Generating Scenarios with Diverse Pedestrian Behaviors for Autonomous Vehicle Testing

Maria Priisalu[1], **Aleksis Pirinen**[2,*], **Ciprian Paduraru**[3,4], **and Cristian Sminchisescu**[1,5]

[1]Lund University, [2]RISE Research Institutes of Sweden, [3]University of Bucharest, [4]Institute of Mathematics of the Romanian Academy, [5]Google Research

Abstract: There exist several datasets for developing self-driving car methodologies. Manually collected datasets impose inherent limitations on the variability of test cases and it is particularly difficult to acquire challenging scenarios, e.g. ones involving collisions with pedestrians. A way to alleviate this is to consider automatic generation of safety-critical scenarios for autonomous vehicle (AV) testing. Existing approaches for scenario generation use heuristic pedestrian behavior models. We instead propose a framework that can use state-of-the-art pedestrian motion models, which is achieved by reformulating the problem as learning where to place pedestrians such that the induced scenarios are collision prone for a given AV. Our pedestrian initial location model can be used in conjunction with any goal driven pedestrian model which makes it possible to challenge an AV with a wide range of pedestrian behaviors – this ensures that the AV can avoid collisions with any pedestrian it encounters. We show that it is possible to learn a collision seeking scenario generation model when both the pedestrian and AV are collision avoiding. The initial location model is conditioned on scene semantics and occlusions to ensure semantic and visual plausibility, which increases the realism of generated scenarios. Our model can be used to test any AV model given sufficient constraints.

Keywords: Autonomous Vehicles, AV Testing, Reinforcement Learning

1 Introduction

Research on autonomous vehicle (AV) models has gained momentum in recent years [1]. There exist both end-to-end AV models which make decisions directly based on visual sensor outputs [1–6], and hierarchical models which require intermediate processing (such as pedestrian detection) of sensor outputs for decision making [7, 8]. To ensure traffic safety, e.g to avoid fatal collisions [9], there is a need to evaluate the various AV models in safety-critical situations. In this paper we consider safety testing of the full pipeline of perceptive AV models – from sensor inputs (e.g. images) to steering. There exist several datasets [10–17] for developing and evaluating AV models, but manually collected data is typically gathered from traffic scenarios that seldom exhibit collision and near-collision scenarios. This shortcoming has lead to recent developments of safety-critical test case generation methods [18–32] for AV models. These existing approaches resort to simulated pedestrians which are not representative of the rich and varied behavior of real pedestrians [33] – either the pedestrian trajectories are handcrafted, or the pedestrian models are trained to behave in unnatural ways (e.g. pedestrian agents which are adversarially trained to collide with vehicles). Thus these methods may provide insufficient insights on how the AV would act in scenarios involving real pedestrians. At the same time, there exist a large number of state-of-the-art pedestrian behavior models [34–49] which learn, from real traffic scenarios, how pedestrians interact with the world.

Different from [18–32], we reformulate the problem of generating challenging scenarios as one of learning the distribution μ of pedestrian initial locations x_0 which are likely to induce collisions between the pedestrian and the AV, for a given pedestrian behavior model π. This reformulation allows the use of state-of-the-art goal driven pedestrian behavior models π in AV test case generation,

*Work partially done while at Lund University.

5th Conference on Robot Learning (CoRL 2021), London, UK.

