Machine Learning for Autonomous Vehicle Pedestrian Intent Prediction

By Dr. Yasushi Wada

Associate Professor of Mechanical Engineering, Tohoku University, Japan

1. Introduction to Autonomous Vehicles and Pedestrian Safety

Pedestrian safety is one of the most important key points for implementing autonomous vehicles in urban environments. The Transportation Research Board (TRB) researchers have defined pedestrian safety as a process to ensure that pedestrians can cross and walk beside roads and streets without encountering excessively high accident and injury risks. Common problems for pedestrian safety include adverse traffic impacts and the difficulty of integrating walking and riding into local laws and regulations. Recognizing pedestrian walking intention is essential for predicting the trajectory of pedestrians and is a critical problem in autonomous vehicle safety. Behavior understanding and predicting the trajectory of pedestrians have received increasing academic attention during the last decade. Nevertheless, most of the existing work used traditional techniques (feature-based) for pedestrian intent prediction from sensor data, which may not provide enough discriminatory power.

Autonomous vehicles have shown increasing capability in decision-making and executing driving tasks in various conditions with machine learning, pattern recognition, deep learning, and the development of high-power embedded systems. The next challenge is the societal acceptance of these autonomous vehicles, which is closely related to the safety and reliability of the vehicle's decision-making. The complexity of the driving environment has increased substantially due to the growing number of autonomous vehicles. Cooperation and coordination intelligence between autonomous vehicles and other road users is an important issue in transitioning to new traffic patterns with maximum safety and efficiency. Predicting the intention of action sequences and representing the most effective road users play a crucial role in achieving these objectives. The coexistence of autonomous vehicles and other road users, such as vehicles, pedestrians, and cyclists, poses a significant challenge in building an intelligent transportation system city with a high level of safety and mobility.

1.1. Overview of Autonomous Vehicles

In 2019, the first level-3 cars were announced by Audi, Mercedes, Tesla, and Volvo. Level-3 was explicitly introduced as a transient level, but it indeed describes two different levels with respect to the overall goal of the autonomous driving systems. On the one hand, we have driverless technologies capable of controlling the vehicle during normal SAE J3016 conditions. These technologies are intended to be used by vehicle occupants. On the other hand, it describes a supervisory system, an automated driving system capable of taking full control of the vehicle in emergencies. The integration of those driverless technologies into the vehicle is, therefore, expected to be the first step towards these technologies operating completely without human supervisors. The next step consists of enabling the autonomous driving systems to perform operations that they are not yet capable of. As previously discussed in the "Levels in Automated Driving" section, human drivers should use these systems only in scenarios where the ADS capabilities are well known and where other human co-drivers are not allowed. Since these types of limitations are not trivial, especially for L3 systems, it is possible that some OEMs have decided to disregard these systems in terms of them being profitable.

According to an economic analysis performed by Intel Corporation, the driverless revolution is expected to create industries with a combined revenue potential of approximately $7 trillion per year. At the same time, up to 585,000 lives a year are expected to be saved by driverless cars. Autonomous vehicles are a type of robot in charge of performing the task of driving in self-driving mode. Autonomous vehicles are designed to augment human driving capabilities. The purpose of these systems is not to replace human drivers but to provide better safety, performance optimization, reduction of operative costs, and enable mobility access for disabled or old people. The first paradigm of autonomous vehicles is similar to that of auto-pilots, assisting people in executing specific tasks without human supervision, like cars in speedway tolls, unmanned aerial vehicles (UAVs) for surveying, and so forth.

1.2. Importance of Pedestrian Intent Prediction

Firstly, pedestrians are the dominant entities for collision threats in urban driving scenarios, especially in situations such as road intersections and pedestrian crosswalks. If an autonomous vehicle is able to predict pedestrian intentions in advance, it can anticipate potential pedestrian actions and analyze their potential threat to take evasive response, such as properly slowing down or stopping the vehicle in advance. On the other hand, if a pedestrian intentionally disables a vehicle, for example, by signaling to cross the road or waving at the vehicle to yield, the vehicle can predict pedestrian intentions and create a safe route, providing sufficient time for the pedestrian to cross the road. Such coordination would improve the traffic flow and promote harmonious driving for mixed traffic scenarios.

It is imperative for autonomous vehicles to not only accurately perceive pedestrian poses and locations with reliable confidence, but also to predict future pedestrian behavior and estimate pedestrian intent given the current situation. The importance of pedestrian intent prediction lies mainly in the following aspects.

2. Key Challenges in Pedestrian Intent Prediction

The main challenge pertaining to pedestrian intent prediction is the intention of individuals. It is often difficult to predict what individuals are going to do next, irrespective of the intent. The ambiguity arises from the complexity of human behavior. The behavior of individuals frequently depends on what others with whom they are sharing the road space are doing. If the goal of a pedestrian is crossed, the pedestrian may behave differently depending on the situation around the crosswalk.

The main contributions of this paper are: 1) A machine learning-based pedestrian intent prediction system that uses only a portion of the observations provided by Autonomous Vehicles (AVs). 2) A notion of effective sensor usage, proposed to identify the most informative sensor information for this task. 3) The deployment of the system on four different AVs, showcasing its adaptability among various implementations.

