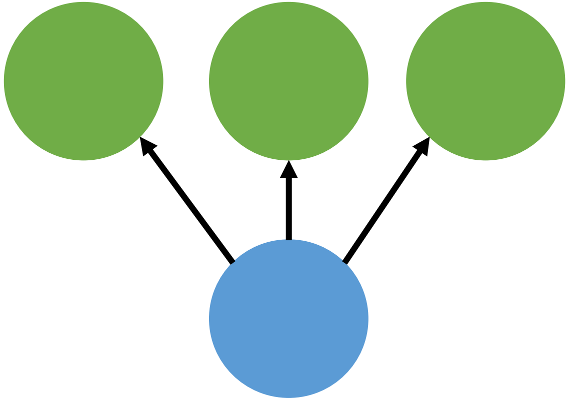


D. Nuss et al. In IV, 2015.

Evidential Theory (DST)

Evidential theory is a decision making strategy that distinguishes **lack of information** from **conflicting information**.

Uninformative prior

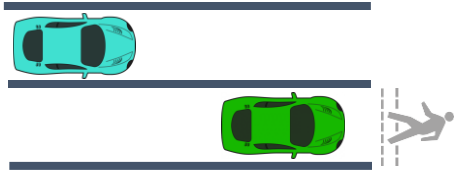


Evidence supporting multiple hypotheses

A. P. Dempster. Classic works of the Dempster-Shafer Theory of Belief Functions, 2008.

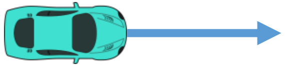
Evidential Theory (DST)

Evidential theory is a decision making strategy that distinguishes **lack of information** from **conflicting information**.



$$m(\{O\}) = 0$$
$$m(\{F\}) = 0$$
$$m(\{O, F\}) = 1$$

Occlusions

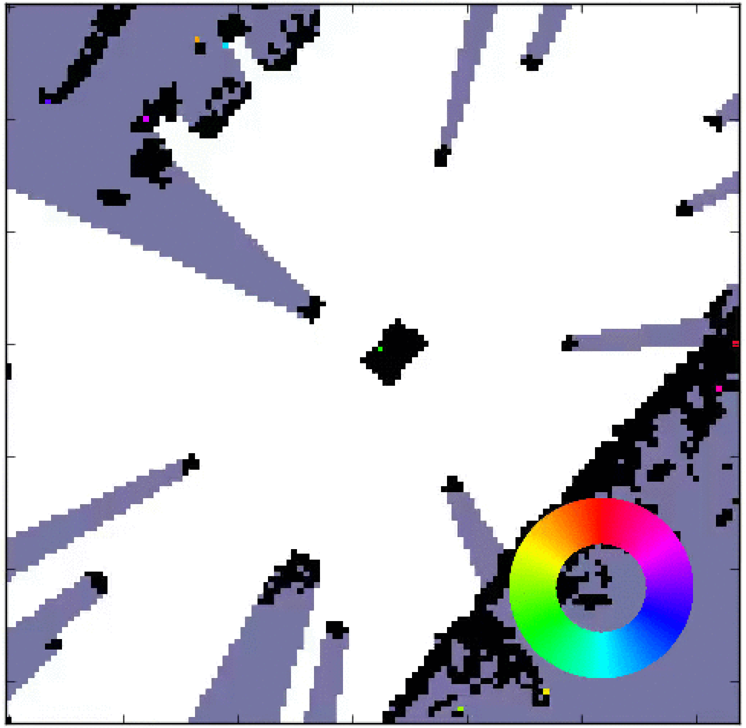


$$m(\{O\}) = 0.5$$
$$m(\{F\}) = 0.5$$
$$m(\{O, F\}) = 0$$

Moving Obstacles

A. P. Dempster. Classic works of the Dempster-Shafer Theory of Belief Functions, 2008.

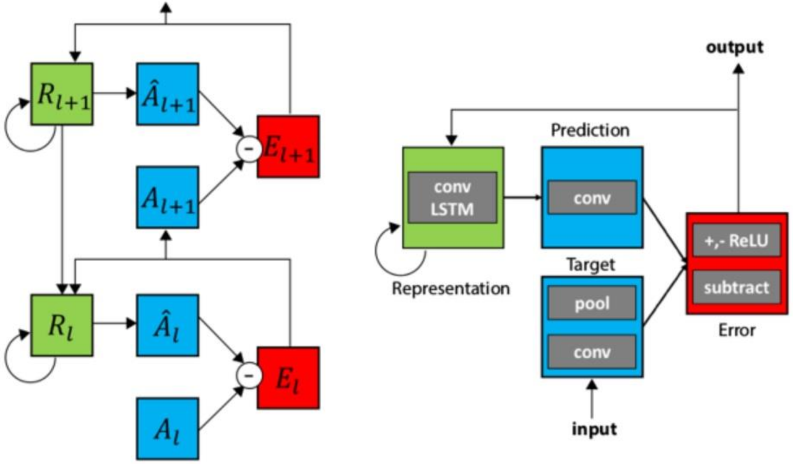
Dynamic Occupancy Grid Maps (DOGMas)



D. Nuss et al. IJRR, 2016.

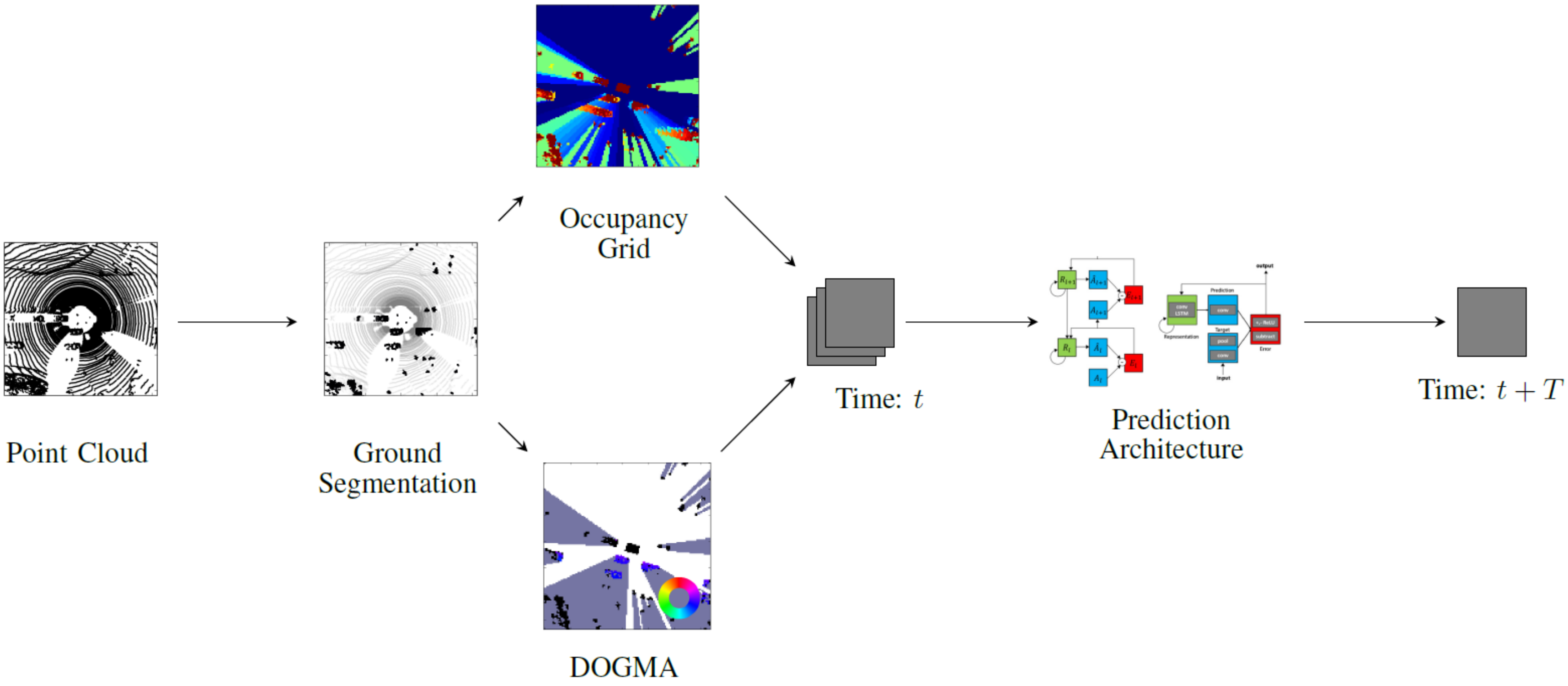
Environment Prediction Architecture: PredNet

- PredNet is a convolutional long short-term memory (ConvLSTM) network for video frame prediction
- Recurrent representation layer to learn spatiotemporal data structure
- Self-supervised



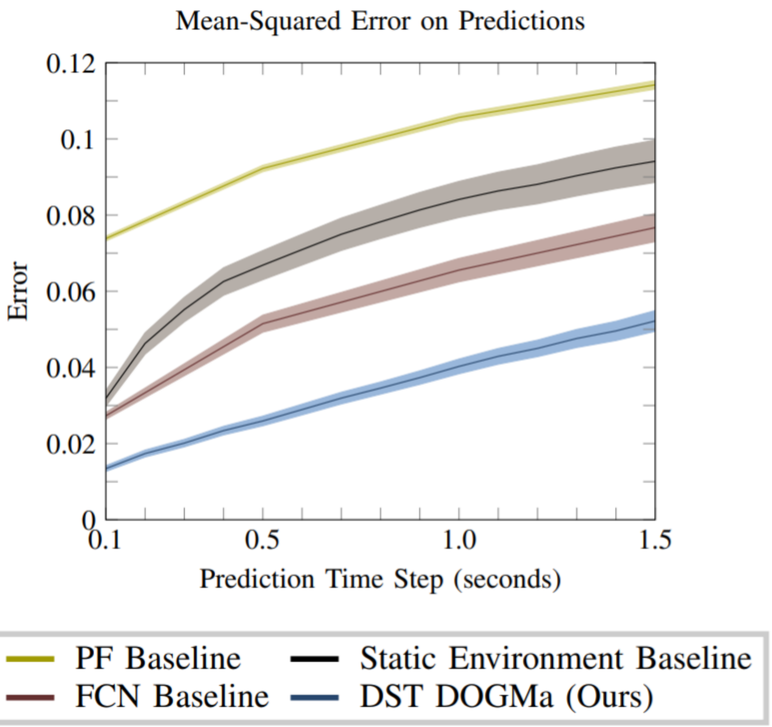
W. Lotter et al. In ICLR, 2017.

