EURECOM Sophia Antipolis

TUM

Modeling Interactions of Autonomous Vehicles and Pedestrians with Deep Multi-Agent Reinforcement Learning for Collision Avoidance

Raphael Trumpp[1], Harald Bayerlein[1], and David Gesbert[2]

I. Problem Definition

Scenario: Interaction of the agents at an unmarked crosswalk (chicken game)

- Heterogenous agents:
  - **Autonomous vehicle (AV):** Level-5; assumed with high quality sensors
  - **Pedestrian:** Attempts to cross the road; limited reliability of state estimations
- **Goal**: Prevent collisions while ensuring smooth traffic flow → develop a pedestrian collision avoidance mitigation (PCAM) system for the AV using deep reinforcement learning (DRL)

II. Related Work

- Chae et al. [1] are the first to develop a PCAM system using deep reinforcement learning (DRL)
- In [2], a PCAM system for multiple pedestrians is proposed using DRL
- **Limitations**: Pedestrian not in focus and no analysis of the influence of uncertainty

III. Contributions

- A deep multi-agent reinforcement learning (DMARL) approach is used as a new perspective on modeling pedestrians
- The influence of observation noise on the agents' performance is evaluated
- Our approach generalizes well over different scenarios and the driving capability is beyond similar works

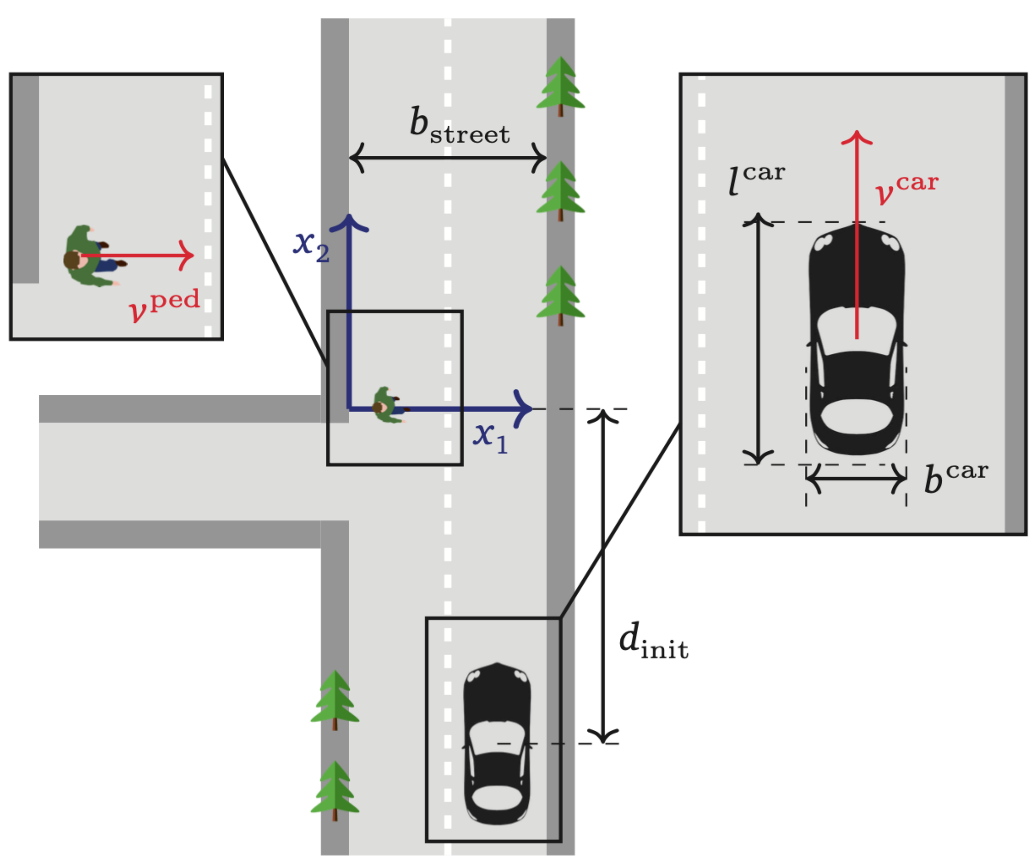

Figure 1: Exemplary driving scenario at an unmarked crosswalk.