This paper proposes a pedestrian intent prediction system intended to help foster safer and more efficient interactions between autonomous systems and traffic participants. The system uses only surrounding vehicle sensors to infer the pedestrian intents, processes this information through a machine learning model, and provides the result to a motion planner. The experimental results demonstrate the effective utilization of surround vehicle sensors to predict pedestrian intents.

Building an autonomous driving system which can make human level decisions requires anticipating potential actions or intents of other traffic participants such as vehicles, pedestrians, and cyclists. Unlike regular drivers, drivers that are equipped with autonomy no longer have the ability to look at the path of the driver's head to predict intent. This is one major barrier to the perception problem for an autonomous vehicle, especially when both pedestrians and cyclists are involved. Autonomous vehicles are expected to use their observations combined with the policy composed of learned reactions to behave naturally in traffic and to avoid causing concern in other traffic participants.

2.1. Variability in Pedestrian Behavior

Proneness to real-time updates to their navigation trajectories just before entering the road due to social conflicts, and the tendency to often choose diagonal paths are examples of pedestrian behavior variation evident in the experiments of Uthmöller et al. and in the case study of Paris, respectively. Variability in pedestrian behavior must be accounted for while modeling pedestrian intent, and we pose the question: How do we address the variety of intentions and the deterministic prediction of pedestrians' main intention in a certain proximity of an intersection? Every attempt to model pedestrian movement prediction and behavior in its entirety requires the development of pedestrian motion models encapsulating reactive and strategic behaviors to be able to fluidly adapt their fate when exposed to dynamic cues. Individual pedestrian models should be able to integrate contextual information to further disambiguate their respective influences and generate behavioral confidence measures. AVs are capable of supporting this development by providing the required information in various formats such as active perception grand truth, through advanced sensor fusion methodologies which aggregate sensed cues into components of higher semantic value, or through decision and planning modules able to selectively stimulate the pedestrian in order to actively update its intent, thus influencing steering, speed, and position. AV data, when shared, holds a large educational value which transcends basic research. Knowledge discovered through AV logging data must be capitalized to improve both AV and pedestrian functionality within urban environments.

Pedestrian behavior is difficult to capture within simple rules, and thus there is significant variability in pedestrian movement characteristics. Factors that can cause variability include environmental factors such as intersections, the number of lanes in a road, lane width, sidewalk width, traffic signalization and zoning type, as well as individual pedestrian factors such as age, gender, weight, height, physical condition, and motility impairment. Contextual variables such as population density, time of day, day of the week, and weather conditions also play a role in influencing pedestrian mobility. Variability in pedestrian movement can be addressed through effective modeling and tracking of reactive and strategic pedestrian behavior to predict their intent, and hence provide reliable input to AV control modules. AVs need to predict pedestrian intent with a high level of confidence, as pedestrians may not adhere to explicit rules due to individual motivation, social conflict, economic reasons, or even unpredictable events such as sudden illnesses.

2.2. Complex Urban Environments

Challenges for the autonomous vehicle will also be different from the suburban challenges in urban environments. The urban autonomous vehicles will continuously interact with other vehicles that might not have autonomous capabilities, interact with numerous non-motorized vehicles such as pedestrians and cyclists. Also, the urban autonomous vehicles need to plan intelligent movements among a variety of spatial and temporal traffic participants - and accommodate dissimilar traffic conventions that may even oppose each other. In very challenging city environments, common misconceptions such as less traffic infrastructure means fewer standards and rules in the behavior of drivers are incorrect.

Another characteristic of urban environments is the high unpredictability and complex interactions among the various agents. Complex urban environments' characteristic behaviors include erratic driving patterns, sudden stops, u-turns without warning, lane switches which appear dangerous to the human eye whereas normal to other agents in the environment. Also, multimodality behavior aspects are specific only in complex urban environments. Examples are jaywalking pedestrians, cyclists, and scooters riding on sidewalks, self-driving vehicles saying "we are waiting" to other pedestrians. These behaviors should be taken into consideration for pedestrian intent prediction.

3. Data Collection and Annotation

The problem and design of this work mean that the input features will be varied and complex, such that increasingly advanced models will be used. The core attributes utilized in constructing ground truth are the pedestrian's courses using initial trajectory vector directions, speed and local turn angles as calculated between positions in relation to public transit stops. Other attributes will be compiled in a database that will simplify inspection of relevant individual occurrences that can later be directly extracted to be fed to the best performing models, such as sidewalk typology and behavioral variables. The pedestrian's intention will, however, be restricted to a maximal set of three distinct actions for each individual instance: approach the curb, vision on either direction of the road perpendicular to the sidewalk, or remain traveling perpendicular to the sidewalk. Accessories like cellphones will also be excluded from the annotation.

Currently, the task of pedestrian intent prediction in the field of autonomous vehicle navigation relies on data that is recorded from custom-built vehicles with arrays of expensive sensors, driven throughout several cities in multiple countries. This is both costly and unscalable. One aim of this work is to conduct experiments to demonstrate that readily available, large-scale datasets from crowd-sourced GPS tracking data and other public datasets, like the New York Department of Traffic and the UMD campus catalog, when processed and properly curated, can also be used to successfully train models for such a task. The predominant goal is to predict the pedestrian's intent far enough in advance, when approaching the road indicates intent and preparedness to cross, and eventually, when actually in the road indicates intent and resolve while crossing.

