A Virtual Reality Framework for Human-Driver Interaction Research: Safe and Cost-Effective Data Collection

Luca Crosato
luca.crosato@northumbria.ac.uk
Northumbria University
Newcastle upon Tyne, United Kingdom

Chongfeng Wei
c.wei@qub.ac.uk
Queen's University Belfast
Belfast, United Kingdom

Edmond S. L. Ho
Shu-Lim.Ho@glasgow.ac.uk
University of Glasgow
Glasgow, Scotland, United Kingdom

Hubert P. H. Shum
hubert.shum@durham.ac.uk
Durham University
Durham, United Kingdom

Yuzhu Sun
y.sun@qub.ac.uk
Queen's University Belfast
Belfast, United Kingdom

ABSTRACT

The advancement of automated driving technology has led to new challenges in the interaction between automated vehicles and human road users. However, there is currently no complete theory that explains how human road users interact with vehicles, and studying them in real-world settings is often unsafe and time-consuming. This study proposes a 3D Virtual Reality (VR) framework for studying how pedestrians interact with human-driven vehicles. The framework uses VR technology to collect data in a safe and cost-effective way, and deep learning methods are used to predict pedestrian trajectories. Specifically, graph neural networks have been used to model pedestrian future trajectories and the probability of crossing the road. The results of this study show that the proposed framework can be for collecting high-quality data on pedestrian-vehicle interactions in a safe and efficient manner. The data can then be used to develop new theories of human-vehicle interaction and aid the Autonomous Vehicles research.

CCS CONCEPTS

• **Computer systems organization** → *Robotic autonomy*; • **Computing methodologies** → **Neural networks**; **Virtual reality**.

KEYWORDS

Virtual Reality, Autonomous Driving, Human-Robot Interaction, Unreal Engine.

ACM Reference Format:
Luca Crosato, Chongfeng Wei, Edmond S. L. Ho, Hubert P. H. Shum, and Yuzhu Sun. 2024. A Virtual Reality Framework for Human-Driver Interaction Research: Safe and Cost-Effective Data Collection. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3610977.3634923

HRI '24, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0322-5/24/03…$15.00
https://doi.org/10.1145/3610977.3634923

1 INTRODUCTION

The development of self-driving cars presents new challenges, especially in how they interact with human drivers or pedestrians. Existing models struggle to integrate the complex behavior of both parties, resulting in overly cautious AV (Autonomous Vehicle) behavior and difficulty for other road users to predict AV intentions [21]. At the same time, AVs will be sharing the road with human road users (HRUs) for the foreseeable future, which could lead to traffic issues if AVs do not understand how to interact with HRUs effectively. Unfortunately, there are currently no complete theories that explain how HRUs and AVs interact with each other. To ensure that AVs blend in smoothly with traffic, they need to model their surroundings and predict the movements of nearby traffic participants.

Pedestrian safety has always been a major concern on the roads. With the rise of AVs, it becomes even more crucial to understand pedestrian behavior to ensure their safety. As stated in [10], pedestrians are at a high risk of accidents, accounting for over 22% of all road traffic fatalities in the European Union in 2013. Moreover, the behavior of pedestrians on the road is highly variable, depending on several factors such as their age, gender, culture, and even mood. For instance, children may be more prone to distraction while elderly individuals may walk more slowly and need more time to cross the road. Therefore, investigating pedestrian behavior and identifying patterns can help AVs adapt to different scenarios and minimize the risk of accidents. By analyzing pedestrian behavior, we can develop better technology that can prevent accidents and protect the most vulnerable road users.

According to research in [13], the kinematics and signalling information of autonomous vehicles (AVs) play a crucial role in influencing pedestrian behavior, particularly since there is no driver involved. Therefore, it is important to identify the specific motion cues or signals that have the most significant impact on pedestrian behavior, as this holds significant research value.

Previous research has generally agreed that the distance or time to collision (TTC) between vehicles and pedestrians is the primary kinematic cue that influences pedestrian behavior [12]. However, a recent study has shown that pedestrians use multiple sources of information from vehicle kinematics instead of relying solely on one cue. The impact of speed, distance, and TTC on pedestrian behavior

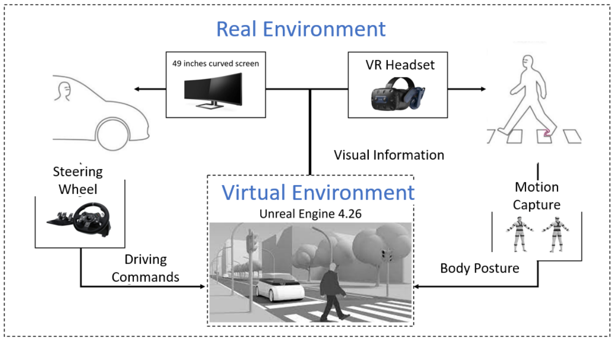


Figure 1: Overview of our VR Environment. The VR environments allow AVs or human-driven vehicles to virtually interact with a human-pedestrian.

is mutually coupled [25]. Moreover, in pedestrian-vehicle interactions, evidence suggests that the driving manoeuvre of the driver like deceleration, plays a critical role in affecting pedestrian behavior. Vehicle movements are linked to pedestrian trust in vehicles, emotions, and influence [17].

With the advancement of machine learning techniques, learning the aforementioned complex behaviors is an intuitive solution. However, acquiring data to study pedestrian movement can be both difficult and costly. Although near-collision events are crucial to developing accurate pedestrian models, analyzing them in the real world can be dangerous. Fortunately, recent advancements in Virtual Reality (VR) technology have enabled the creation of virtual environments that provide a safe and cost-effective means of collecting data for Autonomous Driving (AD) studies [27].

