論文題目 #1

從閱讀他人論文中尋找自己的研究主題

Datasets for Motion Prediction in AD

論文題目:

- A review of trajectory prediction evaluation metrics in AV
- A review of trajectory prediction datasets in AV

Recap of Previous Lecture

- 做研究要有 **動機**

- 多數的交通意外來自車輛駕駛的人為疏失。
- 開發提升安全的緊急煞車系統(AEB)，在車輛即將發生碰撞前，自動煞車以減緩車輛的損傷以及乘客的傷害。

https://www.artc.org.tw/tw/knowledge/articles/13706

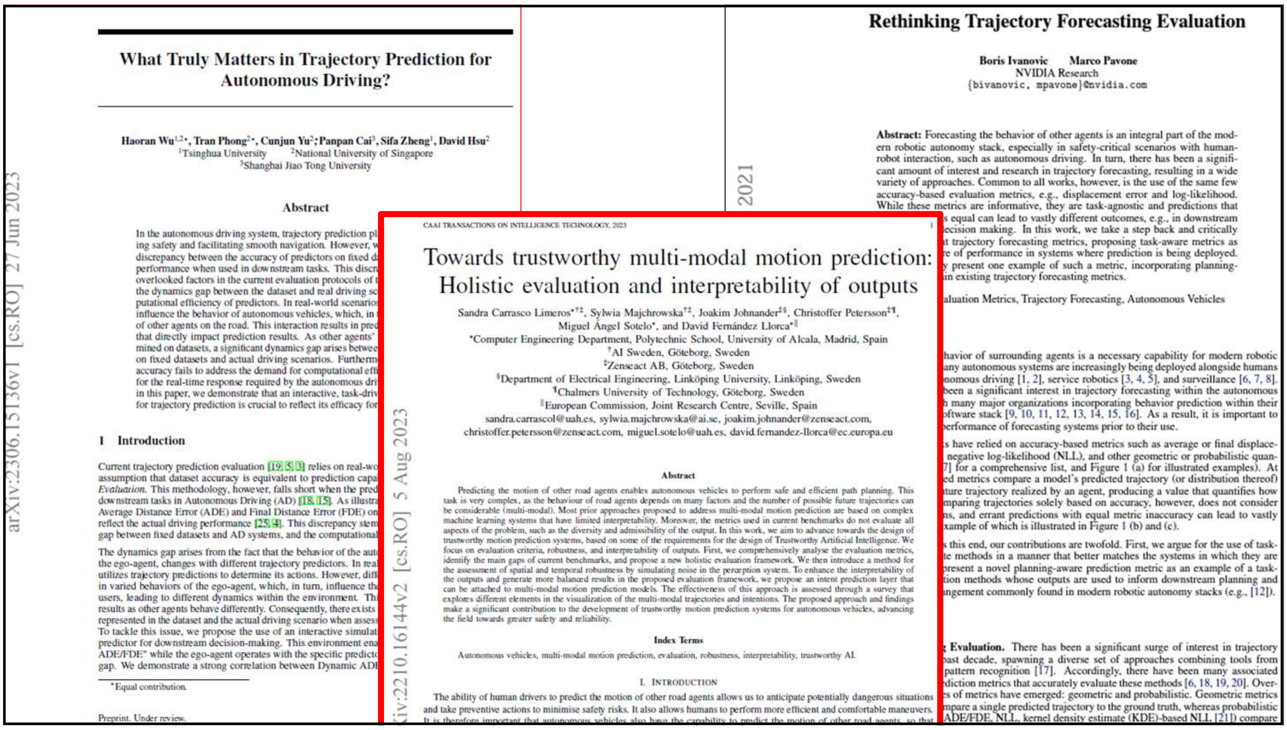
自動緊急煞車系統
(Automatic Emergency Braking)

AEB是什麼？AEB全稱Autonomous Emergency Braking主動煞停系統，主要透過感測器（雷達、鏡頭等）偵測前方目標，透過控制器計算危險程度，當駕駛分心時，造成與前車過近，系統藉由聲響或燈號提醒，甚至主動介入達成煞停目的。

https://c.8891.com.tw/feature/1082

- Trajectory / Intention Prediction

https://www.atssa.com/Blog-News/ATSSA-Blog/atssa-issues-recommendations-for-a-vulnerable-road-users-program



What Truly Matters in Trajectory Prediction for Autonomous Driving?

Haoran Wu[1,2*], Tran Phong[2*], Cunjun Yu[2]; Panpan Cai[3], Sifa Zheng[1], David Hsu[2]
[1]Tsinghua University [2]National University of Singapore
[3]Shanghai Jiao Tong University

arXiv:2306.15136v1 [cs.RO] 27 Jun 2023

Rethinking Trajectory Forecasting Evaluation

Boris Ivanovic Marco Pavone
NVIDIA Research
{bivanovic, mpavone}@nvidia.com

CAAI TRANSACTIONS ON INTELLIGENCE TECHNOLOGY, 2023 1

Towards trustworthy multi-modal motion prediction: Holistic evaluation and interpretability of outputs

Sandra Carrasco Limeros[*†‡], Sylwia Majchrowska[†‡], Joakim Johnander[‡§], Christoffer Petersson[‡¶], Miguel Ángel Sotelo[*], and David Fernández Llorca[*‖]
[*]Computer Engineering Department, Polytechnic School, University of Alcala, Madrid, Spain
[†]AI Sweden, Göteborg, Sweden
[‡]Zenseact AB, Göteborg, Sweden
[§]Department of Electrical Engineering, Linköping University, Linköping, Sweden
[¶]Chalmers University of Technology, Göteborg, Sweden
[‖]European Commission, Joint Research Centre, Seville, Spain
sandra.carrascol@uah.es, sylwia.majchrowska@ai.se, joakim.johnander@zenseact.com, christoffer.petersson@zenseact.com, miguel.sotelo@uah.es, david.fernandez-llorca@ec.europa.eu

arXiv:2210.16144v2 [cs.RO] 5 Aug 2023

Abstract

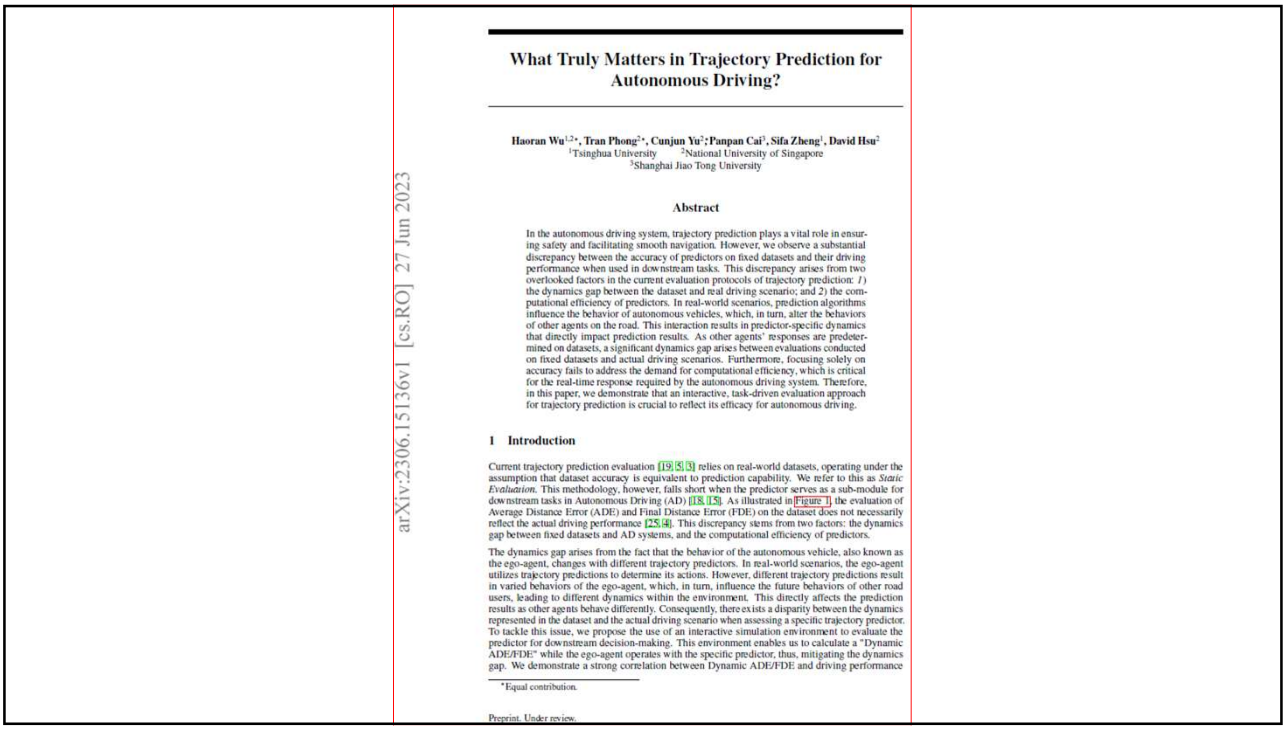
Predicting the motion of other road agents enables autonomous vehicles to perform safe and efficient path planning. This task is very complex, as the behaviour of road agents depends on many factors and the number of possible future trajectories can be considerable (multi-modal). Most prior approaches proposed to address multi-modal motion prediction are based on complex machine learning systems that have limited interpretability. Moreover, the metrics used in current benchmarks do not evaluate all aspects of the problem, such as the diversity and admissibility of the output. In this work, we aim to advance towards the design of trustworthy motion prediction systems, based on some of the requirements for the design of Trustworthy Artificial Intelligence. We focus on evaluation criteria, robustness, and interpretability of outputs. First, we comprehensively analyse the evaluation metrics, identify the main gaps of current benchmarks, and propose a new holistic evaluation framework. We then introduce a method for the assessment of spatial and temporal robustness by simulating noise in the perception system. To enhance the interpretability of the outputs and generate more balanced results in the proposed evaluation framework, we propose an intent prediction layer that can be attached to multi-modal motion prediction models. The effectiveness of this approach is assessed through a survey that explores different elements in the visualization of the multi-modal trajectories and intentions. The proposed approach and findings make a significant contribution to the development of trustworthy motion prediction systems for autonomous vehicles, advancing the field towards greater safety and reliability.

