Deep Learning-based Behavior Prediction for Enhanced Safety in Autonomous Vehicle Environments

By Dr. Małgorzata Pioro-Mianowska

Associate Professor of Computer Science, AGH University of Science and Technology, Poland

1. Introduction

By doing so, vehicles can proactively communicate the state of the current environment to the vehicle's systems, to the driver, to the surrounding infrastructure, and to other vehicles. Predicting the state of the environment, possibly including the intentions of the actors and objects within it, has the potential for significant advancements in enhancing the safety of transportation systems. Collaborative initiatives among auto manufacturers, governments, and academic institutions coordinate activities related to the development, design, and operation of ADAS.

Traditional safety systems, such as vehicle airbags, have been designed to respond to the intrusion of an external stimulus rather than to predict and prevent an impending event. Furthermore, by the time other advanced driver-assistance system (ADAS) sensors are triggered, it is normally too late to reverse the series of actions that led to the crash. Vehicle safety can benefit from real-time information on the state of its environment outside the vehicle and the intended goals of agents within that environment. Thus, the design and implementation of ADAS in the form of cars communicate predictions about the intentions of the actors like pedestrians, bicycles, motorcyclists, and other drivers.

1.1. Background and Motivation

Current formulations for this task predict the full trajectory of the pedestrian or the driver at a single moment in the future. The same actors are embodied, walking in the same position at different amplitudes and never refusing to predict their movement. Our key motivation, in contrast, is safety: the goal of an autonomous vehicle is not under the risk of hitting a pedestrian or another car, and instead, safety-critical behaviors are identified in advance and avoided by the car. Because the autonomy of the vehicle must anticipate the location of all

This work is licensed under CC BY-NC-SA 4.0.

actors in the roadway and generate safe behaviors to avoid critical situations. This is important; the ability of an autonomous vehicle to not reach an intersection is much more valuable than the ability to make decisions when it really arrives, so it helps to coordinate and adapt the autonomous driving schedules to minimize social costs, such as delays and consumption. Furthermore, by predicting behaviors of agents in the environment, the behavior of the autonomous vehicle can be adjusted to reinforce the confidence of the pedestrians and other drivers, making the perceived better behavior. This, in particular, can lead to greater acceptance of autonomous vehicles, improving the design of safety, the possibilities, and the efficacy to integrate the developments into society.

Autonomous vehicles can lead to a significant reduction in the number of traffic-related accidents. The power of AI, in particular, deep learning, could be a solution for this. Artificial Intelligence (AI)-based methods make it possible for cars to interpret their surroundings, predicting the behaviors of other self-driving cars. Image analysis is also essential to find objects and make driving decisions. Recently, deep learning has made great advances in understanding complex structures in images and videos. As a result, deep learning techniques are now widely used to assist autonomous vehicles and generally consider deep learning methods that only deal with object detection. However, many levels of driving decisions remain beyond the scope of deep learning applications. For instance, predicting the intentions of a pedestrian or another driver and their personal behavior intentions are key realistic contexts. In real applications, these delays may lead to jams, accidents, property damage, or injuries.

1.2. Research Objectives

Hence, as a significant research objective, the models developed within will aid significantly in reducing the overall complexity and improve decision-making, while also reducing computational requirements. It is important to note that the work reported is only the first step in the processing pipeline and still requires modules in addition to the models—yet, this work will address some of the most challenging visual perception and prediction tasks, contribute to the AI systems underlying the next level of autonomous vehicle safety. Moreover, the developed models will also be utilized in the interim motion planning phase of the vehicle. Finally, the models developed will investigate different methods to balance trade-offs between model complexity and prediction accuracy and the ways they can be augmented

This work is licensed under CC BY-NC-SA 4.0.

to include further multimodal predictions. Specifically, the models investigated will include variations of recurrent neural networks, convolutional neural networks, and learning long-lived recurrent models and reinforced that accurately predict the density of predictions while staying reliable.

The primary objective of the research reported in this thesis is to develop a set of novel deep learning-based prediction models that can aid advanced control systems in autonomous vehicles. In particular, the goal is to train models on observed tracked pedestrian trajectories in order to accurately predict a diversely defined set of desired future behaviors in order to create computationally efficient models for use in control systems that guide autonomous vehicles in the overall safety and comfort of these types of situations. To date, while numerous methods based on deep learning have yielded accurate predictions for a subset of these behaviors, the creation of high-dimensional predictions is difficult, largely because of the noisy data and numerous modes that the distributions of these behaviors often contain. Instead, at present it is largely up to advanced decision-making modules that sit downstream to navigate the uncertainties that arise.

2. Autonomous Vehicle Technologies

In addition, automated lane-changing controls enable vehicles to change lanes according to user-defined constraints. The behavior prediction system indicated that an automated vehicle would change lanes. Based on this information, the other drivers around the vehicle could easily understand the vehicle movement intent of the automated vehicle. The proposed prediction system used recurrent neural networks (RNN) to predict driver lane-changing behavior. The predicted path was made visible to other drivers by using an LED rear light panel. The display system enabled clear communication of the vehicle's decision about when to move in and when to move out of the center safety envelope. With these advanced driver assistance systems, accident rate reduction results were obtained.

