Deep Learning for Real-time Pedestrian Intention Recognition in Autonomous Driving

By Dr. Paulo Sérgio

Professor of Informatics, University of Minho (UMinho)

1. Introduction to Autonomous Driving

[1]Autonomous driving is entering our lives in the form of commercial vehicles and private cars at an increasing pace. Many advanced features such as automatic parking, adaptive cruise control, and lane-keeping assistance are already available in most modern cars. Also, the first vehicles with conditional (ef.) or supervised automation are already commercially available in limited areas, and fully autonomous vehicles are being tested in many places around the world. At the same time, the possibilities of artificial intelligence and machine learning are constantly growing, partly due to many resources that are made public. In addition, machine learning algorithms that can handle image, audio, and video inputs are becoming more accurate and available. Consequently, it seems like a natural development to use such perception methods in autonomous vehicles to improve and develop the interaction between vehicles, drivers, and other road users [2].[3] One aspect of closely related research is understanding and predicting the intentions of vulnerable road users (VRU) in traffic (Stol, El Aad, & De Winter, 2019a). This is an essential part of the basic traffic tasks scanning, hazard detection, decision-making and driving behavior planning, as summarized by Dewitte et al. (2019). Consequently, there is a distinction between intentions, which need to be considered before driving actions, and actions that are currently observable. Both in detail analysis and in learning the surrounding traffic, the intention of the VRUs is therefore an important piece of information. The intention varies depending on the type of VRU. Cyclists can affect the traffic laterally, either when they overtake the ego-vehicle waiting at an intersection or get ahead of it while approaching. Their intention will hence influence if the vehicle can start or if another VRU conflicts and the ego-vehicle needs to wait. Similarly, the intentions of pedestrians will influence the stop-and-go decision. This is of particular interest for the area of autonomous vehicles and the human-machine interaction (HMI) between the VRUs and the automated vehicle.

1.1. Overview of Autonomous Vehicles

From the previous studies, it is observed that pedestrian generation from sensor data like camera and LiDAR is crucial to improve the pedestrian detection model for autonomous driving. Those models have different difficulty levels. In this paper [4], authors meticulously designed three synthetic data experiments and one mixed experiment to identify the feasibility and superiority of the proposed data generation scheme. Furthermore, data sources override the performance of all baseline detectors and thus, realistic and diverse LiDAR sensor data with different complexity must be considered to improve generalization. More importantly, in practical applications, since real sensor data cannot provide sufficient real pedestrian bounding box annotations, an adaptive synthetic LiDAR sensor data source provides a strategy to generate detectors under the guidance of a real LiDAR sensor model. The main constructing framework of the dataset in this study is the unified model of Carla-Uniform pedestrians and data of detectors are also very practical. In addition, the model is fed with corresponding data in each condition, which will greatly improve the model compared with real or synthetic only data sources.

Real-time pedestrian intention recognition in autonomous driving systems is crucial to prevent traffic accidents and thus protect pedestrians. A real-time pedestrian intention recognition algorithm is proposed in this article [2]. The algorithm can improve the safety of pedestrians and autonomous vehicles. The authors analyze the deep learning need in terms of statistical data dimension in the field of pedestrian intention recognition. The Pedestrian Intention Recognition Network (PIRN) first forwards the learned representation of the frames from the RTEnet into a GRU to obtain temporal context. These context-enriched representations are processed through a Softmax layer to obtain the probability distributions over the pedestrian actions. Performance of the algorithm was tested on the public and self-collected pedestrian external datasets, thereby depicting the robustness and generality of the proposed framework.

2. Pedestrian Intention Recognition in Autonomous Driving

[5] [6]The pedestrian intention recognition problem focuses on extracting deeper information from the pedestrian agent to understand what their potential next actions will be. It relates both to pedestrians' short-term behaviours, such as crossing a road, as well as to their long-term plans, according to their destination and route. Decision-making algorithm should be

capable of recognizing specific pedestrian actions (such as walking, standing, and chatting) and estimating crossing intention probability. Basic research have been conducted mainly in the sanitye community, with different levels of data coming from different modalities, such as LIDAR and based on pure visual modality. Occluded pedestrian action recognition remains a challenging problem, and real-time action recognition methods with high accuracy are an active research topic [7]. Potential Avenues for Future Work • Nowadays, due to the interaction need with artificial agents related to few-shot or open-set recognition, we are foreseeing open-set pedestrian action and crossing prediction problems. Specifically, we presume that a real-time method capable of achieving a good level of prediction even at the first observation would be developed. Future algorithms should be flexible enough to be easily customizable by automotive industry practitioners and researchers considering other stakeholder expectations, such as safety, fairness, legality, and ethics. Finally, it would be very interesting to integrate prediction results in high-level prediction aware planners, to study the potential benefit to safety that re-scheduling of trajectories and speed changes for surrounding agents due to a better anticipation would bring forward.

