



**HAL**  
open science

# Integrating Specialized and Generic Agent Motion Prediction with Dynamic Occupancy Grid Maps

Rabbia Asghar, Wenqian Liu, Lukas Rummelhard, Anne Spalanzani, Christian  
Laugier

## ► To cite this version:

Rabbia Asghar, Wenqian Liu, Lukas Rummelhard, Anne Spalanzani, Christian Laugier. Integrating Specialized and Generic Agent Motion Prediction with Dynamic Occupancy Grid Maps. IROS 2024 - 37th IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 2024, Abu Dhabi, United Arab Emirates. IEEE, pp.1-1, 2024. <hal-04885662>

**HAL Id: hal-04885662**

**<https://hal.science/hal-04885662v1>**

Submitted on 14 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Integrating Specialized and Generic Agent Motion Prediction with Dynamic Occupancy Grid Maps

Rabbia Asghar, Wenqian Liu, Lukas Rummelhard, Anne Spalanzani, Christian Laugier

Université Grenoble Alpes, Inria, Grenoble, France



## Introduction

### Challenges in motion forecasting:

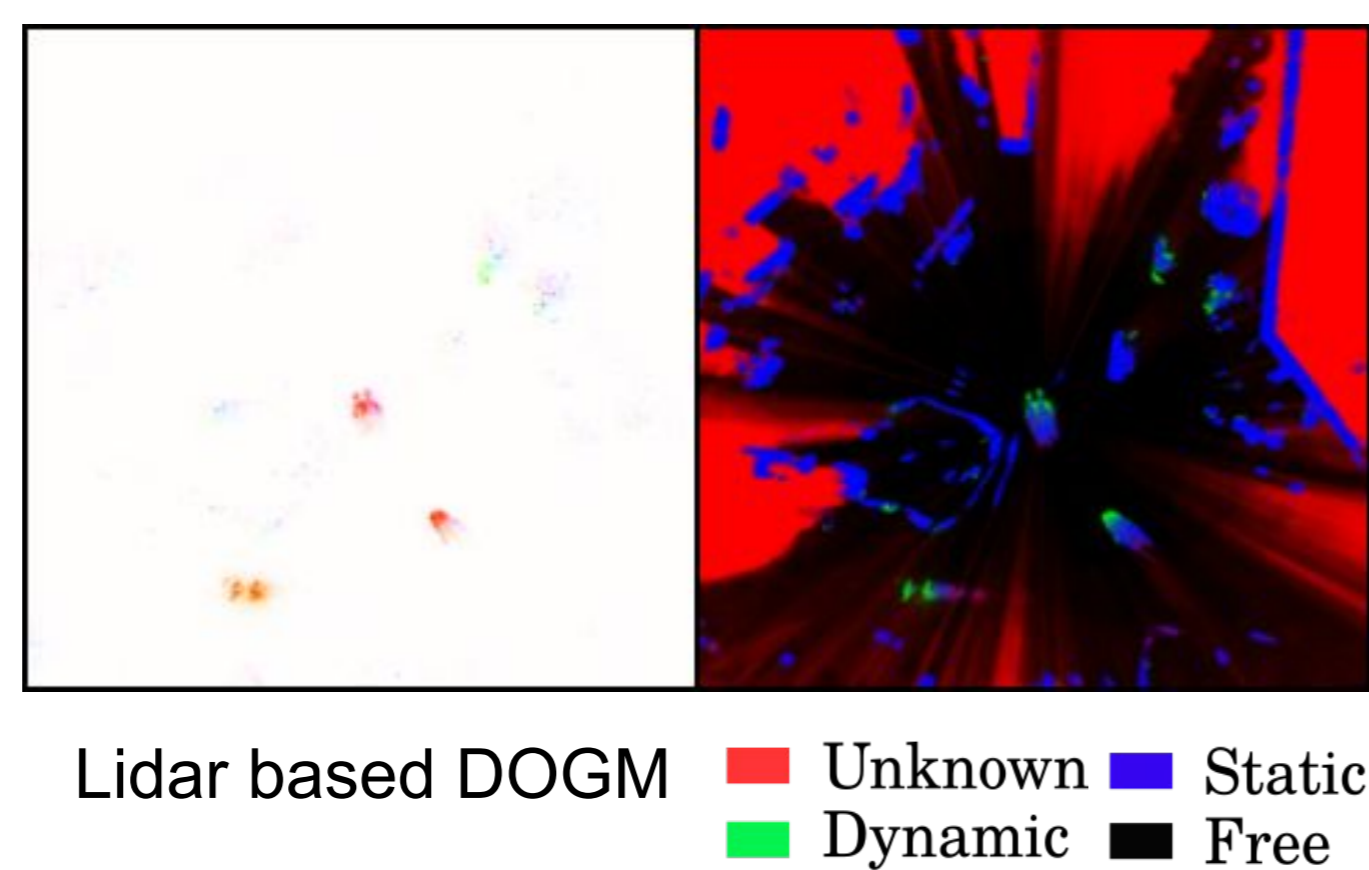
- Partial Scene Understanding: Uncertainty in sensor data, Occlusions
- Complex behavior of agents
- Multiple feasible futures