In this paper, we propose a VR road simulator (Figure 1) to study pedestrian behavior and decision-making. To study the reciprocal interactions between driver and pedestrian, we will have the pedestrian wear a VR headset that immerses them in a virtual environment. A human driver will sit in front of a screen and use a steering wheel and pedals to control the car. While this study focuses on the interactions between a single human driver and pedestrian, we design the system so that it can easily be extended to more cluttered environments with multiple vehicles and pedestrians. We validate the system with a data collection experiment and perform a deep learning based analysis of pedestrian motion.

The contributions of this research work are the following:

- the development of a traffic simulator, with a wireless HMD device (HTC Vive) that allows the users freedom of movement in combination with a motion capture system;
- a framework where pedestrians and user-controlled vehicles can coexist with each other, as well as with Autonomous Vehicles. This framework can be used in the future to aid Autonomous Driving research for validation and testing.

2 RELATED WORK

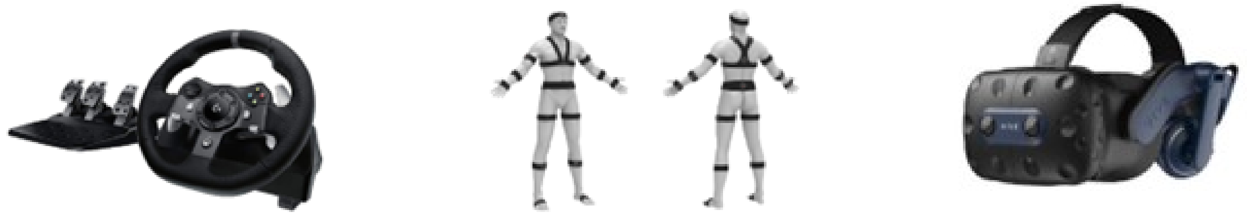
2.0.1 Virtual Reality Systems. Virtual Reality (VR) has several advantages compared to real-world tests in AD. VR allows for a safe and cost-effective study and data collection of interactions. A systematic review from [19] shows that the number of Augmented Reality and Virtual Reality papers with applications in AD has been increasing in the most recent years with an increasing interest in Vulnerable Road Users (VRUs), such as pedestrians or cyclists. VR has several advantages over field data collection for AD research. First of all, safety is ensured thanks to the fact that traffic participants interact within a virtual environment. Collision and near-collision scenarios no longer constitute a problem for VRUs. The environment is entirely controllable, which can be useful for studying specific scenarios or testing different hypotheses. Besides, VR is a more cost-effective way to collect data because it does not require the use of real vehicles and expensive sensors.

A systematic review on VR Studies on AV-Pedestrian Interactions can be found in [26]. Studies [2, 3] have focused on pedestrian gap acceptance behaviour and have shown that no statistically significant differences are present between the virtual and the real environments in pedestrians' intention to cross, demonstrating the efficacy of VR simulations in replicating realistic pedestrian crossing behavior within immersive virtual environments that closely resemble real-world locations. Notably, studies such as [6] have utilized VR environments to investigate communicating features between pedestrians and autonomous vehicles.

VR has also been employed [16] to develop a pedestrian simulator AR-PED that allows the users total freedom of movement without any physical boundary restrictions. [7] Doric et al. developed a pedestrian simulator for research on pedestrian crossing behavior, risk acceptance and as well investigation of the pre-crash phase under defined and reproducible conditions without the risk of harming real human test subjects.

In our work, we have developed an AV-pedestrian simulator using Unreal Engine 4.26 based on HTC Vive Head Mounted Display. Our simulator allows two simultaneous users: a human driven vehicle and a pedestrian. The simulator can be used to carry out research on pedestrian-driver interactions, as well as for testing Autonomous Driving algorithms with humans in the loop in a safe manner. Unlike previous systems, our approach incorporates multiple sensors, enabling real-time streaming of pedestrian positions within the Virtual Reality world. This enhances the realism of interactions between human drivers and pedestrians in our study. The 3D data availability can prove beneficial for future research, particularly in designing advanced Autonomous Vehicle (AV) systems that consider pedestrian poses.

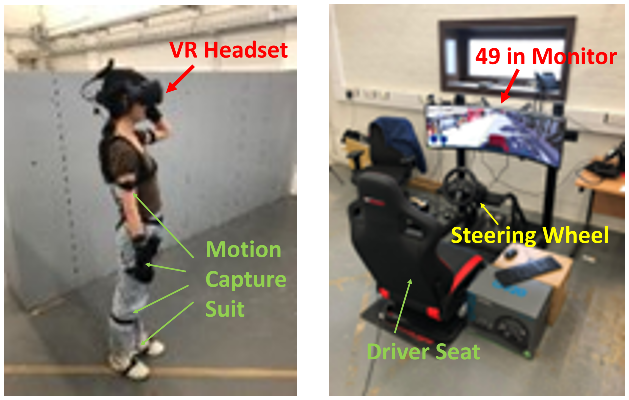
2.0.2 Trajectory Prediction. Pedestrian trajectory prediction is vital for autonomous driving as it enables vehicles to anticipate and react to pedestrian movements, enhancing safety by preventing accidents and ensuring smooth, efficient driving. For instance, [11] proposed a Graph Convolutional Neural Network-based pedestrian trajectory prediction model for generic AV Use Cases. This model used past pedestrian trajectories as the inputs to predict deterministic and probabilistic future trajectories. Other similar models aimed to improve prediction accuracy by considering the social context of interactions. For example, [8] proposed an LSTM pedestrian trajectory prediction model, which considered past trajectories, pedestrian head orientations, and distance to the approaching vehicle as the inputs to the model, as pedestrian head orientations and distance to approaching vehicles may be correlated with pedestrian awareness and perceived collision risks [17].



