Behavior Varied Adversarial Pedestrian Generation for Autonomous Vehicle Testing

Anonymous Author(s)
Affiliation
Address
email

Abstract: There exist several datasets for developing self-driving car methodologies. However, manually collected datasets impose inherent limitations on the variability of test cases and it is particularly difficult to acquire challenging scenarios, e.g. ones which involve collisions with pedestrians. Automatically synthesizing safety-critical scenarios for autonomous vehicles (AVs) can vastly increase the amount of relevant data for training and evaluating such AVs. However, existing approaches for scenario generation rely on handcrafted pedestrian behavior models, which limits the realism of the corresponding scenarios. In this paper we reformulate the problem as learning *where to place* pedestrians such that the induced scenarios are collision prone for an AV. Our pedestrian placement model can be used in conjunction with *any* goal driven pedestrian model – be it hand-designed or learnable – which makes it possible to challenge an AV with a wide range of pedestrian behaviors. Furthermore the proposed pedestrian placement model can be utilized with any goal based AV model given sufficient problem constraints. The pedestrian placement model is further conditioned on both scene semantics and occlusions to ensure semantic and visual plausibility, which increases the realism of the resulting scenarios. Moreover, to our best knowledge we are the first to show that it is possible to learn a collision seeking scenario generation model when both the pedestrian and the AV are collision avoiding. We also provide a discussion on how to extend the results to modern AV models.

Keywords: Autonomous Vehicles, AV Testing, Reinforcement Learning

1 Introduction

Research on autonomous vehicle (AV) models has gained momentum in recent years [1]. With the increase of various AV models the need for safety critical AV testing has increased. Insufficiently tested AVs can cause fatal accidents when deployed in traffic [2]. There exist several datasets [3–10] for developing and evaluating self-driving vehicle methodologies. However, manually collected test data is inherently limited with respect to the variability of test cases. This has motivated the recent development of test case generation methods [11–29]. All of these safety critical test case generation methods model pedestrians as moving entities that do not interact with the environment like humans. Such handcrafted behaviors do not resemble the rich and varied behavior of real pedestrians, and thus these methods risk providing insufficient insights on how the AV would act in corresponding scenarios involving real pedestrians. At the same time there exists a large number of pedestrian behaviour models [30–45] that learn, from real traffic scenarios, how pedestrians interact with the world. In difference to [11–29] we reformulate the problem of generating AV challenging scenarios as a problem of learning the distribution of initial pedestrian locations that induce collisions between the pedestrian and the AV for a given pedestrian behavior model. The reformulation allows the use of modern goal driven pedestrian behaviour models in AV test case generation. There exists three types of pedestrian models; collision-seeking, collision-ignorant and collision avoiding. These three model types arise three different optimization problems in our framework. We are the first to show empirically that a non-trivial solution exists for the problem that arises when the pedestrian model is collision avoiding. In difference to previous work [11–29] we explicitly model scene semantics to

Submitted to the 5th Conference on Robot Learning (CoRL 2021). Do not distribute.

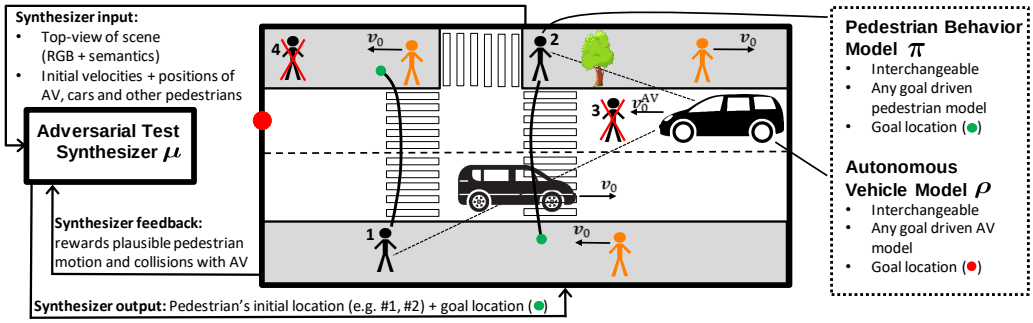

Figure 1: Overview of the proposed safety critical test case generation for AVs. The Adversarial Test Synthesizer (ATS) is trained to place out a pedestrian with behavior model π such that the induced scenario leads to a collision with the AV ρ (car to the right in this case). Four example scenarios are shown. Scenarios #3 and #4 are visually infeasible, as a pedestrian cannot simply appear from nowhere into the line of sight of an AV. Scenario #1: and Scenario #2: are challenging as the pedestrian is close to the AV and it is not in the line of sight of the AV due to the occlusions (with the car and tree respectively).

generate semantically and visually feasible simulation scenarios. The proposed model can also be used for data augmentation.

In traffic an AV will counter pedestrians with various degrees of traffic experience. Different individuals may take risks, follow traffic rules and plan for the AV's motion to different degrees. Independently on the level of traffic experience or agility of the pedestrian, an AV should be able to avoid collisions with the pedestrian when he/she appears from an occluded space. To ensure this AV testing should reflect the true variation of different pedestrian behaviors. Previous work [11–29] either assume that the pedestrian motion can be modelled by a simple constant velocity model or that the pedestrian motion is directly adversarial to the AV. In practice however collisions do not occur just when a pedestrian has a perfectly predictable path (i.e. the constant velocity), or when the pedestrian is suicidal and actively seeks to get killed by the AV (i.e. adversarial pedestrian model). Some collisions occur because pedestrians are distracted, due to occlusions or noise. We wish to alleviate the above assumptions on pedestrian motion in generative AV testing. To do so we learn a distribution μ of initial pedestrian locations that lead to collisions with the AV. In testing the proposed approach should be utilized with as many different pedestrian behaviour models as possible, as an AV should be seeking to avoid collisions with all pedestrians (even suicidal pedestrians).