Environment Prediction Approach



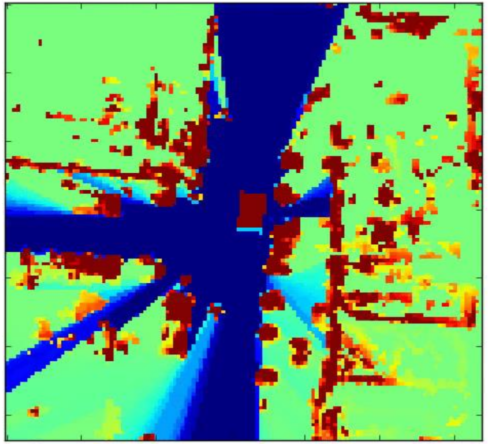
Results

- Data: KITTI LiDAR point clouds
- Grid size: 42.7 m x 42.7 m (33 cm res.)
- Task: predict 1.5 s OGMs ahead from observed 0.5 s

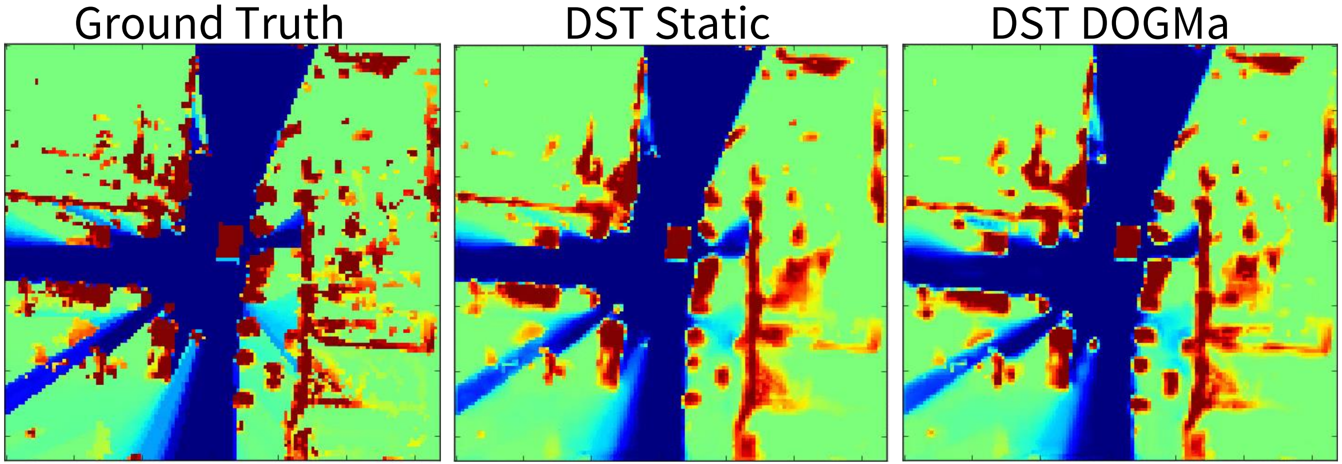


A. Geiger et al. In CVPR, 2012.

Results: Example Ground Truth



Results: Occupancy Grids vs DOGMas



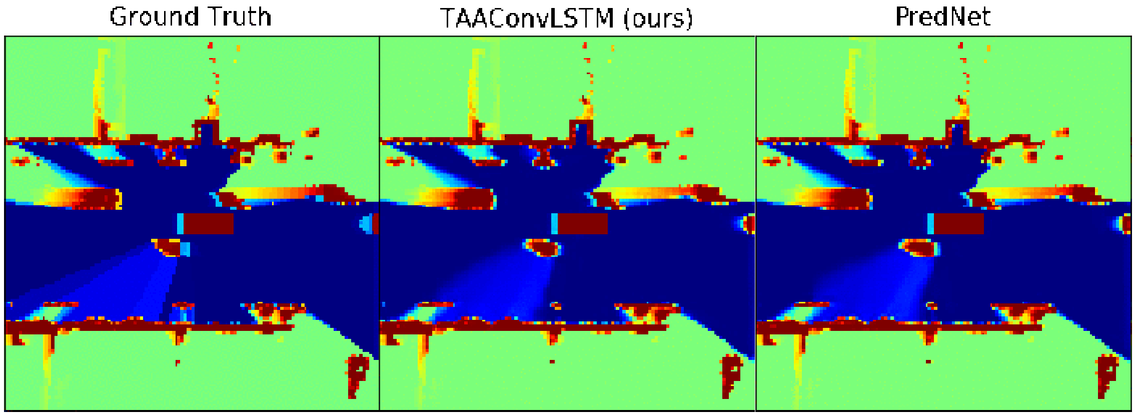
Results: Occupancy Grid vs DOGMa

Mean-Squared Error (MSE) results on test set (lower is better):

T	DST Static ($\times 10^{-3}$)	DST DOGMa ($\times 10^{-3}$)
$0.1\,\mathrm{s}$	13.5 ± 0.7	$\mathbf{13.4 \pm 0.7}$
$0.5\,\mathrm{s}$	26.4 ± 1.3	$\mathbf{25.9 \pm 1.2}$
$1.0\,\mathrm{s}$	40.7 ± 2.0	$\mathbf{40.3 \pm 1.9}$
$1.5\,\mathrm{s}$	52.6 ± 2.7	$\mathbf{52.2 \pm 2.7}$

Attention Improves Long-Term Predictions

- Obstacles disappear with extended time horizon predictions.
  - Use attention to **pay attention** to the obstacles in feature, temporal, and spatial domains.



B. Lange, **M. Itkina**, and M. J. Kochenderfer, "Attention augmented ConvLSTM for environment prediction". In IROS 2021.

Takeaways: Spatiotemporal Environment Prediction

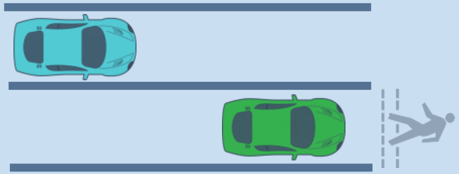
1. Occupancy prediction in cluttered urban settings can be framed as video frame prediction to enable **end-to-end spatiotemporal environment prediction**.

2. ConvLSTM architectures learn a capable internal dynamic representation of the environment from OGM data **without the need for additional dynamic input**.

3. Moving obstacle predictions at **long time horizons** can be improved with the use of **attention**.

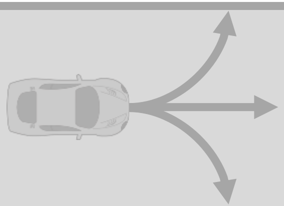
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]

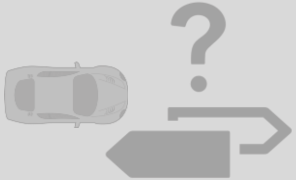


Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]



Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

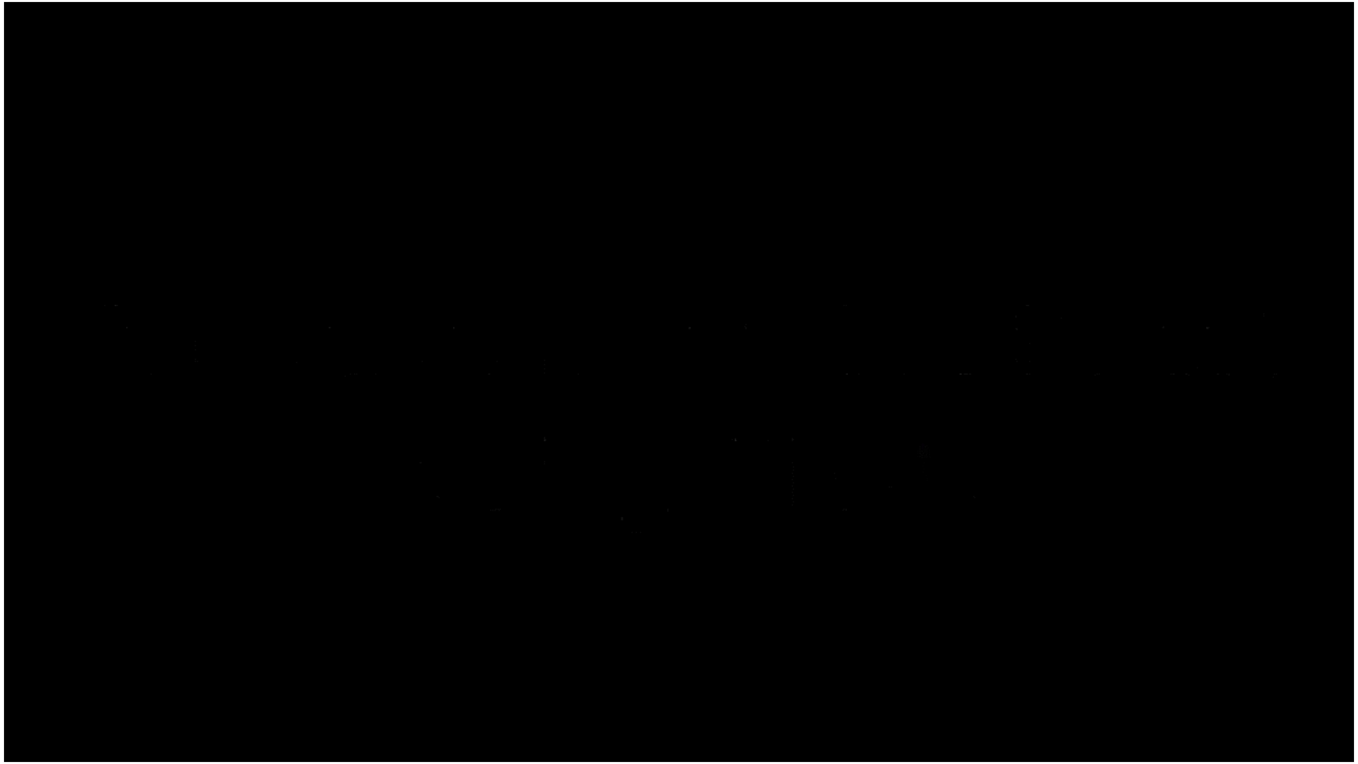
Occlusion Inference Using People as Sensors

Research question:

How can we infer occupancy in ***occluded regions*** *from the* ***observed behaviors*** *of other drivers?*

M. Itkina, Y.-J. Mun, K. Driggs-Campbell, and M. J. Kochenderfer, "Multi-Agent Variational Occlusion Inference Using People as Sensors". In ICRA, 2022.

Motivation: Occluded Pedestrians



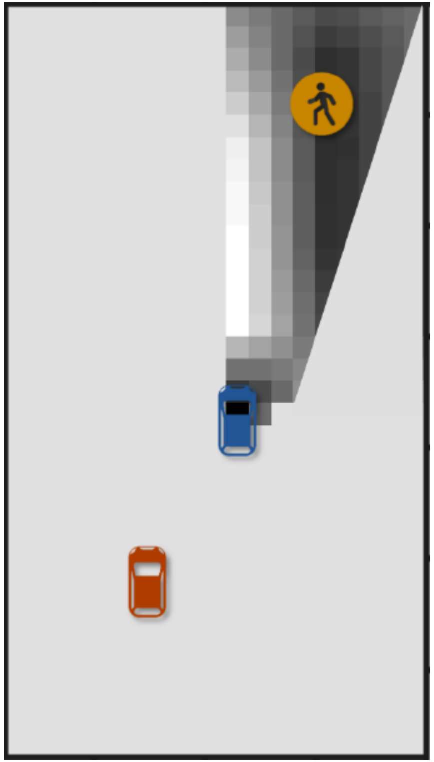
[Credit: https://www.youtube.com/watch?v=2iYTNi2MjlQ]

Occlusion Inference: Robots vs Humans

- Spatial occlusions may cause **overly cautious or dangerous** robot behavior.
- Human drivers make inferences from their observed surroundings to safely progress in driving maneuvers **despite spatial uncertainty**.
- AVs should have similar **occlusion inference** capabilities.