(a) Logitech G920 Steering Wheel and Pedals. **(b) Perception Neuron Motion Capture System.** **(c) HTC Vive Pro 2.**

Figure 2: Hardware components used for capturing data (a and b) and visual feedback to the pedestrian (c).

We consider deep learning based methods for estimating the joint trajectory of both the pedestrian and the car, in order to predict interactions. Deep learning is becoming more and more common in predicting pedestrian trajectories because of its impressive ability to represent data. In particular, the Social-LSTM [1] uses Recurrent Neural Networks (RNNs) to model the trajectory of each pedestrian in combination with a social-pooling operation to consider surrounding agents. Another approach in modeling human-human interaction for pedestrian trajectory prediction is to use graphs, as they can better capture the structure of the scene. Social-BiGAT [9] uses a combination of LSTM to model the trajectory of each pedestrian and Graph Attention Network (GAT) to model their interactions. In our work we decided to use STGCNN network, in order to be able to include multiple agents in the future, we used Social-STGCNN [14], which represents trajectories as a spatio-temporal graph and adapted the network architecture to the problem we are solving.



(a) Pedestrian. **(b) Driver system.**

Figure 3: a) A pedestrian wearing the VR and Motion Capture system, b) the driver setup.

3 METHODOLOGY

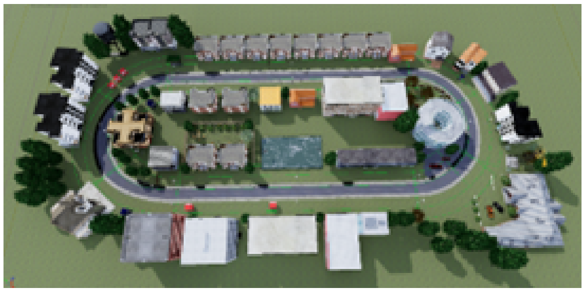
This study introduces a new simulation framework (Figure 1) that allows users to control pedestrians and vehicles in a virtual environment. The framework is powered by the latest advances in computer hardware, software, and networking. The following sections will detail the system design and key components of the system. We also conducted a data collection experiment where we invited participants to evaluate the system and collect their trajectory data. VR allows developers to create realistic simulations of pedestrians in a variety of environments. This allows them to test autonomous vehicles in a safe and controlled environment, and to collect data on how pedestrians interact with autonomous vehicles. This data can then be used to improve the safety and performance of autonomous vehicles. The trajectory data will be analysed with deep learning techniques to develop a trajectory prediction model, which can aid the development of pedestrian simulators as well.

3.1 Design of the Virtual Environment

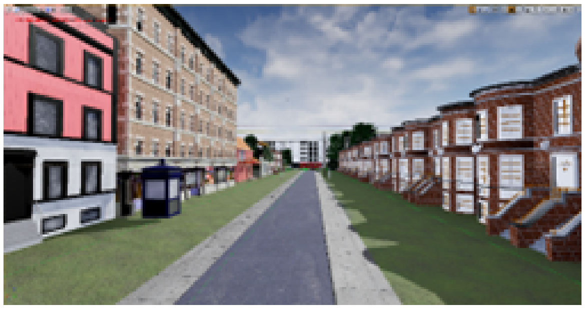
Developing a virtual reality environment requires combining together a lot of different pieces of hardware and software. In this subsection, we will go through the system components and design. The current system allows a pedestrian and user-controlled vehicle access to the virtual environment at the same time and the setup is illustrated in Figure 3.

3.1.1 Hardware Components. The pedestrian hardware components consist of a Perception Neuron Motion Capture Suit (Fig. 2b) and an HTC Vive Pro 2 Virtual Reality headset (see Fig. 2c). The Perception Neuron Motion Capture Suit is a wearable system that tracks the position and orientation of the body's major joints. It obtains a pedestrian posture representation which is transferred to the virtual environment to create a realistic digital representation of the virtual pedestrian (pedestrian avatar). The HTC Vive Pro 2 VR headset is a high-end VR headset that provides a sharp, high-resolution image. It also has a wide field of view, which helps to create a more immersive experience. Figure 3a shows a participant wearing the mocap suit and the VR headset. The VR headset is mounting a wifi-adapter that allows the user to move freely, without being tethered to the desktop computer.

A 49-inch screen is used to display the virtual reality environment to the driver. The driver is controlling the virtual vehicle with a Logitech G920 Steering Wheel and Pedals (see Fig. 2a). The Logitech G920 Steering Wheel and Pedals are a high-quality driving simulator set that provides realistic feedback to the driver. This helps the driver to feel like they are actually driving a real vehicle. A driver seat is also used to make the overall experience more similar to a real vehicle, resembling a driver video game. The driver seat can be adjusted to give a comfortable experience to different drivers. Our current lab allows us to operate in an area of approximately 7 m × 7 m and the current SteamVR Base Station we are using allows an area of up to 10 m × 10 m. This effectively limits the area where



(a) Top view of our map.



(b) Single lane view.

Figure 4: Screenshots of the virtual environment.

the pedestrian can move, thereby impacting the maximum size of the virtual environment in which the virtual pedestrian can operate. For this reason, we focused on a single-road environment in this first study and we will consider multi-lane scenarios when possible.

The VR environment runs on a desktop computer with an NVIDIA GeForce RTX 3060 Ti GPU and 11th Gen Intel(R) Core(TM) i7-11700 @ 2.50GHz GPU. To keep the costs low we connect both the driver and the pedestrian to the same machine but it is also possible to have multiple machines connected in a client-server architecture.