We separate the problem of finding the pedestrian location distribution μ from modelling pedestrian behavior. Pedestrian location distribution in scenes is affected by the scene semantics. For example pedestrians are more likely to reside near or on sidewalks. Further pedestrians are less likely to be observed by a AV when they are occluded from the AV's perspective. Clearly the probability of a collision depends on the pedestrian's location in relation to the AV and the AV's velocity. Following these observations we propose a model in Fig. 1 that learns the pedestrian location distribution μ based on the scene semantics, distance to the vehicle, dynamic occupancy and occlusion map.

We propose to treat both the AV and the pedestrian as black-box models to allow for the testing of any goal driven pedestrian and AV models. The main problem is then to find a pedestrian location distribution μ such that the number of collisions between a black-box AV and a black-box pedestrian is maximal in expectation. The problem can be seen as a black box optimization or as a reinforcement learning problem. Bayesian Optimization scales poorly as the number of parameters increases. From our experience in traffic we can conclude that there is a strong correlation between occluded spaces and high-collision pedestrian locations. Further we expect that pedestrians are more likely to reside on sidewalks than on grass, thus μ is likely to be shaped by both the scene semantics and occlusions. Bayesian Optimization cannot be used because according to our observations μ is inherently discontinuous Fig. 2 because visually feasible pedestrian can only appear from occluded spaces. We use RL to find μ. The Adversarial Test Synthesizer (ATS) Fig. 1 is an agent that places out pedestrians in the given scene. The ATS selects initial locations for the pedestrian according to its policy, the distribution μ. The policy μ is optimized to increase the number of collision. For the agent ATS the uncontrollable external dynamics include the AV, all pedestrians, cars and the scene. We model the policy μ as a heatmap over the scene parametrized by a convolutional neural network.

In our formulation the ATS μ and the AV's model ρ are playing an indirect constrained min-max game. In the problem formulation we make no assumptions about the pedestrian behavior model π. The pedestrian behavior model can be cooperative with the AV ρ or to the ATS μ, or ignorant. The problem then becomes a constrained indirect three-agent minimax game with up to two agents per team. The study of the equilibrium is beyond this paper, however it is clear that if the AV ρ, pedestrian behavior π and pedestrian location distribution μ are unconstrained then the minimax problem has a trivial solution. If the pedestrian is always initialized arbitrarily close to the front of the AV (when the AV's initial velocity is forward) then this will always lead to a collision. If the pedestrian and the AV always stand still or always move in opposite directions, then there are never any collisions. To avoid trivial solutions, sufficient constraints are needed.

2 Related Work

Previous work [14, 16, 17] have studied the generation of pedestrian trajectories $(x_0, \ldots, x_T)$ for AV testing, such that the number of collisions is maximized. Then the optimal pedestrian trajectory $(x_0, \ldots, x_T)$ is also adversarial to the AV ρ. This leads to the pedestrian only behaving in a suicidal manner. This is unnecessarily limiting as the pedestrian behavior could also be helpful in avoiding collisions. In testing we wish to ultimately ensure that the AV can avoid collisions both with adversarial pedestrians but also with collision avoiding pedestrians. In [11] pedestrians are modelled with constant velocity, and are initialized from a set of predefined positions. The AV is retrained in a loop with the test case generator. In [14] existing trajectories are adapted to become adversarial. The suggested method is interesting but requires a varied ground truth dataset and the generated data is dependent on variability of the existing dataset.

Further the study of visual relations in data from the AV's perspective [46–48] has been raised recently. In [46] a visual prior is learnt for where pedestrians and other objects can appear in a camera mounted on a AV. This illustrates that the appearance of pedestrians in scenes should be structured to be visually plausible. A similar problem of realistic object placement in LiDAR scenes is studied in [47]. Finally [48] show that visual cues from an on-board camera can be used to learn walkeable areas in a scene. The work of [46–48] indicate that realistic data contains strong correlations between the scene's semantic structure and the presence and behavior of pedestrians and stationary obstacles of an AV. Given the previous we take to use such semantic cues and occlusions in the location distribution μ.

3 Methodology

Classical existence proof of zero-sum games cannot be applied [49] because the problem at hand is not a zero-sum game as the AV has loss terms evaluating its behavior that are not present in the pedestrian's loss function. To illustrate the underlying problem assume that the models' loss functions can be written as a sum of the expectation of the number of collisions and other loss components. Let the number of collisions between the AV ρ and the pedestrian model be measured by an indicator function I that is 1 if a collision occurs between the AV and the pedestrian, and 0 otherwise. The AV model ρ and the pedestrian location distribution model μ are learnt by minimizing the loss function J^ρ, J^μ respectively. In more detail the arising problem is the following,

$$\min_\rho J^\rho = \min_\rho \left(\mathbb{E}_{\mu,\rho,\pi}[I] + f_\rho(\rho)\right) \text{ s.t. } \rho \in B_\rho \tag{1}$$

$$\min_\mu J^\mu = \max_\mu \left(\mathbb{E}_{\mu,\rho,\pi}[I] - f_\mu(\mu)\right) \text{ s.t. } \mu \in B_\mu, \tag{2}$$

where f_ρ, f_μ are the loss components of J^ρ and J^μ respectively, and B_ρ and B_μ are behavior constraints on ρ and μ respectively. Equations (1) and (2) express the general optimization problem when the pedestrian behavior π is constant or independent of $\mathbb{E}[I]$. If the pedestrian behavior model π is collision avoiding then the problem gains the equation

$$\min_\pi J^\pi = \min_\pi \left(\mathbb{E}_{\mu,\rho,\pi}[I] + f_\pi(\mu)\right) \text{ s.t. } \pi \in B_\pi, \tag{3}$$

where f_π is a loss component of J^π, and B_π the set of behavioral constraints of π. If the pedestrian behavior model π is adversarial then (3) will instead be $\max_\pi \left(\mathbb{E}_{\mu,\rho,\pi}[I] - f_\pi(\pi)\right)$ s.t. $\pi \in B_\pi$. From this it it is clear that the behavior policy π affects the solution of both μ and ρ. Further depending on

the choice of π the set of applied constraints B_ρ, B_μ may need to be adjusted to ensure that none of the models converge to a trivial solution. Previous work have considered the cases of adversarial or constant velocity π. In our experiments we illustrate that with sufficient constraints on μ, π, ρ a non-trivial solution exists for (2) even for collision avoiding pedestrian behavior π, i.e. (3).