3.1. Sensor Technologies for Data Collection

Camera. Generally, subjects detected using a 2D camera can be classified either into various classes such as vehicles or pedestrians, or different activities such as walking or standing still when designing a vision-based pedestrian detection and activity recognition model. Therefore, using a camera sensor is one of the most intuitive and straightforward methods to understand the behavior of a pedestrian. With robust machine learning models, the vision-based methods have shown impressive results using deep learning networks such as convolutional neural networks (CNNs).

To develop and verify pedestrian intent prediction models, it is crucial to collect various types of data. Both 2D and 3D sensor data types are required to consider the pedestrian's various actions and other environmental aspects. Different sensor technologies and types that would be able to provide 2D (image) and 3D data are reviewed and briefly introduced with their pros and cons. Initially, camera sensors, which provide 2D images, are introduced. Next, 3D LiDAR sensors and depth camera sensors, which can provide a 3D point cloud or 3D-depth information, are reviewed. Finally, sensor fusions which combine multi-sensor techniques to further enhance sensing results are introduced.

3.2. Annotation Techniques for Pedestrian Intent

Section 3.2.1 MODE 1: Online Data Collection One cost-effective way to collect annotated obstruction or pedestrian intent data is through the use of web-sourced data such as Google

Maps or Open Data projects created for large cities. This data can be customized to the geographic location in which the pedestrian intent predictor is to be applied. Time of day, day of week, and season should be part of the decision-making process when selecting the frames to be used from such sources. Simulation techniques can be applied to the cartographic locations selected from these online sources, resulting in an environment that can be used to collect both negative pedestrian intent (obstruction) data, as well as positive pedestrian behaviors, to be used in ML model development. This data collection does not leverage AI to create ground truth data. It does allow for customization of the intent data.

The four main annotation approaches that can be considered when addressing pedestrian intent, namely the use of online data, offline data using simulators, ground truth data enhanced with heuristic rules, and the use of ground truth data derived from human annotation, are described in this section. A significant discussion is devoted to the development of a human annotation process that has been developed that allows for the free movement of annotated pedestrians to exhibit normal human intent within a time-constrained ground truth video production regime. This allowed the cost-effective collection of ground truth data on pedestrian intent.

4. Feature Engineering for Pedestrian Intent Prediction

In an active urban intersection environment, pedestrians who express altering kinds of intention behaviors during their spatial and temporal interactions contribute significantly to the safety issues. However, it is difficult to model all such behaviors of those pedestrians in pedestrian intention predictions for the purpose of warning and guidance of the decision-making system in an autonomous driving vehicle. In most previous methods, binary classification via simple deep sigmoid models corresponding to the start and end time of the crossing behavior was proposed in literature, which focuses on volume and frequencies of specially labeled pedestrian-vehicle interaction data rather than the meaning of pedestrian intention behaviors in reality.

Though in real-world pedestrian-vehicle interactions, pedestrians may have varied motion behaviors. They may make a late-stage decision regarding crossing intent. They may also exhibit behaviors during interaction corresponding to waiting, hesitating, preparing to cross, or already starting to cross behaviors. In this work, a generalized model is proposed for predicting pedestrian behavior at each interaction phase that assists the autonomous vehicle decision-maker in making the correct action decision with an action type other than "stop" in static driving conditions around intersections. In addition to the standard robot and pedestrian attributes, the proposed model accounts for multi-modal interaction situations such as current approach behavior during interaction with the vehicle. In this model, nine mutual interaction time gap features representing the interactions at the current and previous situations between the pedestrian and his/her nearby pedestrian as well as the nearest vehicle driving on the road are included.

One of the key challenges in developing robust pedestrian intent prediction models for autonomous vehicle pedestrian interaction is the sparse outcome space. In this work, the aim is to not only predict at every interaction phase the intention of the pedestrian but also to accurately assist the autonomous vehicle in making the correct action decision. Intention prediction has commonly been formulated as a classification problem and involves binary prediction of whether the pedestrian will cross or not at each interaction phase time step. Given the short length at which the situation trajectory is temporally observed, this naturally leads to a sparse classification problem.

4.1. Spatial and Temporal Features

Since the visual and the high-resolution sensor outputs usually are in two or three different resolutions, we fused only their general features, such as their weights and the outputs produced after being forward projected to the camera end-near zones. The performance of each relevant sensor is improved by training an independent retrieval neural network, in order to assess the activation of the hidden layers for each different bi-dimensional convolution or Cropping2D layer of the corresponding sensor-related feature extraction module. This network receives all the sensor-related features and the IoU (Intersection over Union) value between the hand-labeled and the estimated bounding boxes in the image domain provided by a given proposal, considerably speeding up and showing the potential of a reliable joint proposal function between the classes of interest that can directly feed a final hierarchical combination of each proposal type in the bird-eye and image domain. In case of a bootstrap approach, or when we adapt to a vehicle without some of the sensors, we can disable some of these sensor-related features.