Advanced driver assistance systems (ADAS) are safety systems designed to automate, adapt, and enhance vehicle systems for safety and better driving. In this study, several advanced driver assistance systems are proposed for enhancing driving safety. A Vision-Based Turn-Left Assistant (VBTLA) system capable of interacting with pedestrians, cyclists, or other vehicle drivers was proposed. The problem was formulated as an image region detection and tracking problem. A convolutional neural network (CNN) was designed to predict the nearby

This work is licensed under CC BY-NC-SA 4.0.

surrounding environment and to fit user-defined constraints for acceptable waiting time. If the wait time for making a left turn exceeds the limit, the system notifies the traffic control system to request additional time. If additional waiting time is not feasible, the system generates alternative routes.

2.1. Sensor Fusion

At every step, filter the 3D points based on camera calibration. Yield only the 3D points that fall inside the FOV and in front of the camera. Assign 3D points with the same transformed pixel the same label. Assign the class and the confidence score of the original detection to these 3D points. Convert the detection in 3D back to 3D boxes. At every step, filter the detections based on LiDAR specification to discard off-limit properties of the scene. Yield only the 3D points that fall inside the FOV and in front of the LiDAR. Filter 3D boxes prior to returning the results in the LiDAR coordinate frame.

3D objects detected by different sensors are usually converted to a common coordinate frame using sensor-specific calibrated extrinsic transforms, as shown in Figure. For such calibration, two sensors with a common field of view are necessary, and objects could be transformed onto the coordinate frame of the same and a common transform with the reference sensor if desired. Detected objects from the overlapping FOV of the LiDAR and the camera are fused using initial object-to-object association by leveraging geometry relationships as an initial step. Detected 3D objects in 2D space are projected from a camera image. Bounding box detections from two sensors' points of view are converted into a labeled 3D point cloud of the scene in LiDAR coordinates. Transform the labeled 3D points to the image plane in camera coordinates. While a 2D/3D deep learning approach could solve this problem, our approach is at least as fast, simpler, and increases the network performance.

2.2. Computer Vision

Computer vision is the field of study mainly in computer science that focuses on replicating the behavior of human vision into a computer system. The human vision system is usually defined as the complex process of constructing and interpreting visual intelligence to understand the information gained from the world surrounding us. This is a field that can enable a computer to understand and process an image taken using an image sensor, which is a highly fundamental aspect of computer vision implementations. Since image sensory

This work is licensed under CC BY-NC-SA 4.0.

devices often achieve higher resolutions, even less expensive image sensors are capable of taking highly detailed photos of humans, animals, or objects. The amount of information that can be captured from a digital image is significantly more when it is assessed compared to most other biological detecting receptive fields. The image file can indeed attain very large file sizes that are representative of raw digital sensor data if it fails to undergo compression. This is especially true for images that contain very high resolutions or numerous colors. In the last couple of decades, more precise details have been derived from computer vision data, making it a significant expertise in the fields of robotics and image processing.

3. Behavior Prediction in Autonomous Vehicles

This chapter is organized as follows. The need for a high-level social model of drivers in motorway scenarios at present is outlined, suggesting the contribution of deep learning to creating such a model. The target and the conjecture in the work, the difficulties and open questions are set out just before the next section reviews the research literature. Possible architectures are presented in the subsequent section. Such architectures draw upon what is novel in the approach, a case-base of motorway driving behaviors created from asking real drivers by interview how they determine their social behavior over a selection of 4,000 real-life behaviors observed during two half-hours of driving each. A simple case-based approach is then described. Analysis of the results indicates improvements from using nearest neighbor methods, and finally, a deep learning example using a locally recurrent model. Those results, coupled with the need for that local recurrence to 'see' the circumstances whereby drivers derogate, will be the hook-up for that section and local recurrence. The next section is problems, difficulties, and directions for future research.

A crucial feature of driver assistance systems is triggering computers that assess the core driving skill decisions, including overtaking, lane-changing, steering, and maybe taking over control of the vehicle. Abiding by the given rules of the road from a human perspective is considered appropriate social behavior. Finding a social model for the motorway scenario is a requirement for the acceptance of any robotics system. Such social models have been created in many areas of activity, and the specificity of the motorway domain and the need for amendments are well recognized. While several areas of behavioral modeling require techniques from many human science senses or action or physical behavior modeling, which could use perspective human vision techniques, social driver support modeling requires

This work is licensed under CC BY-NC-SA 4.0.

neither. Awareness remains the single casualty of ITS. We could prevail if the systems were able to learn from observing drivers during rides. Sociological studies or soft system methodologies, including participative techniques, include observations of human action. Applied informatics denotes techniques of observing through human vision, such as video cameras and video vans.