3.1 Special Case: Three Reinforcement Learning Agents

We view the problem of learning the pedestrian initial location distribution as a reinforcement learning (RL) problem with three agents; the pedestrian, the AV and the Adversarial Test Synthesizer (ATS) agent. We begin by treating the AV and the pedestrian model with common notation. At timestep $t \in [0, T]$ the pedestrian model and the AV move in the scene by taking actions a_t^π and a_t^ρ respectively, or commonly noted by vector $a_t = (a_t^\pi, a_t^\rho)$. At each timestep the pedestrian's action is sampled from the pedestrian policy $a_t^\pi \sim \pi(.|s_t^\pi)$ conditioned on the pedestrian's observation of the scene s_t^π which includes the AV. In the same manner the AV chooses actions by sampling $a_t^\rho \sim \rho(.|s_t^\rho)$ from its observation s_t^ρ of the scene which includes the pedestrian. The states s_t^π and s_t^ρ contain the pedestrian's and the AV's goal location respectively. The states s_t^π and s_t^ρ are combined in a vector $s_t = (s_t^\pi, s_t^\rho)$. The unknown world model $p(s_{t+1}|s_t, a_t)$ provides the transition probabilities between states from t to $t+1$ conditioned on the actions taken by the pedestrian agent and the AV. The pedestrian model's actions are evaluated by a reward function $r_\pi(s_t, a_t, s_{t+1})$ and the AV's by the reward function $r_\rho(s_t, a_t, s_{t+1})$. In training the policies π and ρ are chosen such that the respective expected discounted cumulative future reward, i.e. the utility is maximized at each state.

It is assumed that the AV's initial location y_0, initial velocity v_ρ^0 and goal location are given. The ATS observes s^μ the top view RGB and semantic images S of the scene, and the constant velocity future predicted occupancy and initial locations of the AV, the external cars and external pedestrians in the scene denoted D. The ATS's state s^μ also contains the product between the occlusion map O from the AV's perspective and μ's prior distribution P. The prior P is a heuristic of μ giving high probability to pedestrian initial locations that are close to the AV and can lead to a collision assuming constant motion v_ρ^0 for the AV. The ATS agent gives the pedestrian agent an initial location x_0 by sampling from the ATS's policy $OP\mu(s^\mu)$ the product of the occlusion map, the prior distribution P and the learnt policy model μ. Without loss of generality we will in §3.1 omit OP from $OP\mu(s^\mu)$ and let μ denote ATS's policy. The pedestrian with an initial location x_0 is given a goal location g^π such that the pedestrian's path to g^π coincides with the AV's. The pedestrian agent is given an initial velocity v_π^0 towards the goal, with a speed such that if the AV and the pedestrian keep a constant velocity then the AV and the pedestrian will collide. Given the pedestrian has an initial location x_0, initial velocity v_π^0 and a goal g^π the states in s_0 can be evaluated through simulation. Our aim is to find the the initial pedestrian location distribution μ, the policy of the ATS agent, that leads to the highest utility for the reward function $r_\mu(s_t, a_t, s_{t+1})$. The reward function $r_\mu(s_t, a_t, s_{t+1})$ attains its highest value when the AV and the pedestrian agent collide.

In our experiments all three of the learnable models ρ, π, μ are modelled by policy gradient models and share the loss function

$$J = \mathbb{E}_{x_0 \sim \mu(s^\mu), a_t^\pi \sim \pi, a_t^\rho \sim \rho, s_t \sim p(.|s_t, a_t)} \left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t^\pi, a_t^\rho, s_{t+1}) \right], \tag{4}$$

where $r = [r_\mu, r_\pi, r_\rho]$. The loss functions' dependence on I is expressed in the different reward functions. To simplify notations let the state, action history $(s_0, a_0, ..., s_T, a_T, s_{T+1})$ of an episode be denoted τ, and the discounted cumulative reward $R = \sum_{t=0}^{T} \gamma^t r(s_t, a_t, s_{t+1})$. We can express (4) as $\mathbb{E}[R] = \int_\tau R(\tau) p_\tau(\tau) d\tau$, where p_τ is the probability density function of τ. Assuming finite sequences (4) can be rewritten revealing the relationship between μ and r,

$$\mathbb{E}[R] = \int_\tau \sum_{t=0}^{T} \gamma^t r(s_t, a_t, s_{t+1}) \mu(x_0) \prod_{k=0}^{t} \pi(a_k^\pi | s_k^\pi) \rho(a_k^\rho | s_k^\rho) p(s_{k+1} | s_k, a_k) \mathrm{d}\tau. \tag{5}$$

From (5) it is clear that μ's one step reward is $R_\mu = \sum_{t=0}^{T} \gamma^t r_\mu(s_t, a_t, s_{t+1})$. We use the the policy gradient method to find μ. If the environment model p is known, we can try to find the closed form solution of the integral (5). Using Bellman equations would then allow for a white-box treatment of the AV and pedestrian dynamics. Otherwise Monte Carlo estimates can be used to solve the problem. Following REINFORCE [50] the AV ρ and the pedestrian location distribution μ can be learnt simultaneously as shown in the supplement.