Index Terms

Autonomous vehicles, multi-modal motion prediction, evaluation, robustness, interpretability, trustworthy AI.

I. INTRODUCTION

The ability of human drivers to predict the motion of other road agents allows us to anticipate potentially dangerous situations and take preventive actions to minimise safety risks. It also allows humans to perform more efficient and comfortable maneuvers.



What Truly Matters in Trajectory Prediction for Autonomous Driving?

Haoran Wu[1,2*], Tran Phong[2*], Cunjun Yu[2]; Panpan Cai[3], Sifa Zheng[1], David Hsu[2]
[1]Tsinghua University [2]National University of Singapore
[3]Shanghai Jiao Tong University

arXiv:2306.15136v1 [cs.RO] 27 Jun 2023

Abstract

In the autonomous driving system, trajectory prediction plays a vital role in ensuring safety and facilitating smooth navigation. However, we observe a substantial discrepancy between the accuracy of predictors on fixed datasets and their driving performance when used in downstream tasks. This discrepancy arises from two overlooked factors in the current evaluation protocols of trajectory prediction: *1*) the dynamics gap between the dataset and real driving scenario; and *2*) the computational efficiency of predictors. In real-world scenarios, prediction algorithms influence the behavior of autonomous vehicles, which, in turn, alter the behaviors of other agents on the road. This interaction results in predictor-specific dynamics that directly impact prediction results. As other agents' responses are predetermined on datasets, a significant dynamics gap arises between evaluations conducted on fixed datasets and actual driving scenarios. Furthermore, focusing solely on accuracy fails to address the demand for computational efficiency, which is critical for the real-time response required by the autonomous driving system. Therefore, in this paper, we demonstrate that an interactive, task-driven evaluation approach for trajectory prediction is crucial to reflect its efficacy for autonomous driving.

1 Introduction

Current trajectory prediction evaluation [19, 5, 3] relies on real-world datasets, operating under the assumption that dataset accuracy is equivalent to prediction capability. We refer to this as *Static Evaluation*. This methodology, however, falls short when the predictor serves as a sub-module for downstream tasks in Autonomous Driving (AD) [18, 15]. As illustrated in Figure 1, the evaluation of Average Distance Error (ADE) and Final Distance Error (FDE) on the dataset does not necessarily reflect the actual driving performance [25, 4]. This discrepancy stems from two factors: the dynamics gap between fixed datasets and AD systems, and the computational efficiency of predictors.

The dynamics gap arises from the fact that the behavior of the autonomous vehicle, also known as the ego-agent, changes with different trajectory predictors. In real-world scenarios, the ego-agent utilizes trajectory predictions to determine its actions. However, different trajectory predictions result in varied behaviors of the ego-agent, which, in turn, influence the future behaviors of other road users, leading to different dynamics within the environment. This directly affects the prediction results as other agents behave differently. Consequently, there exists a disparity between the dynamics represented in the dataset and the actual driving scenario when assessing a specific trajectory predictor. To tackle this issue, we propose the use of an interactive simulation environment to evaluate the predictor for downstream decision-making. This environment enables us to calculate a "Dynamic ADE/FDE" while the ego-agent operates with the specific predictor, thus, mitigating the dynamics gap. We demonstrate a strong correlation between Dynamic ADE/FDE and driving performance

*Equal contribution.

Preprint. Under review.

https://arxiv.org/abs/2306.15136

What Truly Matters in Trajectory Prediction for Autonomous Driving?

Haoran Wu[1,2*], Tran Phong[2*], Cunjun Yu[2], PanPan Cai[3], Sifa Zheng, David Hsu[2]