Selection of the spatial features primarily depends on the sensors available to the perception module and the advanced driver-assistance systems (ADAS) of the vehicle. In its most general form, we can estimate the typical commuting vehicle sensors can observe azimuth (yaw orientation angle), elevation or range, pixel location on LiDAR (polar and/or Cartesian) and mutual-time with pixel location (latitude, longitude, and orientation) on a camera, which work on different wavelengths and are placed on different locations of the vehicle. In this paper, we further describe and extract from the raw stixel description of our pedestrian proposal module both its typical depth-related (as for a LiDAR sensor) and its appearance-related (as for a camera sensor) spatial features. Additionally, we employ hand-crafted position-related and appearance-related spatial features as obtained from the stixel description of the pedestrian estimation module or by applying CNNs to the original camera pixels corresponding to the bounding box of a proposal coming from the pedestrian proposal module. The presentation of these relevant modules is provided on the web.

4.2. Contextual Information Integration

Schönborn and Knoll regarded pedestrian motion validations by testing power guidance and informed the pedestrian of the intention. This method led the pedestrian toward the vehicle preferred crossing interval. Similarly, Laugier et al. used a reward model-quality level for communicating specific, advanced safety tips. However, these methods are not embedded into the controller, and proper moving pedestrian considerations are requested. Jiang and Kovall proposed an intent model by proper feature extraction and feature selection. Then, they tested the combinations of machine learning algorithms to point out the importance of a shared-risk consideration for robot motion planning. However, this early pedestrian intent prediction does not consider multimodal uncertainty, encounters data faults, and requests vehicle exceptions. Smith et al. created a kernel density estimation (KDE) of pedestrian motion through a GMM and a sensor fusion approach to predict their crossings.

We proposed a neural pedestrian intent prediction framework for better understanding pedestrian contextual information by predicting pedestrian future paths with temporal probabilistic modeling. The pedestrian contextual model is embedded into the vehicle using an early motion planner and a subsequent shared-risk consideration. The proposed neural intent predictor can predict pedestrian future trajectories with multimodal uncertainty as well as the potential pedestrian crossing intervals that can provide the vehicle with negatively labeled samples. We design our neural model using a convolutional long short-term memory (ConvLSTM) architecture such that we can predict future pedestrian trajectories with multimodal uncertainty using a probability-of-probability prediction instead of the traditional deterministic training for intention events. As a result, the proposed neural network can output the future multimodal likely pedestrian motion in varying urban road environments.

5. Machine Learning Algorithms for Pedestrian Intent Prediction

5.1 Datasets for Pedestrian Intent Prediction County of Los Angeles and Metro provided the Movement, Activity, and Interaction Safety for Autonomous Vehicles (MAIS-AV) Ped1.0 dataset to the research community, which consists of camera frames and lidar point clouds for all road users. The actual positional and rotational data pertaining to road users, ego-vehicle front present therein can be extracted. The annotation data consists of different levels of details of bounding boxes along with other types of data representations such as segmentation masks, poses and individual joints annotation, etc., depending upon the dataset type and release version. The current study utilized Position-only Pedestrian 2D bounding box annotation data for training machine learning models. Details of this dataset are already described in Table 4 and involved in the previous study using the same dataset.

There are several machine learning algorithms that can be applied to AV pedestrian intent prediction tasks. The problem domain of interest in this chapter is to predict the pedestrian intention to cross a road as a binary outcome as mentioned in Section 2. This section not only presents the details of the dataset used but also highlights the formulations of feature sets as well as machine learning models required. Parameters used for machine learning models employed in this study were adjusted using grid search.

5.1. Supervised Learning Approaches

Logistic Regression and Support Vector Machine: These supervised learning methods are among the most commonly used algorithms in the field of machine learning. For example, linear or polynomial logistic regression classifiers have been used in some state-of-the-art pedestrian intent prediction models. To apply logistic regression to the pedestrian intent prediction classification problem, one has to define and extract some predictive features from the sensory data. Some common features used in this context include pedestrian poses, their relative positions with respect to the ego vehicle, as well as some GIS features. These features then serve as input for the logistic regression model, and their weights are based on the learned logistic function. These weights are learned using maximum likelihood estimation or regularized losses. Support Vector Machine (SVM), like logistic regression, is also a linear classification technique. Its goal is to find a hyperplane that separates the data between two classes, maximize the margin between the hyperplane and the closest data, and then penalize the data that is inside this maximum margin.

Most learning methods for intent inference problem are supervised, which learn predictive models from labeled training data. In these approaches, one selects relevant features from the raw sensory data, processes them using well-known techniques, and uses this processed data to classify the intent. The features of existing state-of-the-art supervised approaches often include pedestrian poses, their relative positions with respect to the ego vehicle, and the scene semantic information. This section describes two widely used supervised learning algorithms and details how the pedestrian intent prediction can be framed as a classification problem.