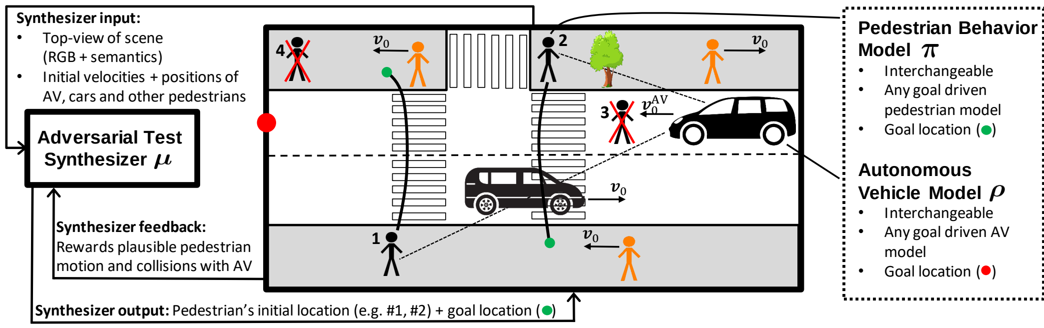

Figure 1: Overview of the proposed safety-critical test case generation model for AVs. The *Adversarial Test Synthesizer (ATS)* is trained to position a pedestrian with behavior model π such that the induced scenario is likely to yield a collision with the AV ρ (car to the right). Four example scenarios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

which means that the AV can be stress tested in more realistic scenarios compared to prior works. There exist three types of pedestrian behavior models – collision seeking, collision ignorant, and collision avoiding – each of which gives rise to a distinct optimization problem in our framework. We are the first to show empirically that a non-trivial solution exists when the pedestrian model is collision avoiding. Different from previous work [18–32], we explicitly model scene semantics to learn to generate semantically and visually plausible traffic scenarios. Our model can be used to augment existing data by adding simulated safety critical pedestrians to real traffic scenarios.

In real traffic an AV can be expected to encounter pedestrians with a range of different behaviors. Some individuals follow traffic rules and plan their movements based on the surroundings; others are inattentive and take risks. Independently of the pedestrian's overall behavior, an AV should be able to avoid collisions with the pedestrian when it appears from an occluded space. To ensure this, AV test scenarios should cover the true variation of different pedestrian behaviors. Previous works [18–32] either assume that the pedestrian motion can be modelled by a simple constant velocity model, or that the pedestrian motion is adversarial to the AV. In reality however, collisions do not occur only when a pedestrian has a perfectly predictable path (e.g. constant velocity), or when the pedestrian is actively seeking to get hit by the AV (e.g. adversarial pedestrian model). Quite the contrary – most collisions occur because pedestrians are distracted, due to occlusions or noise. To alleviate the previous unrealistic assumptions on pedestrian motion in generative AV testing, we separate the problem of *finding* the pedestrian location distribution μ from the *modelling* of the pedestrian behavior. The main problem is then to find a location distribution μ such that the number of collisions between a black-box AV and a black-box pedestrian is maximal in expectation. In AV testing the proposed approach should be used with as many different pedestrian behavior models as possible, as an AV should be seeking to avoid collisions with all (even collision seeking) pedestrians.

The pedestrian location distribution μ, shown in Fig. 1, is conditioned on scene semantics, distance to the AV, as well as a dynamic occupancy and occlusion map. Occlusions can cause an AV to miss a pedestrian (or vise versa) [50] and can thus cause collisions, therefore affecting the shape of μ. Furthermore, μ is likely to be shaped by the scene semantics (e.g., pedestrians are more likely to reside on sidewalks than on grass) [33, 51, 52]. In previous works, test case generation for AVs has been treated as a reinforcement learning problem [20–24, 32] or as a black-box optimization problem solved by bayesian optimization (BO) [18, 21]. BO [53] cannot be used to learn μ as μ is inherently discontinuous – in realistic scenarios pedestrians can only appear from occluded spaces [51] (cf. Fig. 2). Reinforcement learning (RL) on the other hand does not assume that the policy μ is continuous, and avoids the curse of dimensionality (that occurs in classical control and planning methods) in problems, like ours, with large state spaces with unknown world dynamics [54]. We thus propose the *Adversarial Test Synthesizer (ATS)*, an RL agent which positions pedestrians in a given scene (see Fig. 1). It selects initial locations for the pedestrian according to its policy μ, which is optimized to increase the number of collisions. For the ATS agent, the uncontrollable external dynamics include the scene, the AV, and all other pedestrians and cars. We model μ as a heatmap over the scene, parametrized by a deep convolutional neural network. Our pedestrian initial location model μ allows collision seeking scenario generation with any goal driven pedestrian behavior model and any AV model. This allows for more varied and more realistic testing of the AV. We show that near-collision scenario generation with a collision avoiding pedestrian gives rise to a previously unstudied optimization problem in AV testing. We show that this problem has a solution.