3.1.2 Software Components. The VR environment was designed with Unreal Engine (UE) 4.26 software, a powerful game engine used to create realistic and immersive virtual worlds. It provides a wide range of features that are specifically designed for VR development and easily allows the integration and development of VR games. The motion capture raw information is processed by Axis Neuron, the software provided by Perception Neuron. Axis Neuron allows streaming body posture to Unreal Engine 4.26 via TCP/IP connection and also provides an UE 4.26 plugin that can animate pedestrian avatars. The VR headset is also connected to UE 4.26 via SteamVR, a virtual reality platform. One of the main technical issues faced when combining the VR headset with the motion capture system is given by the fact that they use two different reference systems. Both the VR headset and the Perception Neuron can stream pedestrian position information to UE. We decided to rely on the VR headset's head position for mapping pedestrian's location from the real world to the virtual world, and use the motion capture system solely for animating the avatar's body posture. This is because the motion capture unit that is placed on the head often gives inaccurate head orientation information due to interference with the VR headset placed on top of it.

The driving commands can be sent to pre-built vehicles in UE 4.26 directly. However, this does not allow for the change in vehicle dynamic parameters and does not give access to dynamical information such as vehicle acceleration, angular acceleration or friction. To build a customized vehicle with full access to the dynamic model, we simulate the vehicle dynamics with a Python script, which captures the driver commands and updates the vehicle position in UE 4.26. UE 4.26 provides information such as ground surface structure and vehicle model parameters that allow the Python Script to update the vehicle dynamics. Finally, the positions of pedestrians and vehicles are both mapped into the UE 4.26 virtual environment's world Cartesian coordinate system.

3.1.3 Map Design. As already mentioned, we used Unreal Engine 4.26 to develop a virtual urban environment. We are interested in studying interactions with a vehicle and a pedestrian when the pedestrian is attempting to cross the road without any road signs (no zebra crossing or traffic lights). We designed a single-lane environment where with a loop-shaped road structure (see Fig. 4a). The shape of the map will be very useful during the data collection experiments, as it allows the driver to constantly drive without having to restart the simulation after each interaction episode is over. We built a 3.65m wide single-lane road (see Fig. 4b) and added some buildings and trees to make it resemble a realistic road. The road includes a combination of straight and curved sections, and at the turning corners, we added buildings of varying shapes and sizes. This diverse setup allows us to collect data across a range of scenarios. Note that the size of all the meshes in the virtual environment, such as pedestrians, vehicles, buildings, and trees, is adjusted to a 1:1 scale with the real world. This is essential because the surrounding environment significantly influences the driver and pedestrian's perception of their own speed. No obstacles were added near the pavement as we are not interested in studying the effect of occlusion on driver and pedestrian decisions. After setting up the whole simulation system, drivers sitting in front of the driving simulator can see the pedestrian in the virtual traffic environment on the screen while the pedestrian can see the car operated by the driver in the VR headset.

3.2 Trajectory Prediction

Trajectory prediction is important for autonomous driving because it allows the vehicle to anticipate the movements of other traffic participants, such as vehicles, pedestrians, and cyclists. This information is essential for the vehicle to make safe and timely decisions, such as when to brake, change lanes, or accelerate.

The problem can be formulated as follows: we have a group of N agents (vehicles and pedestrians) whose trajectory is observed over a time period T_o, and we want to predict their future trajectories over a time horizon T_p. For each agent n, we represent their trajectory as a set of 2D coordinates (x_n^t, y_n^t) for each time step t. We assume that the distribution of these coordinates follows a bivariate Gaussian distribution, denoted as $p_n^t \sim \mathcal{N}(\mu_n^t, \sigma_n^t, \rho_n^t)$. Our goal is to estimate the parameters of this distribution $(\mu, \sigma, and \rho)$ and minimize the negative log-likelihood to improve the accuracy of

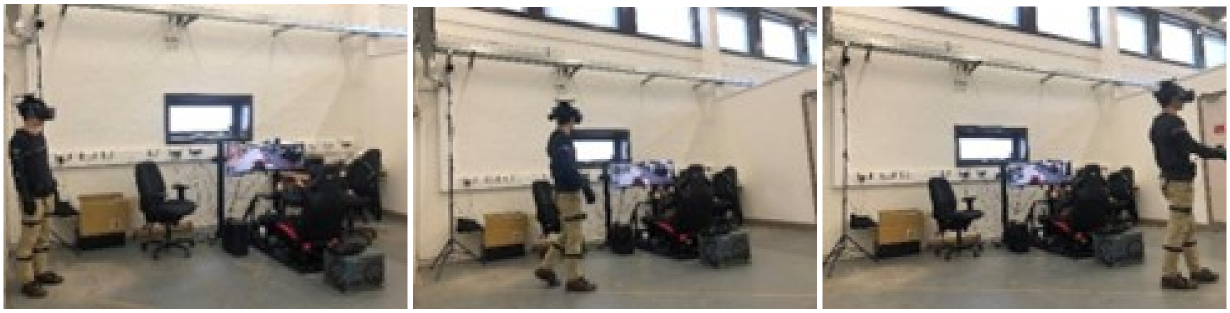


Figure 5: A pedestrian and a driver during a data collection experiment.

our predictions. We denote the predicted trajectory as $\hat{p}_n^t$, which follows the estimated bi-variate Gaussian distribution $\mathcal{N}(\hat{\mu}_n^t, \hat{\sigma}_n^t, \hat{\rho}_n^t)$. The model is trained to minimize:

$$L^n(\mathbf{W}) = -\sum_{t=1}^{T_p} \log\left(\mathcal{P}\left(\mathbf{p}_t^n | \hat{\mu}_t^n, \hat{\sigma}_t^n, \hat{\rho}_t^n\right)\right), \tag{1}$$

where $\mathbf{W}$ are all the trainable parameters of the model, $\hat{\mu}_t^n$, $\hat{\sigma}_t^n$, $\hat{\rho}_t^n$ are the mean, variance and correlation of the distribution.