[1]Tsinghua University

[2]National University of Singapore

[3]Shanghai Jiao Tong University

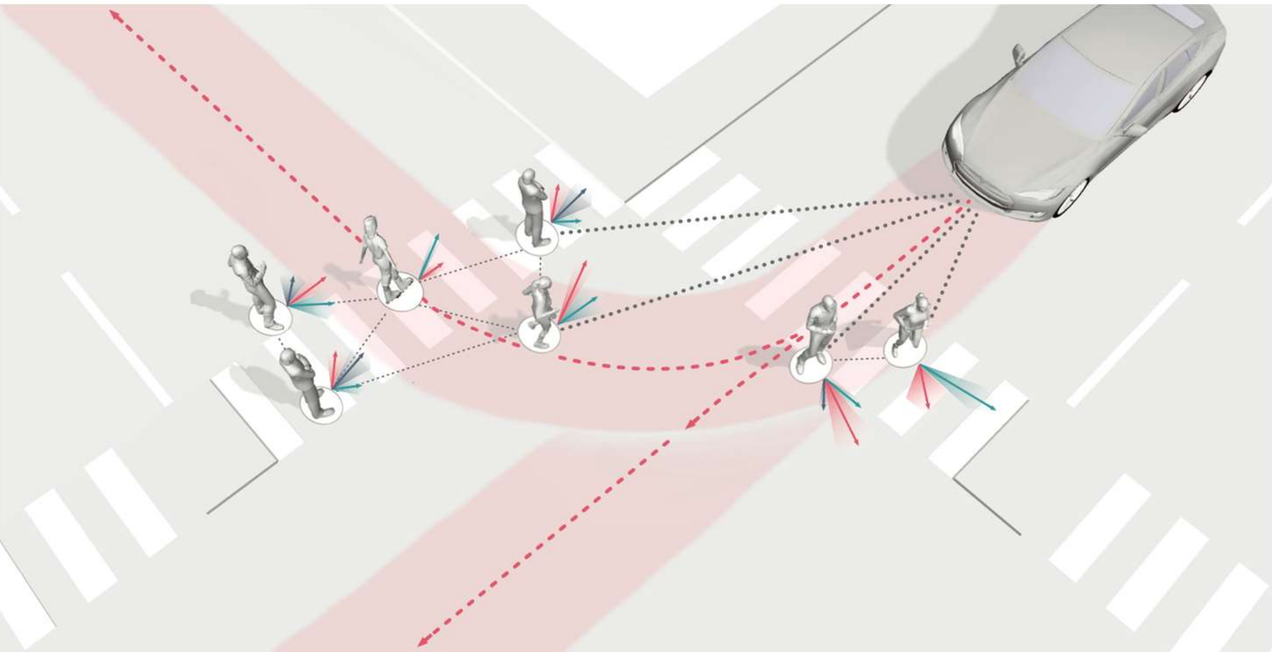
June 2023

https://arxiv.org/abs/2306.15136

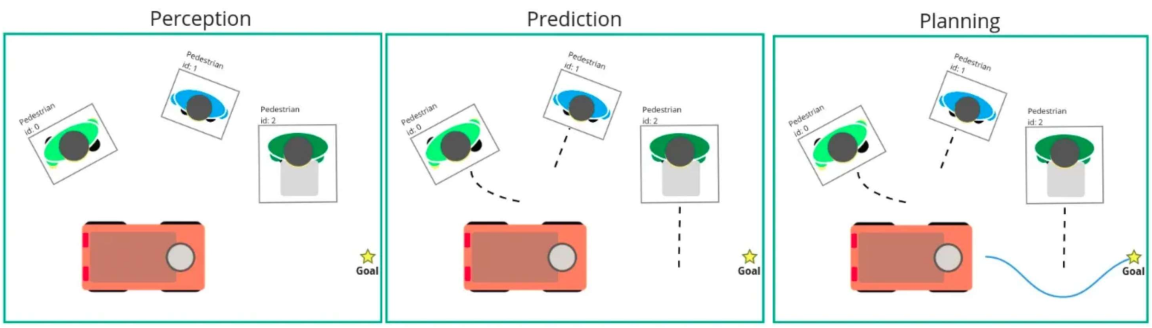
Preprint Under review

Problem Description

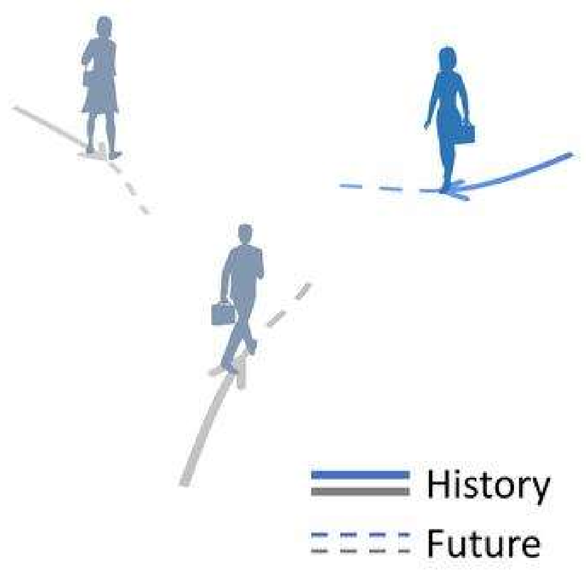
Trajectory Prediction



https://ai.stanford.edu/blog/trajectory-forecasting/



https://shorturl.at/qrvM6t

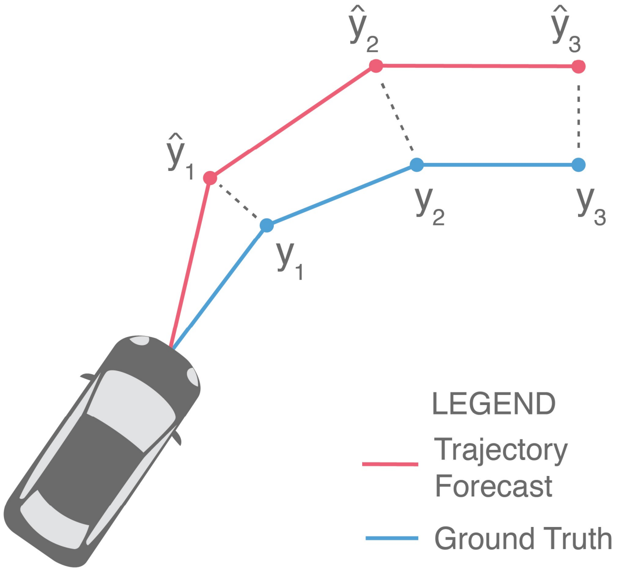


https://paperswithcode.com/task/trajectory-forecasting

Performance Evaluation in Trajectory Forecasting

Average Displacement Error (ADE)
Final Displacement Error (FDE)

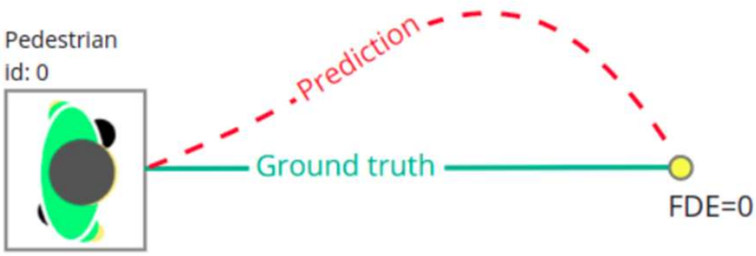
minADE
minFDE



$$ADE(\hat{y}, y) = \frac{1}{T}\sum_{i=1}^{T} ||\hat{y}_i - y_i||^2$$

$$FDE(\hat{y}, y) = ||\hat{y}_T - y_T||^2$$

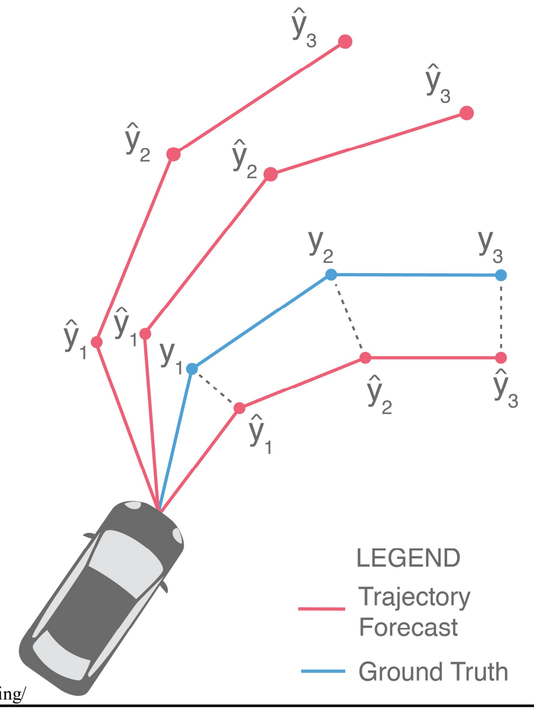
https://ai.stanford.edu/blog/trajectory-forecasting/



https://shorturl.at/qrvM6t

Best-of-N (BoN)

Multi-modal



https://ai.stanford.edu/blog/trajectory-forecasting/

Table 2: Trajectory prediction metrics.

Metric Name	Metric Equation
ADE	$\frac{1}{T}\sum_{i=1}^{T}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2}$
FDE	$\sqrt{(x_T-\hat{x}_T)^2+(y_T-\hat{y}_T)^2}$
minADE	$\min_{k\in K}\frac{1}{T}\sum_{i=1}^{T}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2}$
minFDE	$\min_{k\in K}\sqrt{(x_T-\hat{x}_T)^2+(y_T-\hat{y}_T)^2}$

arXiv:2306.15136v1 [cs.RO] 27 Jun 2023

What Truly Matters in Trajectory Prediction for Autonomous Driving?

Haoran Wu[1,2*], Tran Phong[2*], Cunjun Yu[2]; Panpan Cai[3], Sifa Zheng[1], David Hsu[2]

[1]Tsinghua University [2]National University of Singapore

[3]Shanghai Jiao Tong University

Abstract

In the autonomous driving system, trajectory prediction plays a vital role in ensuring safety and facilitating smooth navigation. However, we observe a substantial discrepancy between the accuracy of predictors on fixed datasets and their driving performance when used in downstream tasks. This discrepancy arises from two overlooked factors in the current evaluation protocols of trajectory prediction: *1*) the dynamics gap between the dataset and real driving scenario; and *2*) the computational efficiency of predictors. In real-world scenarios, prediction algorithms influence the behavior of autonomous vehicles, which, in turn, alter the behaviors of other agents on the road. This interaction results in predictor-specific dynamics that directly impact prediction results. As other agents' responses are predetermined on datasets, a significant dynamics gap arises between evaluations conducted on fixed datasets and actual driving scenarios. Furthermore, focusing solely on accuracy fails to address the demand for computational efficiency, which is critical for the real-time response required by the autonomous driving system. Therefore, in this paper, we demonstrate that an interactive, task-driven evaluation approach for trajectory prediction is crucial to reflect its efficacy for autonomous driving.

1 Introduction

Current trajectory prediction evaluation [19, 5, 3] relies on real-world datasets, operating under the assumption that dataset accuracy is equivalent to prediction capability. We refer to this as *Static Evaluation*. This methodology, however, falls short when the predictor serves as a sub-module for downstream tasks in Autonomous Driving (AD) [18, 15]. As illustrated in Figure 1, the evaluation of Average Distance Error (ADE) and Final Distance Error (FDE) on the dataset does not necessarily reflect the actual driving performance [25, 4]. This discrepancy stems from two factors: the dynamics gap between fixed datasets and AD systems, and the computational efficiency of predictors.