1.1 Related Work

Previous works [21, 23, 24] have studied the generation of full pedestrian trajectories $(x_0, \ldots, x_T)$ for AV testing, such that the trajectory is adversarial to the AV ρ. This leads to the pedestrian only behaving in a suicidal manner. This is unnecessarily limiting as typically it is not only the AV which aims to avoid collisions. Ultimately in testing we wish to ensure that the AV can avoid collisions with adversarial as well as collision avoiding pedestrians. In [18] 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 [21] existing trajectories are adapted to become adversarial. The suggested method requires a varied ground truth dataset and the generated data is dependent on the variability of the existing dataset. Similarly, [25, 26, 28–31] augment existing datasets in a latent or trajectory space, which again requires a large and varied ground truth dataset.

When only testing the vehicle control of a hierarchical AV system, the set of initial locations that cause collisions can be found by the Hamiltonian-Jacobi reachability set [55]. This is not possible in our setup since we consider the full stack of the AV, not only the control problem. Moreover, in our framework the pedestrian is not necessarily adversarial to the AV, and the scene dynamics cannot be described by a differential game. Our proposed approach allows the testing of AV models with pedestrian models that are semantically aware, collision avoiding, goal reaching and articulated. We do not use robust control methods because we utilize explicit pedestrian behavior models.

There are a number of recent studies which explore visual relations in data from the AV's perspective [51, 52, 56]. Makansi et al. [51] learn a visual prior for where pedestrians and other objects can appear from the perspective of a camera mounted on an AV. A similar problem of realistic object placement in LiDAR scenes is studied in [56]. Finally, [52] show that visual cues from an on-board camera can be used to learn walkable areas in a scene. The results of [51, 52, 56] indicate that realistic data contains strong correlations between the scene's semantic structure as well as the the presence and behavior of pedestrians and stationary obstacles. We thus include such semantic cues and occlusions in our proposed pedestrian location distribution model μ, as described in §2.

2 Methodology

In our framework, the ATS μ and the AV model ρ play an indirect constrained minimax game, and no assumptions are made about the pedestrian behavior model π. Thus π can be cooperative with either the AV ρ or ATS μ, or be ignorant with respect to both of these. The problem then becomes a constrained indirect three-agent minimax game with up to two agents per team. The study of the equilibrium [57, 58] is beyond the scope of 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.

To illustrate the minimax problem, assume that the loss functions for the AV ρ, the pedestrian π, and the ATS μ are respectively given by a sum of the expectation of the number of collisions and other loss components. Let the number of collisions between a given AV and pedestrian be measured by an indicator function I that is 1 if a collision occurs and 0 otherwise. The AV model ρ and the pedestrian location distribution model μ are learnt by minimizing the loss functions J_ρ and J_μ respectively,

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

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

where f_ρ and f_μ are loss components of J_ρ and J_μ, respectively. And B_ρ and B_μ describe the model constraints of ρ, and μ respectively. Equations (1) - (2) express the general optimization problem when the pedestrian behavior π is independent of $\mathbb{E}[I]$ (for example constant velocity π). If π is collision avoiding then (1) - (2) together with the following equation describe the general problem

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

where f_π is the loss component of J_π, and B_π describes the constraints on the model π. If the pedestrian behavior model π is adversarial then (3) will be replaced by $\max_{\pi} \left(\mathbb{E}_{\mu,\rho,\pi}[I] + f_{\pi}(\mu)\right)$ s.t. $\pi \in B_\pi$. It is clear that the choice of the behavior policy π changes the optimization problem and affects the solutions of μ and ρ. Depending on the choice of π, the set of applied constraints B_ρ and B_μ may need to be adjusted to ensure that none of the models converge to a trivial solution. Previous works [18–32] have considered the cases where π is adversarial or a constant velocity model. In our experiments we illustrate that with sufficient constraints on μ, π and ρ, a non-trivial solution exists for (2) when π is collision avoiding. The classical existence conditions of a solution of zero-sum game cannot be applied [57, 59] because the problem at hand is not a zero-sum game, as the pedestrian has loss terms f_π that are not present in the AV's loss function J_ρ.