[1] R. Trumpp and H. Bayerlein are with the TUM School of Engineering and Design, Technical University of Munich, Germany, e-mail: {raphael.trumpp, h.bayerlein}@tum.de.

[2] D. Gesbert is with the Communication Systems Department, EURECOM, Sophia Antipolis, France, e-mail: david.gesbert@eurecom.fr.

R. Trumpp and H. Bayerlein were supported by the Chair of Cyber-Physical Systems in Production Engineering at TUM. H. Bayerlein and D. Gesbert were partially supported by the French government, through the 3IA Côte d'Azur project number ANR-19-P3IA-0002, as well as by the TSN CARNOT Institute under project Robots4IoT.

References

1] H. Chae, C. M. Kang, B. Kim, J. Kim, C. C. Chung, and J. W. Choi, "Autonomous braking system via deep reinforcement learning," IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017.

[2] N. Deshpande, D. Vaufreydaz, and A. Spalanzani, "Behavioral decision-making for urban autonomous driving in the presence of pedestrians using deep recurrent Q-network," in 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2020.

[3] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016.

IV. Methodology

Idea: Modelling different levels of intelligent behavior

Level-1	Acting rationally by perceiving the environment
Level-2	Learning and exploring new strategies

Pedestrian models

- Level-1: Action u_t^{ped} is taken at each time step

$$u_t^{\text{ped}} = \begin{cases} \text{walk}, & \text{if } \mathrm{TTC}_t \geq 3\text{s} \\ \text{wait}, & \text{otherwise} \end{cases}$$

- Level-2: **DDQN** [3] model with state observations s_t^{ped}, actions u_t^{ped}, and reward function $\mathcal{R}^{\text{ped}}$

$$\mathcal{U}^{\text{ped}} = \{\text{wait}, \text{walk}\}$$

$$r_{t+1}^{\text{ped}} = -\tau^{\text{ped}} - \begin{cases} \beta^{\text{ped}}, & \text{if collision} = \text{True} \\ 0, & \text{otherwise} \end{cases}$$

AV models

- Level-1: Action u_t^{car} is taken from a best response analysis
- Level-2: **DDQN** [3] model with state observations s_t^{AV}, actions u_t^{AV}, and reward function $\mathcal{R}^{\text{AV}}$:

$$\mathcal{U}^{\text{ped}} = \{-9.8, -5.8, -3.8, 0, 1, 2, 3\} \frac{\text{m}}{\text{s}^2}$$

$$r_t^{\text{AV}} = -\tau^{\text{AV}} - \begin{cases} \beta^{\text{AV}}, & \text{if collision} = \text{True} \\ 0, & \text{otherwise} \end{cases}$$

$$- \begin{cases} \psi^{\text{AV}}, & \text{if } v_t^{\text{AV}} > v_{\text{limit}}^{\text{AV}} \\ 0, & \text{otherwise} \end{cases}$$

Environment models

- Partially-observable Markov decision process (POMDP) with tuple:

$$(\mathcal{S}^{\text{AV}}, \mathcal{Z}^{\text{AV}}, \mathcal{U}^{\text{AV}}, \mathcal{T}, \mathcal{O}, \mathcal{R}^{\text{AV}}, \gamma)$$

- Markov game (MG) with agents $\mathcal{W} = \{w^{\text{AV}}, w^{\text{ped}}\}$ and tuple:

$$(\mathcal{W}, \mathcal{S}, \mathcal{Z}, \mathcal{U}, \mathcal{T}, \mathcal{O}, \mathcal{R}, \gamma)$$

- **Multiplicative noise model to account for uncertainty**

$$\mathcal{O}: z_t = (1 + n_t) \cdot s_t \quad \text{with} \quad n_t \sim \mathcal{N}(0, \alpha^2)$$