The dynamics gap arises from the fact that the behavior of the autonomous vehicle, also known as the ego-agent, changes with different trajectory predictors. In real-world scenarios, the ego-agent utilizes trajectory predictions to determine its actions. However, different trajectory predictions result in varied behaviors of the ego-agent, which, in turn, influence the future behaviors of other road users, leading to different dynamics within the environment. This directly affects the prediction results as other agents behave differently. Consequently, there exists a disparity between the dynamics represented in the dataset and the actual driving scenario when assessing a specific trajectory predictor. To tackle this issue, we propose the use of an interactive simulation environment to evaluate the predictor for downstream decision-making. This environment enables us to calculate a "Dynamic ADE/FDE" while the ego-agent operates with the specific predictor, thus, mitigating the dynamics gap. We demonstrate a strong correlation between Dynamic ADE/FDE and driving performance

*Equal contribution.

Preprint. Under review.

- Motivation
- Problem
- Method
- Result
- Conclusion

In the autonomous driving system, trajectory prediction plays a vital role in ensuring safety and facilitating smooth navigation. However, we observe a substantial discrepancy between the accuracy of predictors on fixed datasets and their driving performance when used in downstream tasks. This discrepancy arises from two overlooked factors in the current evaluation protocols of trajectory prediction: 1) the dynamics gap between the dataset and real driving scenario; and 2) the computational efficiency of predictors. In real-world scenarios, prediction algorithms influence the behavior of autonomous vehicles, which, in turn, alter the behaviors of other agents on the road. This interaction results in predictor-specific dynamics that directly impact prediction results. As other agents' responses are predetermined on datasets, a significant dynamics gap arises between evaluations conducted on fixed datasets and actual driving scenarios. Furthermore, focusing solely on accuracy fails to address the demand for computational efficiency, which is critical for the real-time response required by the autonomous driving system. Therefore, in this paper, we demonstrate that an interactive, task-driven evaluation approach for trajectory prediction is crucial to reflect its efficacy for autonomous driving.

Research Gap

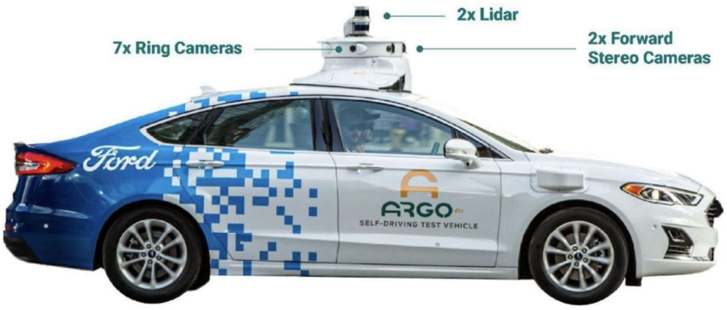
Lack of study/insufficient study

Limitation of previous research

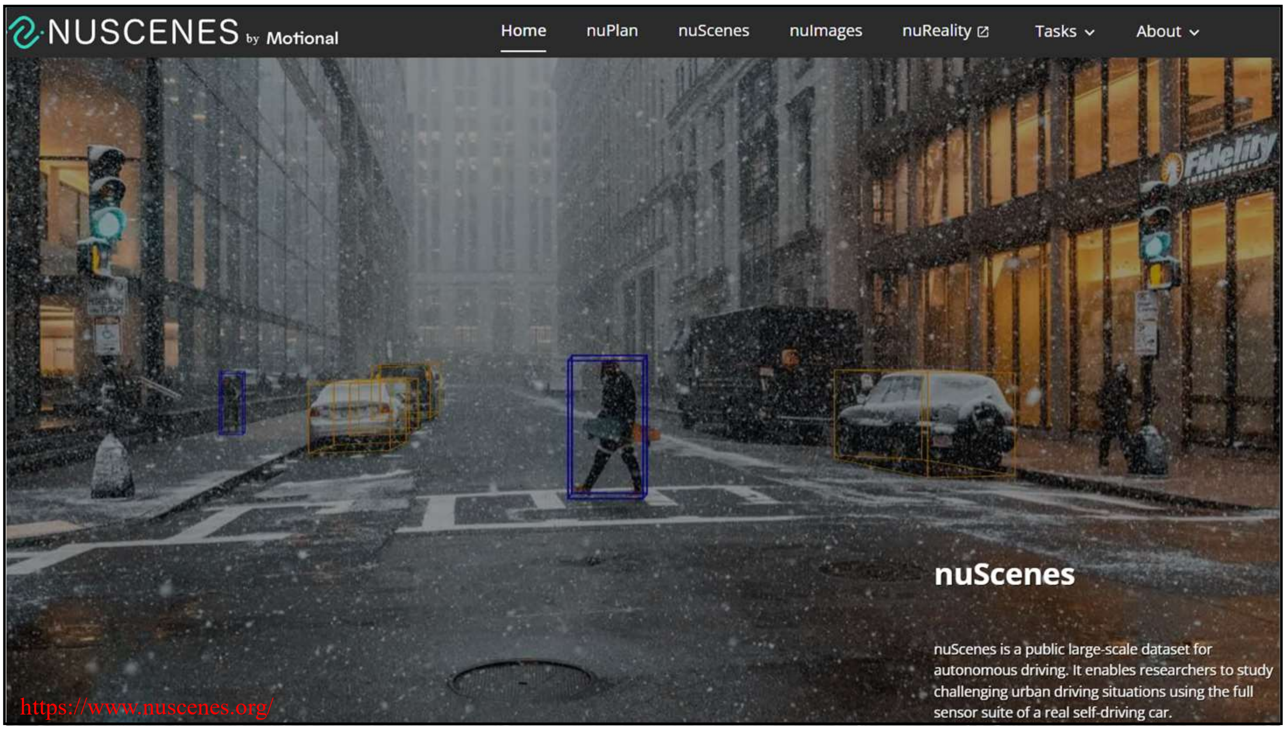
Research Gap of Current Trajectory Prediction

Dataset Accuracy	Prediction Capability
Accuracy on fixed datasets	Driving performance in real-world tasks