2.1 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 ATS. At timestep $t \in \{0, T-1\}$ the pedestrian and the AV move in the scene by taking actions a_t^π and a_t^ρ, respectively; we gather these in a joint vector $a_t = (a_t^\pi, a_t^\rho)$. The pedestrian's action is sampled from the pedestrian policy $a_t^\pi \sim \pi(.|s_t^\pi)$ conditioned on its observation s_t^π of the scene which includes the AV. Similarly, the AV chooses actions as $a_t^\rho \sim \rho(.|s_t^\rho)$ where s_t^ρ is the AV's observation of the scene which includes the pedestrian. The states s_t^π and s_t^ρ respectively contain the pedestrian's and the AV's final goal location. We join s_t^π and s_t^ρ as 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 from state s_t to state s_{t+1} when the pedestrian and the AV take the joint action a_t. The pedestrian's and the AV's actions are evaluated by the reward functions $r_\pi(s_t, a_t, s_{t+1})$ and $r_\rho(s_t, a_t, s_{t+1})$, respectively. The policies π and ρ are trained to maximize the respective expected discounted cumulative future rewards (i.e. the utility) at each state s_t.

We assume that the AV's initial location y_0, initial velocity v_0^ρ and final goal location are given. Before the 0th timestep the ATS observes $s^\mu = (S, D, OP)$, where S is the top view RGB and semantic images of the scene, with constant velocity predicted dynamic occupancy D of the AV(calculated from y_0 and v_0^ρ), external cars and external pedestrians in the scene, and OP is the elementwise product between the occlusion map from the AV's perspective O and μ's prior distribution P . The prior P is a heuristic of μ which assigns high probability to pedestrian initial locations that are close to the AV and that can lead to a collision assuming constant motion v_0^ρ of the AV. The ATS agent samples an initial pedestrian location x_0 from the policy $OP\mu(s^\mu)$, which is the product between OP and the learnable policy μ. To reduce notational clutter we will in §2.1 omit the notation OP from $OP\mu$ and let μ denote the policy of ATS. The pedestrian with an initial location x_0 is given a goal location g^π and velocity v_0^π such that the pedestrian's path to g^π coincides with the AV's assuming both move with constant velocity. After sampling the pedestrian's initial location x_0 we simulate the pedestrian at the location x_0 with velocity v_0^π. Next we can simulate the pedestrian's and the AV's observation of the world s_0^π, s_0^ρ at $t = 0$. Our aim is to find the the initial distribution μ – i.e. the policy of the ATS agent – which leads to the highest utility for the reward function $r_\mu(s_t, a_t, s_{t+1})$, where r_μ attains its highest value when the AV and pedestrian collide.

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

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

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 $\tau = (a_0, s_1, ..., a_{T-1}, s_T)$, and the discounted cumulative reward $R = \sum_{t=0}^{T-1} \gamma^t r(s_t, a_t, s_{t+1})$. We can express (4) as $\mathbb{E}[R] = \int_{s^\mu} \int_{x_0} \int_\tau R(x_0, \tau) q(s^\mu) \mu(x_0|s^\mu) p_\tau(\tau|x_0) d\tau dx_0 ds^\mu$, where p_τ is the probability density function of τ given x_0, and q is the probability density function of s^μ. Let Ω be the set of allowed values for (τ, x_0, s^μ) then for finite T (4) can be rewritten to reveal the relationship between μ and r,

$$J = \int_\Omega q(s^\mu)\mu(x_0|s^\mu) \sum_{t=0}^{T-1} \gamma^t r(s_t, a_t, s_{t+1}) \prod_{k=0}^{t} \pi(a_k^\pi|s_k^\pi)\rho(a_k^\rho|s_k^\rho)p(s_{k+1}|s_k, a_k)\mathrm{d}\tau\mathrm{d}x_0\mathrm{d}s^\mu. \quad (5)$$