- **Varying environment parameters**

$$v_{\text{init}}^{\text{AV}} \sim \mathcal{U}\left(30 \frac{\text{km}}{\text{h}}, 50 \frac{\text{km}}{\text{h}}\right)$$

$$\mathrm{TTC}_{\text{init}} \sim \mathcal{U}(1.0\text{s}, 5.0\text{s})$$

$$v_{\text{walk}}^{\text{ped}} \in \{1.16, 1.38, 1.47, 1.53, 1.55\} \frac{\text{m}}{\text{s}}$$

$$b^{\text{street}} = \{6.0, 7.5\}\text{m}$$

V. Results

- Training: 8,000 episodes; DQN has a replay buffer of 50,000 experiences
- Evaluation: 80% confidence with median of 8 independent runs
- Noise evaluation: $\alpha^{\text{AV}} = 0.05$ and $\alpha^{\text{ped}} = \{0.0, 0.1, 0.2, 0.4, 0.5\}$
- Independent learning scheme for DMARL (semi-cooperative)

Setting-1: Pedestrian with fixed policy

DRL	POMDP	Level-1 pedestrian	Level-2 AV

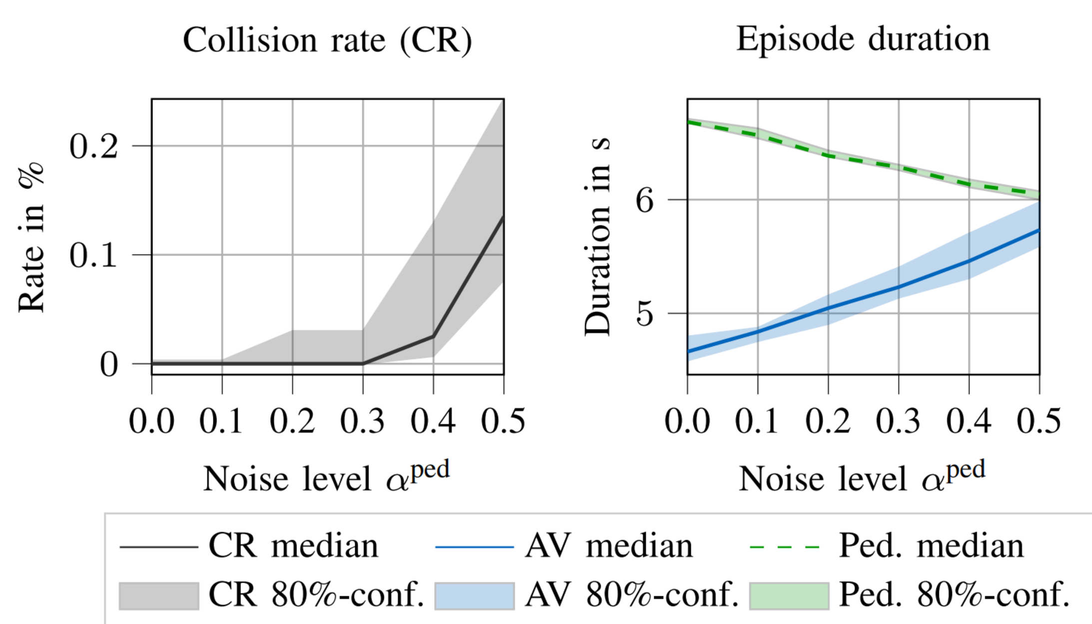

Figure 2: Performance in setting-1 over different noise levels.