3.2.1 Pedestrian Forecasting Architectures. We use the network proposed by [14] for the task of trajectory prediction. The Social-STGCNN model comprises two main components: the Spatio-Temporal Graph Convolution Neural Network (ST-GCNN) and the Time-Extrapolator Convolution Neural Network (TXP-CNN). ST-GCNN performs spatiotemporal convolutions on the graph representation of pedestrian trajectories, extracting compact features representing observed trajectory history. Subsequently, TXP-CNN utilizes these features to predict future trajectories collectively for all pedestrians, earning its name as a Time-Extrapolator due to its role in extrapolating future trajectories through convolution operations.

We have adapted the network architecture proposed in [14] to our problem. The framework is shown in Fig. 6. In particular, the network has been designed to predict pedestrian motion in crowded scenarios. We chose this network architecture because it has shown excellent performance in trajectory prediction tasks and can be easily extended to include more pedestrians and vehicles in the future. Choosing a neural network architecture for its ability to accommodate multiple agents is a strategic decision that prepares for complex scenarios involving multiple entities. This architecture offers flexibility for future scenarios and aligns with the goal of studying multi-agent AV-pedestrian interactions.

Overfitting posed a significant challenge in training on our dataset, prompting the incorporation of dropout and regularization loss in our neural network model. These measures, serving as effective safeguards, enhance model generalization and ensure robust performance on unseen data. We have also changed the number of network layers and conducted ablation studies to better suit our problem. Ablation studies on the network layers and hyperparameters are highlighted in Section 4.2. Finally, we have introduced data augmentation by applying transformations on the input data, including rotations and reflection. This allows to effectively increase the available data without loss of generality.

4 EXPERIMENTS

4.1 Data Collection

We collected interaction trajectories of the pedestrian and the car in our proposed virtual environment. The pedestrian data includes their absolute position with respect to the world Cartesian coordinates of the virtual environment. For the driver, we gathered data on steering, throttle, and brake commands, as well as the car's absolute position. Our primary objective is to predict the future behavior of pedestrians based on the behavior of the car. Unlike other datasets present in previous literature, our dataset contains not only relative position data but also driving commands from the driver. These driving commands, while directly affecting vehicle actions, are always reflected in the vehicle's absolute position with a time delay. Therefore, they provide more valuable and time-effective information for predicting pedestrian behavior. For instance, drivers should be aware that in real pedestrian-driver interactions, if they are about to apply the brakes, pedestrians are more likely to attempt to cross the road.

Fig. 5 shows the process of the data collection experiment in the real world and Fig. 7 shows the corresponding scenes in virtual reality. The pedestrian can spawn at any point along the straight road segments and on both sides of the road. This has two main advantages. Firstly, the driver is unaware of the exact pedestrian spawn location, which adds uncertainty to the scenario. Secondly, the data collected has a bigger variety, which can improve the generalisation of machine learning methods, such as those used in Section 4.2. After setting up the whole simulation system, drivers sitting in front of the driving simulator can see the pedestrian in the virtual traffic environment on the screen while the pedestrian can see the car operated by the driver in the VR headset (Fig. 5).

When the driver starts driving, the pedestrian is randomly spawned on the opposite side of the circular road. As a result, buildings obstruct pedestrians from the driver's view until they make a turn. The pedestrian is unaware of when the car will appear as well. It's worth noting that each time they spawn, pedestrians need to look straight ahead to calibrate the view of the VR and motion capture unit worn on the head. During the recordings, the pedestrian was told to make a crossing decision every time they saw the car coming. Since the relative initial distance between the car and the pedestrian is random, the data has a wider variety of initial Time To Collision (TTC) values. TTC is a measure of how much time it will take for two objects to collide. It is a critical metric for autonomous vehicles,

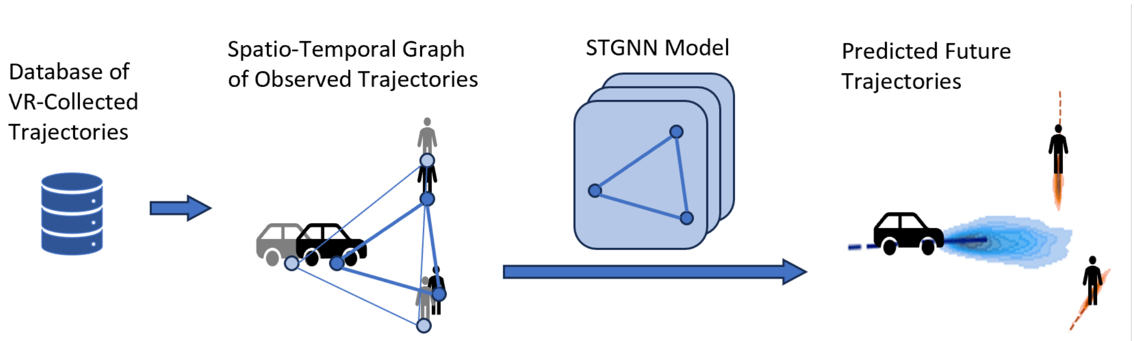

Database of VR-Collected Trajectories
Spatio-Temporal Graph of Observed Trajectories
STGNN Model
Predicted Future Trajectories

Figure 6: Overview of the STGCNN architecture and how it has been integrated in our system. The data collected with the VR equipment is processed and the STGCNN provides predictions of future trajectories.

Figure 7: A screenshot of the map from the road perspective.

ADE/FDE	1	2	3
2	0.57/0.87	0.60/0.96	0.78/0.90
3	0.76/0.97	0.78/0.87	0.69/0.89
5	0.60/0.95	**0.51/0.83**	0.57/0.92
7	0.63/0.83	0.63/0.95	0.76/0.97

Table 1: Ablation study on network layers. The first row indicates the number of ST-GCNN layers. The first column indicates the number of TXP-CNN layers. The metrics are indicated as ADE/FDE in meters for each table entry.