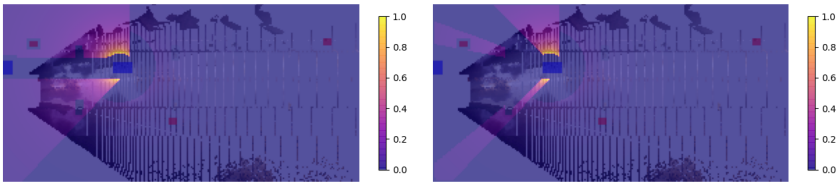

Figure 2: *Left:* A top view image of a sample prior P of μ. In red are other pedestrians, and in blue are cars. The prior implies a higher likelihood of pedestrian initial placement which are close to the AV. *Right:* The same prior after a multiplication with the occlusion map O.

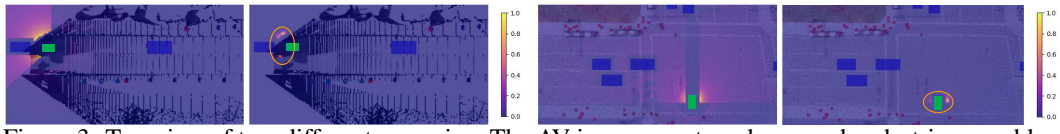

Figure 3: Top-view of two different scenarios. The AV is green, external cars and pedestrians are blue and dark red, respectively. In each of the two examples, the left image shows the prior distribution and the right image shows the final initial distribution. *Left example:* The prior P induces a high likelihood for initializing a pedestrian close to the AV, but the probability map is very smeared out. The final distribution $P\mu$ is much less scattered than P and more peaked close to the AV (indicated also with an external orange ellipsoid, to more clearly show where the probability mass is). *Right example:* We see a similar phenomenon as in the left example, in the dense dataset.

3.2 Adversarial Test Synthesizer

The ATS is a policy gradient agent, with policy μ. Its objective is to provide an initial position x_0 to the pedestrian agent such that the pedestrian collides with the AV. To do so the ATS needs to find locations near the AV where pedestrians and their motion are difficult to detect for the AV. To this end the ATS observes a top view image of the scene, the prior P of μ that depends on the distance to the AV, and the temporal mapping of dynamic objects D. The initial distribution μ depends also on the scene semantics S, as ATS should learn that pedestrians are more likely to reside near certain semantic classes such as pavement. The policy μ is conditioned on the state $s^\mu = (S, D, OP)$ of size $(128 \times 256 \times C)$, where $C = 17$ is the number of channels. Due to the success of neural networks in vision tasks and the visual nature of the input μ is modelled by a two layered convolutional neural network with bilinear interpolations and a softmax output layer (see Fig.1 in supplement). The output of the network $\mu(s^\mu)$ is a heatmap of size (128×256). The heatmap is multiplied by the prior P (see Fig. 2) to avoid sampling x_0 that cannot possibly lead to a collision. To enforce visual feasibility the ATS can be required to sample x_0 only from locations that are occluded for the AV. This can be done by sampling $x_0 \sim OP\mu(s^\mu)$, *i.e.*the product of the occlusion map O, the prior P and $\mu(s^\mu)$.

The ATS's reward r_μ evaluates the pedestrian position x_t at timestep t. Collisions with all external objects, cars and pedestrians are penalized but collisions with the AV are given a positive reward. Steps x_t taken in areas often visited by pedestrians are rewarded. Steps a_t towards the goal g^π are rewarded. The reward r_μ is adapted from r_π §3.3.

3.3 Pedestrian Model

The collision avoiding pedestrian behavior policy π is the goal driven Sematic Pedestrian Locmotion model (CARLA SPL) [42]. The π is a policy gradient agent that is trained by alternatively optimizing π for the maximum likelihood objective of pedestrian trajectory forecasting and for the policy gradient objective of collision avoidance. The reward function r_π of π encourages motion in pedestrian dense areas with the reward term R_{ped} and penalizes collisions with cars (including the AV), other pedestrians and static objects in the reward term R_{coll}. The reward component R_g encourages movement towards the goal location g^π, and R_ϕ penalizes unnaturally large motions. The reward function is further detailed in the supplementary.

The model observes a local crop $S_t(x_t)$ of size $5m \times 5m$ of the semantic labels and RGB top view image of scene S and a local crop $D_t(x_t)$ of the dynamic occupancy map D_t. Further the

state s^π contains a history of past actions and poses taken by the pedestrian in the past $N = 12$ timesteps, the displacement to the closest car d_t^x and the displacement to the goal g^π. In summary $s_t^\pi = (S_t(x_t), D_t(x_t), a_{t-1}^\pi \dots a_{t-N}^\pi, d_t^x, \|x_t - g^\pi\|)$. The policy gradient model takes a step a_t^π consisting of a direction and a speed. The step a_t^π is articulated by the Human Locomotion Network.

Unless otherwise stated the pedestrian models weights are kept constant to not deviate from the learnt pedestrian motion. The CARLA SPL model is trained to avoid collisions with the external cars. The external cars have a lower average speed than highest possible speed for ρ. This implies that π expects ρ to always have the same dynamics as its surrounding cars.

3.4 Autonomous Vehicle Model

The AV model ρ is intentionally simple to illustrate the framework empirically and to avoid making constraining assumptions about the AV. The focus of this work is to show that a collision avoiding pedestrian behavior can be successfully used in autonomous AV test case generation given enough constraints on the problem. The AV is a policy gradient model with the state $s_t^\rho = (\|x_t - y_t\|, d_t, \delta_t)$ at timestep t; where $\|x_t - y_t\|$ is the AV's distance to the pedestrian agent, d_t is the AV's distance to the closest car, and δ_t is the AV's intersection with the sidewalk. The AV's speed c_t is sampled from $\mathcal{N}(\text{sigmoid}(w^T s_t^\rho + b), \sigma_\rho)$, where w, b are learnt weights, and $\sigma_\rho = 0.1$. The sampled speed c_t is then scaled by the maximal speed of 70km/h. The AV's initial position y_0 and direction are chosen randomly among the external cars' constant velocity future trajectories.