Argoverse

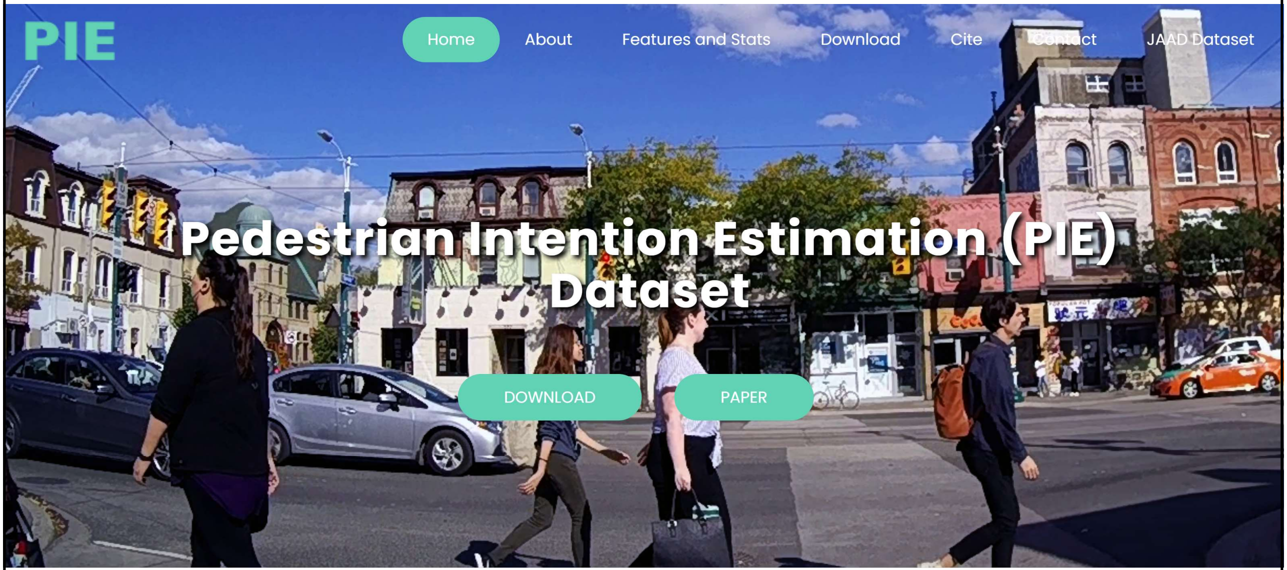


https://www.argoverse.org/index.html



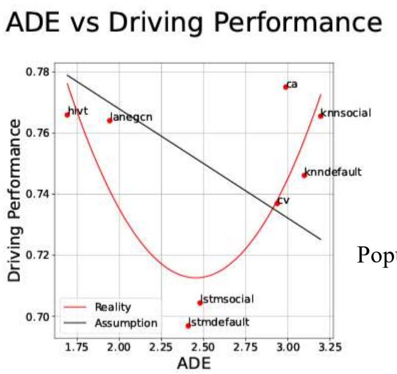
https://www.nuscenes.org/

JAAD

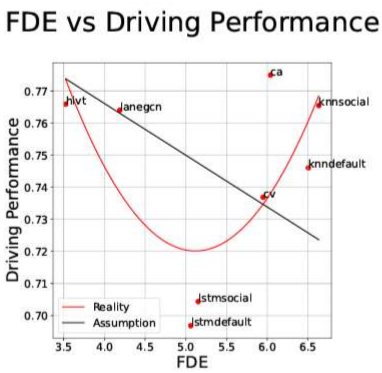


https://data.nvision2.eecs.yorku.ca/PIE_dataset/

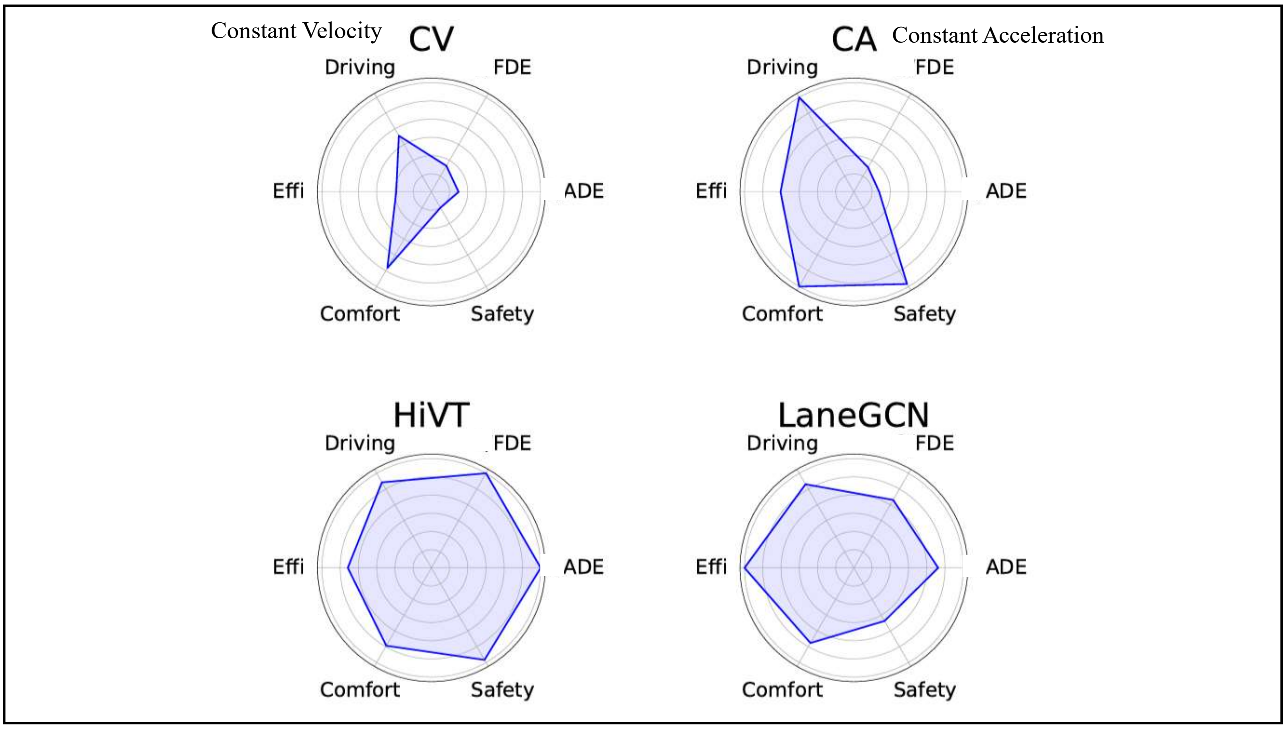
No strong correlation between current prediction evaluation metrics and real-driving performance

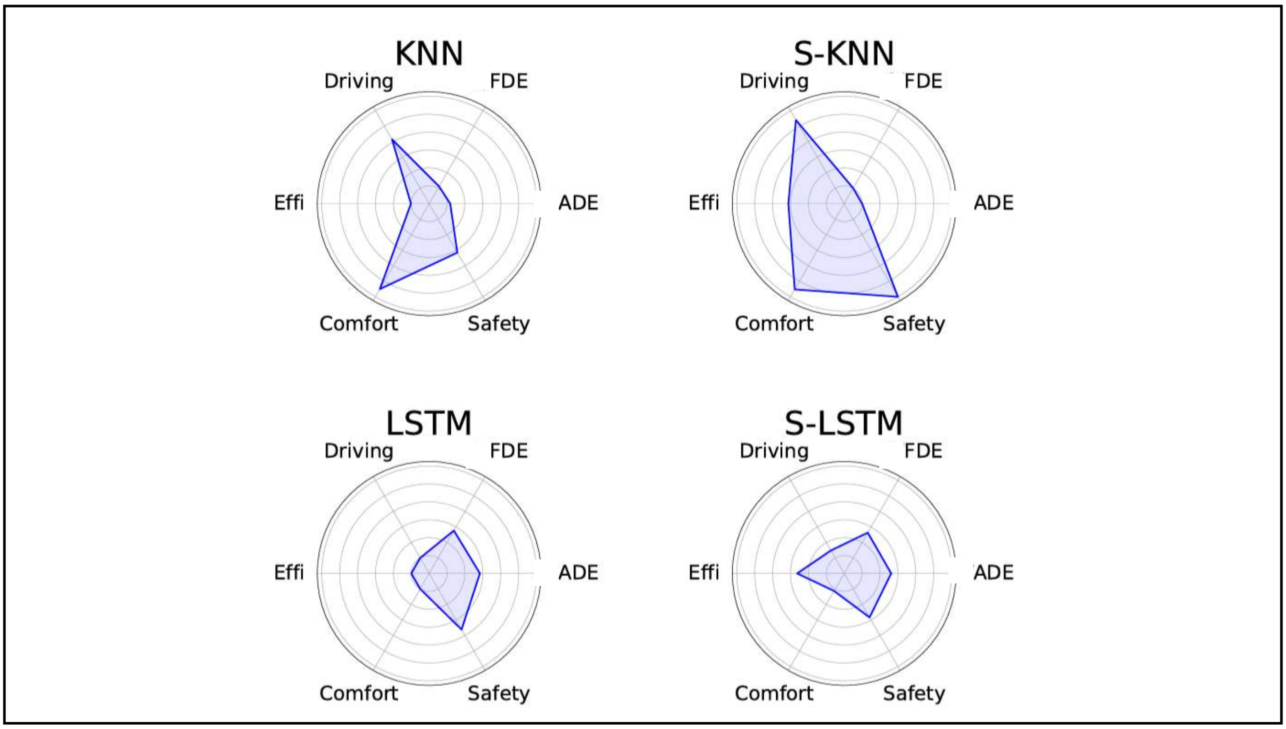


Popular belief (BlackCurve)



Real driving performance (RedCurve)


Constant Velocity
CV
Driving
FDE
ADE
Effi
Comfort
Safety
CA
Constant Acceleration
Driving
FDE
ADE
Effi
Comfort
Safety
HiVT
Driving
FDE
ADE
Effi
Comfort
Safety
LaneGCN
Driving
FDE
ADE
Effi
Comfort
Safety


KNN
Driving
FDE
ADE
Effi
Comfort
Safety
S-KNN
Driving
FDE
ADE
Effi
Comfort
Safety
LSTM
Driving
FDE
ADE
Effi
Comfort
Safety
S-LSTM
Driving
FDE
ADE
Effi
Comfort
Safety

Driving Performance: Two Overlooked Factors

- Dynamic gap between the dataset and real driving scenario
- Computational efficiency of prediction models (predictors)
  - Focus only on accuracy fails to address computational efficiency
  - Balance between computational efficiency and prediction accuracy

Dynamic ADE/FDE

- An **interactive** simulation environment
- Strong correlation between Dynamic ADE/FDE and driving performance

Two Issues

- Limitation of current trajectory prediction evaluation methods
- Task-driven interactive evaluation metrics
  - Effective way to evaluate prediction models for AD by considering dynamics gap

Related Work

- Modeling Approach
  - Physics-based
  - Learning-based
- Output Type
  - Intention
  - Single trajectory
  - Multi-trajectory
  - Occupancy map
- Situational Awareness
  - Unawareness
  - Interaction
  - Scene
  - Map awareness

Ability to incorporate environmental info → Collision avoidance
Efficient driving

Experiment

- 4 model-based, 6 learning-based models with varying output types and situational awareness

Table 1: Selected prediction methods.