ADE/FDE	Car	Pedestrian	Average
No weights	3.10/5.81	0.36/0.55	1.73/3.18
L1	0.93/1.30	0.18/0.30	0.55/0.80
L2	**0.85/1.39**	**0.17/0.25**	**0.51/0.83**
Learnable	0.86/1.38	0.19/0.26	0.53/0.82
MLP	1.13/1.62	0.33/0.65	0.73/1.14

Table 2: Network performance based on different adjacency matrices and comparison with MLP network. No weights refer to an adjacency matrix with ones on the diagonal, L1 and L2 norms are also analysed. The metrics are indicated as ADE/FDE expressed in meters for each table entry.

as it allows them to assess the risk of a collision and take evasive action if necessary.

We invited a total of 16 driver-pedestrian pairs. The participants are people of different ages and genders. The drivers are all people with at least 3 years of driving experience, holding a valid UK driving license. We recorded data for each pair for about an hour, with an average recording time of 30 minutes for each pair. The total is 8 hours of effective trajectory data for driver-pedestrian interactions. The data collected has then been pre-processed for neural network training.

4.2 Experimental Results

We use the network described in [14], consisting of ST-GCNN layers and TXP-CNN layers, with the PReLU activation functions. We choose this network for its excellent capabilities in trajectory prediction tasks and its extendability in the future to include more pedestrians and vehicles. We divide the collected data into three groups: training, validation, and test datasets with roughly a 7:2:1 ratio (i.e. 11 driver-pedestrian pairs for training, 3 pairs for validation and 2 pairs for testing, chosen randomly).

We utilized a training batch size of 32 and employed Stochastic Gradient Descent (SGD) to train the model for 250 epochs, setting the initial learning rate to 0.01 with linear decay. The choice of batch size was influenced by the size of the dataset and compared against network performances obtained with batch sizes of 16 and 64. One of the major problems encountered during training on our dataset was overfitting. To mitigate the risk of overfitting in our neural network model, we have incorporated dropout and regularization loss into our training process. These techniques collectively serve as effective safeguards, enhancing the generalization capabilities of our model and ensuring its performance on unseen data. Ablation studies were also conducted on the number of STGCNN and TXP-CNN layers. As reported in Table 1, the optimal number of layers was found to be 2 STGCNN layers and 5 TXP-CNN layers. A higher number of layers (3 STGCNN) resulted in network performance degradation due to rapid overfitting. On the other hand, a single STGCNN layer might not be enough to capture complex human interactions, therefore resulting in worse performances.

The time horizon for the prediction is set to 2.4s in the future, as this is considered long-term prediction for autonomous driving tasks. The metrics used to evaluate the trajectory prediction model are the Average Displacement Error (ADE) and Final Displacement Error (FDE). We denote with N_T the total number of trajectories in the dataset, T_{obs} the observation time, T_p the prediction horizon. We

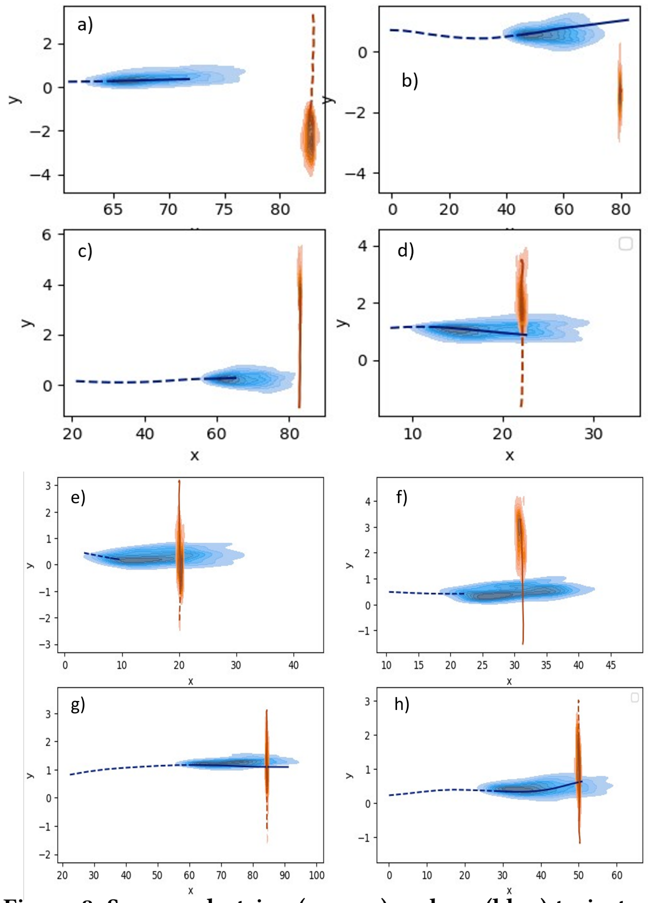


Figure 8: Some pedestrian (orange) and car (blue) trajectories with predictions. Previous trajectory: dashed line, future trajectory ground truth: solid line, predicted trajectory distribution: color density area. A missing dash line or solid line indicates that the corresponding agent is not moving

.

indicate ground truth values and predictions of position coordinates as (x, y) and $(\hat{x}, \hat{y})$ respectively. The ADE is the average distance between the predicted trajectory and the ground truth trajectory over the entire prediction horizon:

$$\text{ADE} = \frac{1}{N_T} \sum_{i=1}^{N_T} \frac{\sum_{t=T_{obs}+1}^{T_{obs}+T_p} \sqrt{(\hat{x}_i^t - x_i^t)^2 + (\hat{y}_i^t - y_i^t)^2}}{T_p - (T_{obs} + 1)} \tag{2}$$