3.1. Traditional Approaches

Beyond the traditional methods, several other deep learning-based behavior prediction works have made anchoring in vehicle driving datasets. Most of those used Long Short-Term Memory (LSTM) or more recent Transformer Networks (TNN). These methods, although requiring large amounts of data for training, are able to automatically and more accurately learn the relationships between the data and the behavior. Our work in this paper takes most of its inspiration from the recent works that have utilized deep learning techniques to learn and predict the nuanced behaviors in vehicle driving corridors.

We segment behavior prediction approaches into two broad categories: traditional methods and deep learning methods. Traditional methods predate deep learning and often rely on elaborate access to spatial features of agents to extract spatial-temporal features or relationships to accurately predict behaviors. These methods are often computationally expensive and are sometimes confusing to design and implement. These are the main disadvantages of models based on traditional methods. Validation of the different relationships between parameters and behavior is particularly difficult and time-consuming, and it is often not guaranteed that the models will generalize well outside of their training scenario.

3.2. Deep Learning Techniques

Indeed, techniques for uncertainty quantification, such as probabilistic and deep generative models, represent important future steps in model development for autonomous driving perception tasks. One successful example of this approach is the dropout technique applied to each unit of the network, which represents Bayesian model averaging of the network for point predictions. Ren et al. successfully employed this technique to estimate the uncertainty on ground-segmentation tasks, follower position estimation, and motion planning prediction in autonomous driving systems. Their technique demonstrated superior results, especially

This work is licensed under CC BY-NC-SA 4.0.

when used for driving planning tasks. In contrast to standard deep generative models that have a limited representation of multimodal predictions, sequence-to-sequence models, such as RNNs, can be effectively adapted to generate multiple hypotheses and thus can be used for obtaining multimodal traffic predictions. In general, all the deep learning techniques mentioned here aim to learn explicit representations and perform efficient and non-linear mapping of input data.

Deep Boltzmann Machines (DBMs) represent a unique approach within the realm of deep learning algorithms. DBMs are multilayered neural networks that consist of multiple layers of stochastic binary units. Although the training of DBMs is slow and challenging, this network architecture has a greater capacity to learn complicated structured data such as those found in vision and speech signal problems. These networks exhibit efficient sample estimation, and they can be trained with or without labels. Recently, Variational Autoencoders (VAEs) have been proposed for learning deep generative models. These algorithms aim to approximate full Bayesian inference, which is intractable because of high-dimensional models. VAEs have a simple architecture that is easy to train, and they exhibit strong generative performance on various complex datasets. Unlike standard deep learning models that are only capable of providing point predictions, RNNs and their potentials can be explored for training non-parametric models for spatial-temporal traffic prediction tasks.

We now identify some of the many emerging deep learning algorithms currently being applied in a diverse set of prediction problems. Convolutional Neural Networks (CNNs) are currently receiving much attention as they achieve superior performance in specific types of computer vision problems. These networks contain multiple convolutional layers to build local feature maps, and they achieve satisfactory performance in relation to translation invariance. CNNs have been proposed for object recognition tasks in the literature and also show promise on a variety of other problems such as human action recognition. Recurrent Neural Networks (RNNs) are also an increasingly popular deep learning technique. Techniques that build on RNNs are the Long Short-Term Memory (LSTM) architecture and the Gated Recurrent Unit (GRU) architecture.

4. Explainable AI in Autonomous Vehicles

In this context, the task of behavior prediction concerns the estimation of the future motion of the actor remaining present when building its driving policy. Occupants of an autonomous

This work is licensed under CC BY-NC-SA 4.0.

vehicle may be getting ready to alight, pedestrians and cyclists have their own behaviors, including looking to be considered, and other vehicles may be about to change lanes or turn at intersections. Additionally, any prediction method used by autonomous vehicle applications must guarantee a high degree of safety. Such systems are working as safety drivers in a wide variety of cases (from traffic situations almost completely under control to extremely critical ones) and provide input to the system's components regarding the planning of future movements and acceleration/deceleration.

Automotive artificial intelligence (AI) is expected to make significant contributions to the development of advanced driver assistance systems (ADAS) and autonomous vehicles. An important task is to predict the driving states of vehicles to enhance safety and reliability. The driving decisions of autonomous vehicles aim to achieve advanced capabilities, combining cautiousness with acceptable speed. Therefore, understanding the system's decisions and predictions is of high importance for the safety assurance of autonomous driving systems and is considered a necessary property for a wide variety of applications as well.

4.1. Importance of Explainability

Although deep learning has achieved unprecedented results and established itself as the current paradigm for any problem that can be described by labeled data, this black-box characteristic has contributed to the assumption that methods of this kind are difficult to interpret. When considering resources that are critical in decision-making such as self-driving vehicles, whether on small- or large-scale tasks, this characteristic makes people distrust models that we should trust, but apparently cannot. In the context of self-driving vehicle behavior parsing, explainability is an essential feature since clearly understanding the models we rely on is crucial for the decision-making process of all underlying software pipelines.