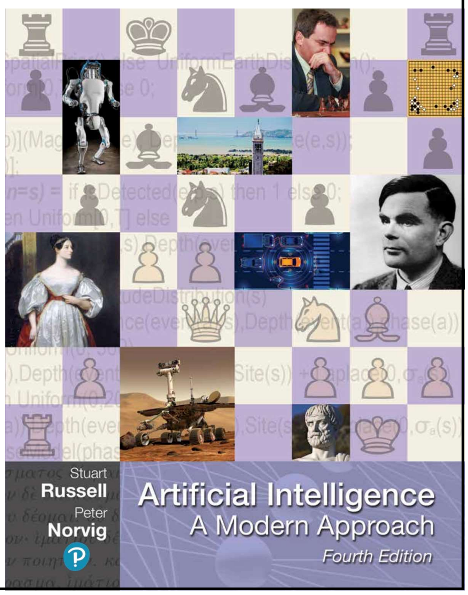
Modeling Approach	Method	Output Type	Interaction Aware	Scene Aware	Map Aware
Model-based	CV [20]	ST	✗	✗	✗
	CA [20]	ST	✗	✗	✗
	KNN [5]	MT	✗	✗	✗
	S-KNN [5]	MT	✓	✗	✗
Data-driven	LSTM	ST	✗	✗	✗
	S-LSTM [1]	ST	✓	✗	✗
	HiVT [27]	MT	✓	✓	✓
	LaneGCN [14]	MT	✓	✓	✓
	HOME [8]	OM	✓	✓	✓
	DSP [26]	MT	✓	✓	✓

RNN
Transformer
GNN
CNN

*Abbreviations: **ST**: Single-Trajectory, **MT**: Multi-Trajectory, **OM**: Occupancy Map.

Planning with Motion Prediction

- State
- Action
- Transition Function
- Planner
- Object function



https://github.com/AdaCompNUS/summit

Simulator

What's SUMMIT?

(a) Singapore-Highway (b) Magic-Roundabout (c) Meskel-Intersection

(d) Singapore-Highway in SUMMIT (e) Magic-Roundabout in SUMMIT (f) Meskel-Intersection in SUMMIT

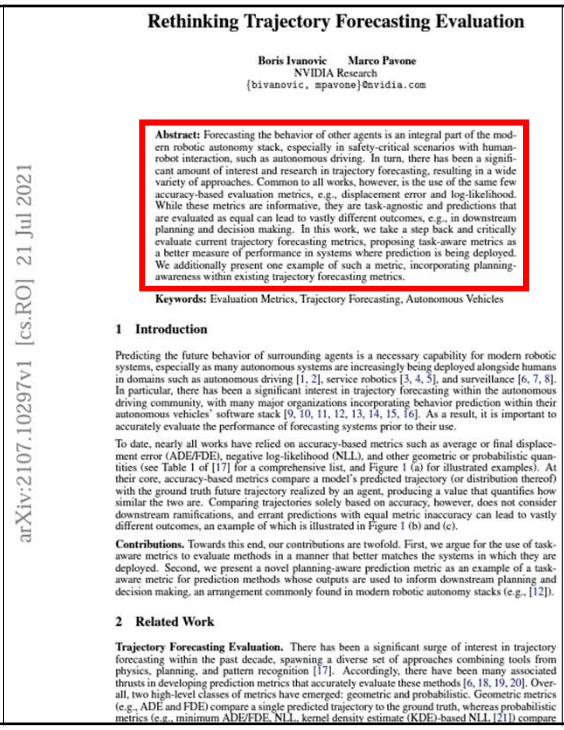
SUMMIT (Simulator for Urban Driving in Massive Mixed Traffic) is an open-source simulator with a focus on generating high-fidelity, interactive data for unregulated, dense urban traffic on complex real-world maps. It works with map data in the form of OSM files and SUMO networks to generate crowds of heterogeneous traffic agents with sophisticated and realistic unregulated behaviors. SUMMIT can work with map data fetched from online sources, providing a virtually unlimited source of complex environments.

Evaluation

- Motion Prediction Performance
  - ADE/FDE, minADE/minFDE
- Driving Performance
  - Safety
  - Comfort
  - Efficiency

Conclusion

- Dynamic Gap
- Computational efficiency of prediction models
- Task-driven interactive evaluation



arXiv:2107.10297v1 [cs.RO] 21 Jul 2021

Rethinking Trajectory Forecasting Evaluation

Boris Ivanovic **Marco Pavone**
NVIDIA Research
{bivanovic, mpavone}@nvidia.com

Abstract: Forecasting the behavior of other agents is an integral part of the modern robotic autonomy stack, especially in safety-critical scenarios with human-robot interaction, such as autonomous driving. In turn, there has been a significant amount of interest and research in trajectory forecasting, resulting in a wide variety of approaches. Common to all works, however, is the use of the same few accuracy-based evaluation metrics, e.g., displacement error and log-likelihood. While these metrics are informative, they are task-agnostic and predictions that are evaluated as equal can lead to vastly different outcomes, e.g., in downstream planning and decision making. In this work, we take a step back and critically evaluate current trajectory forecasting metrics, proposing task-aware metrics as a better measure of performance in systems where prediction is being deployed. We additionally present one example of such a metric, incorporating planning-awareness within existing trajectory forecasting metrics.

Keywords: Evaluation Metrics, Trajectory Forecasting, Autonomous Vehicles

1 Introduction

Predicting the future behavior of surrounding agents is a necessary capability for modern robotic systems, especially as many autonomous systems are increasingly being deployed alongside humans in domains such as autonomous driving [1, 2], service robotics [3, 4, 5], and surveillance [6, 7, 8]. In particular, there has been a significant interest in trajectory forecasting within the autonomous driving community, with many major organizations incorporating behavior prediction within their autonomous vehicles' software stack [9, 10, 11, 12, 13, 14, 15, 16]. As a result, it is important to accurately evaluate the performance of forecasting systems prior to their use.

To date, nearly all works have relied on accuracy-based metrics such as average or final displacement error (ADE/FDE), negative log-likelihood (NLL), and other geometric or probabilistic quantities (see Table 1 of [17] for a comprehensive list, and Figure 1 (a) for illustrated examples). At their core, accuracy-based metrics compare a model's predicted trajectory (or distribution thereof) with the ground truth future trajectory realized by an agent, producing a value that quantifies how similar the two are. Comparing trajectories solely based on accuracy, however, does not consider downstream ramifications, and errant predictions with equal metric inaccuracy can lead to vastly different outcomes, an example of which is illustrated in Figure 1 (b) and (c).

Contributions. Towards this end, our contributions are twofold. First, we argue for the use of task-aware metrics to evaluate methods in a manner that better matches the systems in which they are deployed. Second, we present a novel planning-aware prediction metric as an example of a task-aware metric for prediction methods whose outputs are used to inform downstream planning and decision making, an arrangement commonly found in modern robotic autonomy stacks (e.g., [12]).

2 Related Work

Trajectory Forecasting Evaluation. There has been a significant surge of interest in trajectory forecasting within the past decade, spawning a diverse set of approaches combining tools from physics, planning, and pattern recognition [17]. Accordingly, there have been many associated thrusts in developing prediction metrics that accurately evaluate these methods [6, 18, 19, 20]. Overall, two high-level classes of metrics have emerged: geometric and probabilistic. Geometric metrics (e.g., ADE and FDE) compare a single predicted trajectory to the ground truth, whereas probabilistic metrics (e.g., minimum ADE/FDE, NLL, kernel density estimate (KDE)-based NLL [21]) compare

- Motivation
- Problem
- Method
- Result
- Conclusion

Forecasting the behavior of other agents is an integral part of the modern robotic autonomy stack, especially in safety-critical scenarios with human-robot interaction, such as autonomous driving. In turn, there has been a significant amount of interest and research in trajectory forecasting, resulting in a wide variety of approaches. Common to all works, however, is the use of the same few accuracy-based evaluation metrics, e.g., displacement error and log-likelihood. While these metrics are informative, they are task-agnostic and predictions that are evaluated as equal can lead to vastly different outcomes, e.g., in downstream planning and decision-making. In this work, we take a step back and critically evaluate current trajectory forecasting metrics, proposing task-aware metrics as a better measure of performance in systems where prediction is being deployed. We additionally present one example of such a metric, incorporating planning awareness within existing trajectory forecasting metrics.