2.1. Importance and Challenges

The prediction of pedestrian trajectories is critical when trying to avoid collisions, especially when vehicle dynamics are considered into account. Many authors have proposed methodologies to address the above challenge of the real-time prediction of pedestrian crossing intentions. An opto-geometric tracker is used to track pedestrian trajectories, and road user kinematic and dynamic spatial interaction models are employed to predict pedestrian behaviors. In this study, we propose and evaluate a real-time pedestrian crossing intention prediction system. It is based on an instance of highly efficient image-based human skeleton estimation. We investigate the extent to which the observed pedestrian motion, and in particular, time samples covering approaching pedestrians, are predictive of pedestrian intention. Such a measure can be crucial for real-time pedestrian intention prediction in contrast to various vision-based prediction algorithms known from the literature [4].

The recognition of pedestrians who intend to cross the road is one of the most crucial tasks for safe and efficient driving with autonomous cars [8]. At urban intersections, people are often positioned between obstacles, partially occluded by parked cars or other structures, or walking in groups or pairs at non-right angles. In such conditions, sensors that are mounted

on a car, e.g., cameras or LiDAR, may exhibit limitations in successfully detecting and tracking pedestrians. As a result, in adverse weather conditions, high traffic, or poor sensor visibility, even today's state-of-the-art systems may sometimes miss recognizing pedestrians that are crossing in front of an approaching car. These scenarios might lead to serious collisions that cause injury or even loss of life. By obtaining knowledge of the environment and monitoring pedestrians, the behavior of an individual that has an intention to cross the road can be predicted. This can be achieved by using a predictive algorithm that estimates the future behavior of the pedestrian from their current movement pattern [7].

3. Deep Learning Fundamentals

Deep learning paradigm is the most successful variant of machine learning paradigm for learning meaningful representations of the data. The learning of the different abstraction levels and hidden structure in the data is achieved through optimization of a non-convex loss function (E.g. back-propagation of errors in neural networks). The deep learning paradigm achieves this through training very large models, like deep neural networks, with a large amount of data labelled and unlabelled both [9]. The deep learning paradigm is not new in the community of 202 real time intention prediction but it has shown significantly better performance compared to the traditional machine learning models. To this end, this chapter will first describe fundamental back-propagation through time algorithm, that is used to train a very specific deep learning architecture Recurrent Neural Networks and Long Short Term Memory models. Due to the importance of the interaction of the data that causes significant change in the behavior of the pedestrian this sub-machine will be discussed in significant detail.

In this chapter, we cover the essential theoretical and practical basis of deep learning, which is vital for understanding the whole system, specific components, and the methodology of training deep-learning-based models. This section will go over the necessary deep learning fundamentals as relevant to our proposed method of pedestrian intention recognition. We will describe the key concepts and the state-of-the-art descriptions of various architectures used in respect to pedestrian related solutions [10]. Since training these networks need a significant amount of data and often in autonomous vehicle setting labels are provided after a triggering (braking event in case of K&A data set), active learning has also been proposed in this chapter. Overall, a deep learning-based pedestrian intention recognition system will be

the result of this chapter, which will be implemented as a part of one of the two point of-module described in Section 5.3.

3.1. Neural Networks

Since pedestrian related hazards sometimes occur in dark areas or indoors, some studies have focused on capturing clearly visible pedestrians and expanding datasets for pedestrian intention recognition by generating virtual pedestrians [7]. It is pointed out that some pedestrians can be partially occluded in their data. The data can be very similar to the test dataset while including these instances, and these pedestrians can be classified as 'unsafe to cross' in the test stage. But the system may tell wrong results. Some of the research papers proposed a mixture of dataset real and virtual. In this study, the pedestrian poles were occluded with the virtual humans created using Mixamo website and included in the training set, and the performance was evaluated with test data containing the actual humans.

Although neural networks allow us to make predictions from input data, most of them so far are designed for traditional pattern recognition tasks. Existing pedestrian behavior recognition systems mainly employ the following deep learning architectures [2]. In most cases, single-frame pedestrian intention recognition problems are solved. However, pedestrian intention is progressive and continues while walking. Processing only a single image may not address the intent of the pedestrian. Hassen et al. proposed to combine long short-term memory and convolutional neural networks to recognize pedestrian's intention, distinguishing between pedestrian crossing and not crossing divisions. In their research, Ropero et al. evaluated the performance of a deep belief network for recognizing pedestrian intention based on 4 s of video sequences. In particular, Segu´ı et al. proposed an attention-based Pedestrian Crossing Decision (aPCD) framework that operates in the pedestrian's region of interest and filter out the non-relevant information.

