AIthena

Long-term pedestrian trajectory prediction model

MODEL CARD

Author name/s and affiliation	Michael Stolz, Virtual Vehicle Research GmbH Ilma Okanovic, Virtual Vehicle Research GmbH Jan Ahmetspahic, Virtual Vehicle Research GmbH
Contact details	michael.stolz@v2c2.at, ilma.okanovic@v2c2.at, jan.ahmetspahic@v2c2.at
Link/s	Virtual Vehicle Research GmbH
Affiliation logo	virtual vehicle
Date	01/10/2023

Disclaimer

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

INTRODUCTION

MODEL DESCRIPTION

The model is to perform long-term prediction of detected pedestrians' trajectories based on map information and the history of pedestrians' states. Its intended use is to provide one of the functionalities necessary for traffic navigation of a robo-taxi. Therefore, any subsequent information about the model should be understood within the context of a robo-taxi service.

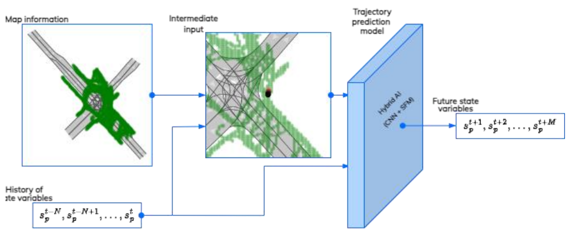

Figure 1. Prediction of future M states based on intermediate map information and history of N states (current state included).

- **Model version:** 0.0
- **Model license:** Open source
- **Citation:**

[1] F.-C. Chou *et al.*, "Predicting Motion of Vulnerable Road Users using High-Definition Maps and Efficient ConvNets," in *IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, 2020, pp. 1655–1662. doi: 10.1109/IV47402.2020.9304564.

[2] H. Cui *et al.*, "Deep Kinematic Models for Kinematically Feasible Vehicle Trajectory Predictions," in *IEEE International Conference on Robotics and Automation (ICRA)x*, Paris, France, 2020, pp. 10563–10569. doi: 10.1109/ICRA40945.2020.9197560.

- **Repository link:** [Long-term pedestrian trajectory prediction model]

MODEL DETAILS

MODEL ARCHITECTURE

- **Model type:** Hybrid AI model consisting of data-based (CNN) and physics-based (SFM) components.
- **Model task:** Regression
- **Number of layers:**
- **Number of nodes:**

- **Inputs:** A 250x250 RGB image (intermediate input) of the map with encoded information about the pedestrians' potential regions of movement and states s_p^t up to a certain time horizon of width N.
- **Outputs:** Future M states of the detected pedestrian.

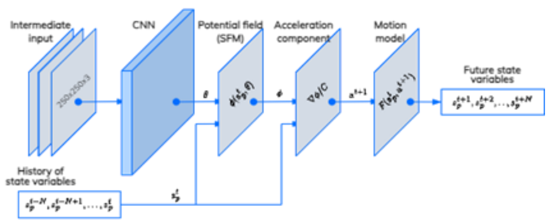

Figure 2. Hybrid AI model architecture consisting of convolutional neural network (CNN) and social force model (SFM)

MODEL DETAILS

TRAINING DATA

- **Dataset description:**
  - The dataset consists of ~500k .png images representing intermediate input generated based on three different intersection locations. Each pedestrian tracked in images has its ground truth states stored in .pkl files ready to be used for loss calculation.
  - The original recordings are part of the publicly available inD dataset. For more information on the inD dataset refer to its accompanying data card.
- **Data splitting:**
  - The intermediate input images were split into training and validation sets by random sampling. However, the link to their corresponding GT trajectory file was preserved.
- **Data augmentation techniques:**
  - Birdseye view of the traffic scene is contained within a 250x250 pixels RBG image.
  - For each timestamp, an image placing the tracked pedestrian at its centre will be generated. The map information will be restricted to 200 m in each direction of the frame axes relative to the frame centre.
  - The encoding of the map and state information is in accordance with the CommonRoad module and schema.

EVALUATION DATA

- **Dataset description:**
  - The dataset consists of ~100k .png images representing intermediate input generated based on the fourth remaining intersection location recorded in the inD dataset. Same as for the training data, the GT trajectories corresponding to the pedestrians tracked in images are stored in .pkl files.
  - The same data preprocessing steps described for the training dataset are performed for the testing dataset as well.