5.2. Unsupervised and Semi-Supervised Learning

Mistakes are made by our proposed pedestrian intent prediction model. However, an essential quality is that it provides an open-ended and structured representation of a pedestrian's desired future maneuvers, even when events are not probable. This simple formulation encourages future work to be more flexible and use future interactions to adapt methods with little difficulty. When characterization of interactions scaling with agent numbers settle in real-world environments, this will be crucial. Countless advantages come from exploring these aspects. In addition to the vehicles and pedestrians, cyclist behaviors will be predictable, and extended formalization will be brought into agent-agent interaction modeling design. This shall reduce interactions-based prediction research's goal risk.

Despite all the complexities involved in forecasting pedestrian intent, predictive models are persistently oversimplified. The discounting of pedestrian VSDs and the complexity of pedestrian interactions lead prediction models to ignore or oversimplify potential maneuvers and belief states.

Furthermore, after the pedestrian has communicated their intent in some way, their actions can be influenced by the vehicle's responses. Consequently, through this feedback loop, responsible VSD behaviors exist within the pedestrian. Therefore, the pedestrians themselves are neither strictly reactive nor deterministic. While relying on an interactive agent model, the automated vehicle's AV predictive ability is frequently assumed deterministic, but inherently, true future states are unknown at the time of prediction. Only the vehicle observer possesses this information.

There is abundant data obtained through historic drives, which includes video and LIDAR data of the forward surrounds of the vehicle, precompiled tracks from pedestrians, and the true future trajectories of pedestrians. Every pedestrian interacting with an automated vehicle has some notion of what they would like to do (i.e., their intended future behaviors). To date, this level of understanding has not fully filtered into the major components of the vehicle's automated prediction methods.

A pedestrian intent prediction system is trained to predict the pedestrian's intended future trajectory. Our prediction method is machine-learning-based, depending on large volumes of historic data collected from countless drives. The likelihood of observing particular pedestrian behaviors is computed to model uncertainty. Additionally, the model uncovers and uses the latent structure of pedestrian movement to gain better predictive performance.

6. Evaluation Metrics for Model Performance

Generalization of Traditional Metrics The accuracy of some unseen data in the autonomous vehicle pedestrian intent prediction problem will always meet a minimum value which is equal to the frequency of the most occurring class. This stems from the objective of many applied problems, that is to predict instances of the minor class as accurately as possible without losing the ability to correctly predict instances of the major class. Algorithm performance is in all practical cases biased towards the majority class. Even though optimizing classification measures is not of direct interest in many topics relating to study or application, they should be effectively used. The precision, recall, and F1-score of both classes should be analyzed together in datasets showing a significant discrepancy. However, in the autonomous vehicle pedestrian intent prediction problem complexity, it is a much more user-friendly method to present precision and recall in a metric that focuses primarily on the positive class such as AUC-ROC and AUC-PRC.

Traditional Metrics The quality of autonomous vehicle pedestrian intent prediction can be estimated by traditional machine learning metrics such as accuracy, precision, recall, F1-score, AUC-ROC, and AUC-PRC. However, in semantically unbalanced data or conditions such as autonomous vehicle collision prediction, metrics need to be fully understood for the correct usage.

6.1. Accuracy and Precision

The first metric observed is that the intent model does not predict stationary persons' stationary activity well enough; 31.04% of frames are directed well to people stopped with an F-score of 75.5. Keep in mind that 43% of the test data pedestrians were walking, 17% were standing or not detected by the word classifier, and 22% of test data units of pedestrian time were labeled as some form of stationary state. Both of the following analyses include all pedestrian frames in the mobility layer. Because of this mobility layer filter only in the coupling mode setting, this filtering has affinity title error consequences, i.e., a missing poor predictions' stationary person can be considered a moving person that was not properly labeled, or stationary persons can get some spurious mobilities as reflective of the coupler's mobility.

In the test data, 22% of pedestrian frames were labeled with intent. What is reported for the intent model being benchmarked are F-score results as overall full-day averages, and statistics by person and by time-of-day. The test data pedestrian annotations are converted to a model performance quality metric details in this section. What follows are various ways that F-score cannot be computed for these models, followed by explanations why and alternative model comparison approaches to substitute. Note that at no time does the intent model's missing intent label predictions make these naive models have F-scores exceeding 50.0.

The next few sections discuss the results in detail, starting with the accuracy and precision of the intent model output. In the training data, 33% of frames were annotated with pedestrian intent. As a result, a naive classification model, which predicts all frames as no-analogy, can have an F-score of only 50.0. The second naive baseline is the annotation work of 85%, the average of the most common frame annotation in the training data. The third naive baseline model's F-score is 57.7. In a frame by frame model in which all persons start in the non-analogy state, moving persons were set as walking and stationary persons were set as stopping. Table III compares the complexity of these naive models with the intent model being benchmarked.

6.2. Recall and F1 Score

When either recall or precision is low, the F1 score is also low, leading to an overall bad classification of the data. One problem that can arise in the comparisons of different machine learning models is when the classes are in an imbalanced state. In this type of situation, accuracy is possibly no longer a valid metric to reference classifier performance. This is because a model can produce accurate predictions by simply predicting only the majority class. In our work, recall and F1 score were used when evaluating the accuracy of models. Any class with a sample count smaller than 1000 units should use these same metrics for evaluation. These scores can demonstrate the performance of different model prediction accuracy in both positive and negative labels.