4. Deep Learning Techniques for Pedestrian Intention Recognition

We observe that one of the typical ways of addressing this issue is to use off-the-shelf convolutional neural networks (CNNs) trained on a large dataset of images for view-dependent pedestrian localization. Partial occlusion can often still muddle the predictions and may lead to wrong predictions in localization on the detected bounding box [10]. More recently, deep object detectors like YOLO have been used to directly predict pedestrian's

poses for pedestrian orientation estimation while detecting pedestrians. Other works simultaneously estimate pedestrian's walking direction and intention in the view of an image using a deep neural network. Furthermore, attention mechanisms have been progressively gaining ground in deep learning, originating in the need to replicate human cognitive processes that focus on the essential areas of an image in the visual attention guidance and thus severely reduce computation cost to recognize the performance [11].

Systems for advanced driver-assistance and automation that are based on onboard perception need accurate and real-time reasoning about the surrounding environment to ensure safety. We observe that most existing works on planetary INTENTION recognition-aware ofnebarlyd, different human intention modes can occur in the same region of the spatial view and it is indeed important to take into account the pedestrian orientation information for intention prediction [12]. It may be particularly challenging to make correct predictions based on partial or heavily occluded information. This may occur frequently due to the line of sight from the vehicle to the pedestrian being obscured by occluding objects, including other pedestrians or vehicles, etc.

4.1. Convolutional Neural Networks (CNNs)

Different adaptations are possible when designing a CNN for intention and activity recognition, and CNNs adapted for these specific tasks are called Deep Learning-based Feature Extractors (DLFEs). Several papers reported how they were able to surpass the standard approach to classification recognition problems by using a DLF and how a plain DL feature extractor is too general for these specific applications [13]. The DLF uses mainly resized and cropped frames for the convo- lutional architecture's loaded images, while in the original paper, all images are resized or centrally cropped to the fixed size during the input and batch resize in CNN training. For example, a specific ad hoc DL adaptation was designed and tested for human actions detection and classification, named Pedestrian IntentionNet. Using DL introduced the requirement of a high amount of training data. This problem, combined with the possible bias incorrectly learned from training data, arises from the introduction of high complexity and the many parameters in the architectures. For this reason some researchers tried to constrain the model complexity.

Deep Convolutional Neural Networks (CNNs) are among the most popular architectures used in computer vision applications, from large-scale image recognition to pedestrian

detection [14]. Different versions and alternations of CNNs [15] have been implemented in a variety of pedestrian intention recognition problems recently. Recently, an intensive research effort has led to the development of powerful techniques for designing CNN architectures. Since their first successes, these architectures have been extended with many side mechanisms and components, including residual connections, multi-scale filters, dilation filters, spatial filters and kinds of temporal filters, and ensemble learning.

5. Real-time Processing in Autonomous Systems

Under this encapsulate, the DRDC-BN is dedicated to the computationally efficient real-time interaction of the pedestrian intents and takes real-time inference and real-time training into account. The adopted measures for the aforementioned elements are 43.24 ms for acknowledgment time in the inference mode and 31.69 ms for five consecutive iterations at the training mode; both, over PASCAL VOC 2012 data with 16 images per second top capacity. Since it is a run-time feature in an autonomous vehicle, the binary label relating noisy images are inversely utilized to speak about the treadmill noisy process of the real-world applications. From the real-time computational perspective, we additionally discuss that such a real-time motion state of the pedestrians might help, to a noticeable degree, the medicos for the injury estimation purposes), tends to be nowadays significantly used across many fields including automotive applications.

Deep learning architectures are complex and computational heavy, leading to the consumption of vast computational resources and power to facilitate their operations [16]. For instance, DAS (DenseNet as an image detector for scanning the potential pedestrians) net [17] uses 276,138 parameters in their network Afterward, it would imply to millions of floating points operations in the DAS interpretation to produce some final inputs for each pedestrian-vehicle pair. Regardless of how robust the pedestrian intention interpretation of an architecture is, the 'real-time' feature is a crucial requirement that must be fulfilled. For high-valued automotive applications such as autonomous driving systems (ADS), a significant necessity is to utilize the most handsome consequence in a real-time manner [18]. Therefore, the RDR (real world) scenarios raise challenges in designing real-time architectures especially those based on deep learning. A crucial consideration for this direction is the ability to design such architectures creating more compact and faster models. Procuring such characteristics not only alleviates the great demand for the computational resources and power but also it is

appealing for the minimization of the D/C (Data to Control) latency which is a necessary characteristic for the safety-critical automotive applications. Although the practical difference between the terms 'real-time' and 'low-latency' is well defined, they can be used interchangeably with a very minimal difference.