Existing prediction methods using Occupancy Grid Maps (OGMs) primarily focus on agent-agnostic scene prediction while agent-specific predictions provide specialized behavior insights with the available semantic information.

## Scene Representation

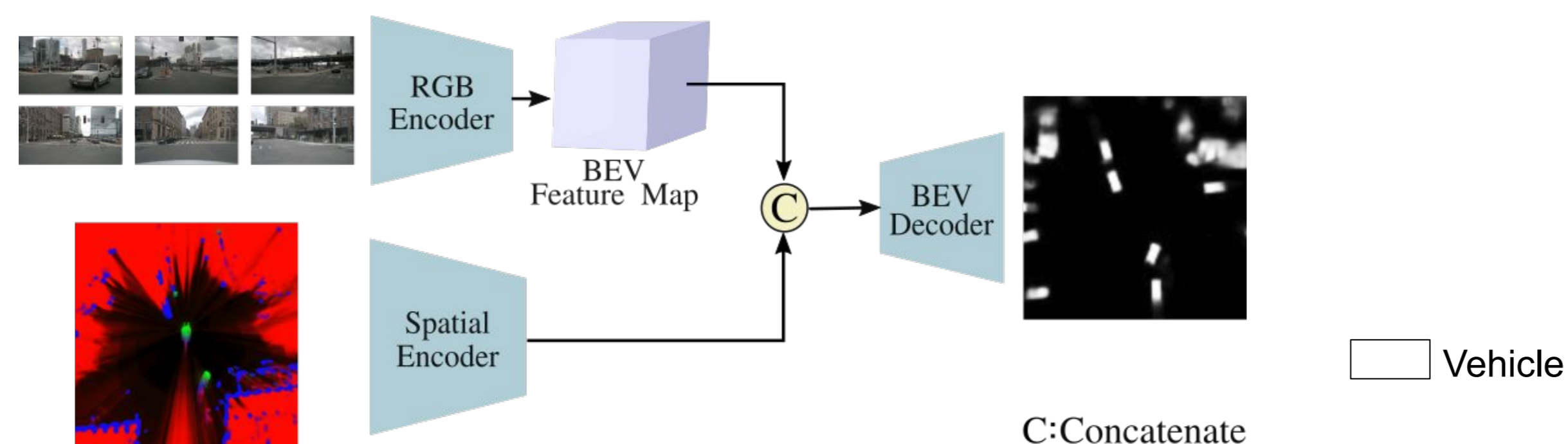
### Dynamic Occupancy Grid Maps (DOGMs)

- Bird's-eye-view grid
- Efficient sensor fusion
- Real-time computation
- No confidence thresholding
- No semantic information



### Semantic Information

- BEV features encoded from Camera Images
- Fused with Occupancy State Grid



## Contribution

- A **Multi-Task Framework** that predicts segmentation grids, occupancy grids and flow grids to represent scene evolution
- **Novel Approach** to combine the Agent Specific Behaviour Prediction and Agent Agnostic Motion Prediction