In addition to accuracy, other performance metrics are important when evaluating imbalanced data. One of these popular metrics is known as recall. Recall is a measure of what fraction of actual positive cases were identified correctly. The formula for recall can be found in equation 6. Recall fluctuates from 0 to 1. An F1 score addresses the problem of accuracy in cases where recall or precision might no longer be adequate and incorporates these into one score. F1 Score is the harmonic mean of recall and precision.

7. Real-World Applications of Pedestrian Intent Prediction

As a simpler approach for small-scale scenarios that might not have high redundancy nor allow for the investment in new technologies, we proposed some small-scale solutions for pedestrian intent prediction. These solutions can show interesting preliminary data for different stakeholders and show that the subject can be a recurring part of the conversation around traffic safety in the near future.

Automated driving systems are gradually starting to become a part of people's lives in urban areas. Small fleets of autonomous taxi shuttles are tested in restricted environments, and precautions are established for safety measures. However, some challenges before the widespread use of fully autonomous cars have not been fully tackled, especially those involving perception problems around the vehicle with diverse activities and challenging pedestrian behavior, such as group activities or aiming to cross the street. These scenarios could benefit to a high degree from innovations in different levels of the sensor stack, control algorithms, and path planning routines.

In this paper, we explore the advantages of implementing ML solutions that make use of pedestrian intent prediction. We preview a pioneer work that formally addressed this concept and propose improvements and new solutions that are feasible to make it easier for adoption. We validate these solutions in well-established environments and suggest next steps to tackle real-world challenges.

7.1. Case Studies in Urban Environments

Since this is an ongoing area of research, we are not yet fully aware of all of the problems involved. For now, we are considering specific scenarios with handled pedestrian intent prediction with the assumption that other active agents share this capability. These cases also include an intersection traversal or a crosswalk, and typically a person standing near the roadway. Optionally, in the future, we may have to consider construction barriers, older individuals who may cross so slowly the autonomous vehicle can move before the person is finished leaving the intersection, or steeper curbs (prohibiting wheelchair access to that street corner).

While substantial work on autonomous vehicles considers the complexity of traveling through an urban environment, we wanted to investigate this complex environment. Autonomous cars in a city have many different types of entities that it may be interested in - at different times we may want to consider other vehicles, pedestrians, bicyclists, police, fire departments, or an ambulance. This leads to a problem of dividing an environment into important areas at all times and having 3D maps of the surroundings. Urban driving is severely asymmetric for the occupants of different vehicles. For example, a delivery truck has a person inside and probably won't move for hours in order to deliver packages. This creates a long-term, predictable trap of the truck's future behavior; very different from a car with a distracted driver that will (hopefully) have coffee and have a different plan.

7.2. Integration with Autonomous Vehicles

There are several challenges for the implementation of autonomous vehicles. One important challenge is pedestrian detection, which exhibits several shortcomings, including inadequate performance under different conditions. Even if the pedestrian is identified after a certain point, current machine learning techniques also lack the ability to determine subtle body language, making it difficult to predict a pedestrian's intent. In this paper, we attempt to address the intent prediction problem by proposing a geometry-independent, real-time ensemble architecture to improve robustness. Our work assumes pedestrians travel on the right side of US roads, as required by traffic regulations. Our pipeline is composed of two machine learning models: a frontend to detect a pedestrian's posture and a backend to determine his or her intent.

Robust sensor technology and advanced control systems are critical components for developing safe and efficient autonomous vehicles. However, a primary challenge for these systems is the accurate prediction of pedestrian intent. To address this problem, we propose a machine learning solution. Our architecture is composed of an ensemble of models. The primary model uses a convolutional neural network to detect a pedestrian's spatial posture at any moment in time. Depending on the scene, the secondary model consists either of recurrent networks to learn sequential intent features or 3D convolutions for volumetric feature extraction. Using real-world data, we evaluate our solution and subsequently define performance metrics to select the proper secondary model for our ensemble. Our high-performance ensemble significantly improves over current state-of-the-art machines, making it a viable choice when integrating pedestrian prediction into autonomous vehicles for better safety.

8. Ethical and Legal Considerations

The development of autonomous vehicle technology that will affect U.S. Native communities has gone largely unnoticed in public statement form, media, and academic and corporate developments. Native communities, however, stand to reason the new technology in important—and sometimes unique—ways given their outstanding ability to control public roads within their territory. These Native Americans are responsible for developing this technology in conjunction with stakeholders, applying firms, individuals, and other entities. It provides governments with the prospect to prevent Native Americans from realizing these potential advantages and to deliberate when using the technology. They must plan in consultation with and with individual and collective consent from their tribal members to ensure that autonomous vehicles approved for use on public roads will benefit Native American individuals, families, and communities.

As the application of machine learning to autonomous vehicles becomes a reality, some important ethical and legal considerations arise. These issues may not be paramount when developing autonomous vehicles for use in large urban centers, for instance, but they are essential in their development to be applied to native communities. Key ethical and legal questions that affect the implementation of autonomous vehicles in the field of native communities include access control, data privacy and ownership, and the remedy of harm caused by accidents involving these vehicles. Several of the guidelines put forward are: heavy involvement of indigenous communities in the development of the technology, use and protection of an individual and the collective importance of their knowledge system. These involve designing consultation and collective consent frameworks and potentially co-managing the project.