Even though relatively simple machine learning models may be used to solve certain problems satisfactorily, their explanation, and therefore the understanding from a human perspective, is usually only possible for a limited number of concepts depending on the complexity of the model. The lack of explainability on artificial intelligence models and their decision-making processes is one of the main drawbacks and certainly one of the main reasons why such models, despite achieving state-of-the-art results and being quite robust, are not yet widely adopted.

This work is licensed under CC BY-NC-SA 4.0.

4.2. Techniques for Explainable AI

Conversely to XAI from the traditional setting of computer vision and natural language processing tasks, there is a lack of incorporated approaches in driver object behavior prediction in dynamic driving scenarios. Currently used data-driven approaches for behavior prediction in the research domain of autonomous driving are based on deep learning (map the input to the output) or RNN (relies on time-series data). While these two types of approaches give promising results, the decision made by the model is a prediction (extrinsic evaluation). In fact, the model itself often contains no explanation why it makes this decision (intrinsic evaluation).

Explainability (XAI) is described in literature as the ability of an intelligent system to explain its decision processes and processes to a human collaborator using human-understandable terms. Major concerns related to AI applications in autonomous vehicles are the demonstrateability and interpretability of AI while selecting complex patterns in many situations of public life. In the case that the predictions from AI do not match the expectations of the targeted user base (human drivers, bystanders...), it is crucial that AI is always able to present a simplified underlying decision process for dispute resolution. The adoption of XAI paves the way for the user to oversee AI effectively.

5. Threat Analysis in Autonomous Vehicle Networks

In this section, traffic analysis in AVEC is closely examined to identify potential sources of disturbance that it may address and validate. We have thoroughly analyzed susceptible adversarial potential threats that may exploit the security of the AVEC network's highly collaborative vehicle capabilities. These threats are also validated through minimal simulation studies. With these risk analysis findings, a thorough overview of the AVEC implementation is directed to guide the proper design of AVEC algorithms and the construction itself, addressing the possible adversarial attempts of the attacking network mentioned. Finally, the flow of these examined results and review proposals are discussed in the conclusion, and final perspectives are provided.

While enhancements in vehicle automation contribute to greater road safety, autonomous vehicles are also more interconnected than traditional cars. They communicatively interact to form networks that require both safety guarantees and additional security measures. The

This work is licensed under CC BY-NC-SA 4.0.

ability of autonomous vehicles to communicate and share data allows individuals to disrupt the network. A motivated attacker could use a vehicle connected in autopilot to isolate it, start a jamming transmitter, or use hacking mechanisms to manipulate the network's functioning. Such disruptive behavior often involves more than one car in a network with multiple vehicles. Prior literature has focused on driving autonomous vehicles within the environment, often in accordance with the U.S. Environmental Protection Agency (EPA). Three relative dynamics levels based on automated driving functions combine Cooperative Adaptive Cruise Control (CACC) and the traffic state in which the AV interacts.

5.1. Types of Threats

There are several types of threats to autonomous vehicles: attackers could try to deceive the car (e.g., with fake traffic signs, faked lines on the road surface, or deceptive clothing), criminals with limited means (e.g., unfriendly individuals with clubs or other weapons) could try to generate dangerous situations where no detected obstacle leads to successful triggering an evasion reaction of the car that does not perform well due to car dynamics. Another threat scenario is that things occur that are never seen by a development team during the lifetime of the car, for example, someone brings a police line up sign that forces the car to recognize it as the one and only correct stop sign. Clearly security measures (e.g., tamper-proof sensors) and redundancy can help in many of these situations, but it will not be possible to anticipate all possible threat scenarios and solve all of them in advance. Hence, there will always be uncertainty in any situation, and it remains an open question as to how the car intelligence should react when some objects in a scene are classified but suddenly some unknown objects appear. Inhibition or slowing down until objects that are unknown are classified correctly might be a good solution sometimes, while hard decisions that are always taken whenever necessary are reasonable as well.

5.2. Challenges in Threat Detection

Anticipating the actions of traffic actors is often referred to as threat detection. Essentially, at almost all times, the vehicle parking's presence or movement along the traveled path is expected for threat detection. Potential collision detections help the motion planner in motion prediction and path planning decisions. Therefore, for robust detection, it should aid with possible motion patterns. These should take the extent of the possibilities into account and not

This work is licensed under CC BY-NC-SA 4.0.

be overly conservative, allowing for collision avoidance decisions to be made in a timely manner.

The complexity of driving situations (e.g., occlusions, non-rigid objects, and various object appearances), high mobility of road actors, and varied spatial arrangements make the task of autonomous vehicle (AV) path planning and decision making challenging. As a consequence, almost all current AV studies overestimate what the AVs can process, represent, learn, and use, regarding the understanding of the dynamic environment.