The AV ρ is assumed to have a constant direction and the policy gradient model controls the speed of the AV. Speed control can be enough to avoid collisions, as the AV can stop or accelerate to avoid a collision. Extending the AV's model to direction changes complicates the learning as the AV receives two conflicting objectives: to move to a goal location further ahead and to avoid collisions. The research on AVs deals with balancing such conflicting objectives, and in the future we aim to replace the minimal AV model with a state of the art AV model. Without introducing additional constraints a state of the art AV model may outperform the ATS in early training leading to 0 gradients.

The reward function r_ρ penalizes the AV for collisions with cars, people and static objects. The AV is penalized for driving on the sidewalk proportionally to the AV's overlap with the sidewalk. To motivate the AV to move, a positive reward is given at the end of the episode for the distance travelled $\|y_0 - y_T\|$. The full reward function r_ρ is given in the supplementary.

4 Experiments

We experiment on a dataset gathered from CARLA[51]. Training data is collected from Town 1 and consists of 100 training and 50 validation scenes. The test set consists of 37 scenes from Town 2. For each scene a 3d reconstruction of RGB and semantic segmentation is created from a AV's perspective. We also introduce a smaller more object-dense dataset consisting of 4 different simulations of 5 scenes, gathered from a drone's perspective. We will refer to this as the Dense CARLA dataset (D). In all experiments the scenes $51m \times 25.6m$ are voxelized into $20cm$ cube voxels. All of the tested ATS models are evaluated with a *base AV model*. During initial experimentation it was noted that the AV model ρ had trouble learning collision avoidance without an initializer μ. The *base AV model* is trained on two scenes of the dataset for 200 epochs with an initializer that is trained on the training dataset for 10 epochs. A trajectory length of $T = 30$ are used to train the AV model, and a trajectory length of $T = 100$ is used to train the pedestrian initial distribution models. Each scene is evaluated for 10 epochs with a sequence length $T = 100$. During testing the pedestrian and AV models perform the mode and mean actions respectively. The ATS model is sampled from. The models are evaluated three times with different random seeds and the average and the standard deviation (stdev) of the three runs is reported. The following metrics are reported,

- #. collisions - number of collisions the AV model ρ has with pedestrians on average.
- Avg distance - average distance in meters travelled by the AV during an episode.
- π-entropy - entropy of the pedestrian policy during the length of an episode.

In Table 1 *left* the proposed pedestrian initial distribution model $OP\mu$ from §3.2 generates more than twice as many collisions (std=0.01) as sampling x_0 from the priors P and the occlusion-masked prior OP. This confirms that μ learns and improves beyond the initial prior distribution, and that $OP\mu$ produces more collisions than the hand-designed heuristics P and OP. The baseline *Random*

Table 1: *Left:* The proposed $OP\mu$ generates more collisions with the *base AV model* than the baselines; the heuristics the priors P and OP and the random initialization from occluded spaces *Random* O. *Right:* An ablation studying the effect of the prior and dataset density during the training of μ shows that the μ is robust to changes in the prior during training as $OP - \mu$ and $P - \mu$ trained with the priors OP and P respectively, have indistinguishable collision rates (stdev is 0.02).

	Random O	Prior P	Prior OP	$OP\mu$
#. collisions	0.05	0.09	0.09	**0.23**
Avg distance	8.9	8.6	8.7	**7.5**
π-entropy	0.81	0.60	0.64	**0.21**

	$P - \mu$	$OP - \mu$	$P\mu$-D
#. collisions	**0.23**	**0.22**	0.19
Avg distance	7.5	7.8	7.7
π-entropy	**0.21**	0.23	0.29

Table 2: Changing the collision avoiding *SPL* for a distracted *SPL*+ϵ or an adversarial *SPL A.* does not change the number of collisions. The adversarial *STPN* model that is trained simultaneously with μ has higher entropy than the *SPL* models because the adversarial *STPN* model has no motion priors. The constant velocity *HLN* has the lowest collision frequency.

	Collision avoiding SPL	Distracted SPL+ϵ	Adversarial SPL A.	Adversarial STPN	Constant velocity HLN
#. collisions	**0.23**($\pm$0.01)	**0.22**($\pm$0.01)	**0.22**($\pm$0.02)	0.17($\pm$0.03)	0.13($\pm$0.01)
Avg distance	**7.5**($\pm$0.6)	**7.7**($\pm$0.5)	**7.5**($\pm$0.4)	8.2($\pm$0.5)	8.9($\pm$0.2)
π-entropy	0.21($\pm$0.01)	0.22($\pm$0.01)	0.10($\pm$0.01)	0.51($\pm$0.01)	**0**

O the random initialization of pedestrians from occluded spaces with 360°field of view has the lowest collision frequency. This is likely because occluded spaces may be far from the AV. The proposed $OP\mu$ has a lower π-entropy than the prior OP suggesting that $OP\mu$ has learnt to initialize the pedestrian such that the pedestrian's direction of movement is as predictable as possible. With low π-entropy μ has more control over π's trajectory. To the *left* in Fig. 3 the prior P and the corresponding scene's $P\mu$ distribution are visually compared. The $P\mu$ has learnt decisively to initialize the pedestrian near the AV, and with a higher probability towards the sidewalk than the road.