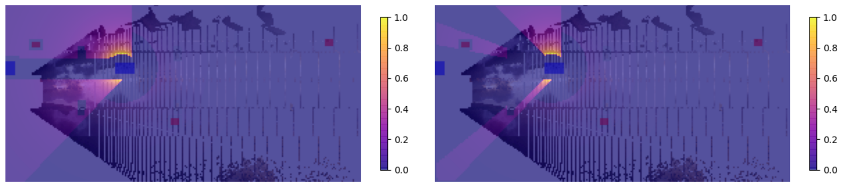

Figure 2: *Left:* A top view image of a sample prior P of μ. In red are other pedestrians, and in blue are cars. The prior implies a higher likelihood of pedestrian initial placement which are close to the AV. *Right:* The same prior after a multiplication with the occlusion map O.

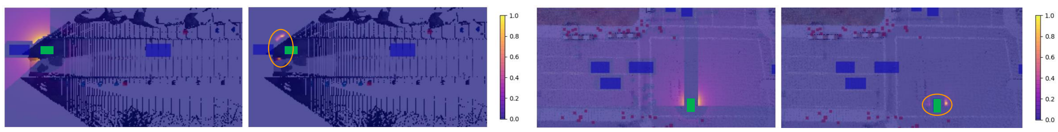

Figure 3: Top-view of two different scenarios. The AV is green, external cars and pedestrians are blue and dark red, respectively. In each of the two examples, the left image shows the prior distribution and the right image shows the final initial distribution. *Left example:* The prior P induces a high 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 (see supplement).

From the above it is clear that μ's one step reward is $R_\mu = \sum_{t=0}^{T-1} \gamma^t r_\mu(s_t, a_t, s_{t+1})$. We use REINFORCE [60] to find μ. The models ρ and μ can be learnt simultaneously as shown in the supplement. If the environment model p is known, we can try to find the closed form solution of (5). Using Bellman equations would then allow for a white-box treatment of the AV and the pedestrian.

2.2 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 gets an input consisting of 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^μ of size $(128\times256\times C)$, 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 bi-linear 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 at timestep t the pedestrian's behavior at position x_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 §2.3.

2.3 Pedestrian Model

The collision avoiding pedestrian behavior policy π is the goal driven Semantic Pedestrian Locomotion model (CARLA SPL) [46]. 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 (see supplement) 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 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 the highest possible speed for ρ. This implies that π expects ρ to always have the same dynamics as its surrounding cars.

2.4 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 collision avoiding pedestrian behavior models 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 allow directional 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. Replacing the current AV model with a state-of-the-art AV model requires additional constraints to avoid 0 gradients in early training (as the trained AV model may outperform the untrained ATS).

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.

3 Experiments

We experiment on a dataset gathered from CARLA[61]. 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. In all experiments the scenes $51m \times 25.6m$ are voxelized into $20cm$ cube voxels. All of the tested ATS models are evaluated and trained with the *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 for 200 epochs with a μ that is trained on the training dataset for 10 epochs. A trajectory length of $T = 30$ is used to train the AV model, and $T = 100$ is used to train the pedestrian initial distribution models. Each scene is evaluated for 10 episodes with $T = 100$. During testing the pedestrian and AV models perform the mode and mean actions respectively. The action of the ATS model is sampled. The models are evaluated with three different random seeds and the average and the standard deviation (stdev) of the three runs are reported. The reported metrics are

– #. collisions - number of collisions the AV model ρ has with pedestrians on average.
– π-entropy - entropy of the pedestrian policy during the length of an episode.