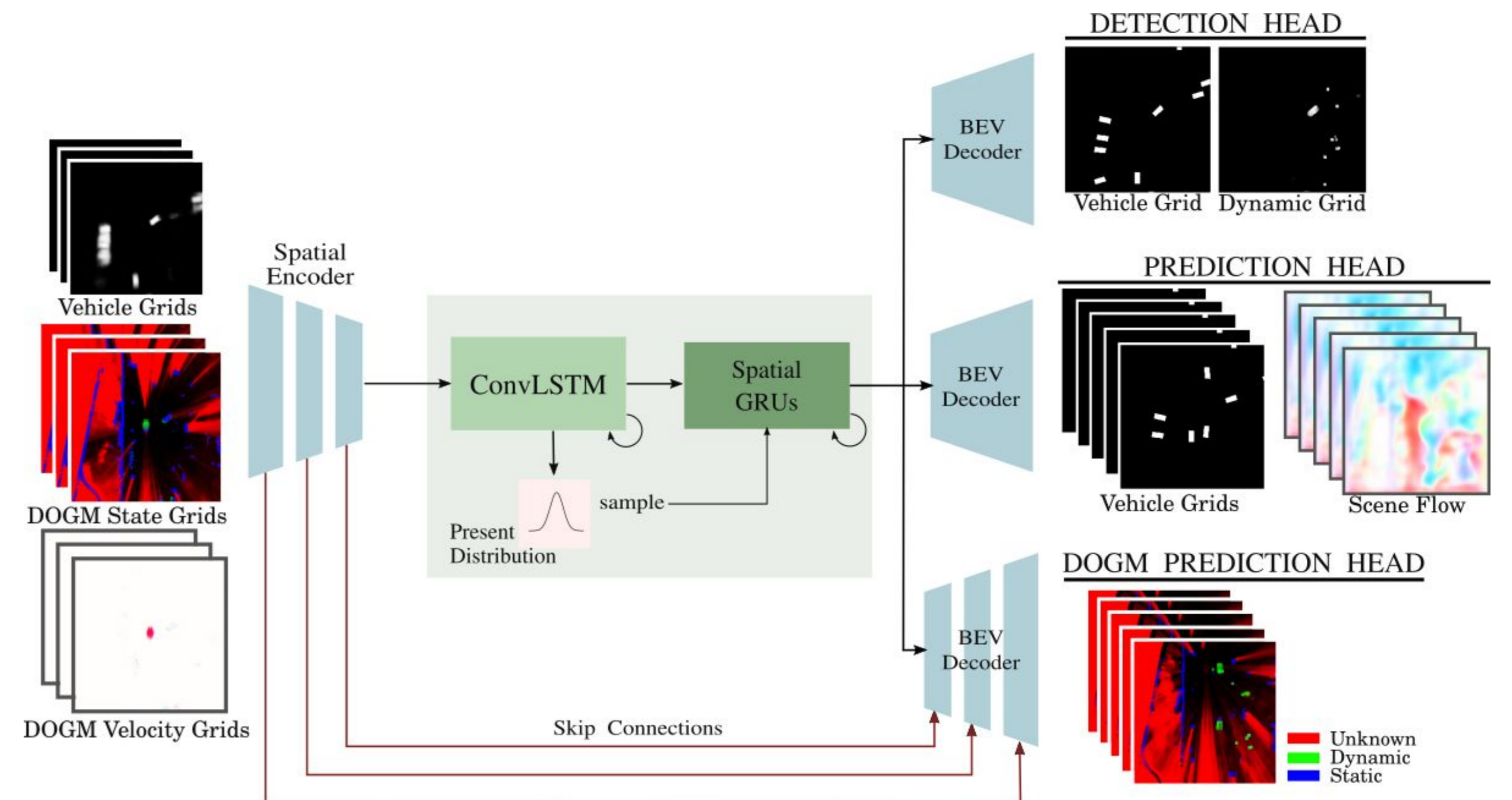
## Quantitative Results

Experiments are conducted on Nuscenes & Lyft datasets with grid dimension of 60x60m. For training, the grid images are resized to 240x240 pixels. In total, NuScenes dataset has 23K training and 5K validation sequences. Lyft dataset has 7K training and 2K validation sequences. Evaluation of vehicle semantic grid prediction compared to video prediction methods PredRNN [3] and PredNet [4].