People as Sensors (PaS)

- Use observed driver behaviors as additional 'sensor' measurements in an OGM.
- Limitations:
  - Single 'sensor' vehicle
  - Limited to a two-vehicle crosswalk
  - Does not account for the multimodality of the occluded occupancy
  - Real-time capability not investigated



O. Afolabi et al. In IROS, 2018.

Occlusion Inference Contributions

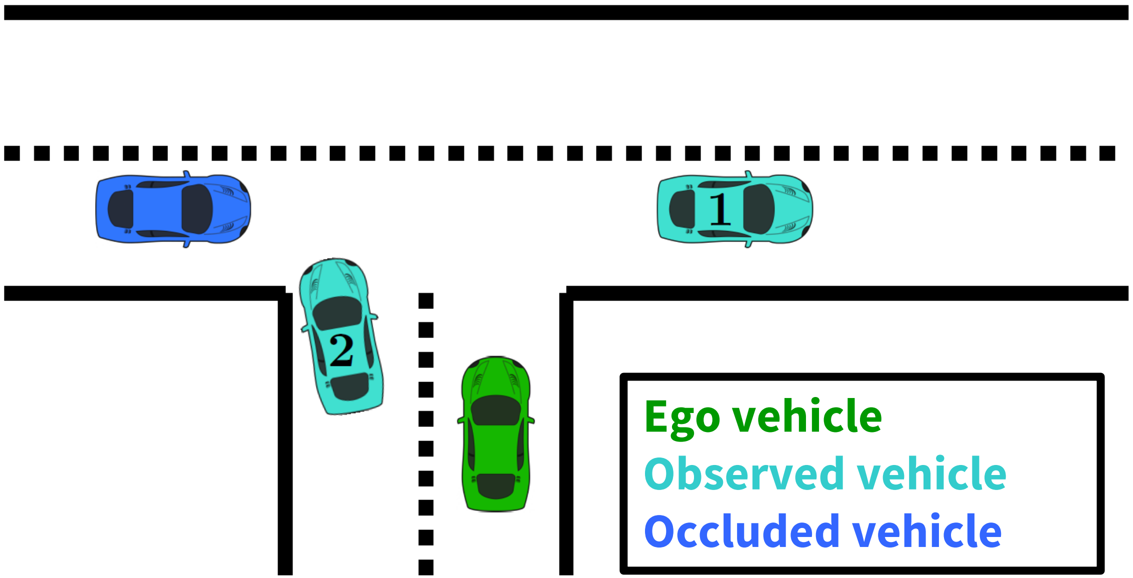
1. Propose a **multi-agent** People as Sensors occlusion inference paradigm using evidential sensor fusion.

2. Learn a driver sensor model end-to-end using a conditional variational autoencoder (CVAE) to capture **multimodality** in the distribution.

3. Demonstrate **real-time capable** performance on a cluttered, real-world intersection in the INTERACTION dataset.

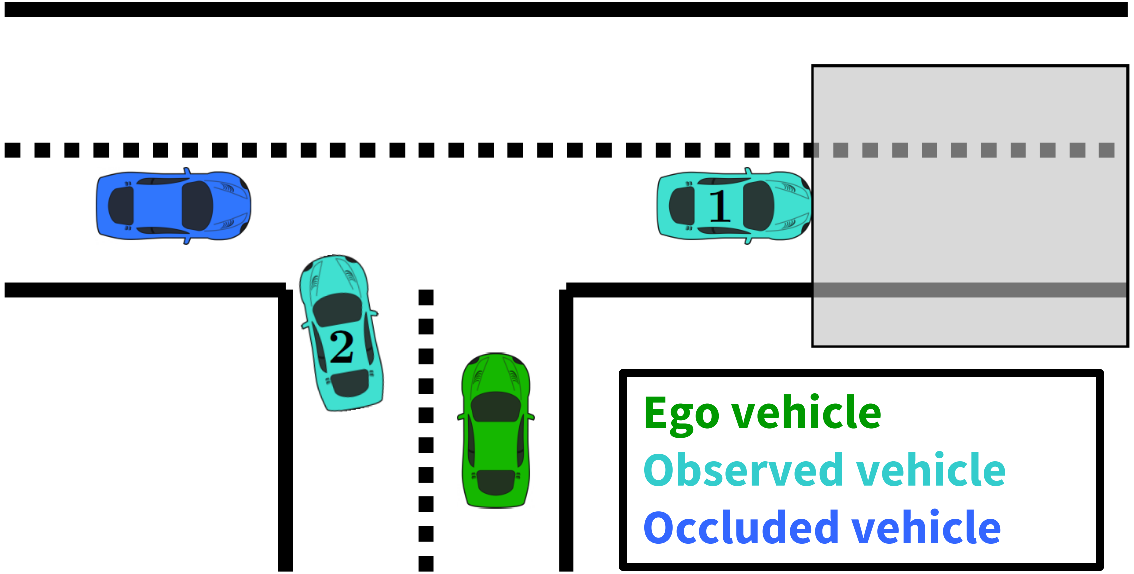
Occlusion Inference Approach Overview

We consider a general scene around an ego vehicle with observed and occluded road agents.



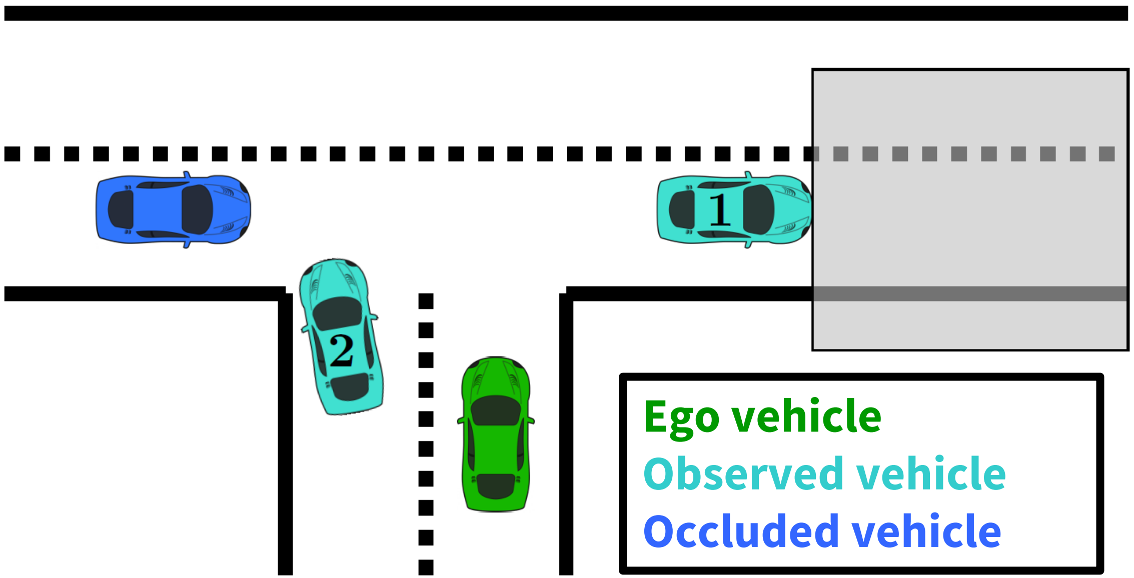
Occlusion Inference Approach Overview

We need a mapping from **driver trajectory** to the **environment** ahead.



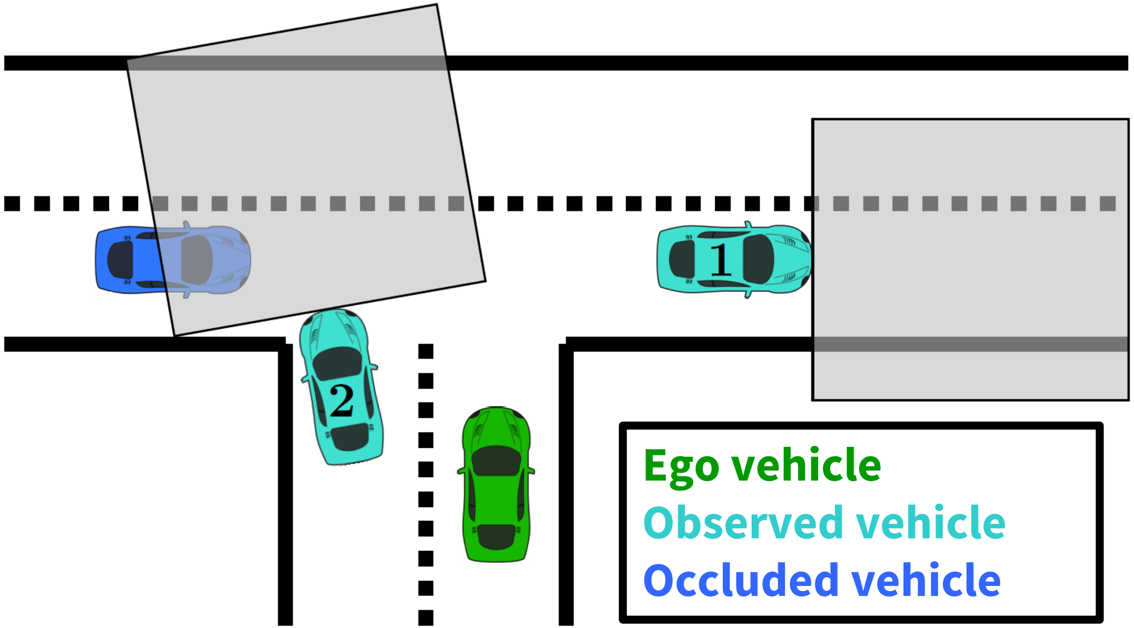
Occlusion Inference Approach Overview

The occupancy estimate is incorporated into an **ego vehicle's map**.



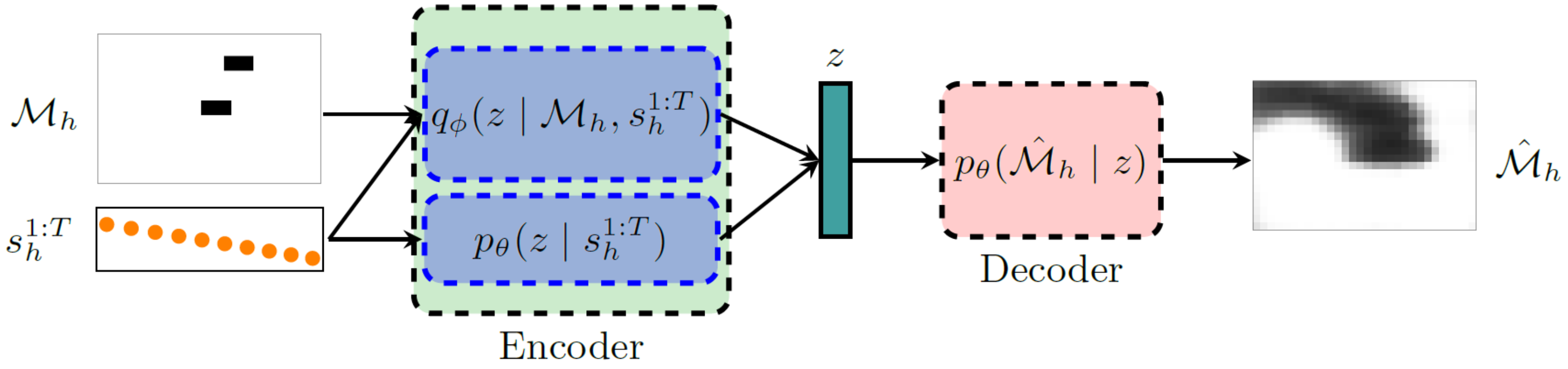
Occlusion Inference Approach Overview

This procedure can be repeated for **multiple** drivers.