Autonomy Stack

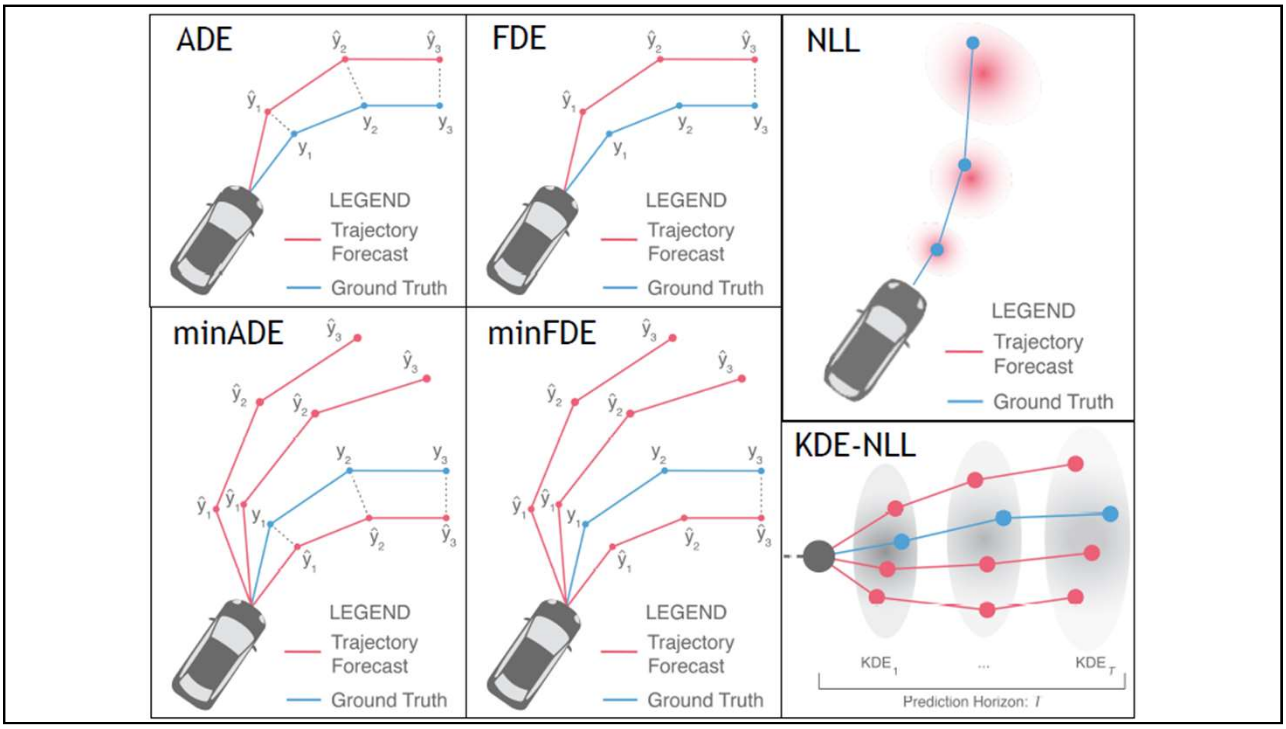
- Detection, Tracking, Prediction, Planning

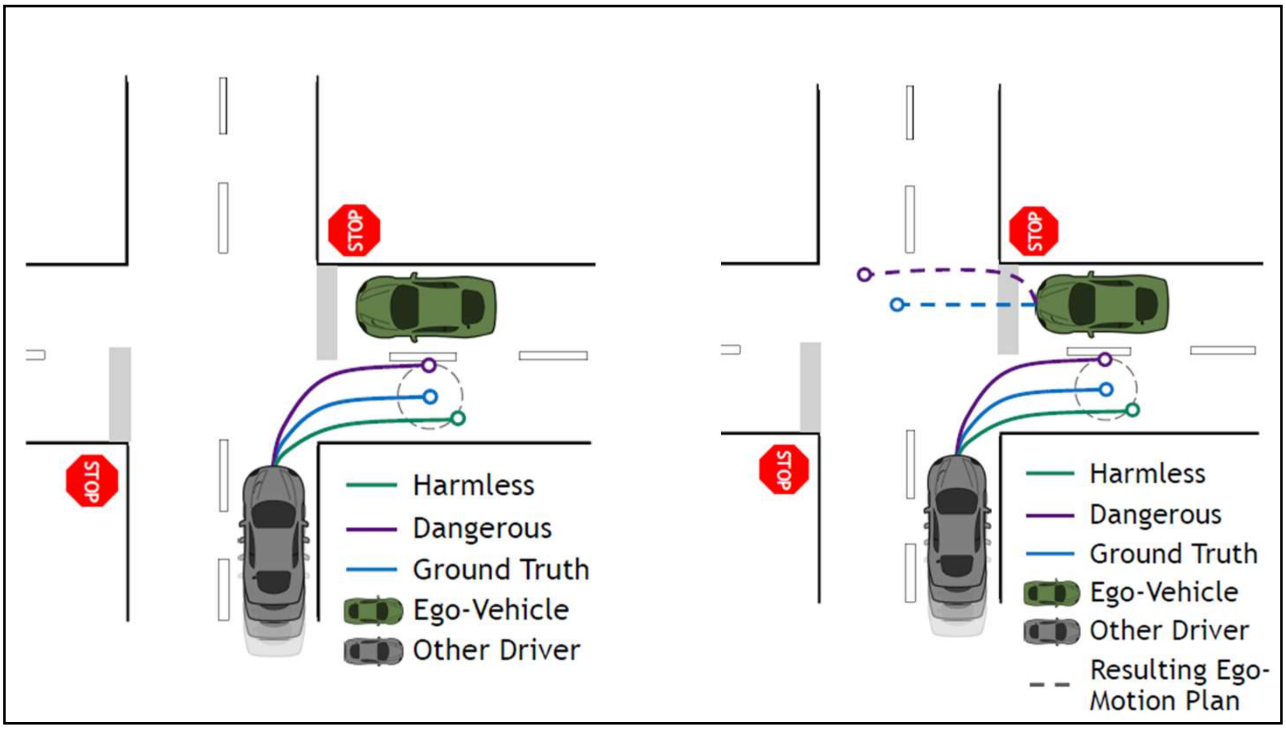
Contributions

- Proof of Concept
- Task-aware metrics
- Planning-aware prediction metric

Related Work

- ADE/FDE
- minADE/minFDE
- Negative log-likelihood (NLL), Kernel density estimate (KDE)-based NLL


ADE
FDE
NLL
minADE
minFDE
KDE-NLL
LEGEND
Trajectory Forecast
Ground Truth
KDE_1
...
KDE_T
Prediction Horizon: T


STOP
Harmless
Dangerous
Ground Truth
Ego-Vehicle
Other Driver
Resulting Ego-Motion Plan

Research Gap

- Existing metrics evaluate the performance of trajectory forecasting methods in isolation.
- Handling perception uncertainty
- Integrating prediction and planning

a closed-loop

- Prediction errors are asymmetric
  - Predictions with the same metric accuracy may lead to vastly different outcomes,

Related Work

- Task-aware Metrics
  - Planning KL Divergence (PKL)

[30] J. Philion, A. Kar, and S. Fidler. Learning to evaluate perception models using planner-centric metrics. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2020.

Proposed Task-aware Prediction Metrics

- Able to capture asymmetries in downstream tasks
  - Weigh prediction accuracies based on planning influence
  - Learn planning cost function
- Task-aware and method agnostic

planning-informed (PI)

$$\boxed{\text{PI-Metric} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} f(a, |\nabla_{\hat{\mathbf{s}}^{(t:T)}} c|) \cdot \text{Metric}(\hat{\mathbf{s}}_a^{(t:T)}, \mathbf{s}_a^{(t:T)})}$$

- Computational feasible
- Interpretable

Datasets

nuScenes, 2019

Lyft, 2020

Waymo, 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, 2019.

J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska. One thousand and one hours: Self-driving motion prediction dataset. In *Conf. on Robot Learning*, 2020.

One Thousand and One Hours: Self-driving Motion Prediction Dataset

John Houston Guido Zuidhof Luca Bergamini Yawei Ye
Long Chen Ashesh Jain Sammy Omari Vladimir Iglovikov Peter Ondruska
Lyft Level 5
level5data@lyft.com

arXiv:2006.14480v2 [cs.CV] 16 Nov 2020

Figure 1: An overview of the released dataset for motion modelling, consisting of 1,118 hours of recorded self-driving perception data on a route spanning 6.8 miles between the train station and the office (red). The examples on the bottom-left show released scenes on top of the high-definition semantic map that capture road geometries and the aerial view of the area.

Abstract: Motivated by the impact of large-scale datasets on ML systems we present the largest self-driving dataset for motion prediction to date, containing over 1,000 hours of data. This was collected by a fleet of 20 autonomous vehicles along a fixed route in Palo Alto, California, over a four-month period. It consists of 170,000 scenes, where each scene is 25 seconds long and captures the perception output of the self-driving system, which encodes the precise positions and motions of nearby vehicles, cyclists, and pedestrians over time. On top of this, the dataset contains a high-definition semantic map with 15,242 labelled elements and a high-definition aerial view over the area. We show that using a dataset of this size dramatically improves performance for key self-driving problems. Combined with the provided software kit, this collection forms the largest and most detailed dataset to date for the development of self-driving machine learning tasks, such as motion forecasting, motion planning and simulation.