In Table 1 *right* it can be seen that seen that masking the prior of μ with an occlusion map does not affect the number of collisions generated. This suggests that curriculum learning may be used to enforce larger changes to the prior P to eventually facilitate 360°field of view occlusion masks (for realistic LiDAR data testing). A visual comparison of P and OP can be seen in Fig. 2. The proposed μ is robust to dynamics changes in the training and testing as seen in the small drop in the number of collisions when comparing $P\mu$-D and $P - \mu$ in Table 1 *right*. The model $P\mu$-D trained on the denser dataset leads to fewer collisions than the base model P and has a higher π-entropy than $P\mu$, likely because the π is less controllable by μ in denser traffic. The distribution of $P\mu$-D is visualized in Fig. 3 to the right on a scene from the denser dataset.

In Table 2 the following pedestrian behavior policies are used to train μ,

- *Collisin avoiding SPL* - the goal reaching collision avoiding pedestrian model described in §3.3
- *Distracted SPL*+ϵ - a distracted SPL pedestrian. With a 0.3 probability at each timestep the pedestrian will not notice the AV for $m \sim Poisson(2)$ timesteps.
- *Adversarial SPL A.* - an adversarial SPL agent. The *SPL* model that is finetuned with the R^μ reward. *SPL A.* is trained simultaneously with μ (see supplementary Algoritm 1 with $\alpha_\rho = 0$).
- *Adversarial STPN A.* - an adversarial agent that has Semantic Trajectory Policy Network architecture [42] i.e. the SPL architecture without the Human Locomotion Network (HLN). The *STPN A.* is trained from random weights simultaneously with μ to maximize the R^μ reward. Note that *STPN A.* is not trained to maximize the negative log- likelihood of pedestrian trajectories like the *SPL* models, and it is the only model without locomotion.
- *Constant velocity HLN* - constant velocity motion articulated by [52]. The agent moves towards the goal with a speed drawn from a Gaussian with $\mu = 1.23\text{ms}^{-1}$ and $\sigma = 0.3$ (same as [53]).

The models *SPL*, the distracted *SPL*+ϵ and the adversarial pedestrian policy *SPL A.* are on-par, showing that μ can learn to control the collision avoiding *SPL* as well as a more distracted pedestrian or a pedestrian behavior model that is finetuned to collide with the AV, thus showing no significant disadvantage to utilizing a collision avoiding pedestrian behavior in the presence of enough constraints. Even though the *CV* model is the most controllable, the initializer trained to control *CV* results in the lowest collision rate because the AV has an easy time avoiding collisions with the CV HLN. On

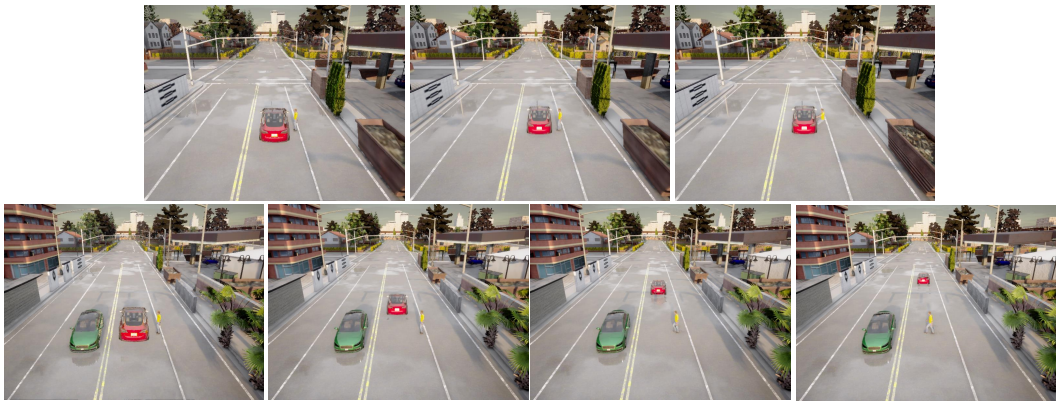

Figure 4: Sample trajectories of the *Simultaneous-*μ, ρ model, sub-sampled at 5 frames from frame 0. *First row:* The AV changes speed thus causing the pedestrian to incorrectly estimate the AV's motion and walk into the AV. *Second row:* The pedestrian waits for the AV to pass before crossing the road.

the other hand the *STPN A.* model has the highest entropy as the model has no motion priors and can get hit by the AV by zig-zagging in the middle of the road. The μ trained on *SPL*+ϵ distracted pedestrian could be expected to have a higher collision rate than *SPL*, as π has a noisier estimate of the AV's position. Unfortunately μ does not learn to utilize this unnaturally unstructured (and thus unpredictable) noise. This illustrates the need for a realistic noise model. Natural noise in the pedestrian's observation of the AV could be expected to be more structured, for example high noise levels are expected near occluded spaces. The sample trajectories in Fig. 4 of simultaneously trained μ and ρ show a collision prone initialization, and an initialization that does not lead to a collision. Further details, results and visualizations are provided in the supplement.

5 Conclusion and Future Work

We have proposed a general framework that is capable of stress testing the collision avoidance of AVs with a wide range of pedestrian behaviors. Our empirical evaluations showed that a goal driven pedestrian model with any behavior relative to collisions can be used in this framework. This is a significant result, as it was not before obvious that a collision avoiding pedestrian model could be used to generate collisions with a collision avoiding AV. We propose the Adversarial Test Synthesizer (ATS) agent that learns the pedestrian initial locations distribution that lead to collisions with the AV. The ATS is modelled by a neural network μ observing the top view image of the scene, the scene semantics, the occupancy of dynamic objects and the prior, and outputting a distribution of pedestrian initial locations. We showed in the experiments that μ can learn to utilize a collision avoiding pedestrian behavior model that is trained on ground truth pedestrian data and has human locomotive dynamics. This work shows that generative AV tests can utilize state of the art pedestrian motion models instead of the typically used handcrafted behavior models. Verifying AVs with state of the art pedestrian forecasting models decreases the statistical difference between the tested and real pedestrian behaviors and thus hopefully will aid in avoiding real life AV crashes.