6. Real-time Analysis and Explanations

Two variant architectures are proposed to mitigate this increased prediction time. In the first, dual state decoder (DSD), the capabilities of the proposed model are broken into two components: the first carrying out the visual encoding in parallel with input structure information, and a second, state encoding, that carries out the rest of the required state processing. The principal advantage of this approach is that the intense parallelism used during inference now applies to the majority, or even all, of the task. The second approach, Behavior Complexity aware Context (Behc), allows the proposed model to have the required complexity for more difficult situations but simplifies the prediction of simpler cases, exploiting faster, more lightweight processing techniques. Small, low-resolution inputs are assumed to be ones which may be used in simple situations and are therefore quickly processed, relying on the minimized size of the network to make faster predictions. Time and monitoring resources can thus be directed towards more complex cases that are likely to need more careful reasoning, achieving lower expected total latency. All data, no matter the likely type, passes through the state encoder as usual. Little extra processing is actually required.

One common critique of the use of deep learning in critical application areas is that they cannot provide timely results. In the context of the proposed model architecture, it is imperative to have not only a reliable prediction, but also one that is delivered within a small enough time frame that it can be used to reliably avoid collisions. Previous work suggests that while the DSD variant of the proposed model takes longer than the equivalent DSD-base model to return a prediction, these predictions are slightly more accurate than those of the DSD base model, and certainly provide the level of accuracy needed for liability and confidence estimates. Specifically, a Behc variant of the proposed model is significantly faster than the corresponding Behc-base mode. In both cases, the uncertainty benefits derived from

This work is licensed under CC BY-NC-SA 4.0.

DSD remain consistent, ensuring that system confidence can remain high in potentially life-threatening situations.

6.1. Real-time Data Processing

There are four categories that should be defined explicitly: (1) Client Application; (2) Pubsp; (3) Data Processing Layer; (4) DataSource DCU. They have descriptions separately as follows: (1) The client application has another format in print, including the feeder and display module. The worker module should have one thread for data publishing, where transporting data to the pubsub layer. The feeder and viewer are binding to the CUDA context for efficient communication to the video decoder and extender. (2) Data collecting unit (pubsp and data preprocessing server, as shown in Fig. 10): The Data Source Device Computing Unit (DCU) should be running as pubsp instances and computation unit. The data computation and collating can be dealt with based on the CUDA context to improve efficiency. They will send all information to a submodule called Data Processing Layer as soon as the data is ready using both OpenCV and GStreamer as the dependent libraries.

The stream data collected from each piece of hardware provides valuable real-time information for our algorithm. Although this information has its fixed format, it might deviate from the initial true value. Usually, these items have a fixed structure, like timestamp, device ID, and dataset. The fields of the collated dataset can be either raw or processed. For the raw data, it is maintained based on the original format. For processed data, it should be computed from real hardware or potentially further processing. A raw example can be found in LCASPub-1393. Thus, for information in each device, it can be any data, like the steps shown in Figure 9. As these can happen in the streaming state, the data needs to be processed in a real-time manner. Please remember that the timestamp is required in controlling the order between data.

6.2. Explanation Generation Techniques

To develop explanations which are controllable and thus usable for downstream tasks, score-based sampling allows generating instances along the convex hull of the training data, which has been compared to other metric-based techniques before. Explanation generation for safety enhancement might be helpful in severe accidents which involve self-learning models. However, in such situations, it might also be necessary to assure that each model in a vehicle

This work is licensed under CC BY-NC-SA 4.0.

fleet makes decisions in a safe way, so low Efact models which reveal the underlying reasoning can be applied to identify learning deficiencies in the models.

Visualizations or summary statistics of the feature importance can be employed for transparent models, which make the reasoning explicit, as shown in the section on decision support. While we have not yet observed explanations for vision models in the area of autonomous driving, literature contains research on these for applications of activity recognition. In such work, those explanation generation techniques which make the reasoning explicit are usually based on feature maps recorded during model development (and corresponding maps of the features for the final prediction), occlusion to attain feature importance, or saliency analysis.

Explanation generation has the goal of explaining actions a model takes when solving a problem. For behavior models, this may mean generating information about the upcoming behavior of other traffic participants, which is of great use in advanced driver assistance systems. Frequently, a good explanation is informative and understandable for the human decision-maker, which means that either the explanation itself actually reveals the reasoning behind an AI model's decision or that the explanation mirrors the decision process in a way that allows the human to understand why a certain decision was made.

7. Case Studies and Applications

Recently, some propose to use deep learning approaches for behavioral prediction for various traffic participants such as pedestrians, bicyclists, and drivers, noting that Zadeh et al. developed dense trajectories and convolutional neural networks (CNNs) to predict the future motion of pedestrians and cyclists in urban traffic. However, our proposed method can generate.