In Table 1 *left* the proposed pedestrian initial distribution models $OP\mu$ and $P\mu$ from §2.2 generate more than twice as many collisions (std=0.01) as sampling x_0 from the priors P and the occlusion-

Table 1: *Left:* The proposed $OP\mu$ and $P\mu$ generate more than twice the collisions compared to 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 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, and tested with the prior OP, have indistinguishable collision rates (stdev 0.01).

	Random O	Prior P	Prior OP	$OP\mu$	$P\mu$
#. collisions	0.06	0.10	0.09	**0.22**	**0.24**
π-entropy	0.85	0.60	0.65	**0.25**	**0.24**

Testing prior OP	$P\mu$	$OP\mu$
#. collisions	**0.21**	**0.22**
π-entropy	0.29	**0.25**

Table 2: Collision rates of the $P\mu$ model trained with collision avoiding, distracted, collision seeking and constant velocity pedestrians. The pedestrian model does not affect the collision rate of the proposed μ, as long as the pedestrian model is not the constant velocity model.

	Collision avoiding SPL	Distracted SPL+ϵ	Adversarial SPL A.	Adversarial STPN	Constant velocity HLN
#. collisions	**0.21**($\pm$0.02)	**0.22**($\pm$0.03)	**0.22**($\pm$0.01)	**0.19**($\pm$0.01)	0.11($\pm$0.02)
π-entropy	0.29($\pm$0.01)	0.25($\pm$0.02)	0.029($\pm$0.001)	0.53($\pm$0.03)	**0**

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* 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 *left* the models $OP\mu$ (i.e. μ trained and tested with the prior OP) and $P\mu$ (i.e. μ trained and tested with prior P) showed no significant difference. Showing that a 90°view occlusion map does not significantly affect μ. Further applying the occlusion mask O only in testing does not affect the number of collisions, as seen when comparing $P\mu$ to $OP\mu$ in Table 1 *right*. This suggests that curriculum learning may be used to enforce larger changes to the prior P to facilitate 360°field of view occlusion masks (for LiDAR data). A visual comparison of P and OP can be seen in Fig. 2.

In Table 2 the following pedestrian behavior policies are used to train μ,
- *Collision avoiding SPL* - the goal reaching collision avoiding pedestrian model described in §2.3
- *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 Algorithm 1 with $\alpha_{\rho} = 0$).
- *Adversarial STPN A.* - an adversarial agent that has the Semantic Trajectory Policy Network architecture [46] 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 the number or collisions ($R_{STPN} = I$ from §2). *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 CV* - constant velocity motion articulated by [62]. The agent moves towards the goal with a speed drawn from a Gaussian with $\mu = 1.23\text{ms}^{-1}$ and $\sigma = 0.3$ [63].

The models $P\mu$ trained with the collision avoiding *SPL*, the distracted *SPL+*ϵ, the adversarial finetuned pedestrian policy *SPL A.* and the adversarial *STPN A.* (most similar to previous work) are on-par, showing that μ can learn to control the collision avoiding *SPL* as well as an adversarial pedestrian model. The collision seeking *STPN A.* does not outperform the collision avoiding *SPL* likely due to *STPN A.*'s high entropy that makes *STPN A.* hard to control for μ. *STPN A.* has no motion priors and can get hit by the AV with motions that have a low likelihood in real pedestrian trajectories, such as zigzagging in the middle of the road. 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*. 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.