8.1. Privacy Concerns

If the researchers can obtain such coupling information, it would be possible to check the model's false positive under real-world conditions after deploying the model on the vehicle. Inevitably, many research methodologies use monitoring people context data and monitoring spaces, requiring logistical assistance. Use of these data is almost assured through collaboration with public agencies or private actors. However, the assessment processes must respect the usual concerns related to the processing of personal data: notification of pedestrians is necessary when personal data are to be processed. Furthermore, an autonomous vehicle provider should be transparent about the impact on privacy and individual freedoms. In the testing phase, privacy concerns about automated processing based on location data should be addressed through impact assessments. Goals should be met by appropriate solutions, such as automated video editing techniques or anonymization, simulation, etc.

One of the main concerns in evaluating the pedestrian intention prediction model is the privacy of the pedestrians who provide the data. Although most of the currently available data has been collected with consent, allowing the use of the data for research purposes, there is still the possibility of recognition or personal identification based on what the pedestrian is carrying, where they are walking, and their activities, even though the face of the pedestrians is blurred. For the sake of verification, the provider of the pedestrian trajectory could be given a unique identifier, unless they are walking in a private space or the data is not highly sensitive. The presence of such identifying information means that the dataset is, to some extent, directly linkable to a real person. For example, the trajectory that U17AgeFemale is walking with a backpack in front of DreamWorld has a strong correlation with a real person walking in front of a particular shop called DreamWorld wearing a backpack.

8.2. Liability and Responsibility

This is not simply for vehicle drivers and passengers, but it includes them as well. While some claim that reducing the need for licensing the ability to operate an AV reduces the importance of automotive law, so long as one refuses to participate in the community of special machines, then that choice raises important public policy considerations. Indeed, a world in which no ordinary machines appeared on any ordinary road would probably have some value or importance to our quality of life; autonomous vehicles are not immune to the same kind of property liability concerns on the grander scale that regular vehicles are today. Indeed, with the public attention on issues raised by software issues, they tend to come into incredibly sharp focus.

We dove into modal risk measures and potential designs for considering them, but deeper than this is the central question of liability for a harm caused by an AD or an AV more generally. A similar question of responsibility for a non-appearance made at a pedestrian crosswalk, or risk caused by a pedestrian crossing outside of an intersection with the walk signal, is already present today. The introduction of AVs into the transportation ecosystem simply increases the importance of properly considering these questions. Today, when considering all of the benefits of vehicle automation, especially in conjunction with pervasive networking, one of the most exciting is the potential to greatly increase the safety and quality of human life as it is lived day-to-day in our noise-filled and danger-threatened urban environments.

9. Future Directions in Pedestrian Intent Prediction

Many of the proposed algorithms use various sensors that aid in the inference of pedestrian intent. These sensors include laser scanners, radar, LIDAR, cameras, GPS, among others. They provide various prerequisites for intent prediction. However, the data captured from these different sensors needs to be synchronized, rectified, and mapped onto the same spatial-temporal profile. This is necessary to have a wholesome view of the pedestrian's states, provide accurate prediction, and improve the readiness of autonomous vehicles to model various feasible outcomes. This is instead of having to provide a deterministic outcome, which might be suboptimal or cause loss of life.

In recent years and with the advent of improvements in neural network architectures, the focus has been on improving tracking, segmentation, modeling of the pedestrian, and scene understanding problems. These improvements provide a better deep understanding of road users while sacrificing the physiological factors of gait or head nod. These factors are vital concepts that need to be inferred from the road users' states.

In this chapter, we presented the state-of-the-art computer vision algorithms for pedestrian intent prediction. These algorithms have evolved from framing the problem as a regression or classification task to proposing more advanced and interpretable models.

9.1. Advancements in Sensor Technologies

Recent years have seen substantial advancements in surround sensor technologies, especially in low-cost, large-scale sensor manufacturing and in the reduced time and cost of data acquisition and labeling processes. These advancements are changing the trajectory of the development of autonomous vehicles. First, the number of sensors deployed in the US has jumped from single to multiple sensors. Manufacturers are becoming more keenly aware of the need for information redundancy and operational safety, and are moving toward sensor fusion. Second, levels of autonomy have shifted from single to multiple hierarchical levels. On the one hand, thanks to efforts for setting standard benchmarks and improving performance, reducing inference time and parameters, and introducing hardware acceleration, artificial intelligence computing hardware remains an economically viable, embedded mainstream solution.

This chapter presents an autonomous vehicle pedestrian intent prediction system trained on camera images. The system uses the convolutional neural network VGG16 for image feature extraction and a recurrent neural network for temporal fusion. The system detects and tracks pedestrians in an early detection stage. This stage is followed by multiple pedestrian detection and tracking filtering algorithms, each of which employs multiple sensor inputs for driving environment perception. The detected and tracked pedestrians are assigned consistent IDs so that data association is achieved. This is necessary as sensor inputs are utilized for training the feature-extraction and temporal-fusion networks. Finally, each pedestrian is observed by the trained feature-extraction network, and the features of the pedestrian-related bounding box image are used as inputs for data prediction.