This paper presents a deep learning-based framework for behavioral intention prediction to enhance safety in autonomous vehicle environments. The framework is described alongside its use in two typical case studies, and extensive evaluation results are presented. The proposed method demonstrated high performance, especially when handling multi-modal predictions, and the result can be useful in constructing ethical decision-making mechanisms to optimize decision-making in extreme scenarios. The rapid advance in perception, prediction, and control capabilities has demonstrated promising performance in handling

This work is licensed under CC BY-NC-SA 4.0.

typical traffic scenarios, as a result of powerful neural networks, large-scale data-driven algorithms, and large-scale motion planning systems. However, to efficiently optimize the systems and be able to make fair and ethical decisions in extreme cases, human-level understanding of the behavior or intentions of agents (e.g., drivers, pedestrians, and bicyclists) in complex traffic scenarios is also required.

7.1. Previous Studies on Behavior Prediction

Furthermore, the lateral/cross-/longitudinal behaviors of different agents at different locations are often influenced by distinct information, such as the physical appearances of the agent, the semantic scenes apart from the agent, and the potential interactional significance among the agents and the vehicle. The segmentation-based approaches applied semantic segmentation to the video frames and then used different methods to construct the future motion prediction based on the segmentation maps, time-dependent representations, low-level features, or high-level features. Although promising results were achieved, these methods suffer from a potential misinterpretation to the activation regions within the agent's segmentation map and significant computational cost for real-world applications due to the requirement of running the previously trained deep neural network to infer the segmentation map from each input video frame.

Traditional methods use discrete models that output the future location, speed, and direction of every agent given their past observations; these methods are computationally intensive and their outputs lack the continuity of motions. Researchers trained parametric models to directly predict the motions of the agents using their past and current (or future) coordinates; these models are computationally efficient and are less sensitive to errors in intermediate steps of motion prediction; however, it is difficult for them to capture the multimodal behavior in the future motion since the future motion is non-deterministic conditioned on the present scene context.

For general path and behavior planning of autonomous vehicles, it is often essential for them to be able to predict the future behaviors of the agents in the surrounding environments such as pedestrians and traffic participants. Predicting when and what will happen next in the environment is a prerequisite for autonomous driving, since it allows the vehicle to react proactively and in a way that is comprehensible to other traffic participants.

This work is licensed under CC BY-NC-SA 4.0.

7.2. Applications in Real-world Autonomous Vehicle Systems

Autonomous driving is one of the most pursued technologies that will allow single on-board users, as well as reducing traffic jams and fostering better energy management. However, to be fully integrated into different driving environments, autonomous vehicles (AVs) should understand the road scene and the road user intentions, thus increasing mobility efficiency and safety.

This paper proposes vision-based behavior prediction of other road users in the context of autonomous vehicle environment understanding. A recurrent deep learning architecture is applied to the problem of trajectory prediction, which fuses scene context using an attention mechanism. The proposed methodology is used for an intersection practical scenario where uncertainty is high on the future lane-level maneuver of pedestrians. The future trajectory prediction and the scene understanding in general of road users are jointly validated on real-world datasets under different occlusion levels and make part of a real-world autonomous vehicle cooperating with an infrastructures-to-vehicle developed traffic light assistant.

8. Ethical and Legal Implications

In its own right, computer science should pay attention to the legal and ethical questions posed in creating technology and the principles and values embedded in the products it develops. Our work highlights several social dilemmas inherent in the design of technology (e.g., the choice between maintaining social status quo at the cost of excluding vulnerable populations, including differently abled individuals, or the risk cost of incorrectly placed preferences) and we are likely to see more technologies admitting similar trade-offs in the future. These dilemmas cannot be reduced to technical solutions, but they are very much about technology. Researchers should openly discuss these and related challenges when they become aware of them. Looking forward, perhaps professional codes of conduct for computer scientists should include reflections on what values shape the ideas of computer scientists and how these values can be reconciled with the different norms and expectations of society.

Self-driving cars and the related technology systems, like those described in this work, involve issues that require ethical and legal consideration. These problems have been discussed in detail, and we do not pretend to resolve them at present. However, it is important to recognize what questions current trends in technology pose and to identify ways of addressing these

This work is licensed under CC BY-NC-SA 4.0.

issues. Our perspective comes from computer science, a community that now has a substantial role in shaping the systems that society uses and enough influence to take recognition of its responsibility.

8.1. Privacy Concerns

In essence, to guarantee the development and proliferation of technologies, particularly those designed for beneficial societal impacts, they must ensure privacy in the design, production, and usage of such technologies which has also been observed to be relatively low in terms of available privacy protection practices that the technology follows.

While the developments of ubiquitous technology keep presenting their unique characteristics and contributions to society, particularly in terms of different industries, they also raise some concerns given the privacy invasions noted. Many of such technologies have been found to allow for the leakage of personal and sensitive information into the public domain. Due to this, there have been increasing concerns for the privacy of people with regards to the utilization of such systems. Ensuring privacy is paramount to the development and sustainability of technologies that involve the participation of people. This is because, regardless of the tremendous benefits that these technologies can bring to the table, the failure to implement technologies in ways that protect the privacy of the individuals using them can stunt their usage, leading to reactions from public and societal pressure thus deflating their potential impacts.