5.1. Hardware and Software Considerations

In order to design a pedestrian recognition model that can be applied to vehicles, a more detailed consideration is made for the practical application of pedestrian detection and recognition. From the perspective of software, we compare the over-detection rate of the Linear Base Template (LBP)-AdaBoost, Histogram Oriented Gradient (HOG)-Support Vector Machine (SVM) that are more suitable for micro-computers, as well as Faster-RCNN, Single-Shot Multibox Detector (SSD) and Device YoloV on the pascal VOC dataset. The over-detection rate of the pedestrian detection algorithm is compared with the detection rate of the pedestrian detection algorithm. From the aspect of hardware, the research on the light and sound features or quick characteristics of specific objects. [6]datasets such as the VOC pedestrian dataset, the INRIA pedestrian dataset, etc., are used for comparative analysis. Thus, the overall system framework is proposed, whether considering pedestrian + vehicle light dynamic recognition or vehicle dynamic recognition. The effects are ideal.

[19] [20] Most state-of-the-art methods for pedestrian intention recognition in recent years have focused on algorithms or mathematical models. However, little attention has been paid to the considerate considerations for practical design. The actual application of such systems is usually on vehicles, requiring such systems to consider the characteristics of various sensors and processors, as well as real-time performance and low power consumption. The following will elaborate on the design details of the entire perception module, including software algorithm design and hardware selection. By comparing traditional detection algorithms and feature extraction algorithms from three different perspectives, the performance and performance of the final model can be comprehensively compared. The Intel NUC6i5SYK microcomputer with an i5 CPU and 8GB memory is selected for the perception module. It is a relatively mature low-power consumption and relatively high-performance. Dual core i5 processor. Mobile detection for recognition algorithm here, the reason for choosing YOLOv3-LITE is that it is relatively mature, has good recognition accuracy and ensures real-time detection in the vehicle environment.

6. Datasets for Pedestrian Intention Recognition

[21] [4] A common dataset used for pedestrian detection and tracking is the Caltech Pedestrian dataset. This dataset was one of the early datasets published specifically for pedestrian detection, and it lists the behavior and location of pedestrians, and the size of bounding boxes around the pedestrians. A more recent pedestrian detection dataset is the KITTI benchmark. There are 12 sequences in total, and it is tested on different scenarios such as near-pedestrian, far-pedestrian, occlusion, crowded area pedestrian. The KITTI dataset tracks each pedestrian. The pedestrian tracks are provided as bounding boxes in each frame. Another dataset that is commonly used is the Mot16 dataset, which is a dataset about pedestrian detection and re-identificati. There is a relatively new dataset that is MOT20, which is similar to that of Mot16 and it is simplified that there lacks long range detection and there is no identity clue with known detection.[22] There are three new datasets for pedestrian intention recognition in autonomous vehicles. The TRAF dataset is a new, large, real-world traffic dataset recorded in Ann Arbor, Michigan, which contains a large number of motorized and non-motorized vehicles and pedestrians in realistic urban traffic scenes. The videos were recorded at various intersections with different traffic densities, and diverse weather conditions and lightings. We chose the Ann Arbor area to capture the diversity of traffic scenarios, and other than pedestrians, it contains a variety of buses, trucks, cars, cyclists, and motorcycles. The TRAF dataset was recorded in northern Kentucky at 15 road intersections across suburban and urban environment. In the TRAF dataset, there are pedestrian trajectories in high density traffic scenarios, and there are 3190 traffic light sequences annotated with pedestrian signal status. The TRAF dataset also provides personal demographic attributes of 899 pedestrians.

6.1. Kitti Dataset

More realistic predictions by real-time pedestrian intention recognition for autonomous driving. One solution is the usage of artificial neural networks in the form of deep learning, as these already can provide good predictions when well trained and are so hop f un easy to generalize f individuals in form of humans in different regions, not sufficiently covering global differences. Working on the Kitti dataset and real-time demonstrations in a 1:10 scaled autonomous car help to guide attempts towards practicability [23]. Of great interest is that deep learning models outperform individuals and are capable to approximate holism in human decision making.

The Kitti dataset is very popular in the 3d object detection community and is the most widely used benchmark and evaluation metric for the domain [24]. It involves mainly 3D, multi-modal LiDAR and images for data and 3D orientated object bounding boxes and 2D bounding boxes for annotations.Information of interest can be the metadata for every image that tells us which sensor setup was used to obtain the image, for example, which was the hardware on the vehicle or the type of the camera.Linear and rotational velocity for ground truth orientation would be helpful to base orientational errors on. Orientation of the object is additionally needed to calculate the overall bounding box orientation error.