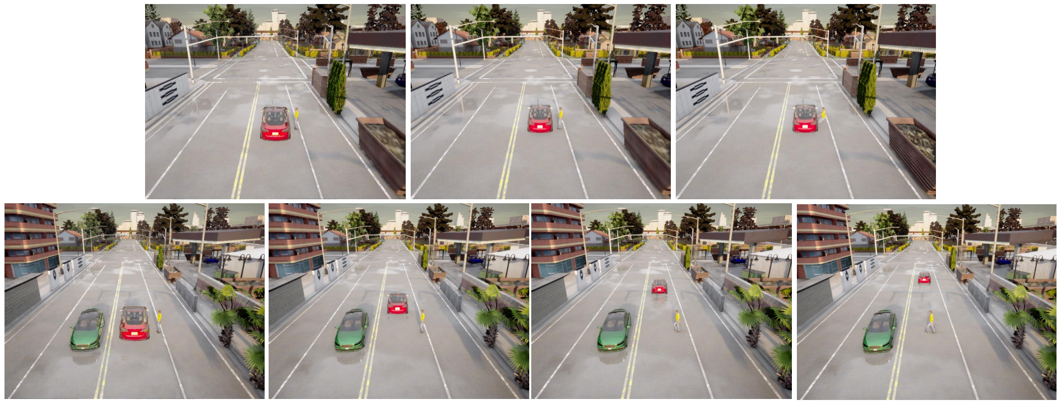

Figure 4: Sample trajectories of the *Simultaneous-μ, ρ* model, sub-sampled at 5 frames from frame 0. *First row:* The AV changes speed thus causing the pedestrian to incorrectly estimate the AV's motion and walk into the AV. *Second row:* The pedestrian waits for the AV to pass before crossing the road.

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. Some sample trajectories of the simultaneously trained μ and ρ (see supplement) are shown in Fig. 4 - a collision prone initialization, and an initialization that does not lead to a collision because the AV speeds away.

4 Conclusions and Future Work

We are the first to utilize state-of-the-art pedestrian forecasting models in generative AV testing. We have proposed a general framework that is capable of stress testing the collision avoidance of AVs with a wide range of pedestrian behavior models. In practice we wish to ensure that an AV can avoid collisions with all pedestrians (intoxicated, law-obedient, children etc.), and thus should test the AV with as many different pedestrian behaviors as possible. Our empirical evaluations show that a goal driven pedestrian model with any behavior can be used in this framework. This is a significant result, as no prior work has shown that a collision avoiding pedestrian model can be used to generate collisions with a collision avoiding AV. To achieve this, we have proposed the *Adversarial Test Synthesizer (ATS)* which, given any goal driven pedestrian model, learns the pedestrian initial location distribution μ that maximizes the expected number of collisions with a given AV. The ATS is modelled by a neural network which receives as input the top view image of the scene, the scene semantics, the occupancy of dynamic objects, and outputs a distribution μ over pedestrian initial locations. We have shown that μ can learn to adversarially position a collision avoiding pedestrian model that has been trained on ground truth pedestrian data and obeys human locomotive dynamics. Our work, for the first time, shows that generative models of AV test scenarios can utilize state-of-the-art pedestrian motion models instead of the typically used models which do not resemble real pedestrian motion. Stress testing AVs with state-of-the-art pedestrian forecasting models decreases the statistical difference between tested and real pedestrian behaviors, which could reduce the likelihood of 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 AV models. 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, finding 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, or by setting a limit on the pedestrian's maximal distance to the AV. Alternatively, the pedestrian model can be constrained by adjusting the noise level of the pedestrian's internal prediction of the AV's future motion, 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.

Acknowledgments

This work was supported by the European Research Council Consolidator grant SEED, CNCS-UEFISCDI PCCF-2016-0180, and the Swedish Foundation for Strategic Research (SSF) Smart Systems Program. We would also like to thank the anonymous reviewers for their useful comments.

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] F. Ye, S. Zhang, P. Wang, and C.-Y. Chan. A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles. *arXiv preprint arXiv:2105.14218, accepted in IEEE Transactions on intelligent transportation systems.*, 2021.

[3] A. Sadat, M. Ren, A. Pokrovsky, Y.-C. Lin, E. Yumer, and R. Urtasun. Jointly learnable behavior and trajectory planning for self-driving vehicles. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3949–3956. IEEE, 2019.

[4] W. Zeng, S. Wang, R. Liao, Y. Chen, B. Yang, and R. Urtasun. Dsdnet: Deep structured self-driving network. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 156–172. Springer, 2020.