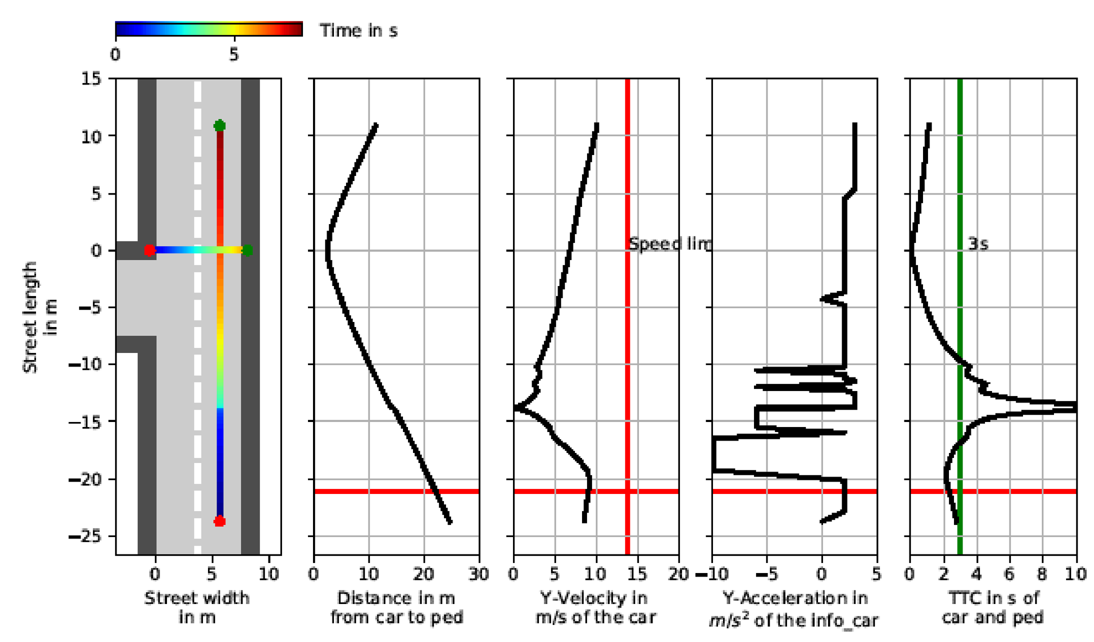

Figure 3: Agent behavior in setting-1 with $\alpha^{\text{ped}} = 0.3$.

Setting-2: Pedestrian with fixed policy

DMARL	MG	**Level-2 pedestrian**	Level-2 AV

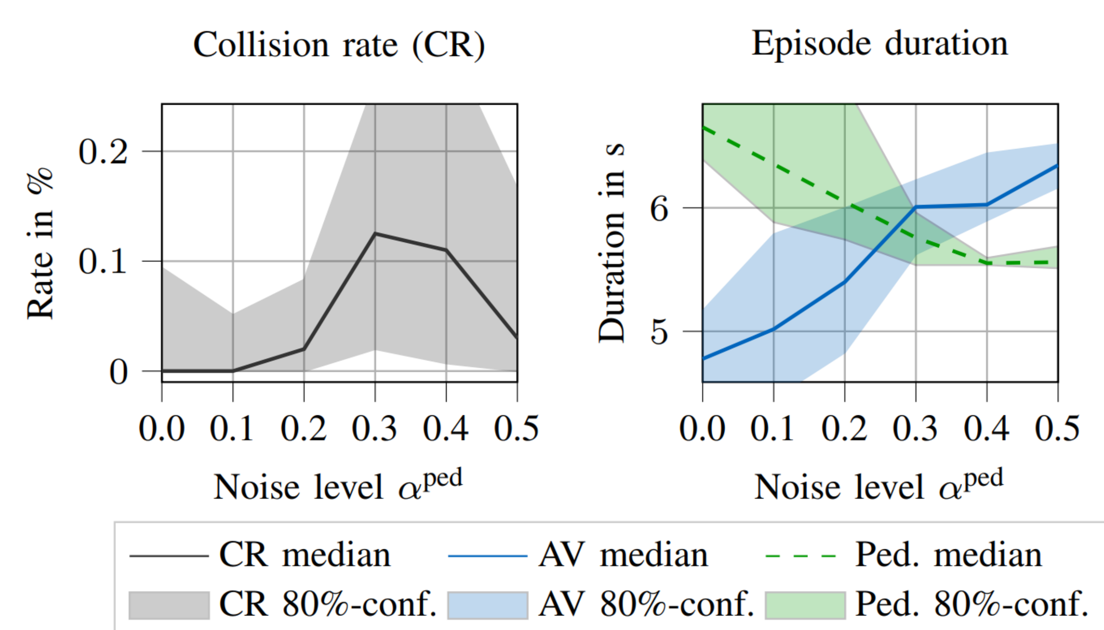

Figure 4: Performance in setting-2 over different noise levels.