We have shown that a learnable pedestrian initial location distribution μ exists for stress testing a basic AV model. Ultimately we wish to extend the result to state of the art AVs. Since the model μ treats the AV and the pedestrian agent as black-boxes, μ can be trained to adjust to the dynamics of a more sophisticated AV as is. However, to find a non-trivial solution will require a careful readjustment of the choice of sufficient but realistic constraints. The problem can be constrained spatially by tight streets, occlusion dense scenes, lack of space due to traffic density and by setting a limit on the pedestrian's maximal distance to the AV. Alternatively the pedestrian's model can be constrained by adjusting the noise level of the pedestrian's internal prediction of the AV's future motion, or the noise level of the pedestrian's observation of the AV, or the pedestrian's goal, dynamics or personality. Similarly the noise in the AV's observation of pedestrians and other traffic participants and their motion should be high enough to lead to collisions. As seen in the experiments, unstructured noise cannot be utilized by μ, thus careful modelling of the noise of the chosen AV's observations is required. In future work the pedestrian's internal prediction of the AV's motion could be impaired with a psychologically or physiologically inspired noise process.

References

[1] Z. Zhu and H. Zhao. A survey of deep rl and il for autonomous driving policy learning. *arXiv preprint arXiv:2101.01993*, 2021.

[2] P. Kohli and A. Chadha. Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash. *CoRR*, abs/1805.11815, 2018. URL http://arxiv.org/abs/1805.11815.

[3] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In *CVPR*, 2019.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The KITTI dataset. *I. J. Robotics Res.*, 32(11):1231–1237, 2013. doi:10.1177/0278364913491297. URL https://doi.org/10.1177/0278364913491297.

[5] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov. Scalability in perception for autonomous driving: Waymo open dataset, 2019.

[6] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, and A. Torralba. Semantic understanding of scenes through the ade20k dataset. *IJCV*, 2018.

[7] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020.

[8] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *ICCV*, 2019.

[9] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang. The apolloscape dataset for autonomous driving. In *CVPR Workshops*, 2018.

[10] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, et al. Lyft level 5 av dataset 2019. urlhttps. *level5. lyft. com/dataset*, 2:5, 2019.

[11] Y. Abeysirigoonawardena, F. Shkurti, and G. Dudek. Generating adversarial driving scenarios in high-fidelity simulators. In *International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019*, pages 8271–8277. IEEE, 2019. doi:10.1109/ICRA.2019.8793740. URL https://doi.org/10.1109/ICRA.2019.8793740.

[12] A. Hamdi, M. Mueller, and B. Ghanem. SADA: semantic adversarial diagnostic attacks for autonomous applications. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 10901–10908. AAAI Press, 2020. URL https://aaai.org/ojs/index.php/AAAI/article/view/6722.

[13] P. Gupta, D. Coleman, and J. Siegel. Towards safer self-driving through great pain (physically adversarial intelligent networks). *ArXiv*, abs/2003.10662, 2020.

[14] J. Wang, A. Pun, J. Tu, S. Manivasagam, A. Sadat, S. Casas, M. Ren, and R. Urtasun. Advsim: Generating safety-critical scenarios for self-driving vehicles. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[15] D. Karunakaran, S. Worrall, and E. Nebot. Efficient falsification approach for autonomous vehicle validation using a parameter optimisation technique based on reinforcement learning, 2020.

[16] D. Karunakaran, S. Worrall, and E. M. Nebot. Efficient statistical validation with edge cases to evaluate highly automated vehicles. In *23rd IEEE International Conference on Intelligent Transportation Systems, ITSC 2020, Rhodes, Greece, September 20-23, 2020*, pages 1–8. IEEE, 2020. doi:10.1109/ITSC45102.2020.9294590. URL https://doi.org/10.1109/ITSC45102.2020.9294590.

[17] W. Ding, B. Chen, M. Xu, and D. Zhao. Learning to collide: An adaptive safety-critical scenarios generating method. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2020, Las Vegas, NV, USA, October 24, 2020 - January 24, 2021*, pages 2243–2250. IEEE, 2020. doi:10.1109/IROS45743.2020.9340696. URL https://doi.org/10.1109/IROS45743.2020.9340696.

[18] N. Tsoi, M. Hussein, J. Espinoza, X. Ruiz, and M. Vázquez. SEAN: social environment for autonomous navigation. In M. Obaid, O. Mubin, Y. Nagai, H. Osawa, Y. Abdelrahman, and M. Fjeld, editors, *HAI '20: 8th International Conference on Human-Agent Interaction, Virtual Event, Australia, November 10-13, 2020*, pages 281–283. ACM, 2020. doi:10.1145/3406499.3418760. URL https://doi.org/10.1145/3406499.3418760.

[19] C. B. Schwab Benedikt and T. Kolbe. Spatio-semantic road space modeling for vehicle–pedestrian simulation to test automated driving systems. *Sustainability*, 3799, 2020.

[20] W. Ding, B. Chen, B. Li, K. J. Eun, and D. Zhao. Multimodal safety-critical scenarios generation for decision-making algorithms evaluation. *IEEE Robotics and Automation Letters*, 6(2):1551–1558, 2021. doi:10.1109/LRA.2021.3058873.

[21] P. Parashar, A. Cosgun, A. Nakhaei, and K. Fujimura. Modeling preemptive behaviors for uncommon hazardous situations from demonstrations. *CoRR*, abs/1806.00143, 2018. URL http://arxiv.org/abs/1806.00143.

[22] M. Wen, J. Park, and K. Cho. A scenario generation pipeline for autonomous vehicle simulators. *Human-centric Computing and Information Sciences*, 10, 12 2020. doi:10.1186/s13673-020-00231-z.