Keywords: Dataset, Self-driving, Motion prediction

Name	Size	Scenes	Map	Annotations	Task
KITTI [1]	6h	50	None	3D bounding boxes	Perception
Oxford RobotCar [8]	71h	100	None	-	Perception
Waymo Open Dataset [9]	10h	1000	None	3D bounding boxes	Perception
ApolloScape Scene Parsing [10]	2h	-	None	3D bounding boxes	Perception
Argoverse 3D Tracking v1.1 [2]	1h	113	Lane center lines, lane connectivity	3D bounding boxes	Perception
Lyft Perception Dataset [3]	2.5h	366	Rasterised road geometry	3D bounding boxes	Perception
nuScenes [11]	6h	1000	Rasterised road geometry	3D bounding boxes, trajectories	Perception, Prediction
ApolloScape Trajectory [12]	2h	103	None	Trajectories	Prediction
Argoverse Forecasting v1.1 [2]	320h	324k	Lane center lines, lane connectivity	Trajectories	Prediction
Ours	1,118h	170k	Road geometry, aerial map, crosswalks, traffic lights state, ...	Trajectories	Prediction, Planning

Table 1: A comparison of various self-driving datasets available today. Our dataset surpasses all others in terms of size, as well as level of detail of the semantic map (see Section 3).

P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, S. Zhao, S. Cheng, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2020.

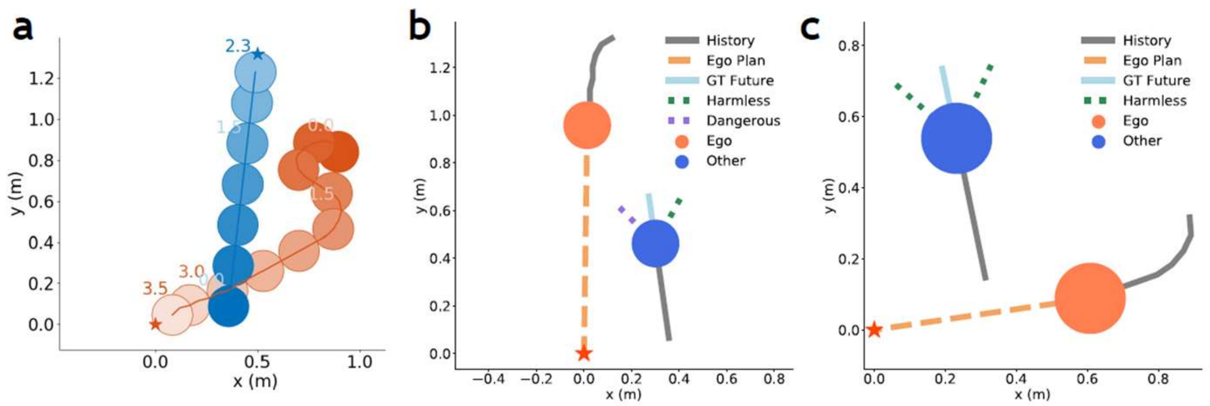


Figure 2: (**a**). An ego-vehicle (orange) maneuvers to the origin while avoiding other agents (blue), lighter colors occur later. Importantly, our method is able to distinguish between metrically-equal errant predictions in a planning-aware manner. In a head-on scenario (**b**), planning sensitivities are much higher for predictions that veer into the ego-vehicle's path (purple dashed) compared to those that steer away (green dashed). Further, when it is unlikely that an agent would influence the ego-vehicle's plan (**c**), our method yields small planning sensitivities for all predictions.

Conclusion

- Task awareness can be injected into existing metrics
- Enabling task-aware evaluation for other components (detection, tracking)

CAAI TRANSACTIONS ON INTELLIGENCE TECHNOLOGY, 2023 1

Towards trustworthy multi-modal motion prediction: Holistic evaluation and interpretability of outputs

Sandra Carrasco Limeros[*†‡], Sylwia Majchrowska[†‡], Joakim Johnander[‡§], Christoffer Petersson[‡¶],
Miguel Ángel Sotelo[*], and David Fernández Llorca[*‖]
[*]Computer Engineering Department, Polytechnic School, University of Alcala, Madrid, Spain
[†]AI Sweden, Göteborg, Sweden
[‡]Zenseact AB, Göteborg, Sweden
[§]Department of Electrical Engineering, Linköping University, Linköping, Sweden
[¶]Chalmers University of Technology, Göteborg, Sweden
[‖]European Commission, Joint Research Centre, Seville, Spain
sandra.carrascol@uah.es, sylwia.majchrowska@ai.se, joakim.johnander@zenseact.com,
christoffer.petersson@zenseact.com, miguel.sotelo@uah.es, david.fernandez-llorca@ec.europa.eu

Abstract

Predicting the motion of other road agents enables autonomous vehicles to perform safe and efficient path planning. This task is very complex, as the behaviour of road agents depends on many factors and the number of possible future trajectories can be considerable (multi-modal). Most prior approaches proposed to address multi-modal motion prediction are based on complex machine learning systems that have limited interpretability. Moreover, the metrics used in current benchmarks do not evaluate all aspects of the problem, such as the diversity and admissibility of the output. In this work, we aim to advance towards the design of trustworthy motion prediction systems, based on some of the requirements for the design of Trustworthy Artificial Intelligence. We focus on evaluation criteria, robustness, and interpretability of outputs. First, we comprehensively analyse the evaluation metrics, identify the main gaps of current benchmarks, and propose a new holistic evaluation framework. We then introduce a method for the assessment of spatial and temporal robustness by simulating noise in the perception system. To enhance the interpretability of the outputs and generate more balanced results in the proposed evaluation framework, we propose an intent prediction layer that can be attached to multi-modal motion prediction models. The effectiveness of this approach is assessed through a survey that explores different elements in the visualization of the multi-modal trajectories and intentions. The proposed approach and findings make a significant contribution to the development of trustworthy motion prediction systems for autonomous vehicles, advancing the field towards greater safety and reliability.

Index Terms

Autonomous vehicles, multi-modal motion prediction, evaluation, robustness, interpretability, trustworthy AI.

I. Introduction

The ability of human drivers to predict the motion of other road agents allows us to anticipate potentially dangerous situations and take preventive actions to minimise safety risks. It also allows humans to perform more efficient and comfortable maneuvers. It is therefore important that autonomous vehicles also have the capability to predict the motion of other road agents, so that they can apply predictive planning approaches and therefore behave in a more human-like manner.

However, predicting future actions and motions of traffic participants is a very complex task, as the behaviour of road agents is influenced by many different variables and interactions [1], [2]. Furthermore, despite the fact that traffic environments are well structured (e.g. street layout, traffic rules), the number of possible future trajectories for each past trajectory for each agent can be considerable, whether for pedestrians, cyclists or vehicles. That is, the problem is multi-modal in nature.

In order to handle this complexity, most of the computational approaches proposed to address multi-modal motion prediction rely on very complex machine learning models which are far from being interpretable. These models are not at human scale and suffer from the characteristic of opacity (i.e., black-box models). Besides, there is no consensus on the most important

- Motivation
- Problem
- Method
- Result
- Conclusion

Predicting the motion of other road agents enables autonomous vehicles to perform safe and efficient path planning. This task is very complex, as the behavior of road agents depends on many factors and the number of possible future trajectories can be considerable (multi-modal). Most prior approaches proposed to address multi-modal motion prediction are based on complex machine learning systems that have limited interpretability. Moreover, the metrics used in current benchmarks do not evaluate all aspects of the problem, such as the diversity and admissibility of the output. In this work, we aim to advance towards the design of trustworthy motion prediction systems, based on some of the requirements for the design of Trustworthy Artificial Intelligence. We focus on evaluation criteria, robustness, and interpretability of outputs. First, we comprehensively analyze the evaluation metrics, identify the main gaps of current benchmarks, and propose a new holistic evaluation framework. We then introduce a method for the assessment of spatial and temporal robustness by simulating noise in the perception system. To enhance the interpretability of the outputs and generate more balanced results in the proposed evaluation framework, we propose an intent prediction layer that can be attached to multi-modal motion prediction models. The effectiveness of this approach is assessed through a survey that explores different elements in the visualization of the multi-modal trajectories and intentions. The proposed approach and findings make a significant contribution to the development of trustworthy motion prediction systems for autonomous vehicles, advancing the field towards greater safety and reliability.