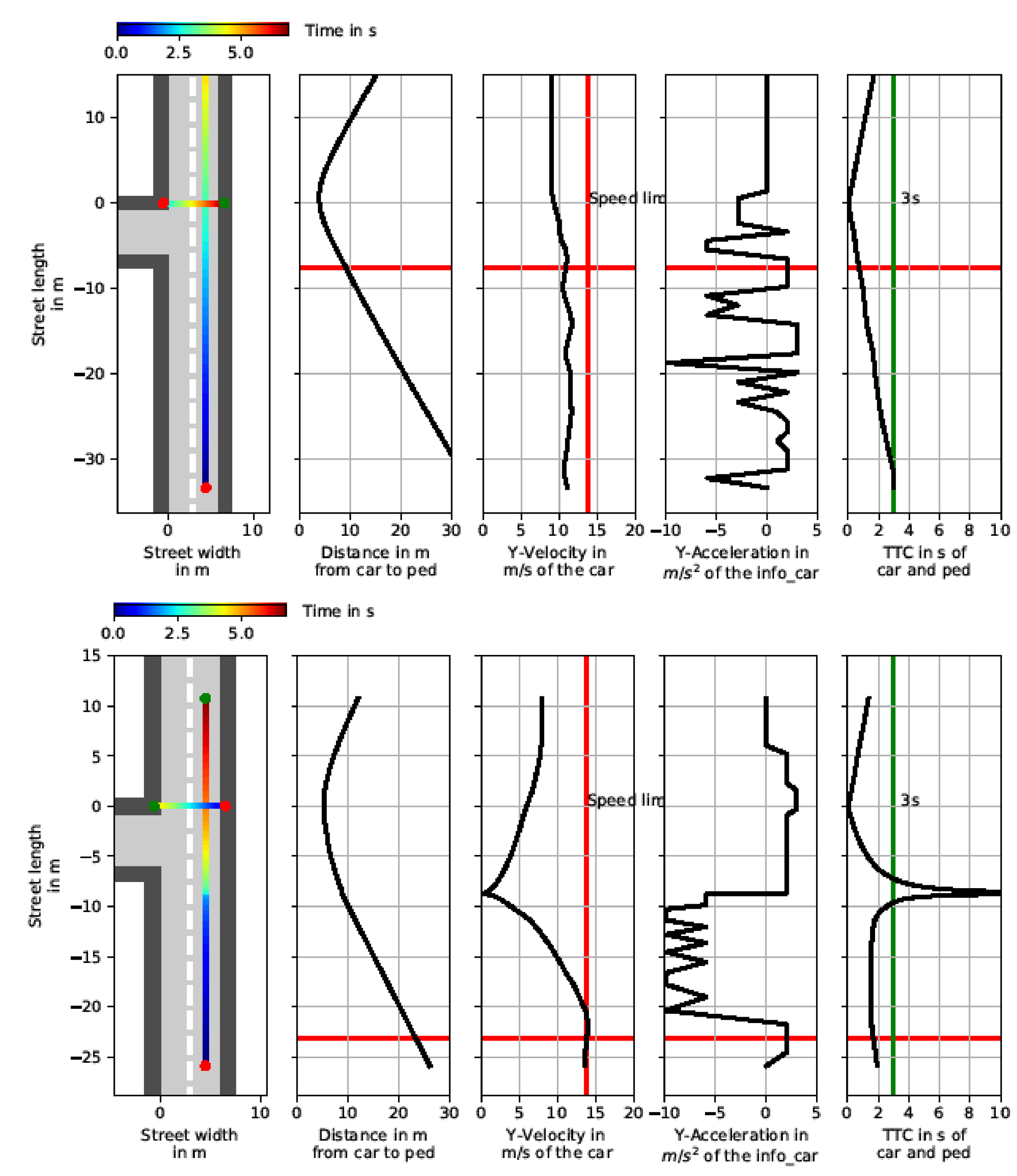

Figure 5: Agent behavior in setting-2 with $\alpha^{\text{ped}} = 0.1$.