Behavior Varied Adversarial Pedestrian Generation for Autonomous Vehicle Testing

Anonymous Author(s)
Affiliation
Address
e-mail

Abstract: There exist several datasets for developing self-driving car methodologies. However, manually collected test datasets impose inherent limitations on the variability of test cases, and it is particularly difficult to acquire sufficiently challenging scenarios. In order to overcome these constraints, we propose to automatically generate difficult scenarios in urban environments, by learning the placements of pedestrians such that the induced scenarios are highly challenging for a given autonomous vehicle (AV). In difference to previous methods we separate the pedestrian behavior model from the scenario generation network, which allows for testing the car model with various pedestrian behaviors. We show empirically that it is possible to train a test case scenario generation model with collision avoiding pedestrian model.

Keywords: Autonomous Vehicles, AV Testing, Reinforcement Learning

1 Introduction

Research in autonomous car models has gained momentum in the recent years [1]. With the increase of various autonomous models the need for extensive evaluation of different autonomous models has increased. The lack of varied safety-critical testing can lead to safety issues and can cost lives in applications [2]. There exists several datasets [3, 4, 5, 6, 7, 8, 9, 10] for developing and evaluating self-driving car methodologies. However, manually collected test data has inherent limitations on the variability of test cases. Further it is particularly difficult to manually acquire sufficiently challenging scenarios without undesired safety risks. This has lead to the development of generative test case generation methods [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]. All of the above methods 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 car. 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 car, i.e. adversarial pedestrian. Some collisions occur because pedestrians are distracted, due to occlusions or noise. We wish to alleviate the assumption on pedestrian motion and study if we can also generate test cases where the car and pedestrian collide even though both attempt to avoid one another. To do so we learn a distribution of initial pedestrian locations $\mu$ that lead to collisions with the car. To our knowledge we are the first to study the effects of different pedestrian behaviours $\pi$ on the distribution of pedestrian locations $\mu$.

Separating the problem of pedestrian location distribution $\mu$ from modelling pedestrian behavior $\pi$ allows us to in detail look at the pedestrian location distribution $\mu$. The problem contains some structure that we hope to learn. Pedestrian distribution in scenes is affected by the natural scene semantics. For example pedestrians are more likely to preside near or on sidewalks and on crosswalks. Further pedestrians are less likely to be observed by a car when they are occluded from the car’s perspective. Finally clearly the probability of a collision depends on the pedestrian’s location in relation to the car and the car’s speed. Following these observations we propose a model in Fig. 1 that learns the pedestrian location distribution $\mu$ conditioned on the scene semantics $S$, occlusion map $O$, temporal mapping of dynamic objects $D$ and a distance dependent prior $P$ to the car.
We propose to treat both the car and the pedestrian as goal driven black-box models to allow for the testing of any pedestrian and car models. The general problem is then to find a pedestrian location distribution \( \mu \) such that the number of collisions between a black-box car and a black-box pedestrian are 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. This in practice means that a parametric distribution needs to be fitted to the pedestrian location distribution \( \mu \). We have strong reason to believe that the optimal pedestrian distribution \( \mu \) is not easy to model since we expect a strong correlation between occluded spaces and high-collision pedestrian locations. To avoid any assumptions about \( \mu \)'s shape we instead use the reinforcement learning approach. The pedestrian and the car are then seen as a part of the unknown model dynamics, and the pedestrian location distribution \( \mu \) is the learnable stochastic policy. This allows us to model \( \mu \) as a convolutional neural network \( \mu_{\Theta} \).

In our formulation the pedestrian location distribution \( \mu \) and the car motion \( \rho \) are playing an indirect constrained min-max game. The pedestrian behavior \( \pi \) can be adversarial to the car or the pedestrian location distribution \( \mu \), or even not directly cooperative to either side. The problem then becomes a three agent mini-max game with up to two agents per team. Viewing traffic as a multi-agent game is a well studied topic within multi agent reinforcement learning [30]. We extend the concept by viewing the pedestrian location distribution \( \mu \) as an additional agent. The study of the equilibrium is beyond this paper however it is clear that if the car \( \rho \), pedestrian behavior \( \pi \) and pedestrian location distribution \( \mu \) are unconstrained then the minimax problem has a trivial solution. If the pedestrian is always initialized arbitrary close to the front of the car then this will always lead to a collision. If the pedestrian and the car always stand still or move in opposite directions, then there are never any collisions. To avoid trivial solutions sufficient amount of additional constraints need to modelled.