[23] A. Demetriou, H. Alfsvåg, S. Rahrovani, and M. Chehreghani. A deep learning framework for generation and analysis of driving scenario trajectories. *ArXiv*, abs/2007.14524, 2020.

[24] D. Nishiyama, M. Y. Castro, S. Maruyama, S. Shiroshita, K. Hamzaoui, Y. Ouyang, G. Rosman, J. A. DeCastro, K. Lee, and A. Gaidon. Discovering avoidable planner failures of autonomous vehicles using counterfactual analysis in behaviorally diverse simulation. In *23rd IEEE International Conference on Intelligent Transportation Systems, ITSC 2020, Rhodes, Greece, September 20-23, 2020*, pages 1–8. IEEE, 2020. doi:10.1109/ITSC45102.2020.9294715. URL https://doi.org/10.1109/ITSC45102.2020.9294715.

[25] B. Schütt, T. Braun, S. Otten, and E. Sax. Sceml: A graphical modeling framework for scenario-based testing of autonomous vehicles. In *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, MODELS '20, page 114–120, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450370196. doi:10.1145/3365438.3410933. URL https://doi.org/10.1145/3365438.3410933.

[26] W. Ding, M. Xu, and D. Zhao. CMTS: A conditional multiple trajectory synthesizer for generating safety-critical driving scenarios. In *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*, pages 4314–4321. IEEE, 2020. doi:10.1109/ICRA40945.2020.9197145. URL https://doi.org/10.1109/ICRA40945.2020.9197145.

[27] Z. W. Sun X., Zhang Y. Building narrative scenarios for human-autonomous vehicle interaction research in simulators. *Advances in Simulation and Digital Human Modeling*, 1206, 2020.

[28] M. Koren and M. J. Kochenderfer. Efficient autonomy validation in simulation with adaptive stress testing. In *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, Auckland, New Zealand, October 27-30, 2019*, pages 4178–4183. IEEE, 2019. doi:10.1109/ITSC.2019.8917403. URL https://doi.org/10.1109/ITSC.2019.8917403.

[29] D. Lee, S. Liu, J. Gu, M. Liu, M. Yang, and J. Kautz. Context-aware synthesis and placement of object instances. In S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 10414–10424, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/c6969ae30d99f73951cb976b88a457af-Abstract.html.

[30] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39(8): 895–935, 2020.

[31] P. Dendorfer, A. Osep, and L. Leal-Taixe. Goal-gan: Multimodal trajectory prediction based on goal position estimation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.

[32] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *CVPR*, 2018.

[33] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. In *NeurIPS*, 2019.

[34] L. Zhang, Q. She, and P. Guo. Stochastic trajectory prediction with social graph network. *CoRR*, abs/1907.10233, 2019. URL http://arxiv.org/abs/1907.10233.

[35] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[36] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[37] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F. Li, and S. Savarese. Social LSTM: human trajectory prediction in crowded spaces. In *CVPR*, 2016.

[38] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. In *CVPR*, 2017.

[39] W. Luo, B. Yang, and R. Urtasun. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *CVPR*, 2018.

[40] T. Zhao, Y. Xu, M. Monfort, W. Choi, C. Baker, Y. Zhao, Y. Wang, and Y. N. Wu. Multi-agent tensor fusion for contextual trajectory prediction. In *CVPR*, 2019.

[41] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In *CVPR*, 2019.

[42] M. Priisalu, C. Paduraru, A. Pirinen, and C. Sminchisescu. Semantic synthesis of pedestrian locomotion. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.

[43] P. Kothari, S. Kreiss, and A. Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[44] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *European Conference on Computer Vision*, pages 507–523. Springer, 2020.

[45] Y. Liu, Q. Yan, and A. Alahi. Social nce: Contrastive learning of socially-aware motion representations. *arXiv preprint arXiv:2012.11717*, 2020.

[46] O. Makansi, Ö. Çiçek, K. Buchicchio, and T. Brox. Multimodal future localization and emergence prediction for objects in egocentric view with a reachability prior. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 4353–4362. IEEE, 2020. doi:10.1109/CVPR42600.2020.00441. URL https://doi.org/10.1109/CVPR42600.2020.00441.

[47] S. Manivasagam, S. Wang, K. Wong, W. Zeng, M. Sazanovich, S. Tan, B. Yang, W. Ma, and R. Urtasun. Lidarsim: Realistic lidar simulation by leveraging the real world. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 11164–11173. IEEE, 2020. doi:10.1109/CVPR42600.2020.01118. URL https://doi.org/10.1109/CVPR42600.2020.01118.

[48] J. Sun, H. Averbuch-Elor, Q. Wang, and N. Snavely. Hidden footprints: Learning contextual walkability from 3d human trails. In A. Vedaldi, H. Bischof, T. Brox, and J. Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XVIII*, volume 12363 of *Lecture Notes in Computer Science*, pages 192–207. Springer, 2020. doi:10.1007/978-3-030-58523-5_12. URL https://doi.org/10.1007/978-3-030-58523-5_12.

[49] E. Smolyakov. Necessary and sufficient conditions for the existence of a saddle point in parametrized programmed differential games. *Cybernetics and Systems Analysis*, 33(5): 686 – 693, 1997. ISSN 10600396. URL http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edo&AN=ejs12272799&site=eds-live&scope=site.

[50] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 1992.

[51] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. Carla: An open urban driving simulator. In *CoRL*, 2017.

[52] D. Holden, T. Komura, and J. Saito. Phase-functioned neural networks for character control. *ACM Trans. Graph.*, 36(4):42:1–42:13, 2017. doi:10.1145/3072959.3073663. URL https://doi.org/10.1145/3072959.3073663.

[53] S. Chandra and A. K. Bharti. Speed distribution curves for pedestrians during walking and crossing. *Procedia-Social and Behavioral Sciences*, 104:660–667, 2013.