9.2. Incorporation of Deep Learning

Deep learning is a type of machine learning that uses a multi-layered structure of artificial neural networks (ANNs) to map input compatible networks. Types of ANNs that can be employed for the mapping part of deep learning include fully connected, convolutional, and recurrent neural networks, based on the multi-layered structure and the type of input. Networks that have a non-linear activation function are used to map input into networks.

Data are used to learn the parameters to transform input to the corresponding network-based outputs during the training phase. Optimizers with back-propagation-based error gradients are used to iteratively update and fine-tune the parameters. These parameters are employed in a given task that corresponds to the learned models for mapping new data based on trained data. The identified task can be performed in a specific application. Although the generalization issue is not solved, deep learning is based on large labeled data, such as image data. Upon providing labeled examples, the training part and other related deep learning endeavors will be successful.

This section addresses the incorporation of machine learning for use in the proposed framework. A detailed review of deep learning, convolutional neural networks (CNNs), and Long Short-Term Memory networks (LSTMs) is provided. These are both crucial, updated, and cutting-edge machine learning advancements which are often used in various applications, including vehicle pose detection. As with vehicle pose detection, recommended approaches for the implementation of deep learning for pedestrian pose detection are considered. It should be noted that an illustrative example of convolutional LSTM is demonstrated on top view overhead imagery of pedestrian detection.

10. Conclusion

We believe that by considering the pedestrian intent in an explicit way, there will be a positive impact on the development in the field of autonomous driving. Furthermore, this work opens other perspectives on which pedestrian intent could be learned. It also brings to the focus the need to create a standard dataset annotated with pedestrians' intent to analyze the performance of models across different use cases and methods. Our work also points to better costmap modeling for driving decision-making policies.

We conducted a series of experiments to present the payoff of our approach on a perspective of SVR prediction task of sensor data. Finally, we showed the improved models' ability to estimate motion, reduce uncertainty, and generalize better for unforeseen conditions or inputs by integrating pedestrians' pose and action. However, in more complex scenarios with dense crowds, similar networks designed like the state-of-the-art grid encoder-decoder model and lattice LSTM model utilizing the spatial context could be implemented to extend this work.

In this chapter, we presented a longitudinal pedestrian path prediction approach based on a learnable costmap that captures the pedestrian intent explicitly. The proposed pipeline indicated successful results and can be used for active decision-making in autonomous driving vehicles. Furthermore, the visualization using class activation maps shows that the learnable costmap networks focus on natural features used by humans such as sidewalks and roadways when predicting pedestrian intent, giving confidence that the created costmaps are learning effectively.

Reference:

1. Perumalsamy, Jegatheeswari, Bhargav Kumar Konidena, and Bhavani Krothapalli. "AI-Driven Risk Modeling in Life Insurance: Advanced Techniques for Mortality and Longevity Prediction." *Journal of Artificial Intelligence Research and Applications* 3.2 (2023): 392-422.
2. Karamthulla, Musarath Jahan, et al. "From Theory to Practice: Implementing AI Technologies in Project Management." *International Journal for Multidisciplinary Research* 6.2 (2024): 1-11.
3. Jeyaraman, J., Krishnamoorthy, G., Konidena, B. K., & Sistla, S. M. K. (2024). Machine Learning for Demand Forecasting in Manufacturing. *International Journal for Multidisciplinary Research*, *6*(1), 1-115.
4. Karamthulla, Musarath Jahan, et al. "Navigating the Future: AI-Driven Project Management in the Digital Era." *International Journal for Multidisciplinary Research* 6.2 (2024): 1-11.
5. Karamthulla, M. J., Prakash, S., Tadimarri, A., & Tomar, M. (2024). Efficiency Unleashed: Harnessing AI for Agile Project Management. *International Journal For Multidisciplinary Research*, *6*(2), 1-13.
6. Jeyaraman, Jawaharbabu, Jesu Narkarunai Arasu Malaiyappan, and Sai Mani Krishna Sistla. "Advancements in Reinforcement Learning Algorithms for Autonomous Systems." *International Journal of Innovative Science and Research Technology (IJISRT)* 9.3 (2024): 1941-1946.

7. Jangoan, Suhas, Gowrisankar Krishnamoorthy, and Jesu Narkarunai Arasu Malaiyappan. "Predictive Maintenance using Machine Learning in Industrial IoT." *International Journal of Innovative Science and Research Technology (IJISRT)* 9.3 (2024): 1909-1915.

8. Jangoan, Suhas, et al. "Demystifying Explainable AI: Understanding, Transparency, and Trust." *International Journal For Multidisciplinary Research* 6.2 (2024): 1-13.

9. Krishnamoorthy, Gowrisankar, et al. "Enhancing Worker Safety in Manufacturing with IoT and ML." *International Journal For Multidisciplinary Research* 6.1 (2024): 1-11.

10. Perumalsamy, Jegatheeswari, Muthukrishnan Muthusubramanian, and Lavanya Shanmugam. "Machine Learning Applications in Actuarial Product Development: Enhancing Pricing and Risk Assessment." *Journal of Science & Technology* 4.4 (2023): 34-65.