7. Evaluation Metrics for Pedestrian Intention Recognition

Pedestrian intention recognition is used to grasp the pedestrian's state, velocity, as well as the proceeding acceleration in closer future – which is then communicated to the autonomous vehicle so that that awareness of for instance critical relative stopping distance situations is visible. [7]

The evaluation metrics comparing our proposed method and the latest state-of-the-art methods of pedestrian intention recognition are mean of average precision (mAP@0.5, mAP@0.7, and mAP@0.05), mean Recall, mean False Positive per Image (mFPI), mean Wall Time for inference, average part score, average segmentation score METvsHET, and average time for point-cloud file generation. 3D multi-modal point clouds for a typical urban scenarios are utilized in the analysis and three different types of straightforward lightweight networks are used for analysis; two Convolutional Recurrent Networks (CRN) ConvLSTM and possibly proposed customised 3D-YOLO. As a result, it can be seen from the Table 1 that volume halved point cloud equivalent 3D visual recognition (VPE-3D) of pedestrian standing still or walking speed

Pedestrian intention recognition is the basis for intelligent driving systems to understand the pedestrians' states and behaviors. As a result, the model's recognition accuracy should be prioritized. To evaluate the performance of the different models, this section will use commonly used metrics in pedestrian detection and crossing prediction tasks, such as Average Precision (AP) with different Intersection over Union (IoU) thresholds, and Receiver Operating Characteristics (ROC) curve. However, an improvement of pedestrian indication recognition should also lead to an improvement in pedestrian detection, thus in this part as well, starting from the establishment of the deep learning model, average precision evaluation

is obtained in the Caltech dataset, and then object-level colored heat map and part colored heat map analysis are performed on the model with optimized crossing recognition [25].

7.1. Intersection over Union (IoU)

In our ST-GCN + YOLO model, especially crafted schemes are needed to achieve good IoU measurements for both the spatial and temporal dimensions. For the spatial edge convolutions with 3 × 3 kernels, 3 convolution layers involve spatial information of the bounding boxes in a frame. L2 normalization after every spatial convolution layer enables more accurate learning based on object masks rather than spatial coordinates. In other words, edge convolution layers, compared to kernel transforms incorporated in the TCN, are effective in capturing this information and obtaining a higher IoU measurement. The combination of the hand-crafted Marginal Photo cross-Entropy loss and our ST-GCN feature hand-crafted loss allow our method to consider the critical time interval between the query frame and bounding box ground truth clip of the current and the future neighbourhood in relation to the actual duration of every actual query frame. By this way, ST-GCN and YOLO in the location of experiments and the geometric mean of detection IoUs is compared under different configurations.

'Intersection over Union' (IoU) measurement is a crucial criterion to evaluate object detection models in comparison to ground truth bounding boxes. It provides a geometrical measure on the intersection and union of the predicted and the actual regions [26]. Due to its inclusion of both the spatial information and the size of the bounding boxes in IoU measurements, model learning approaches, such as mean square, focal, or Geometric Cross Entropy loss, leverage it to learn object detection models as proposed in YOLOv2 and Single Shot MultiBox Detector (SSD). For the temporal domain, the classification results are also measured using IoU since the start and the end of the predicted action frames with respect to actual clips are crucial for optimal attention in feature map or loss computation [27].

8. Applications of Pedestrian Intention Recognition in Autonomous Driving

There are several numerous literature aiming a lot of vision design in intelligent systems. Convolutional Neural Networks (CNNs) play a critical role in establishing impact deep learning in computer vision. In intelligent transportation systems, attention mechanism in convolutional neural network plays a critical role. It effectively helps convolutional neural

network models focus on important object areas while suppressing irrelevant regions with useless details. In this way, it not only enhances the accuracy of object detection but also protects high-speed processing of traffic data in autonomous driving. Thus, under the current scenario and environment of connected vehicles and smart cities, most of the approaches and methods for intelligent transportation's large-scale processing of traffic data has been effectively and wisely developed and improved. Tell the story via machine learning regarding the processing of traffic data in intelligent transportation involves various well-known, trained, and popular models of AI to efficiently assist the intelligent transportation system and smart cities in recognizing various objects, identifying their dynamics, and assuring their real-time separation in order to assist traffic monitoring systems in reducing traffic congestion and accidents in cities.

[2] [10]Pedestrian intention recognition is the basis of human–computer interaction (HCI) in autonomous driving. Deep learning methods have shown promising results in recognizing pedestrian intention with an acceptable precision rate. The application of deep learning for vision sensor utilized in autonomous driving contributes to holistic pedestrian intention recognization. Moreover, there is a thorough investigation of the object detection algorithm trained on the LiDAR object detecting dataset in Dattroff. Despite all these studies regarding deep learning methods and their association with pedestrian intention recognition, applications of pedestrian intention recognition in autonomous driving stay incomplete. The efficient integration with sensor data, particularly the fusion of multiple sensors, is imperative for real-time implementation of pedestrian intention recognition in real-world application. Majority of the published paper by far focus on the AI methods for the enhancement of sensor detection of pedestrian intention. The main contribution of this comprehensive review article is to provide a holistic picture of the applications of current pedestrian intention recognition studies and the adopted multiple sensor fusion architectures.