To our knowledge we are the first to provide a proof of concept study of the effects of the pedestrian behavior policies \( \pi \) on the pedestrian location distribution \( \mu \) in the domain of test case generation for autonomous vehicles. To do so we utilize the semantic pedestrian motion generation model of [31]. In experiments we articulate the studied pedestrian policies by [32] to ensure that plausible human dynamics are proposed by all studied pedestrian behaviours \( \pi \). We note that our results provide some proof of concept insights to the qualities of the pedestrian behavior model \( \pi \) on test case generation. In the future work we hope to extend the experiments with realistic pedestrian detection noise modelling for different sensors, with state of the art car models \( \rho \), and by extending the model to all of the observable traffic entities.

2 Related Work

Previous work [14, 16, 17] have studied the generation of pedestrian trajectories \( \eta \) such that the number of collisions is maximized. Where \( \eta \) includes implicitly the pedestrian position distribution \( \mu \) and the pedestrian behavior \( \pi \). When the objective \( J \) is maximized with respect to \( \eta \) then both \( \pi \) and \( \mu \)
We adopt a reinforcement learning (RL) perspective and study the problem as a sequential problem, where at each timestep the different agents make decisions conditioned on their current observation $s_t$ with constant velocity, and are initialized from set positions. The AV is retrained in a loop with the test case generator. [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 autonomous vehicle’s perspective [33, 34, 35] has been raised recently. In [33] 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 [34]. Finally [35] show that visual cues from an on-board camera can be used to learn walkable areas in a scene.

3 Methodology

To illustrate the relation between the different models’ loss functions, let the number of collisions between the car model $\rho$ and the pedestrian model be measured by an indicator function $I$ that is 1 if a collision occurs between the car and the pedestrian. In the proceeding, we will assume that all models’ loss functions can be composed as an expectation over the number of collisions and other loss components. The car model $\rho$ is learnt by optimizing a loss function $J^\rho$, that can be assumed to maximize the $E[I]$ under some desired behavioral constraints $B_\rho$. The pedestrian spatial distribution $\mu$ and $\rho$ have a loss function $J^\mu$ that maximizes $E[I]$ under some behavioural constrains $B_\mu$.

Collectively the problem can then be expressed as

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

$$\max_{\mu} \mathbb{E}_{\mu, \rho, \pi} [I] + f_\mu(\mu) \text{ s.t. } \mu \in B_\mu$$

where $f_\mu, f_\rho$ contain any other terms of $J^\mu$ and $J^\rho$ respectively. Equations 1 express the general optimization problem of $\rho$ and $\mu$ when the pedestrian behaviour $\pi$ is constant or independent of $E[I]$.

If the pedestrian behaviour model $\pi$ is collision avoiding then the problem gains a third equation that minimizes $E[I]$. If the pedestrian behaviour model $\pi$ is adversarial then the third equation will maximize $E[I]$. From this it is clear that the behaviour policy $\pi$ affects the solution of both $\mu$ and $\rho$.

Further depending on the choice of $\pi$ the set of applied constraints $B_\pi$, $B_\mu$ may need to be adjusted to ensure that none of the models converge to a trivial solution. Previous work has considered the case of adversarial or constant velocity pedestrian behaviour policy $\pi$. In our experiments we illustrate that with sufficient constraints on $\mu, \pi, \rho$ a non-trivial solution exists even for collision avoiding $\pi$.

3.1 Special Case: Three Policy Gradient Agents

We adopt a reinforcement learning (RL) perspective and study the problem as a sequential problem, where at each timestep the different agents make decisions conditioned on their current observation $s_t$ of the current state of the world. The actors take actions $a_t$, which affect the state of the world in the next timestep – moreover, in training a reward $r$ may be given as feedback from the environment.