TABLE I  
VEHICLE SEMANTIC GRID PREDICTION

NuScenes Dataset			
Network	soft-IoU (↑)	soft-IoGT (↑)	soft-IoGT-Dynamic (↑)
PredRNN	0.397	0.599	0.444
PredNet	<b>0.440</b>	0.512	0.177
Ours	0.365	<b>0.670</b>	<b>0.541</b>
Lyft Dataset			
Network	soft-IoU (↑)	soft-IoGT (↑)	soft-IoGT-Dynamic (↑)
PredRNN	0.446	0.594	0.502
PredNet	<b>0.493</b>	0.564	0.382
Ours	0.398	<b>0.627</b>	<b>0.533</b>

## Methods

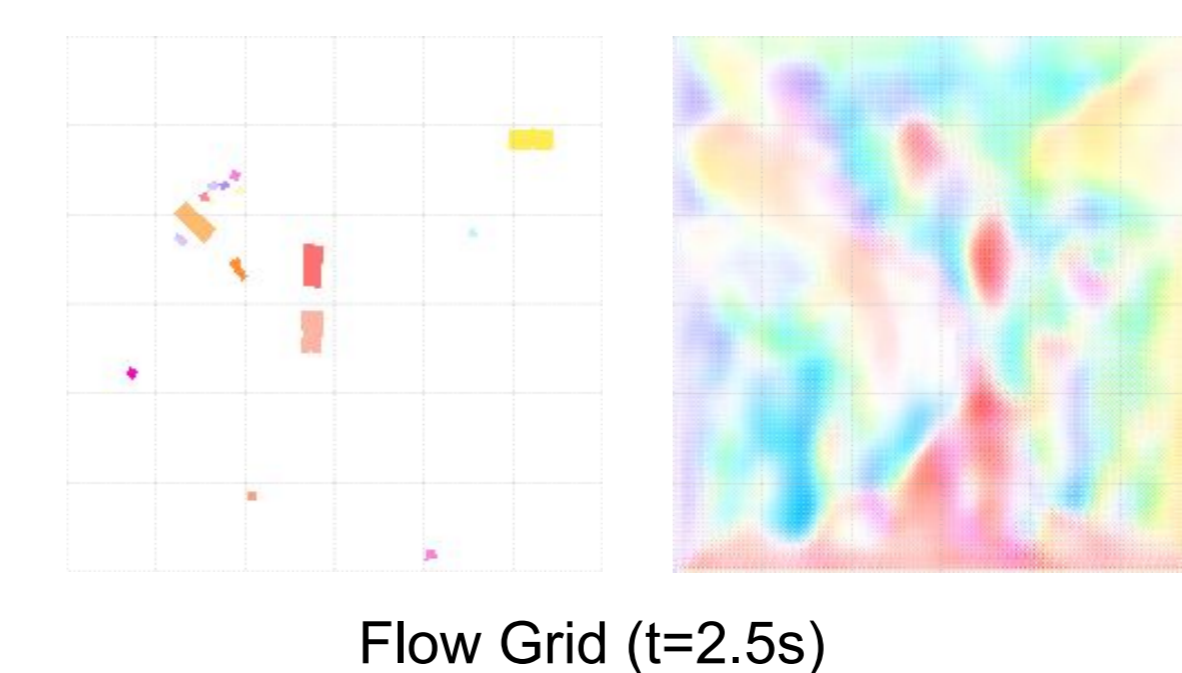
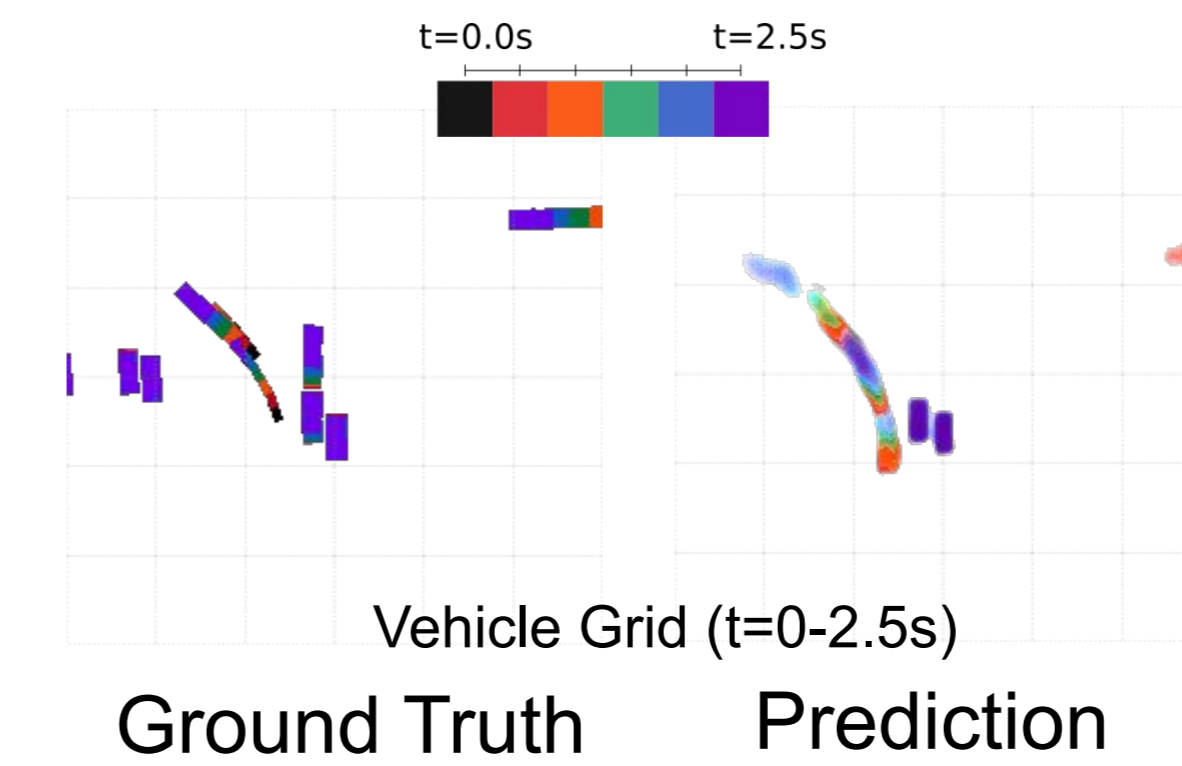
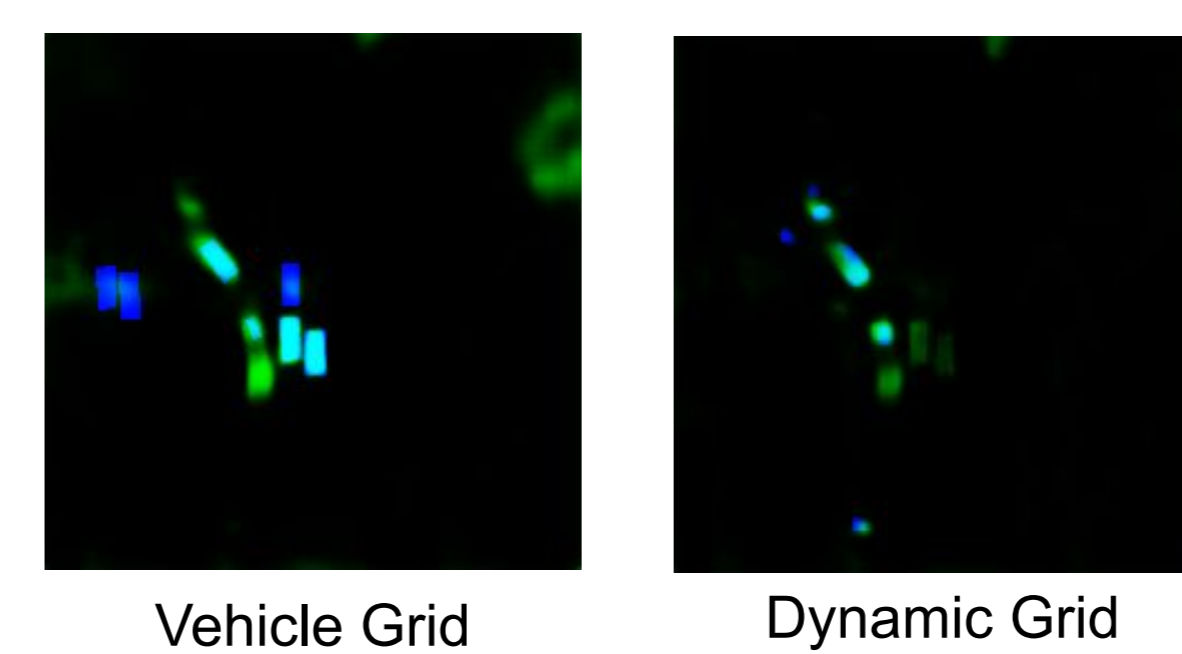
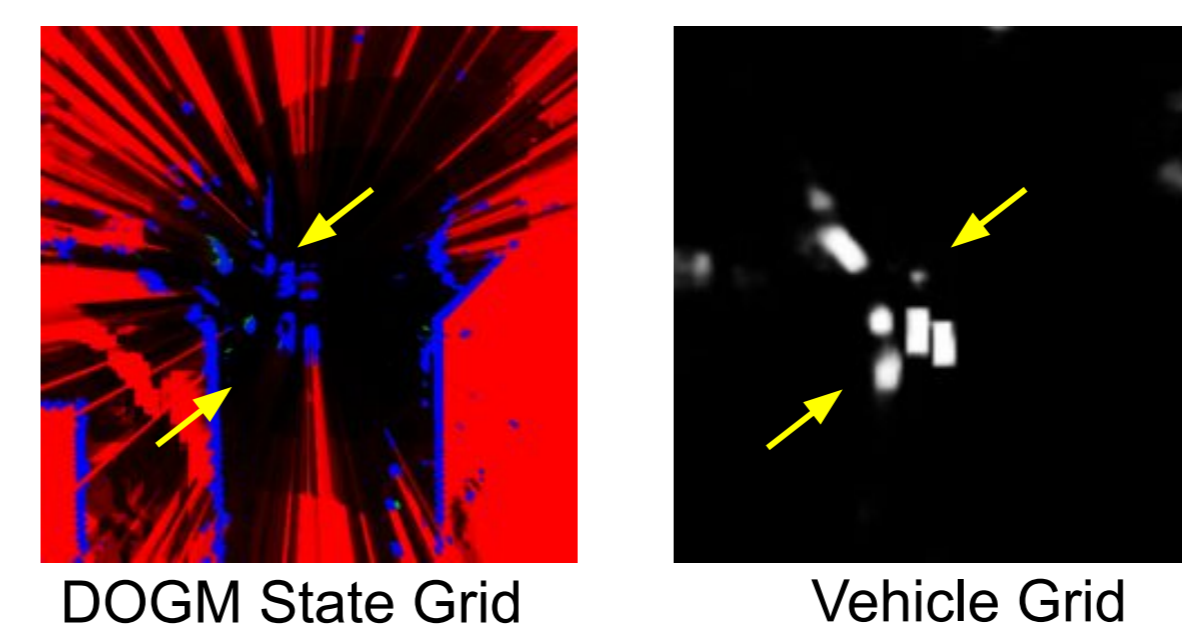