Driver Sensor Model: Conditional Variational Autoencoder (CVAE)

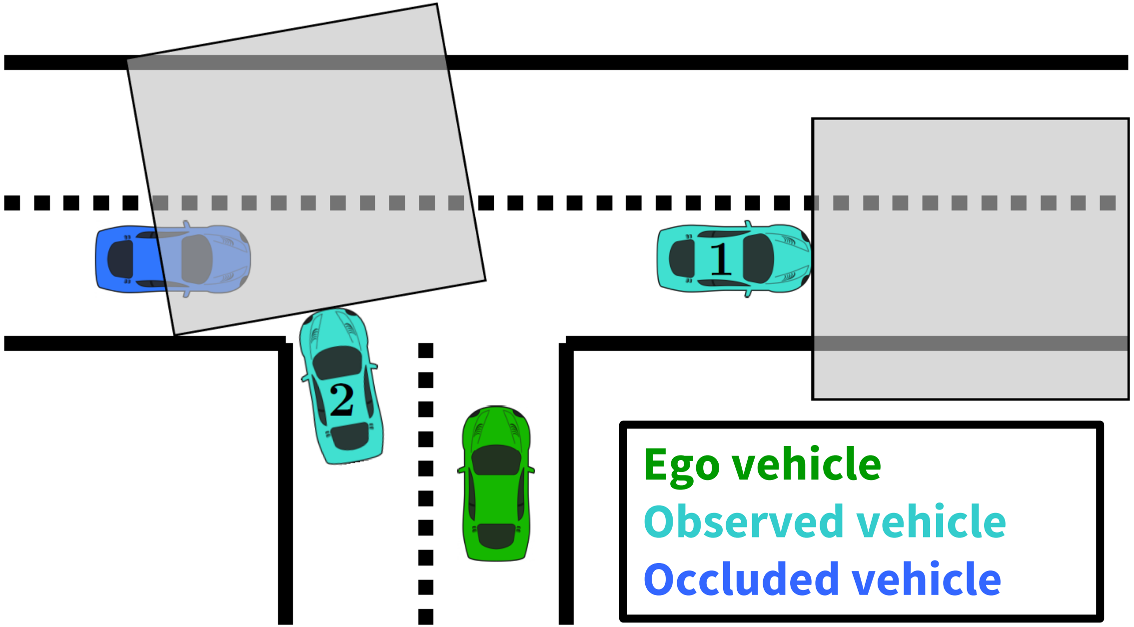
Learn a multimodal mapping from driver trajectory to the OGM representation of the environment ahead of the driver.



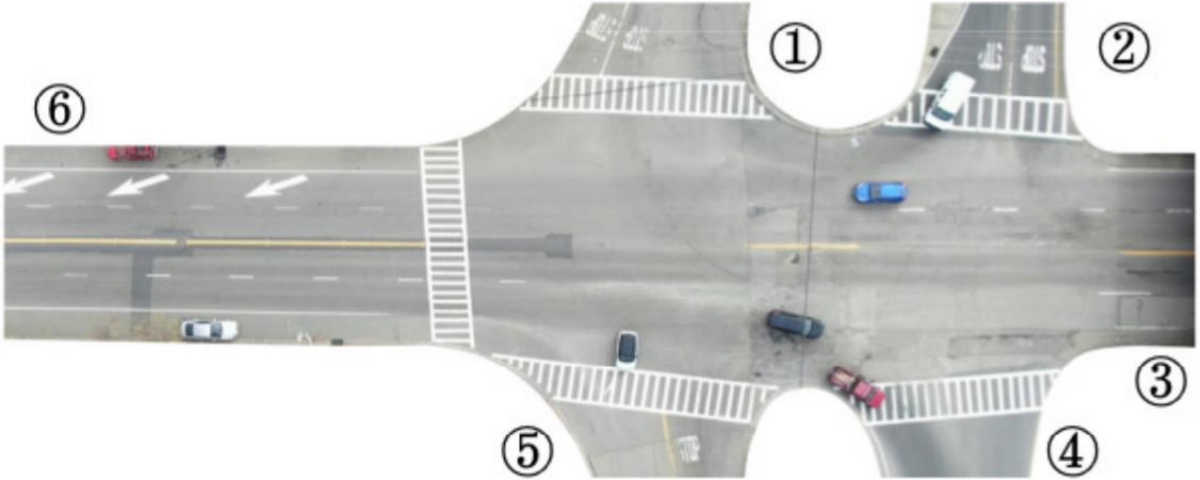
- $s_h^{1:T}$: one second of driver state information $[p_x, p_y, \theta, v_x, v_y, a_x, a_y]$.
- M_h: occupancy grid ahead of the sensor driver (1 m resolution, 20 m x 30 m).
- K = 100 discrete latent classes.

Evidential Sensor Fusion for Occlusion Inference

Multi-agent sensor fusion: combine inferred OGMs from multiple human drivers visible to the ego vehicle using evidential theory.



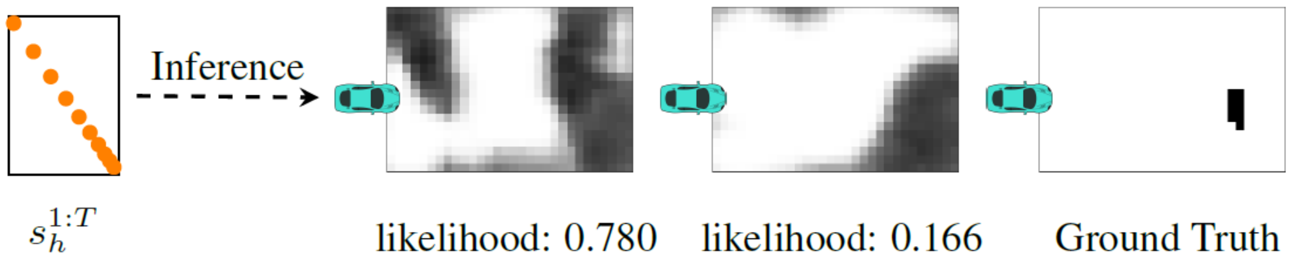
Experiments: INTERACTION Dataset



DR_USA_Intersection_GL

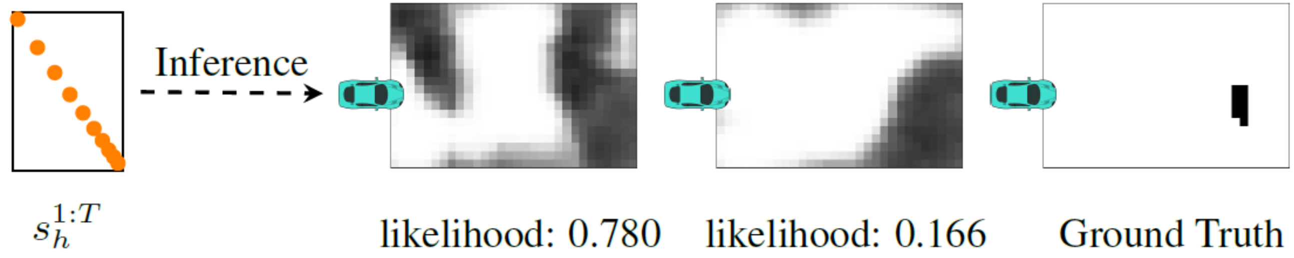
W. Zhan et al. Arxiv, 2019.

Results: Driver Sensor Model



(a) Deceleration maneuver.

Results: Driver Sensor Model



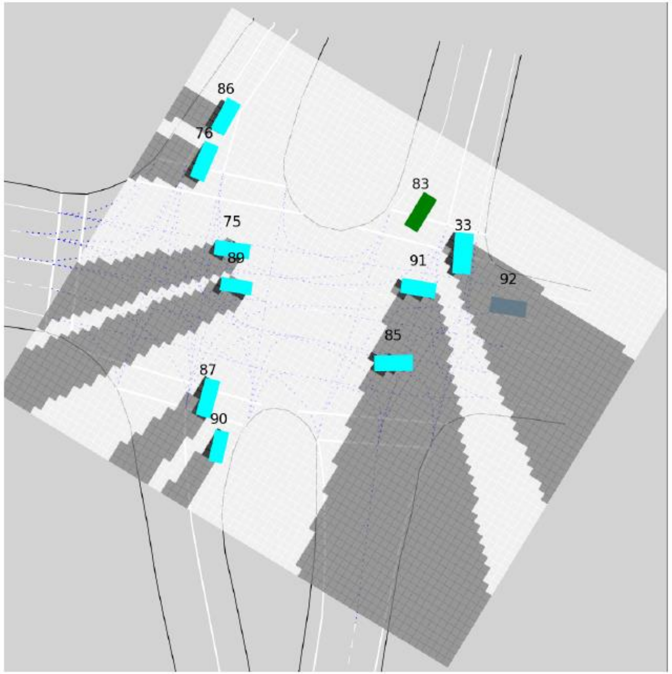
(a) Deceleration maneuver.

(b) Constant speed.

Results: Driver Sensor Model

Method	Occupied	Free	Overall	Occupied	Free	Overall
		Acc. ↑			Top 3 Acc. ↑	
K-means PaS [4]	0.512	0.463	0.465	N/A	N/A	N/A
GMM PaS	0.494	0.439	0.440	0.611	0.562	0.557
Ours	**0.619**	**0.601**	**0.601**	**0.819**	**0.773**	**0.764**
		MSE ↓			Top 3 MSE ↓	
K-means PaS [4]	**0.188**	0.206	0.206	N/A	N/A	N/A
GMM PaS	0.192	0.211	0.211	0.157	0.166	0.169
Ours	0.293	**0.192**	**0.194**	**0.145**	**0.113**	**0.124**
		IS ↓			Top 3 IS ↓	
K-means PaS [4]	0.205	0.020	0.225	N/A	N/A	N/A
GMM PaS	0.209	0.026	0.235	0.185	0.015	0.209
Ours	**0.187**	**0.017**	**0.204**	**0.115**	**0.007**	**0.130**

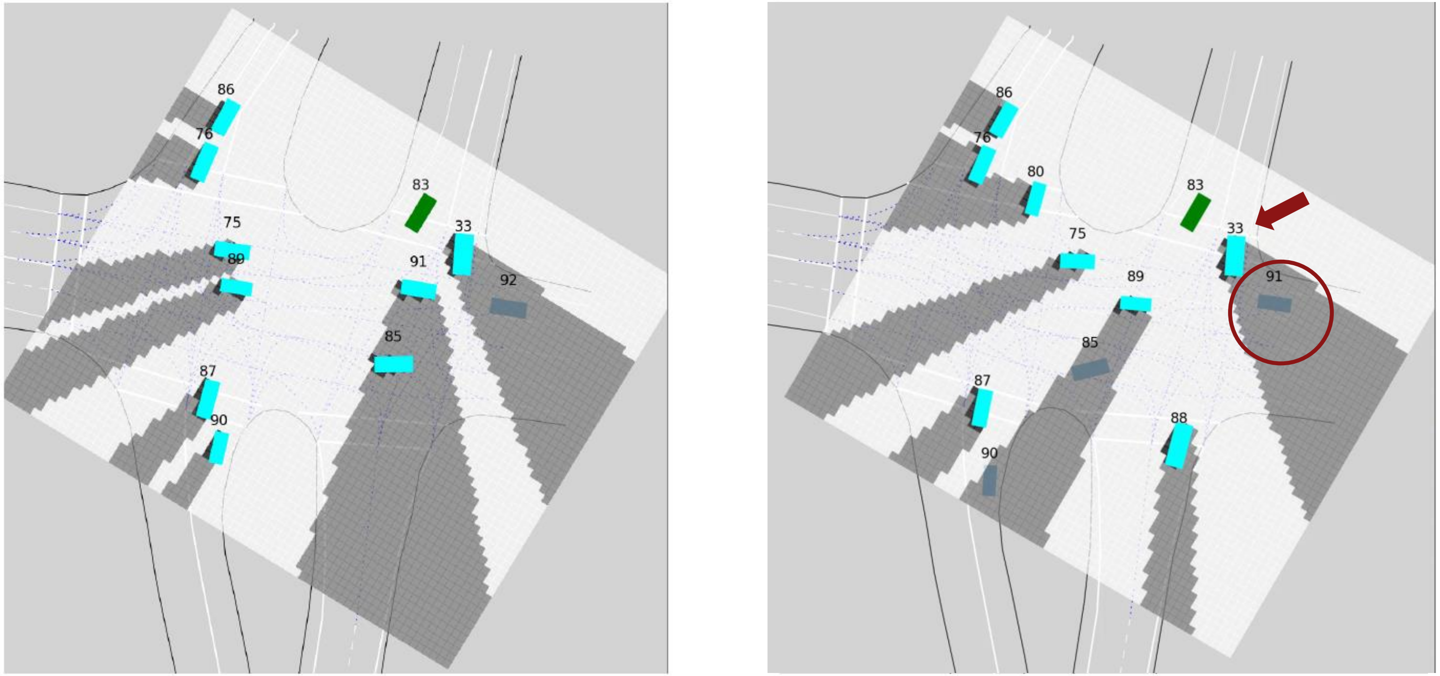
Results: Evidential Sensor Fusion



- All visible drivers for at least **1 s** are considered as driver sensors.
- Inferred OGMs are incorporated into the ego grid only in **occluded regions**.
- **Real-time capable**: 55 Hz fusion update.