INTENDED USE

USE CASES AND USERS

- **Intended use cases:**
  - Trajectory prediction of vulnerable road users (VRU), specifically pedestrians.
  - Complex dynamic traffic situations common for urban roads and road segments such as intersections.
  - ODD: high visibility, absence of extreme weather conditions, smooth and even roads, no occlusion of pedestrians, urban inner-city roads.
  - The model aids in efficient traffic navigation of the robo-taxi all while ensuring the safety of all its primary (passengers) and secondary users (detected pedestrians).
- **Target audience or users of the model:**
  - The primary users of the model are pedestrians requiring a taxi service within a city in Germany or Austria.
  - The primary users do not require specialised expertise. However, all users should be aware that the robo-taxi relies on the traffic regulations that the users are expected to adhere to.
- **Users to be analysed by the model:**
  - The secondary users are the pedestrians whose trajectories will be predicted by the model.

LIMITATIONS

- **Limitations or restrictions of the intended use:**
  - The model can only be used with inputs formatted according to the specifications outlined in the data card.
  - The scenarios depicted in the input images must align with the assumptions stated in the ODD.
  - The adopted traffic participant class cannot be generalized to the excluded traffic participants.
  - Ethical limitations could be caused by the lack of representation of disabled individuals or children in the training dataset.
  - The model will not be relying on any information that is not already publicly available and will not be transmitting inputs or outputs beyond the scope of the system.
- **Guidance on how to address limitations or restrictions:**
  - Extension of the ODD of the future versions of the model.
  - Inclusion of new classes of traffic participants requiring the extension or generalisation of expert knowledge embedded in the model.
- **Identification of potential biases:**
  - Uneven representation of subclasses of pedestrians participating in traffic should be addressed at the dataset generation and preprocessing stage.

EVALUATION AND PERFORMANCE

METRICS

- **Performance metrics:**
  - The accuracy measure will be based on the error between the ground truth and the predicted trajectory of each detected pedestrian.
  - Performance measures that quantify how well the predicted trajectories comply with the physical laws incorporated into the model. Additionally, the comparison will be done with respect to the motion models of varying complexity.
- **Relevant thresholds or criteria:**
  - The model's performance will be compared to the results stated in relevant literature including those proposing solutions solely data- or physics-based.
- **Evaluation method:**
  - The chosen evaluation method is cross-validation.

RESULTS

- **Results of the performance evaluation:**
- **Performance on specific subsets of dataset:**
- **Performance metrics interpretation:**

ETHICAL CONSIDERATIONS

FAIRNESS

- **Fairness metrics:**
  - Equal opportunity will be used to measure individual pedestrian subclass fairness*.
  - Statistical parity difference will be used to measure group fairness.
- **Mitigation of potential biases:**
  - Mitigation of mentioned biases should be addressed at the dataset preprocessing stage.

*The fairness term is understood within its broader context encompassing ethical concerns.

PRIVACY

- **Data collecting and storing:**
  - Once the perception system detects a pedestrian estimating their current state, such data is transferred to the GDB.
  - With every iteration of trajectory prediction, the GDB will be queried for the map information and the pedestrian's state values.
  - It is intended to store only the data that will increase the accuracy of the trajectory prediction. Therefore, only the information up to a certain time horizon will be stored and utilized for prediction.
  - Data will not be shared with third parties and will only be used within robo-taxi's navigation system.

ACCOUNTABILITY

- **Model's decisions transparency:**
  - To increase explainability, a hybrid of data- and physics-based models was defined. Using such a model implies that the embedded physical laws enforce learning compliance with them.
- **Reviewing and auditing:**
  - A method allowing users to express their concerns should be defined at the system level.
- **Limitations or uncertainties:**
  - It is not expected of users to have the expertise necessary to understand physical laws. However, the model still depends on the knowledge it acquires through intermediate input, which could be the source of the "black box" behaviour observed in traditional ML models.

USAGE AND LIMITATIONS

USAGE AND RECOMMENDATIONS

- **Model usage instructions:**
- **Input data format and preprocessing:**
  - PNG format will be used for the enriched map image, while the pedestrian state information (numerical) will be stored in a JSON format.
- **Output interpretation:**
  - Pedestrians' trajectory is predicted according to a defined precision threshold sampled at a specified sampling rate.
  - In the robo-taxi's navigation system, it is intended that the output of the prediction model be used for decision-making. After a decision on the future action is made, a control system will force the vehicle to continue navigating through the traffic.

LIMITATIONS AND RISKS

- **Technical limitations:**
- **Ethical or legal limitations and restrictions:**
- **Guidance on how to address limitations or restrictions:**
- **Risks identification:**