The proposed solution consist of the following components:

- A goal driven car model $\rho(s_t^\rho)$ with a model specific state $s_t^\rho$ at timestep $t$;
- A pedestrian behavior model $\pi(s_t^\pi, g^\pi)$ with a state $s_t^\pi$ at timestep $t$, and a goal location $g^\pi$;
- A learnable distribution $\mu(S, D, O, P)$ over pedestrian initial locations;
- External world dynamics and noise $p(s_{t+1}|s_t)$, where $s_t$ is the current state of the world, including static objects, the autonomous vehicle, the pedestrian, as well as any external traffic participants (e.g. other cars and pedestrians). Any additional external dynamics or noise could be included here, such as the car’s dynamics.

To simplify notation we will use $s_t$ to denote the current time step’s global state (therefore including $s_t^\rho, s_t^\pi$). Similarly we denote by $a_t$ a vector containing the action taken by the car $a_t^\rho$ and the action taken by the pedestrian $a_t^\pi$ the at timestep $t$. 

3
In our experiments all three of the learnable models $\rho, \pi, \mu$ are modelled by a policy gradient model and share the loss function

$$J = \mathbb{E}_{s_0, a_0^p, a_0^r, \tau, \rho, s_{t+1}} \left[ \sum_{t=0}^{T-1} \gamma^t r(s_t, a_t^\pi, a_t^\rho, s_{t+1}) \right],$$

where $r = [r_\mu, r_\pi, r_\rho]$ is a reward vector containing the reward $r_\mu$ of the pedestrian distribution policy $\mu$, the reward $r_\pi$ of the pedestrian behaviour policy $\pi$ and the reward $r_\rho$ of the car policy $\rho$. The loss functions’ dependence on $I$ is expressed in the different reward functions. To simplify notations let the state, action history $(s_0, a_0, \ldots, s_T, a_T, s_{T+1})$ of an episode be denoted $\tau$, and the discounted cumulative reward $R = \sum_{t=0}^{T} \gamma^t r(s_t, a_t, s_{t+1})$. We can express (3) as $\mathbb{E}[R] = \int p_\tau(\tau) R(\tau) d\tau$, where $p_\tau$ is the probability density function of $\tau$. Assuming finite sequences allows the model to be rewritten showing the relation between $\mu$ and the rewards attained during an episode,

$$\mathbb{E}[R] = \int \sum_{t=0}^{T} \gamma^t r(s_t, a_t, s_{t+1}) \mu_\theta(s_0) \prod_{t=0}^{T} \pi(a_t^p|s_t) \rho(a_t^r|s_t) p(s_{t+1}|s_t, a_t) d\tau.\tag{4}$$

From (4) it is clear that $\mu$ can be seen as a policy gradient model with the one step reward $\sum_{t=0}^{T} \gamma^t r(s_t, a_t, s_{t+1})$. Now if all of the environment dynamics are known, we could try to find the closed form solution of the integral (4). Using Bellman equations would then allow for a white-box treatment of the car and pedestrian dynamics. Otherwise Monte Carlo estimates can be used to solve the problem. Following REINFORCE [36] the car $\rho$ and the pedestrian distribution $\mu$ can be learnt simultaneously as shown in the supplement. Further even the pedestrian behaviour policy can be learnt simultaneously with $\rho$ and $\mu$ if $\pi$ can be learnt only by sampling from $\mu$, $\rho$ and $\pi$.

### 3.2 Pedestrian Initial Spatial Distribution

In our experiments the pedestrian initial distribution $\mu$ is modelled as a policy gradient agent. Its objective is to provide an initial position $x_0$ to the pedestrian model such that the pedestrian collides with the AV $\rho$. This is achieved by performing roll outs sampling initial states $x_0$ for the pedestrian from $\mu$ and by sampling pedestrian and car moves from $\pi$ and $\rho$ respectively. The roll-outs are evaluated by a reward function $r_\mu$ that rewards collisions between the AV with position $y_0$ and the pedestrian $x_t$, where $t \in [0, T]$ is the timestep. The model $\mu$ observes the current scene $(S, D, O, P)$ of size $w \times h$ containing the AV $\rho$ at position $y_0$ and outputs a distribution of size $w \times h$ over the input scene. The pedestrian’s initial position $x_0 \sim P_\mu(S, D, O, P)$ is considered to be the action taken by the policy gradient agent $\mu$. Since the initializer $\mu$ cannot control the actions of the behaviour policy $\pi$ further the model $\mu$ does not receive a new state in response to the chosen action, only a reward $r_\mu$.