The FDE is the distance between the predicted trajectory and the ground truth trajectory at the end of the prediction horizon, averaged across all trajectories. It can be expressed as:

$$\text{FDE} = \frac{1}{N_T} \sum_{i=1}^{N_T} \sqrt{(\hat{x}_i^{T_{obs}+T_p} - x_i^{T_{obs}+T_p})^2 + (\hat{y}_i^{T_{obs}+T_p} - y_i^{T_{obs}+T_p})^2} \tag{3}$$

We compare the network performance against an MLP network with temporal convolution layers. We also conduct ablation studies on the adjacency matrix function of the STGNN network. We consider no-weights, the reciprocal of L1 distance, the reciprocal of L2 distance, and a learnable adjacency matrix. Given the simplicity of our prediction task, consisting of trajectories of only two agents, we do not see any considerable improvements with the learnable adjacency matrix. No-weights refers to an unweighted adjacency matrix. We see the best results using the L2 distance, which captures the spatial relationships between the agents. In particular, this result confirms the intuition that the further away the agents are from each other, the less the mutual influence is. Table 2 shows the network performance on the task with the adjacency matrix related to the studies and the comparison with other prediction models. So far, the deep neural network analysis demonstrates that the network is capable of predicting pedestrian and car future trajectories and outperforms other prediction models both in ADE and FDE. Our best model achieves a performance of an ADE of 0.17 m and FDE 0.25 m for the pedestrian prediction, and a performance of 0.85 m/1.39 m for the vehicle. This is due to the fact that the velocities of the two agents are much different from each other. We are looking forward to collecting more data and including skeleton data to release an open-source dataset for deep learning that is based on VR collected data.

Fig. 8 shows some sample trajectories for our prediction task. The pedestrian (orange) is crossing in the vertical direction, whereas the car (blue) is moving in the horizontal direction. The predicted future trajectory distribution is represented with a coloured density. Our method's trajectory predictions work well, showing that it's a good fit for our problem. The STGCNN layers ensure that the interactions between the driver and the pedestrian are learnt by the network and correctly predict their behaviour in most scenarios. The sample trajectory predictions show that the method effectively predicts the probability of future positions for the car and the pedestrian.

5 CONCLUSION AND DISCUSSIONS

In this paper, we introduced a VR environment and data collection for pedestrian trajectories, which ensures safety and limited costs for the experiment. The simulator can be used in many different Autonomous Vehicles research in the area of AV/driver interactions with pedestrians, testing of AV control algorithms, pedestrian behaviour prediction and safety. We then analysed the collected data with a motion prediction system based on deep learning which demonstrates that the system can used to predict trajectories for pedestrians that are not present in the dataset.

While our current virtual environment setup and trajectory prediction are designed for one driver and one pedestrian, they can be immediately extended for the study of multi-drivers and multiple pedestrians. On the VR side, this is done simply by introducing more virtual vehicles and allowing multiple users to be captured at the same time. To further enhance the performance, we may look into multi-agent virtual reality pipelines [15] to effectively develop the VR system, and employ intelligent AI models to coordinate multiple virtual vehicles [24]. Similarly, on the trajectory prediction side, our method can incorporate a dynamic graph structure to consider the nearest road users. To further improve the accuracy of the prediction, we may explicitly model agent behaviours [23] and employ diffusion models for trajectory representation learning [5].

Future research directions will have to include multi-sensor data, as we have only focused on pedestrian-car 2D trajectories following the setup of existing work [14], thereby neglecting camera or pedestrian pose for making predictions. In our dataset, we also capture the 3D body movement of the pedestrian, as including them may be useful for improving crossing probability estimation [18]. We also explore a more affordable pipeline in capturing human behaviour, as motion capture system may not be easily scalable to a large number of pedestrians. We will look into depth camera-based system that can capture 2.5D images, with existing research showcasing capturing multiple users [22], as well as RGB camera-based systems combined with 3D pose estimation algorithms for extracting multi-human skeletal movement [4].

Gap-acceptance studies for multiple-lane scenarios are also possible future research directions, especially by employing AI-driven vehicles in the scene [28]. This is an easy way to test how pedestrians would behave in multiple-lane scenarios [20], which has not been researched extensively in the literature, due to the costs of setting up such experiments. Despite some initial work of capturing drivers' and pedestrians' behaviour in such scenarios [29], the is still much room for research particularly focusing on capturing realistic behaviours and effective behaviour modelling.

Another future research direction that the authors are looking to explore is the comparison between VR-generated facts and real-world data, called distribution mismatch. The advantages obtained via VR (cost-effectiveness and safety) can only be relevant if the interactions between drivers and pedestrians in VR align with real-world outcomes. We will explore opportunities to broaden the applicability of our work beyond pedestrian-car interactions to encompass a more comprehensive examination of various road user interactions, enhancing the overall impact of our research.

ACKNOWLEDGMENTS

This project is supported in part by EPSRC (Ref: EP/X031012/1).

REFERENCES

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 961–971.

[2] Austin Valentine Angulo, Erin Robartes, Xiang Guo, T Donna Chen, Arsalan Heydarian, and Brian L Smith. 2023. Demonstration of virtual reality simulation as a tool for understanding and evaluating pedestrian safety and perception at midblock crossings. *Transportation Research Interdisciplinary Perspectives* 20 (2023), 100844.

[3] Rajaram Bhagavathula, Brian Williams, Justin Owens, and Ronald Gibbons. 2018. The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62. SAGE Publications Sage CA: Los Angeles, CA, 2056–2060.

[4] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2021. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 1 (jan 2021), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

[5] Ziyi Chang, George Alex Koulieris, and Hubert P. H. Shum. 2023. On the Design Fundamentals of Diffusion Models: A Survey. *arXiv* (2023). https://doi.org/10.48550/arXiv.2306.04542 arXiv:arXiv:2306.04542 [cs.LG]

[6] Shuchisnigdha Deb, Daniel W Carruth, and Christopher R Hudson. 2020. How communicating features can help pedestrian safety in the presence of self-driving vehicles: virtual reality experiment. *IEEE Transactions on Human-Machine Systems* 50, 2 (2020), 176–186.