[5] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl. Learning by cheating. In *Proceedings of Conference on Robot Learning (CoRL)*, pages 66–75. PMLR, 2020.

[6] M. Toromanoff, E. Wirbel, and F. Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7153–7162, 2020.

[7] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on intelligent vehicles*, 1 (1):33–55, 2016.

[8] Q. Dai, X. Xu, W. Guo, S. Huang, and D. Filev. Towards a systematic computational framework for modeling multi-agent decision-making at micro level for smart vehicles in a smart world. *Robotics and Autonomous Systems*, 144:103859, 2021.

[9] P. Kohli and A. Chadha. Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash. *In Proceedings of Future of Information and Communication Conference (FICC) 2019, Lecture Notes in Networks and Systems, vol 69*.

[10] 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 *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

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

[12] 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.

[13] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, and A. Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision (IJCV)*, 2018.

[14] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[15] 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 *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.

[16] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang. The apolloscape dataset for autonomous driving. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.

[17] 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. *level5. lyft. com/dataset*, 2:5, 2019.

[18] Y. Abeysirigoonawardena, F. Shkurti, and G. Dudek. Generating adversarial driving scenarios in high-fidelity simulators. In *Proceedings of International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019*, pages 8271–8277. IEEE, 2019.

[19] A. Hamdi, M. Mueller, and B. Ghanem. SADA: semantic adversarial diagnostic attacks for autonomous applications. In *Proceedings of Thirty-Fourth Conference on Artificial Intelligence (AAAI)*, pages 10901–10908. AAAI Press, 2020.

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

[21] 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. *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

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

[23] 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.

[24] 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.

[25] 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.

[26] P. Parashar, A. Cosgun, A. Nakhaei, and K. Fujimura. Modeling preemptive behaviors for uncommon hazardous situations from demonstrations. *CoRR*, abs/1806.00143, 2018.

[27] M. Wen, J. Park, and K. Cho. A scenario generation pipeline for autonomous vehicle simulators. *Human-centric Computing and Information Sciences*, 10, 12 2020.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

[36] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[37] 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.

[38] L. Zhang, Q. She, and P. Guo. Stochastic trajectory prediction with social graph network. *CoRR*, abs/1907.10233, 2019.

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

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

[41] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F. Li, and S. Savarese. Social LSTM: human trajectory prediction in crowded spaces. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[42] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[43] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[44] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[45] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[46] 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.

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

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

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

[50] X. Ren, T. Yang, L. E. Li, A. Alahi, and Q. Chen. Safety-aware motion prediction with unseen vehicles for autonomous driving. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[51] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4353–4362. IEEE, 2020.

[52] 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, *In Proceedings of European Conference on Computer Vision (ECCV)*, volume 12363 of *Lecture Notes in Computer Science*, pages 192–207. Springer, 2020.

[53] J. Mockus. On bayes methods for seeking an extremum. *Avtomatika i Vychislitelnaja Technika*, 3:53–62, 1972.

[54] D. Bertsekas. *Reinforcement and Optimal Control*. Athena Scientific, 2019.

[55] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin. Hamilton-jacobi reachability: A brief overview and recent advances. In *Proceedings of the IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 2242–2253. IEEE, 2017.

[56] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11164–11173. IEEE, 2020.

[57] K. Zhang, Z. Yang, and T. Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control*, pages 321–384, 2021.

[58] Q. Dai, X. Xu, W. Guo, S. Huang, and D. Filev. Towards a systematic computational framework for modeling multi-agent decision-making at micro level for smart vehicles in a smart world. *Robotics and Autonomous Systems*, 144:103–859, 2021. ISSN 0921-8890.

[59] 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.

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

[61] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. Carla: An open urban driving simulator. In *Proceedings of Conference on Robotic Learning (CoRL)*, 2017.

[62] D. Holden, T. Komura, and J. Saito. Phase-functioned neural networks for character control. *ACM Trans. Graph.*, 36(4):42:1–42:13, 2017.

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