The prediction architecture takes inspiration from **video prediction**, **conditional variational approach** and **skip connection architecture**.

Loss function:

$$\mathcal{L} = \mathcal{L}_{BCE_{seg}} + \mathcal{L}_{BCE_{warped}} + \mathcal{L}_{L1_{flow}} + \mathcal{L}_{MSE_{ogm}} + \mathcal{L}_{KL}$$

## Qualitative Results



## Conclusions

DOGMs, semantic grids, and flow grids provide complementary and redundant scene information, collectively enhancing prediction results, learning patterns and behaviours based on the observable static scene and motion of dynamic agents. Evaluation on real-world dataset show improved ability to predict vehicle motion compared to baseline video prediction methods with DOGMs.

## References

- [1] L. Rummelhard, et al. "A cross-prediction, hidden-state-augmented approach for dynamic occupancy grid filtering," in ICARCV, 2022.
- [2] A. Hu, et al. "FIERY: Future instance segmentation in bird's-eye view from surround monocular cameras," in ICCV, 2021.
- [3] Wang, Y., et al. "Predrnn: A recurrent neural network for spatiotemporal predictive learning" in IEEE TPAMI, 2022
- [4] Lotter, W., et al. (2016). Deep predictive coding networks for video prediction and unsupervised learning. arXiv preprint arXiv:1605.08104.