[7] Igor Doric, Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, Sebastian Wittmann, Matheus Zimmermann, and Thomas Brandmeier. 2016. A novel approach for researching crossing behavior and risk acceptance: The pedestrian simulator. In *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications.* 39–44.

[8] Arash Kalatian and Bilal Farooq. 2022. A context-aware pedestrian trajectory prediction framework for automated vehicles. *Transportation research part C: emerging technologies* 134 (2022), 103453.

[9] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezatofighi, and Silvio Savarese. 2019. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. *Advances in Neural Information Processing Systems* 32 (2019).

[10] Loreta Levulytė, David Baranyai, Edgar Sokolovskij, and Ádám Török. 2017. Pedestrians' role in road accidents. *International Journal for Traffic and Transport Engineering* 7, 3 (2017), 328–341.

[11] Kunming Li, Stuart Eiffert, Mao Shan, Francisco Gomez-Donoso, Stewart Worrall, and Eduardo Nebot. 2021. Attentional-GCNN: Adaptive pedestrian trajectory prediction towards generic autonomous vehicle use cases. In *2021 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 14241–14247.

[12] Régis Lobjois and Viola Cavallo. 2007. Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task. *Accident analysis & prevention* 39, 5 (2007), 934–943.

[13] Gustav Markkula, Ruth Madigan, Dimitris Nathanael, Evangelia Portouli, Yee Mun Lee, André Dietrich, Jac Billington, Anna Schieben, and Natasha Merat. 2020. Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science* 21, 6 (2020), 728–752.

[14] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. 2020. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 14424–14432.

[15] Alejandra Ospina-Bohórquez, Sara Rodríguez-González, and Diego Vergara-Rodríguez. 2022. A Review on Multi-agent Systems and Virtual Reality. In *Distributed Computing and Artificial Intelligence, Volume 1: 18th International Conference*, Kenji Matsui, Sigeru Omatu, Tan Yigitcanlar, and Sara Rodríguez González (Eds.). Springer International Publishing, Cham, 32–42.

[16] Daniel Perez, Mahmud Hasan, Yuzhong Shen, and Hong Yang. 2019. AR-PED: A framework of augmented reality enabled pedestrian-in-the-loop simulation. *Simulation Modelling Practice and Theory* 94 (2019), 237–249.

[17] Amir Rasouli and John K Tsotsos. 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE transactions on intelligent transportation systems* 21, 3 (2019), 900–918.

[18] Amir Rasouli and John K. Tsotsos. 2020. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (2020), 900–918.

[19] Andreas Riegler, Andreas Riener, and Clemens Holzmann. 2021. A systematic review of virtual reality applications for automated driving: 2009–2020. *Frontiers in human dynamics* 3 (2021), 689856.

[20] Keya Roy, Nam Hong Hoang, and Hai L. Vu. 2022. Modeling Autonomous Vehicles Deployment in a Multilane AV Zone With Mixed Traffic. *IEEE Transactions on Intelligent Transportation Systems* 23, 12 (2022), 23708–23720.

[21] Adarsh Jagan Sathyamoorthy, Utsav Patel, Tianrui Guan, and Dinesh Manocha. 2020. Frozone: Freezing-free, pedestrian-friendly navigation in human crowds. *IEEE Robotics and Automation Letters* 5, 3 (2020), 4352–4359.

[22] Hubert P. H. Shum and Edmond S. L. Ho. 2012. Real-Time Physical Modelling of Character Movements with Microsoft Kinect. In *Proceedings of the 2012 ACM Symposium on Virtual Reality Software and Technology* (Toronto, Ontario, Canada) *(VRST '12)*. ACM, New York, NY, USA, 17–24.

[23] Jianhua Sun, Yuxuan Li, Hao-Shu Fang, and Cewu Lu. 2021. Three steps to multimodal trajectory prediction: Modality clustering, classification and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 13250–13259.

[24] S. Suo, S. Regalado, S. Casas, and R. Urtasun. 2021. TrafficSim: Learning to Simulate Realistic Multi-Agent Behaviors. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 10395–10404.

[25] Kai Tian, Gustav Markkula, Chongfeng Wei, Yee Mun Lee, Ruth Madigan, Natasha Merat, and Richard Romano. 2022. Explaining unsafe pedestrian road crossing behaviours using a Psychophysics-based gap acceptance model. *Safety Science* 154 (2022), 105837.

[26] Tram Thi Minh Tran, Callum Parker, and Martin Tomitsch. 2021. A review of virtual reality studies on autonomous vehicle–pedestrian interaction. *IEEE Transactions on Human-Machine Systems* 51, 6 (2021), 641–652.

[27] Shouwen Yao, Jiahao Zhang, Ziran Hu, Yu Wang, and Xilin Zhou. 2018. Autonomous-driving vehicle test technology based on virtual reality. *The Journal of Engineering* 2018, 16 (2018), 1768–1771.

[28] Zhili Zhang, Songyang Han, Jiangwei Wang, and Fei Miao. 2023. Spatial-Temporal-Aware Safe Multi-Agent Reinforcement Learning of Connected Autonomous Vehicles in Challenging Scenarios. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. 5574–5580.

[29] Fengjiao Zou, Jennifer Ogle, Weimin Jin, Patrick Gerard, Daniel Petty, and Andrew Robb. 2023. Pedestrian Behavior Interacting with Autonomous Vehicles during Unmarked Midblock Multilane Crossings: Role of Infrastructure Design, AV Operations and Signaling. arXiv:2303.17717 [cs.RO]