8.1. Collision Avoidance Systems

In our research, we also considered collision kinematics to avoid false alarms and to generate more complex descriptions during processing [6]. Moreover, researchers are advise drivers to modify the positions of the acceleration pedal and the brake pedal; for example, if a pedestrian is crossing the road 3 s from the planned time. This system can significantly improve the safety of teenagers and drivers with different skill levels.

Two buffer sensors and two main sensors are used here to improve the detection accuracy [13]. As shown in Figure 23, the buffer sensors are also two LiDARs, one is set at a lower position than the main LiDAR, which can potentially detect pedestrians in a close distance; the other is set at a lower lateral position to the right of the car, with a wider field of view, to aid the detection of forward pedestrians. We also use two ultrasonic sensors and one rear-view camera for back information. Finally, three categories of detection information are fed into a fusion module, from which the final projected position on the road plane and the detailed caution suggestion are obtained. For robust IoT device security, see Shaik, Mahammad, et al. (2018) on RBAC implementation.

The final piece of our framework to consider is the collision avoidance reaction module [4]. The depth map from the SegNet module as well as the detected pedestrian/vehicle information is fed into this module. The reaction module provides 3 types of estimated parameters, which are 1) the projected position of the pedestrian/vehicle on the road plane, 2) the distance and relative velocity from the ultrasonic sensors, and 3) caution suggestion to the central controller. A general architecture of the reaction module is illustrated in this. The model receives the raw LiDAR (Light Detection And Ranging) point cloud which captures the 3D information of potential obstacles.

9. Ethical Considerations in Autonomous Driving

Ethical considerations must be weighed in advocating the use of pedestrian movement intentions to support decision-making algorithms of autonomous vehicles. It is essential to guarantee pedestrian rights (e.g., right of way) and safety, and that pedestrians can confidently read the intentions of autonomous vehicles (integration of intentions into human communication strategies: signs, sounds, etc.). Furthermore, the possible invasion of privacy by detecting pedestrian movement intentions must be considered. An explicit cost-sensitive perception module, based on cues that are most relevant from an ethical point of view, is a feasible way to address ethical constraints in pedestrian movement intention recognition. Ethical considerations have already been integrated into decision making with respect to active vehicle safety systems, dead-lock situations when priority rules fail, and legal provisions about algorithm components. These ethical considerations, such as car ethics, important properties, and cost functions should be transferred to the context of pedestrian movement intention recognition to include pedestrians as responsible agents for behavior,

add a cooperative and respectful human–machine interaction, and thus implement the connected and automated road transport systems with the goal of increasing road safety.

Endowing autonomous vehicles with automated interaction can improve driving safety and human–machine relationships. The cognitive basis of pedestrian intention recognition should be integrated in new developments in pedestrian behavior models and improved perception models of autonomous vehicles [28]. Future research should consider the needs for more refined distinctions of classes of intention and more frequent scenario reparation for more naturalistic studies in future.

Pedestrian movement intentions in the AD context: The importance with respect to ethical considerations [29], decision making perspectives, attention allocation and human response [30].

9.1. Bias and Fairness

This module's structure is following an encoder-decoder-based architecture, where the encoder builds a representation of the traffic participant image, while the decoder transforms this into the questions that the attention module can respond to. The attention module's goal is to draw attention over the whole latent representation to understand the most probable next action that same scene will contain. After training over the image data the landmarks (spots of attention) in the pretrained model, were retained, and bypassed the decoder to use this latent feature space structure to improve the model's prediction. Main focus were paid to the accuracy metric of the intention class, including strategies on how to change the system so a human can thrive more with the system based on the analysis of equality between the demographic attributes of humans. With safety prediction scores performances, F1 score was also put to the test. Solutions were produced for the distribution of this metric for improvement in a real world scenario. [4] In future, the study aims to combine this part of recogniser and the pedestrian detector part to produce a full system working under real-time scenarios, as well as on a larger variety of data.

Vehicle perception is important in development of the autonomous vehicles, as it is responsible for detecting the different types of traffic participants from its environment. Because the autonomous vehicles are to interact with the pedestrians on the main road, it needs to recognise the intentions of pedestrians so that the vehicle can decide what actions to

take, for example slowing down so that the vehicle can avoid posing danger to the pedestrians. In this paper, a model for real-time pedestrian intention recognition has been implemented. The model leverages the concept of attention mechanism for focusing on the behaviour of pedestrian in a video frame, which has been extracted by utilising a state-of-the-art object detection algorithm for processing the videos [31].