Results: Evidential Sensor Fusion

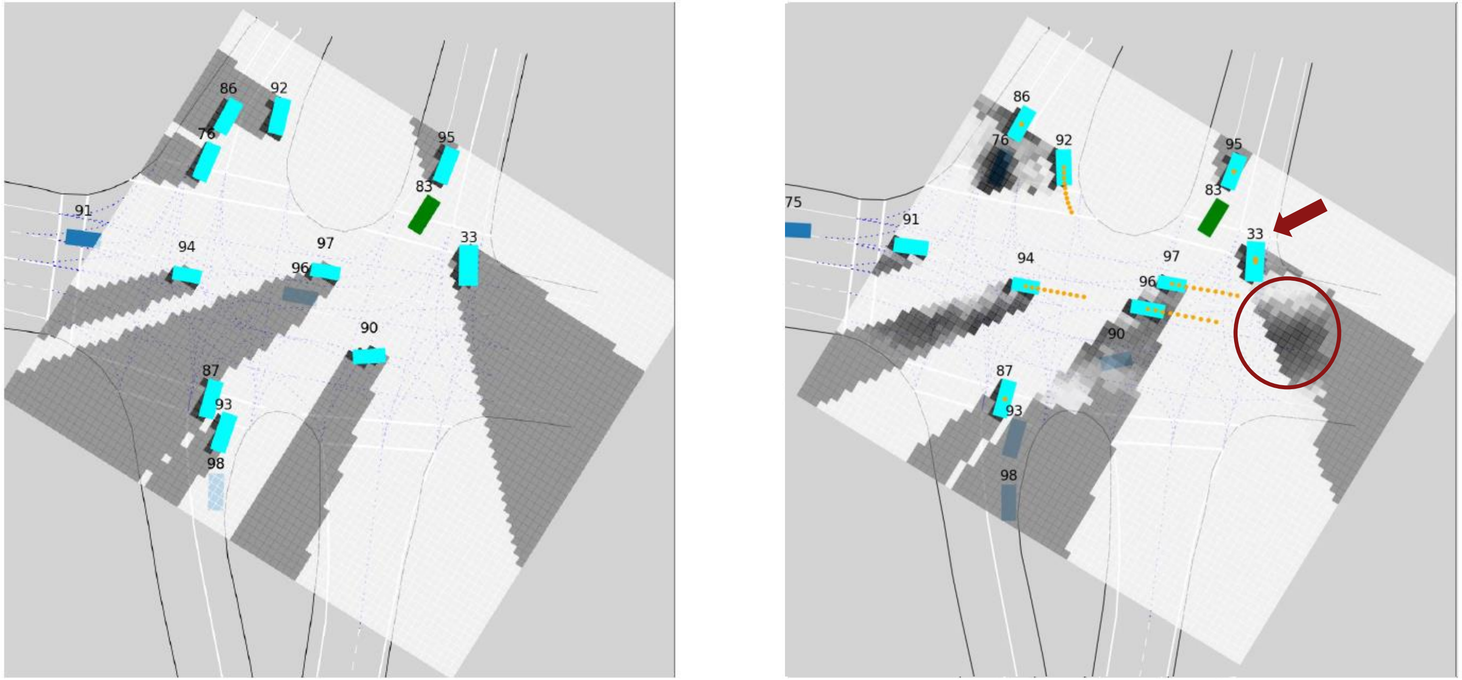
A stopped vehicle is an indicator of occupied space ahead.



(a) Stopped vehicle.

Results: Evidential Sensor Fusion

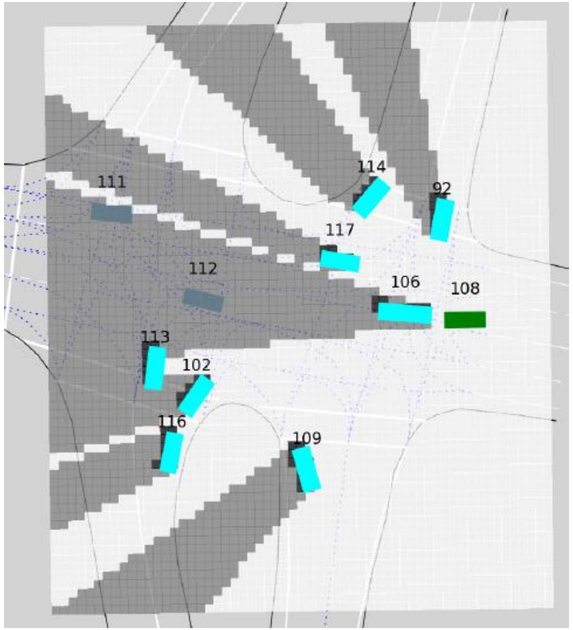
An accelerating vehicle is an indicator of free space ahead.



(b) Acceleration maneuver.

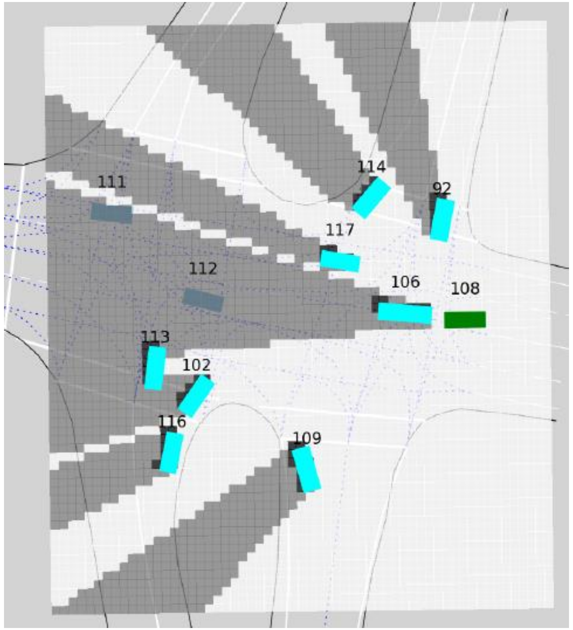
Results: Evidential Sensor Fusion

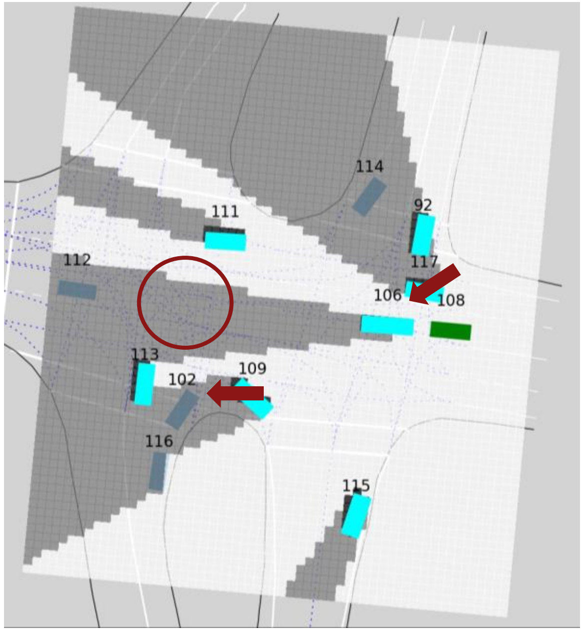
Multiple observed vehicle behaviors will increase certainty of occlusion inference.



Results: Evidential Sensor Fusion

Multiple observed vehicle behaviors will increase certainty of occlusion inference.





(c) Multiple stopped vehicles.

Takeaways: Occlusion Inference using People as Sensors

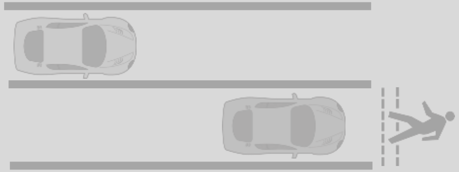
1. We can use sensor fusion techniques to scale People as Sensors to **multi-agent** settings.

2. Capturing **multimodality** is important when learning a mapping from observed behaviors to the environment.

3. Our occlusion inference approach scales to cluttered, real-world settings and achieves **real-time capable** performance.

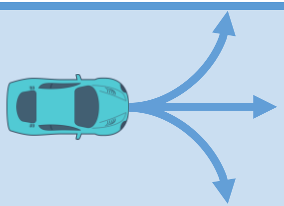
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]

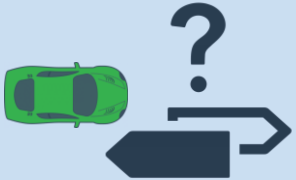


Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]



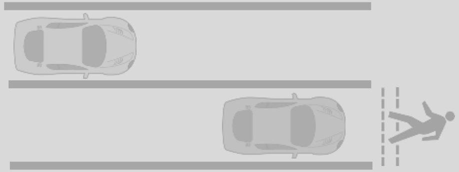
Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

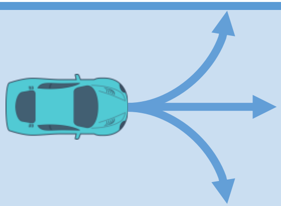
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]



Occluded occupancy inference using people as sensors [ICRA 2022*]

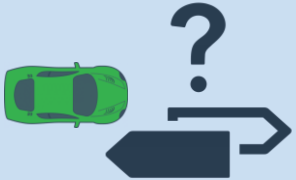
Spatiotemporal perception in human environments



Aleatoric Uncertainty

Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]

Epistemic Uncertainty



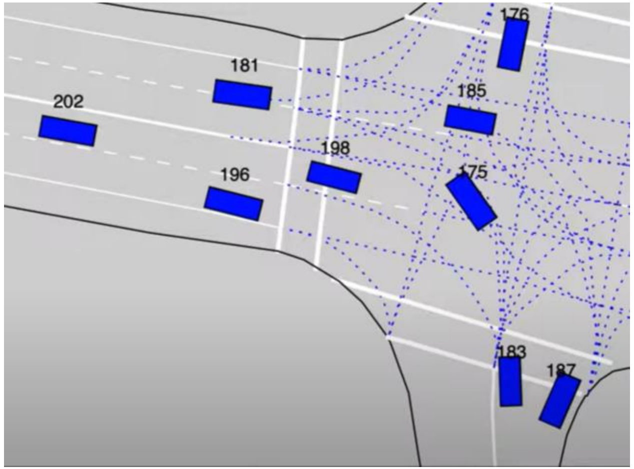
Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

Environment Representations in Robotics

Object-centric (continuous):
e.g., Landmark approaches



Data from the INTERACTION Dataset.

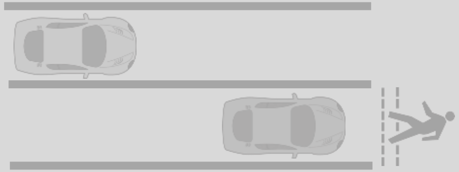
Map-centric (discrete):
e.g., Occupancy grid maps (OGMs)

Data from the KITTI Dataset.