8.2. Regulatory Frameworks

This section also discusses regulations and guidelines released from different international and local legislation such as SAE International, United Nations Economic Commission for Europe, and the US National Highway Traffic Safety Administration. The development of fully autonomous vehicles could have so many benefits. It could lead to the reduction of crashes (more than 90% of crashes are due to human errors), lower traffic congestion, and more efficient daily commutes. However, there are many challenges and concerns in developing such technology. These challenges are part of the motivation that converts autonomous vehicles into a hot area of research nowadays. Furthermore, the autonomous vehicles legislation through the entire vehicle operation on the current infrastructure represents a barrier and a non-standard process.

This work is licensed under CC BY-NC-SA 4.0.

The research in this paper could be relevant in different contexts and different application domains. Some of these application domains are governed by a set of general rules and guidelines, e.g. the self-driving vehicle can govern within the context of its transportation management. However, in general, the automotive sector is regulated by codes, guidelines, and standards, which are elaborated by specific bodies, institutions, or task forces. The technical discussion about the frameworks and standards that must rule the deep learning-based scene prediction and behavior prediction in level 3 to 5 of autonomous driving is presented in this section. To visualize the entire environment at all times, autonomous vehicles are equipped with multiple sensors such as LIDAR, radar, and cameras.

9. Conclusion and Future Directions

Future research for this project includes inferring under different sensor input modalities such as aggregated voxel point clouds, different proposal generation and prediction modules. Different perception-independent trajectory encoding techniques, such as using an open-set classifier, would be tested to improve model robustness against prediction mistakes. The perception-independent global model, which has a higher-quality proposal generation algorithm, a better 3D object trajectory encoding, and a powerful spatial-gap model, can contribute to both improved accuracy and increased diversity and create a more competitive DRAM. Consequently, the local model receives vision data for only highly ambiguous situations, reducing complexity and increasing the robustness of the full-fledged prediction model. More studies will be conducted on the effect of DLISFN2 transformers on the capability of attention weight calculation and its impact on prediction performance. With enhancements in the motion prediction and uncertainty estimation model, the perception-independent model can evolve to decision methodologies and traffic light recognition components. Additionally, extracting more useful information from the current detections and current ground-truth than just the current locations and their class, number of vehicle and traffic-light detections, can improve vision-to-decision modules. The global and local signal model to simulate an ego-collision probability model can be further studied to identify potential violations made by decision algorithms. Configuration differences between the prediction module and the perception module, such as the DRAM training portion of the phase that uses false negative and false positive cases to mimic prediction mistakes and reduce the number of inputs transferred to the autonomous vehicle decision pipeline, will be further

This work is licensed under CC BY-NC-SA 4.0.

explored. Both local decision policy models and global cascaded models, designed from the point of view of scenarios, will calculate potential differences.

We present a deep learning-based architecture for behavior prediction of other agents on the road for safe and efficient decision-making in autonomous vehicles. To preserve temporal information and exploit simplistic model characteristics, we adopt a temporal convolutional network with several improvements to enhance model discriminative power. Specifically, we use SPNet (1D-ConvNet + Long-Short Term Memory) instead of a simple 1D temporal convolution to encode temporal information more effectively and use transformers to focus on selective features and their spatial relations. We further introduce a novel loss function, based on prediction margins, that distinguishes between perceptually-similar classes and reduces false negatives. The state-of-the-art performance on the nuScenes dataset confirms that our model attributes a lot of importance to spatial feature attention for every agent and emphasizes the relevance of spatial feature interactions in understanding the behaviors of each agent. Additionally, the model demonstrates excellent performance on the nuScenes validation set, inferring at 39Hz on local motion energy calculation during test time, showcasing its potential for real-time decision making on-board for autonomous vehicles.

9.1. Summary of Key Findings

In this study, we have investigated the use of a DNN to understand and forecast the interactions of human drivers and autonomous vehicles in a road intersection scenario. We have trained and tested the model on a large-scale dataset generated from VR-based simulation environment. Results showed impressive performance on the visual signals data, which suggests that we can build an effective driver prediction model using only the visual data, sans explicit driver behavior, although to improve model robustness and generalization, multimodal action modeling could be incorporated in language inputs. The effectiveness of the system implies that it can be used as a substitute behavior prediction module both on rideshare and personal transport autonomous cars.

We believe that our work has two key implications for the AV and ITS research communities. The first is the development and validation of a highly accurate, deep learning-based behavior prediction system that can significantly improve the safety, predictability, interaction, and acceptance of AVs in the environment. The second is the creation of a large-scale dataset that will help others in academia, industry, and government to develop and validate their own AV

This work is licensed under CC BY-NC-SA 4.0.

safety systems. We feel that our methods are widely applicable to research that would benefit from action decision space models. Future work should include larger-scale and multimodal input modeling of the behavior decision space for even improved robustness, redundant inputs, more comprehensive uncertainty modeling, alternate architectures and boarding of complex models, and comparisons of advanced deep driving system models for autonomous vehicles.