10. Conclusion and Future Directions

Future work will focus on incrementally optimized pedestrian intention recognition methods for mixed and virtual reality datasets. In the human–machine interaction, Stuttgart Grid, the VE$TIVAL dataset, and VMR datasets will be considered. The proposed method will be applied to Virtual Reality and Mixed Reality differently and using convolutional neural networks only in Virtual Reality. The proposed method assumes that a pedestrian can participate in performing a range of behaviors. However, it is noticeable that mixed and virtual realities are defined primarily by the behavior of the pedestrian toward the observer (the sensors in the autonomous vehicle). Rendering the superclass of all conduct possibilities in this context might increase the achievable performance without the risk of overfitting or overengineering systems for the pedestrian intention recognition as used in the presented method. In addition, An approach to evaluate pedestrian intention recognition performance for both real-time pedestrian intent potential and mixed/virtual reality datasets will be developed and validated [32].

Pedestrian intention recognition is crucial in autonomous driving systems to ensure safe interactions between autonomous vehicles and pedestrians. In summary, this paper proposed a real-time pedestrian intention recognition method based on the Vehicular Mixed Reality (VMR) system. Using a publicly available real-world dataset, the method was evaluated for pedestrian intention recognition. The VMR applies LSTM-based CV models enriched by HMM. For the four crosswalk-related pedestrian intent estimation tasks, the proposed method achieves improvements in AUC ROC (up to 3.04%). The results demonstrate that the proposed method outperforms the best-performing baselines in all tasks, including the most recent MARS benchmark [5].

10.1. Summary of Key Findings

Pedestrian detection and recognition directly rely on the quality/sparsity of sensor data. Autonomous driving systems for extracting pedestrian crossing intention should first include reliable sensing technologies such as global positioning system (GPS), inertial measurement unit (IMU), lidar, and radar to efficiently capture pedestrians and their behavior. LiDAR sensor has become a primary technology for detecting nearby obstacles and humans in an advanced driver assistance system (ADAS), as it provides reliable range data, is robust when it comes to the different appearance of pedestrians, and performs equally well in all lighting conditions, including the night. They try to extract the minimal features required for safe and efficient driving, which may not be suitable for analyzing the complexity in pedestrian-vehicle interaction, and therefore used for our intention recognition. The proposed study paid attention not only to the person's final decision to go or not go, but also to model fine-grained interaction patterns among pedestrians, other vehicles, different crosswalk regions, and their implicit behaviors, to map it to the whole decision-making process [8].

A prototype of a real-time advanced pedestrian intention recognition methodology which enables the improvement of the effectiveness of autonomous vehicles is presented in this chapter. Rather than simply extracting potential and risky pedestrian locations around the ego vehicle, the objective has been to anticipate a pedestrian's intention at a particular time horizon. To that end, nine features representing the dynamic state of the pedestrian have been annealed. The proposed sequence-to-sequence model, which systematically models the temporal contextual information, can achieve excellent results both in terms of accuracy and real-time applicability [33].

References:

[1] Tatineni, Sumanth. "Federated Learning for Privacy-Preserving Data Analysis: Applications and Challenges." *International Journal of Computer Engineering and Technology* 9.6 (2018).

[2] Shaik, Mahammad, et al. "Granular Access Control for the Perpetually Expanding Internet of Things: A Deep Dive into Implementing Role-Based Access Control (RBAC) for Enhanced Device Security and Privacy." *British Journal of Multidisciplinary and Advanced Studies* 2.2 (2018): 136-160.

[3] 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.

[4] P. Jabłoński, J. Iwaniec, and W. Zabierowski, "Comparison of Pedestrian Detectors for LiDAR Sensor Trained on Custom Synthetic, Real and Mixed Datasets," 2022. ncbi.nlm.nih.gov

[5] A. Ranga, F. Giruzzi, J. Bhanushali, E. Wirbel et al., "VRUNet: Multi-Task Learning Model for Intent Prediction of Vulnerable Road Users," 2020. [PDF]

[6] H. Zhang, Y. Liu, C. Wang, R. Fu et al., "Research on a Pedestrian Crossing Intention Recognition Model Based on Natural Observation Data," 2020. ncbi.nlm.nih.gov

[7] H. Fu, L. Sun, Y. Shen, and Y. Wu, "SDR-GAIN: A High Real-Time Occluded Pedestrian Pose Completion Method for Autonomous Driving," 2023. [PDF]

[8] E. Moreno, P. Denny, E. Ward, J. Horgan et al., "Pedestrian Crossing Intention Forecasting at Unsignalized Intersections Using Naturalistic Trajectories," 2023. ncbi.nlm.nih.gov

[9] M. Mobaidul Islam, A. Al Redwan Newaz, and A. Karimoddini, "A Pedestrian Detection and Tracking Framework for Autonomous Cars: Efficient Fusion of Camera and LiDAR Data," 2021. [PDF]