W. Zhan et al. Arxiv, 2019.
A. Geiger et al. In CVPR, 2012.

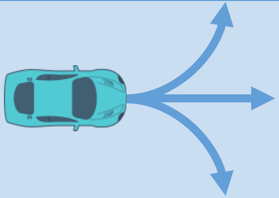
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]



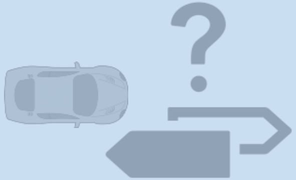
Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Aleatoric Uncertainty

Epistemic Uncertainty



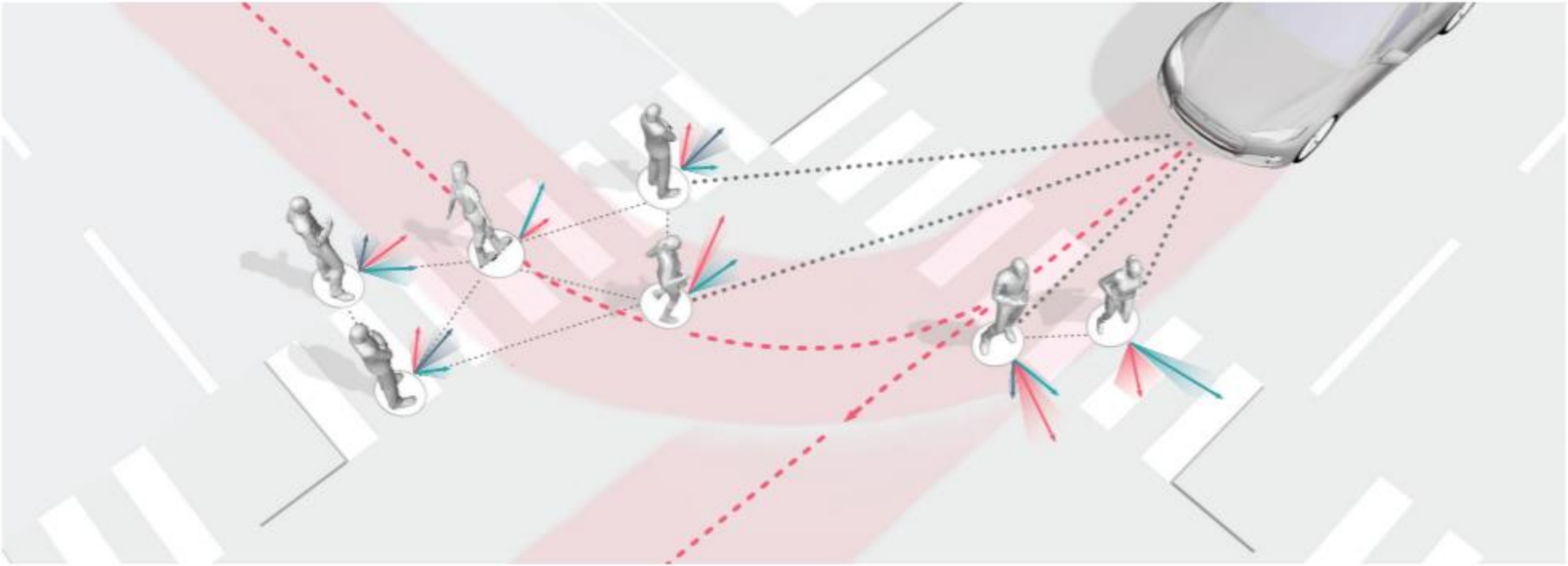
Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]

Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

CVAEs with Discrete Latent Spaces

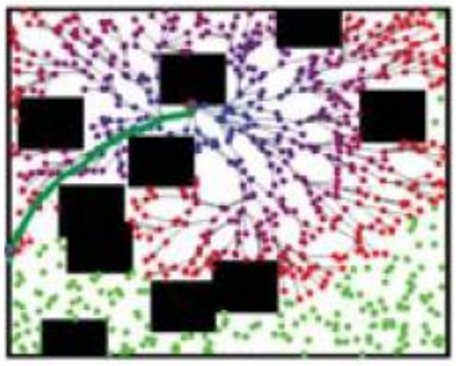


Trajectory prediction

T. Salzmann et al. In ECCV, 2020.

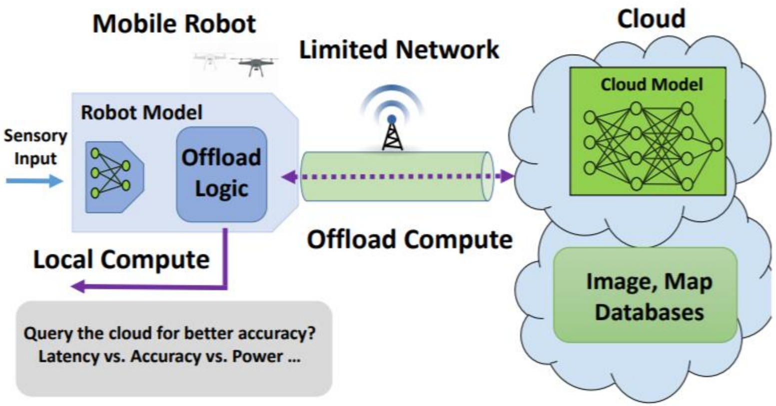
CVAEs with Discrete Latent Spaces

Problem: Discrete latent spaces need to be **sufficiently large** to capture the complexities of real-world data, rendering downstream tasks **computationally challenging**.



Robot motion planning

B. Ichter et al. In ICRA, 2018.



Multi-robot information sharing

S. Chinchali et al. In RSS, 2019.

Stanford University

Sparse, Multimodal Latent Sample Spaces for Trajectory Prediction

Research question:

Can we extract an ***efficient latent sample space*** *without retraining the network or compromising network accuracy?*

M. Itkina, B. Ivanovic, R. Senanayake, M. J. Kochenderfer, and M. Pavone. "Evidential sparsification of multimodal latent spaces in conditional variational autoencoders". In NeurIPS, 2020.

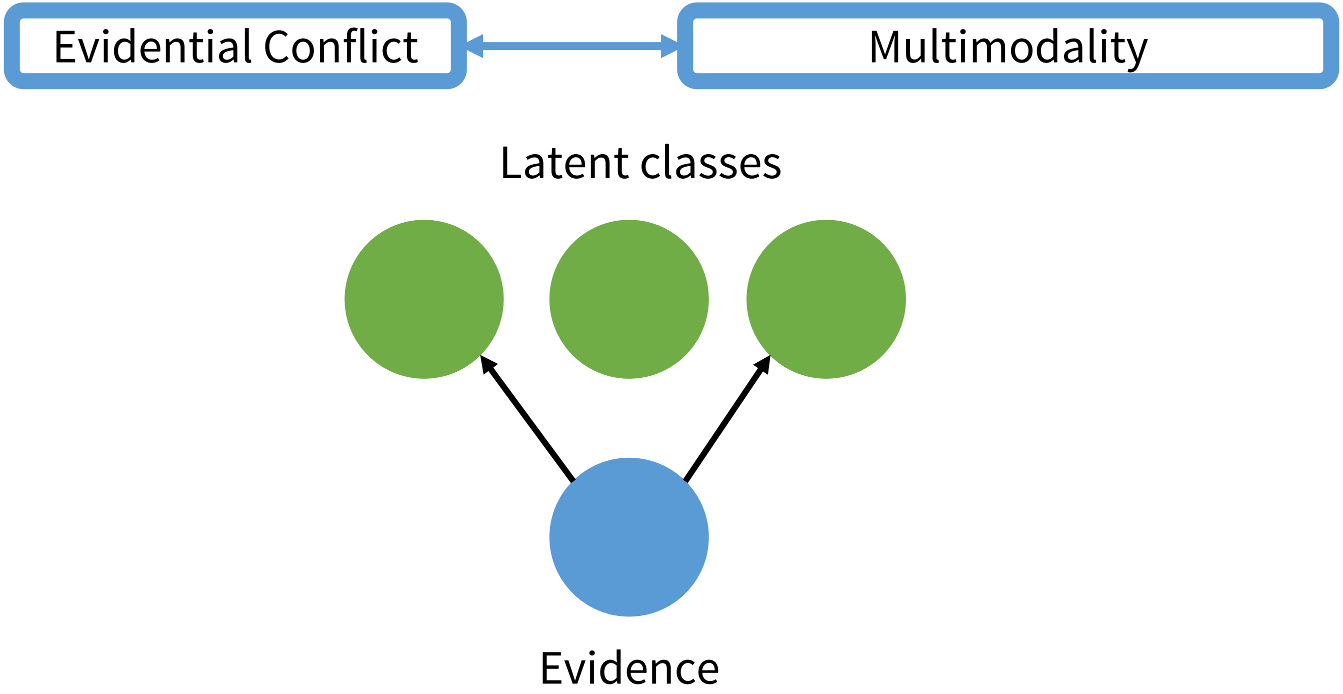
Evidential Sparsification Contributions

1. Introduce a novel method for **post hoc evidential sparsification** of the discrete latent space of a CVAE network.

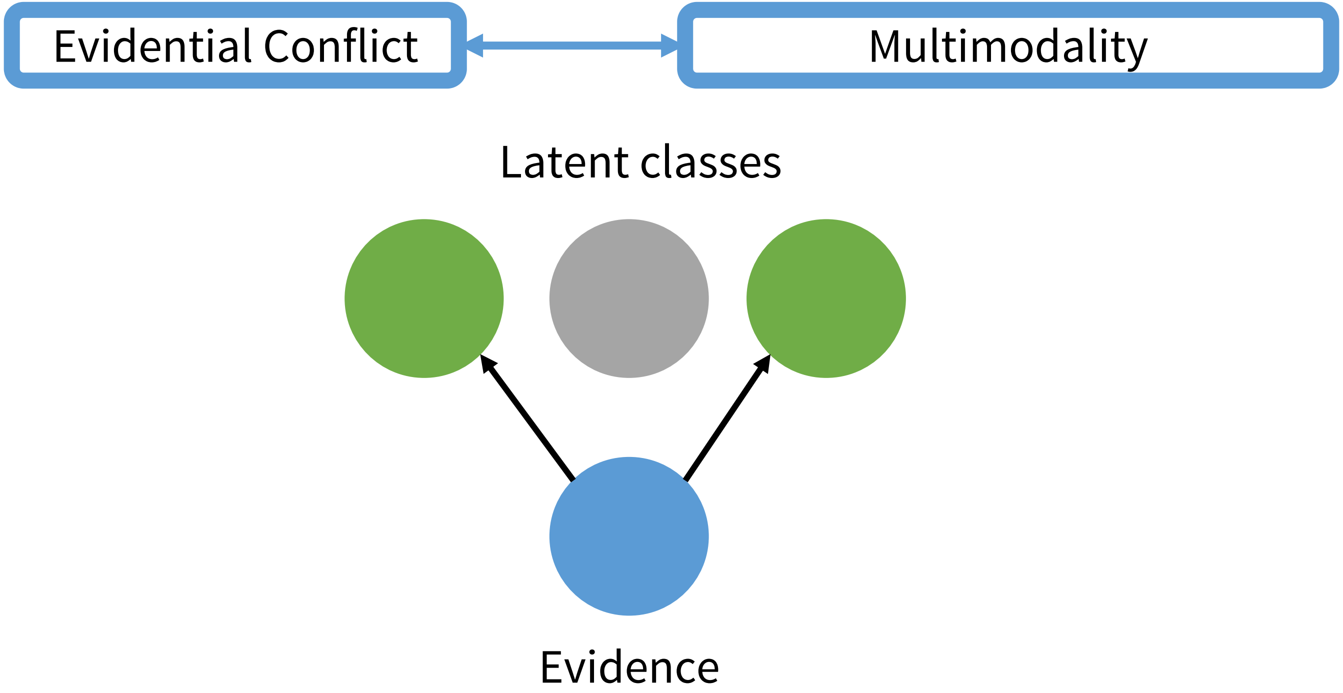
2. Automatically balance latent distribution **sparsity** and **multimodality** without hyperparameter tuning.