9.2. Future Research Directions

Before network implementation or knowledge distillation, it is not clear that the non-linearity in data is large enough to warrant the use of deep networks, which are expensive and time-consuming to develop during the model design phase especially, and are a bit tricky to train, test, and debug than shallow networks, which are also sufficient to work on this task. A bigger dataset to train a more complex network or a different setup to allow additional sensor fusion, as the task is complex, so that increasing available data or combining multiple modals might be useful for driving prediction. In conclusion, this paper presents two prediction models based on feature-level multi-task learning and sensor-level multi-input/out end-to-end learning using a deep learning framework. The AE and DEC prediction results for both models are visualized and compared using the road data generated from both simulation software and real-world open dataset. The performance of all four models is evaluated based on the accuracy of AE and DEC predictions using different metrics at different prediction time offsets.

Finally, some potential future research of the proposed DL-based prediction approaches will be discussed. The inclusion of more complex DL modules, such as multi-input/out DL models, and model regularization (methods with higher penalty for weight magnitude) can be investigated. A dataset with more complex vehicle behavior such as lane change, different types of pedestrian behaviors (e.g., running, walking, and stopping), traffic light conditions (e.g., green, red, and yellow), and stop signs can be collected. Heterogeneous sensor data such as LiDAR data can also be checked to improve prediction performance. Increasing available data and adding more information to the model as well as a possible combination of multiple modals.

10. References

This work is licensed under CC BY-NC-SA 4.0.

1. M. Bojarski et al., "End to End Learning for Self-Driving Cars," 2016.

2. Y. LeCun et al., "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

3. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Adv. Neural Inf. Process. Syst., vol. 25, pp. 1097-1105, 2012.

4. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 779-788, 2016.

5. Tatineni, Sumanth. "Recommendation Systems for Personalized Learning: A Data-Driven Approach in Education." *Journal of Computer Engineering and Technology (JCET)* 4.2 (2020).

6. Vemori, Vamsi. "Human-in-the-Loop Moral Decision-Making Frameworks for Situationally Aware Multi-Modal Autonomous Vehicle Networks: An Accessibility-Focused Approach." *Journal of Computational Intelligence and Robotics* 2.1 (2022): 54-87.

7. Venkataramanan, Srinivasan, Ashok Kumar Reddy Sadhu, and Mahammad Shaik. "Fortifying The Edge: A Multi-Pronged Strategy To Thwart Privacy And Security Threats In Network Access Management For Resource-Constrained And Disparate Internet Of Things (IOT) Devices." *Asian Journal of Multidisciplinary Research & Review* 1.1 (2020): 97-125.

8. Tatineni, Sumanth. "An Integrated Approach to Predictive Maintenance Using IoT and Machine Learning in Manufacturing." *International Journal of Electrical Engineering and Technology (IJEET)* 11.8 (2020).

9. Vemoori, V. "Towards Secure and Trustworthy Autonomous Vehicles: Leveraging Distributed Ledger Technology for Secure Communication and Exploring Explainable Artificial Intelligence for Robust Decision-Making and Comprehensive Testing". *Journal of Science & Technology*, vol. 1, no. 1, Nov. 2020, pp. 130-7, https://thesciencebrigade.com/jst/article/view/224.

This work is licensed under CC BY-NC-SA 4.0.

10. M. Kuderer, S. Gulati, and W. Burgard, "Learning Driving Styles for Autonomous Vehicles from Demonstration," Proc. IEEE Int. Conf. Robot. Autom., pp. 2641-2646, 2015.
11. S. Ross, G. Gordon, and D. Bagnell, "A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning," Proc. 14th Int. Conf. Artif. Intell. Stat., pp. 627-635, 2011.
12. Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798-1828, Aug. 2013.
13. M. R. Endsley, "Design and evaluation for situation awareness enhancement," Proc. Hum. Factors Ergon. Soc. Annu. Meet., vol. 32, no. 2, pp. 97-101, 1988.
14. D. Pomerleau, "ALVINN: An Autonomous Land Vehicle in a Neural Network," Adv. Neural Inf. Process. Syst., vol. 1, pp. 305-313, 1989.
15. A. Dosovitskiy et al., "CARLA: An Open Urban Driving Simulator," Proc. 1st Annu. Conf. Robot Learn., pp. 1-16, 2017.
16. M. Bojarski et al., "Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car," arXiv:1704.07911, 2017.
17. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.
18. I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," Adv. Neural Inf. Process. Syst., vol. 27, pp. 3104-3112, 2014.
19. K. Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," Proc. EMNLP, pp. 1724-1734, 2014.
20. D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv:1312.6114, 2013.

This work is licensed under CC BY-NC-SA 4.0.