[10] B. Ilie Sighencea, R. Ion Stanciu, and C. Daniel Căleanu, "A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction," 2021. ncbi.nlm.nih.gov

[11] J. Cao, C. Song, S. Peng, S. Song et al., "Pedestrian Detection Algorithm for Intelligent Vehicles in Complex Scenarios," 2020. ncbi.nlm.nih.gov

[12] M. Azarmi, M. Rezaei, H. Wang, and S. Glaser, "PIP-Net: Pedestrian Intention Prediction in the Wild," 2024. [PDF]

[13] J. Lorenzo, I. Parra, F. Wirth, C. Stiller et al., "RNN-based Pedestrian Crossing Prediction using Activity and Pose-related Features," 2020. [PDF]

[14] T. Zhang, Z. Han, H. Xu, B. Zhang et al., "CircleNet: Reciprocating Feature Adaptation for Robust Pedestrian Detection," 2022. [PDF]

[15] R. Giuliano, F. Mazzenga, E. Innocenti, F. Fallucchi et al., "Communication Network Architectures for Driver Assistance Systems," 2021. ncbi.nlm.nih.gov

[16] D. Yang, H. Zhang, E. Yurtsever, K. Redmill et al., "Predicting Pedestrian Crossing Intention with Feature Fusion and Spatio-Temporal Attention," 2021. [PDF]

[17] K. Saleh, M. Hossny, and S. Nahavandi, "Real-time Intent Prediction of Pedestrians for Autonomous Ground Vehicles via Spatio-Temporal DenseNet," 2019. [PDF]

[18] C. Zhang, R. Li, W. Kim, D. Yoon et al., "Driver Behavior Recognition via Interwoven Deep Convolutional Neural Nets with Multi-stream Inputs," 2018. [PDF]

[19] Y. Zhang, A. Zhou, F. Zhao, and H. Wu, "A Lightweight Vehicle-Pedestrian Detection Algorithm Based on Attention Mechanism in Traffic Scenarios," 2022. ncbi.nlm.nih.gov

[20] J. R. Peters, "Singly Generated Radical Operator Algebras," 2023. [PDF]

[21] H. Kataoka, Y. Satoh, Y. Aoki, S. Oikawa et al., "Temporal and Fine-Grained Pedestrian Action Recognition on Driving Recorder Database," 2018. ncbi.nlm.nih.gov

[22] R. Trabelsi, R. Khemmar, B. Decoux, J. Y. Ertaud et al., "Recent Advances in Vision-Based On-Road Behaviors Understanding: A Critical Survey," 2022. ncbi.nlm.nih.gov

[23] F. Camara, N. Bellotto, S. Cosar, F. Weber et al., "Pedestrian Models for Autonomous Driving Part II: High-Level Models of Human Behavior," 2020. [PDF]

[24] F. Manfio Barbosa and F. Santos Osório, "Camera-Radar Perception for Autonomous Vehicles and ADAS: Concepts, Datasets and Metrics," 2023. [PDF]

[25] D. Tian, Y. Han, B. Wang, T. Guan et al., "A Review of Intelligent Driving Pedestrian Detection Based on Deep Learning," 2021. ncbi.nlm.nih.gov

[26] M. Ahmed, K. Azeem Hashmi, A. Pagani, M. Liwicki et al., "Survey and Performance Analysis of Deep Learning Based Object Detection in Challenging Environments," 2021. ncbi.nlm.nih.gov

[27] F. Piccoli, R. Balakrishnan, M. Jesus Perez, M. Sachdeo et al., "FuSSI-Net: Fusion of Spatio-temporal Skeletons for Intention Prediction Network," 2020. [PDF]

[28] Z. Fang, D. Vázquez, and A. M. López, "On-Board Detection of Pedestrian Intentions," 2017. ncbi.nlm.nih.gov

[29] R. Chan, R. Dardashti, M. Osinski, M. Rottmann et al., "What should AI see? Using the Public's Opinion to Determine the Perception of an AI," 2022. [PDF]

[30] C. F. Wu, D. D. Xu, S. H. Lu, and W. C. Chen, "Effect of Signal Design of Autonomous Vehicle Intention Presentation on Pedestrians' Cognition," 2022. ncbi.nlm.nih.gov

[31] S. Mahmud Khan, M. Sabbir Salek, V. Harris, G. Comert et al., "Autonomous Vehicles for All?," 2023. [PDF]

[32] W. Morales Alvarez, F. Miguel Moreno, O. Sipele, N. Smirnov et al., "Autonomous Driving: Framework for Pedestrian Intention Estimationin a Real World Scenario," 2020. [PDF]

[33] P. J. Navarro, C. Fernández, R. Borraz, and D. Alonso, "A Machine Learning Approach to Pedestrian Detection for Autonomous Vehicles Using High-Definition 3D Range Data," 2016. ncbi.nlm.nih.gov