3. Achieve a **significant reduction** in the discrete latent space of a CVAE for trajectory prediction **without loss of network performance**.

Key Insight: Evidential Conflict as a Proxy for Multimodality

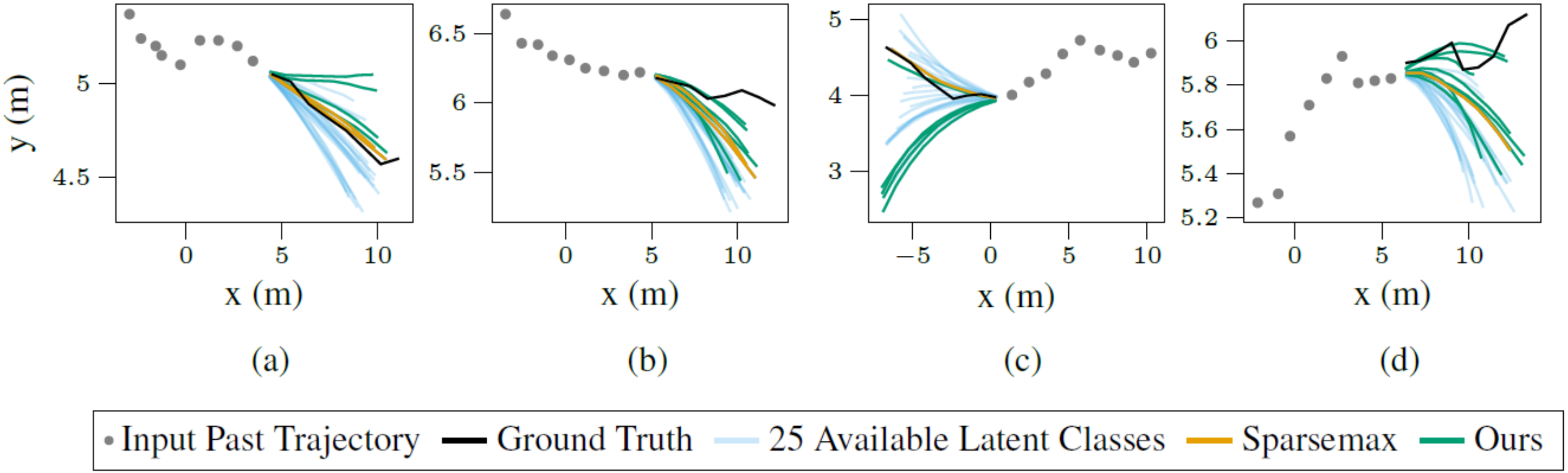


Key Insight: Evidential Conflict as a Proxy for Multimodality



Sparse Latent Sample Spaces for Trajectory Prediction

- Trajectron++ with 25 latent classes, trained on ETH Pedestrian Dataset.
- Our distribution maintains performance of softmax distribution, while achieving more than a **50% reduction** in the latent sample space size.



T. Salzmann et al. In ECCV, 2020.
S. Pellegrini et al. In ICCV, 2009.
A. Martins and R. Astudillo. In ICML, 2016.

Ev-Softmax: Training-Time Extension

- We derived a simple closed form of our post hoc sparsification method:

$$\textsc{EvSoftmax}(\mathbf{v})_k \propto \mathbb{1}\{\underbrace{v_k}_{\text{Logit}} \geq \overline{v}\}e^{v_k}$$

where $\overline{v} = \frac{1}{K}\sum_{i=1}^{K} v_i$.

- During training time, a modification can be used:

$$\textsc{EvSoftmax}_{\text{train},\epsilon}(\mathbf{v})_i \propto (\mathbb{1}\{v_i \geq \overline{v}\} + \epsilon)e^{v_k}$$

P. Chen, **M. Itkina**, R. Senanayake, and M. J. Kochenderfer. “Evidential Softmax for Sparse Multimodal Distributions in Deep Generative Models”. In NeurIPS, 2021.

Takeaways: Evidential Sparsification Contributions

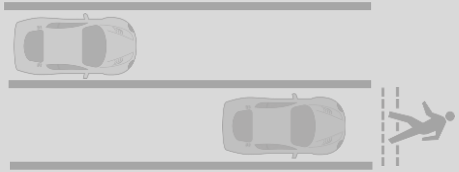
1. Showed the discrete latent space of a CVAE can be **sparsified post hoc** using evidential theory **without sacrificing performance**.

2. The proposed approach automatically balances latent distribution **sparsity** and **multimodality** without hyperparameter tuning.

3. Our method's simplified form can also be used **directly during training**.

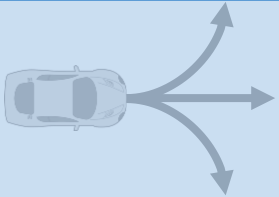
Contributions

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]



Occluded occupancy inference using people as sensors [ICRA 2022*]

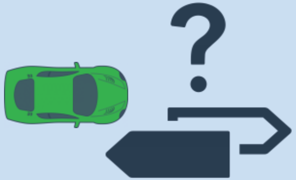
Spatiotemporal perception in human environments



Aleatoric Uncertainty

Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]

Epistemic Uncertainty



Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.

Why is epistemic uncertainty important?

- Neural networks are unreliable on **out-of-distribution (OOD)** inputs.
- For AVs, it is **safety critical** that the prediction module can provide a level of confidence in its output to a downstream planner.
- Existing methods:
  - Applied to **simple classification tasks** or **small regression problems**.
  - Require explicit **OOD data during training**.
  - Uncertainty estimate is **not interpretable**.

A. Malinin and M. Gales. In NeurIPS, 2018.
M. Sensoy et al. In AAAI, 2020.

Interpretable Self-Aware Neural Networks for Robust Trajectory Prediction

Research question:

*How do we meaningfully and tractably capture **epistemic uncertainty** for trajectory prediction?*

M. Itkina and M. J. Kochenderfer. "ISAP: Interpretable Self-Aware Neural Networks for Robust Trajectory Prediction". In Progress.

Epistemic Uncertainty Estimation

- Subjective logic links evidential theory with **Dirichlet distributions**.
- **Evidential deep learning** computes parameters for a second-order distribution (e.g., Dirichlet) over discrete encodings to estimate epistemic uncertainty.
- Explicit training with OOD examples can be avoided with the use of **deep generative models**.

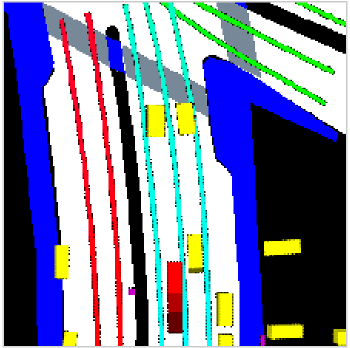
Jøsang, Audun. “Subjective logic”. Springer, 2016.
A. Malinin and M. Gales. In NeurIPS, 2018.
M. Sensoy et al. In AAAI, 2020.
B. Charpentier et al. In NeurIPS, 2020.

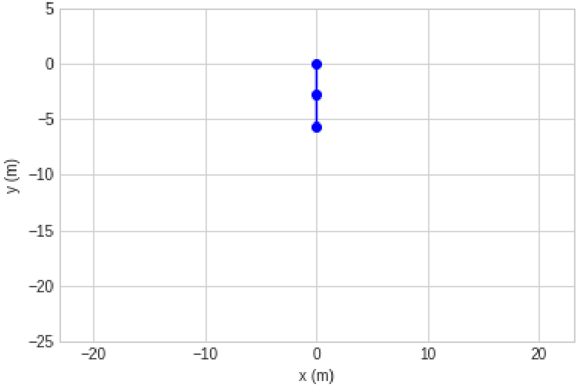
Epistemic Uncertainty Estimation Contributions

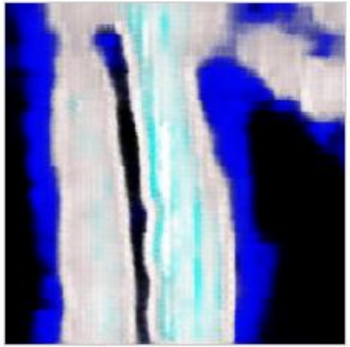
- Investigate a novel application of **evidential deep learning** methods to epistemic uncertainty quantification in trajectory prediction tasks.

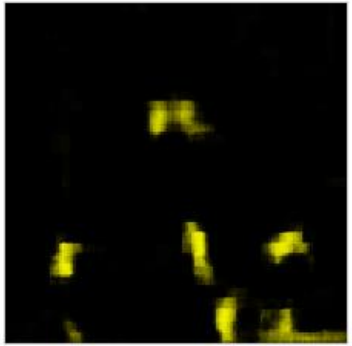
- Define a notion of **interpretability** by distributing the uncertainty over **low-dimensional latent variables** corresponding to past agent behavior, environment map, and social context.

Interpretable Self-Aware Trajectory Prediction









Input Representation

$Dir(\alpha_{agent})$ $Dir(\alpha_{map})$ $Dir(\alpha_{social\ context})$

$Dir(\alpha)$

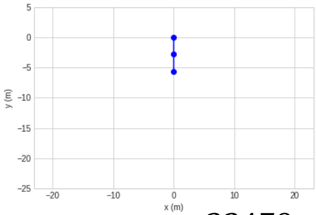
Stanford University

Preliminary Results

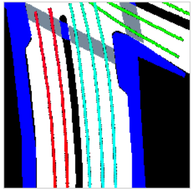
- We test our approach on a modified nuScenes dataset for OOD detection.
  - Slower input trajectories: in distribution
  - Faster input trajectories: OOD distribution

- $\uparrow \alpha_0$ indicates in distribution, $\downarrow \alpha_0$ indicates OOD.

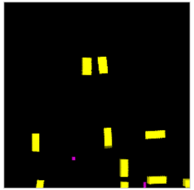
in distribution



$\alpha_{0,agent} = 22479$

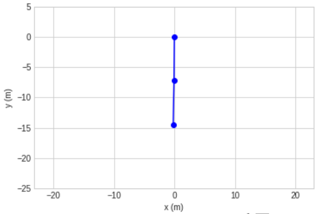


$\alpha_{0,map} = 157$

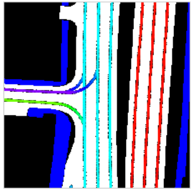


$\alpha_{0,social\ context} = 4270$

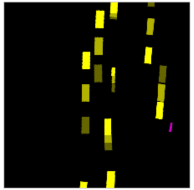
OOD



$\alpha_{0,agent} = 67$



$\alpha_{0,map} = 94$

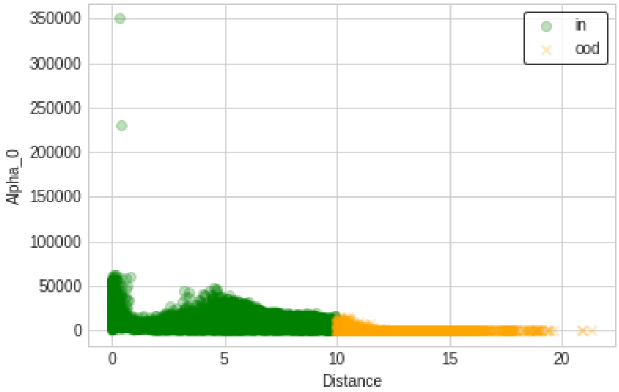


$\alpha_{0,social\ context} = 4004$

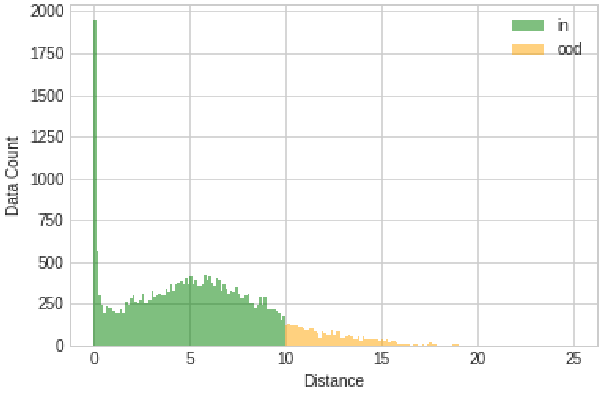
Preliminary Results

- We test our approach on a modified nuScenes dataset for OOD detection.
  - Slower input trajectories: in distribution
  - Faster input trajectories: OOD distribution
- Learned Dirichlet parameters match data distribution trends.

Learned $\alpha_{0,agent}$



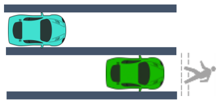
Ground Truth Data Distribution



Contributions

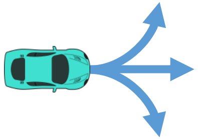
Predict the spatiotemporal environment end-to-end from a map-centric (i.e., occupancy grid) representation of the world using recurrent representation learning.

[ITSC 2019*, IROS 2021, ICRA 2021]



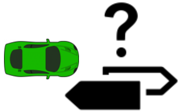
Infer occluded occupancy in multi-agent settings by modeling observed human behaviors as sensor measurements for the environment occupancy.

[ICRA 2022*]



Sparsify the latent sample space in deep generative models for trajectory prediction, while maintaining multimodality, using evidential theory.

[NeurIPS 2020*, NeurIPS 2021]



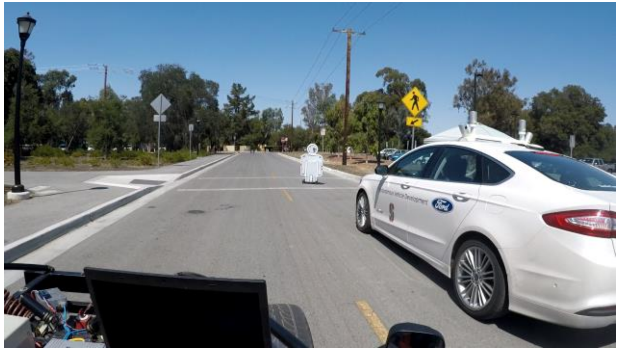
Model epistemic uncertainty for trajectory prediction using deep generative models and second-order distributions over low-dimensional, interpretable latent variables.

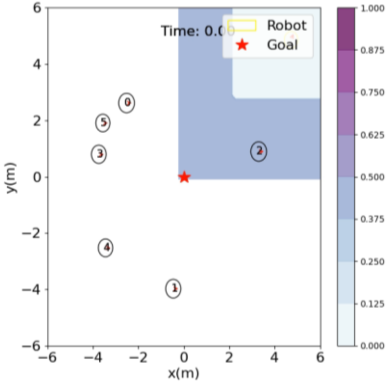
[In Progress*, ITSC 2021]

* M. Itkina is the first author. Otherwise second author.

Future Directions

- Adaptive stress testing for map- versus object-centric perception systems
- Integration of uncertainty-aware perception with planning/control
  - Model predictive control with environment prediction in-the-loop on hardware
  - Occlusion-aware reinforcement learning for crowd navigation
  - Handling epistemic uncertainty output from a perception system





Mun et al., In Progress, 2022.

Publications

M. Itkina, K. Driggs-Campbell, and M. J. Kochenderfer, “Dynamic environment prediction in urban scenes using recurrent representation learning”. In Intelligent Transportation Systems Conference (ITSC), 2019.

B. Lange, **M. Itkina**, and M. J. Kochenderfer, “Attention augmented ConvLSTM for environment prediction”. In Intelligent Robots and Systems (IROS), 2021.

M. Toyungyernsub, **M. Itkina**, R. Senanayake, and M. J. Kochenderfer, “Double-prong ConvLSTM for spatiotemporal occupancy prediction in dynamic environments”. In International Conference on Robotics and Automation (ICRA), 2021.

M. Itkina, B. Ivanovic, R. Senanayake, M. J. Kochenderfer, and M. Pavone. “Evidential sparsification of multimodal latent spaces in conditional variational autoencoders”. In Advances in Neural Information Processing Systems (NeurIPS), 2020.

P. Chen, **M. Itkina**, R. Senanayake, and M. J. Kochenderfer. “Evidential Softmax for Sparse Multimodal Distributions in Deep Generative Models”. In Advances in Neural Information Processing Systems (NeurIPS), 2021.

M. Itkina, Y.-J. Mun, K. Driggs-Campbell, and M. J. Kochenderfer, “Multi-Agent Variational Occlusion Inference Using People as Sensors”. In International Conference on Robotics and Automation (ICRA), 2022.

J. Nitsch, **M. Itkina**, R. Senanayake, J. Nieto, M. Schmidt, R. Siegwart, M. J. Kochenderfer, and C. Cadena, “Out-of-Distribution Detection for Automotive Perception”. In Intelligent Transportation Systems Conference (ITSC), 2021.

References

K. Korosec. "Waymo suspends robotaxi service except for its truly driverless vehicles". TechCrunch, 2020.

C. Gartenberg. "Boston Dynamics robots take over the dance floor in latest video". The Verge, 2020.

E. Demaitre, "Toyota Research Institute shows service robot prototypes in virtual open house". The Robot Report, 2020.

A. Pandey and D. R. Parhi. "Multiple mobile robots navigation and obstacle avoidance using minimum rule based ANFIS network controller in the cluttered environment". International Journal of Advanced Robotics and Automation, 2016.

J. Long, E. Shelhamer, and T. Darrel, "Fully convolutional networks for semantic segmentation". In Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

W. Lotter, G. Kreiman, and D.D. Cox, "Deep predictive coding networks for video prediction and unsupervised learning". In International Conference on Learning Representations (ICLR), 2017.

G. Postica, A. Romanoni, and M. Matteucci. "Robust moving objects detection in lidar data exploiting visual cues". In Intelligent Robots and Systems (IROS), 2016.

D. Nuss, M. Thom, A. Danzer, and K. Dietmayer. "Fusion of laser and monocular camera data in object grid maps for vehicle environment perception". In International Conference on Information Fusion (FUSION), pages 1-8, 2014.

D. Nuss, S. Reuter, M. Thom, T. Yuan, G. Krehl, M. Maile, A. Gern, and K. Dietmayer. "A random finite set approach for dynamic occupancy real-time application". The International Journal of Robotics Research, 37.8, pages 841-866, 2016.

A. P. Dempster. "A generalization of Bayesian inference". Classic works of the Dempster-Shafer Theory of Belief Functions, pages 73–104, 2008.

References

A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite”. In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone. “Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data”. In European Conference on Computer Vision (ECCV), 2020.

L. Kaiser, S. Bengio, A. Roy, A. Vaswani, N. Parmar, J. Uszkoreit, and N. Shazeer. “Fast decoding in sequence models using discrete latent variables”. In International Conference on Machine Learning (ICML), pages 2390–2399, 2018.

A. Vahdat, W. Macready, Z. Bian, A. Khoshaman, and E. Andriyash. “DVAE++: Discrete variational autoencoders with overlapping transformations”. In International Conference on Machine Learning (ICML), pages 5035–5044, 2018.

B. Ichter, J. Harrison, and M. Pavone. “Learning sampling distributions for robot motion planning”. In International Conference on Robotics and Automation (ICRA), pages 7087–7094, 2018.

S. Chinchali, A. Sharma, J. Harrison, A. Elhafsi, D. Kang, E. Pergament, E. Cidon, S. Katti, and M. Pavone. “Network offloading policies for cloud robotics: A learning-based approach”. In Robotics: Science and Systems (RSS), 2019.

S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. “You’ll never walk alone: Modeling social behavior for multi-target tracking”. In International Conference on Computer Vision (ICCV), pages 261–268, 2009.

A. Martins and R. Astudillo. “From softmax to sparsemax: A sparse model of attention and multi-label classification”. In International Conference on Machine Learning (ICML), pages 1614–1623, 2016.

O. Afolabi, K. Driggs–Campbell, R. Dong, M. J. Kochenderfer, and S. S. Sastry. “People as sensors: Imputing maps from human actions”. In International Conference on Intelligent Robots and Systems (IROS), pages 2342–2348, 2018.

References

K. Sohn, H. Lee, and X. Yan. “Learning structured output representation using deep conditional generative models”. In Advances in Neural Information Processing Systems (NeurIPS), pages 3485–3491, 2015.

W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, and M. Tomizuka. “INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION dataset in interactive driving scenarios with semantic maps”. Arxiv, 2019.

K. Lee, K. Lee, H. Lee, and J. Shin. “A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks”. In Advances in Neural Information Processing Systems (NeurIPS), pages 7167–7177, 2018.

Jøsang, Audun. “Subjective logic”. Springer, 2016.

A. Malinin and M. Gales. “Predictive Uncertainty Estimation via Prior Networks”. In Advances in Neural Information Processing Systems (NeurIPS), pages 7047–7058, 2018.

M. Sensoy, L. Kaplan, F. Cerutti, and M. Saleki. “Uncertainty-Aware Deep Classifiers Using Generative Models”. In Conference on Artificial Intelligence (AAAI), 2020.

C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. “Weight Uncertainty in Neural Networks”. In International Conference on Machine Learning (ICML), volume 37, pages 1613–1622, 2015.

B. Charpentier, D. Zügner, and S. Günnemann. "Posterior Network: Uncertainty Estimation without OOD Samples via Density-Based Pseudo-Counts." In Advances in Neural Information Processing Systems (NeurIPS), pages 1356-1367, 2020.

H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. "nuScenes: A multimodal dataset for autonomous driving." In Conference on Computer Vision and Pattern Recognition (CVPR), pages 11621-11631. 2020.

Sponsors

Ford

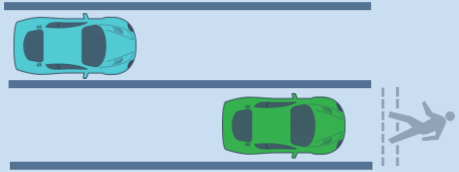
TOYOTA
RESEARCH INSTITUTE

DAIMLER

WAYMO

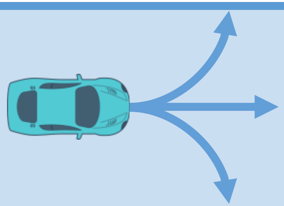
Thank you!

Environment prediction in cluttered scenes [ITSC 2019*, IROS 2021, ICRA 2021]

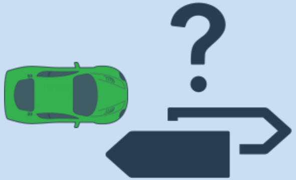


Occluded occupancy inference using people as sensors [ICRA 2022*]

Spatiotemporal perception in human environments



Sparse, multimodal distributions for trajectory prediction [NeurIPS 2020*, NeurIPS 2021]



Epistemic uncertainty estimation for trajectory prediction [In Progress*, ITSC 2021]

Uncertainty-aware machine learning models for trajectory prediction

* M. Itkina is the first author. Otherwise second author.