To avoid sampling from locations that could not possibly lead to a collision we introduce a prior for the $\mu$. Given that the pedestrian has a maximal speed of $3 m^{-1}$ there exists a cone of points $h$ from which the pedestrian can reach the car’s constant velocity trajectory. The prior for the points in $h$ is $||x - y_0||^{-1}$ where $y_0$ is the initial position of the car. The prior $P(x)$ is $0$ within the braking distance $v_\mu^2/(2g \times 0.8)$ of the car assuming dry road conditions ($0.8$ as friction coefficient). This is to avoid sampling from the trivial initializations within the car’s braking distance, thus leading to an inevitable collision. The points $x$ that are on the constant velocity estimate of the car’s trajectory receive a $0$ prior. Finally for all other points the prior is $1/||x-y_0||^2$. The prior encourages the pedestrian agent to be initialized close to the car, to encourage collisions. Without a prior a large number of samples of $x_0$ may be needed to encounter a collision. For a sample prior see Fig. 2.

Note that in order for a collision to occur when the pedestrian behaviour model $\pi$ is collision avoiding, the pedestrian should be initialized close enough to the AV such that collision avoidance will not be trivial. Alternatively the pedestrian’s observation of the car should be noisy enough for the pedestrian to miss the exact location of the car or to incorrectly internally predict the car’s future motion. Lastly additional spatial constraints such as street width, external obstacles and traffic agents can be used to enforce a smaller movement space for the pedestrian, leaving the pedestrian fewer options to avoid a collision with the car. Finally to allow the initializer to control the pedestrian enough to ensure collisions we utilize a goal $g^\tau$ driven pedestrian behaviour model. In the presented experiments the pedestrian’s goal is set to be the reflection of the pedestrian’s position in the car’s trajectory. This goal location enforces the pedestrian and car trajectories to cross. In the supplementary material we present also a model for learning the goal positions $g^2$ of the pedestrian. Further temporal or
Figure 2: Left: A top view image of a sample prior $P$ of $\mu$. In red are other pedestrians, and in blue are cars. Note how the prior implies a higher likelihood of pedestrian initial placement which are close to the AV (as these correspond to challenging scenarios). Right: The same prior after a multiplication with the occlusion map $O$.

Figure 3: Top-view of two different scenarios. The AV is colored in green, while other cars and pedestrians are colored in blue and dark red, respectively. In each of the two examples, the left part shows the prior distribution and the right part shows the final initial distribution. Left example: The prior induces a high likelihood for initializing a pedestrian close to the AV (as is expected, since that corresponds to a challenging scenario), but the probability map is very smeared out. Comparing with the final initial distribution, the probability map is much less scattered 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.

Behavioral constraints could be added to the pedestrian behavior model to allow $\mu$ more control of the pedestrian behavior after initialization.

Pedestrians can be initialized in occluded spaces to ensure that the generated pedestrians are visually plausible when added to existing data. The prior $P$ can be used to enforce pedestrian initialization in occluded spaces. This can be done by replacing the prior $P$ with $OP$, where $O$ is the occlusion map from the AV’s perspective. In the experiments it is assumed that the AV observes front-camera images with a 90-degree view. If the AV instead has no direct blind spots then the occlusion map needs to be blurred before adding to the prior to ensure non-zero gradients when training $\mu$. The pedestrian distribution model $\mu$ observes the scene as $s^\mu = (S, D, PO)$. Here $S$ contains the top view RGB image and semantic labels of the scene (possibly constructed from a reconstruction). The same semantic labels are used as in the pedestrian model §3.4. The dynamic mapping $D$ contains the constant velocity predictions of the external cars and pedestrians. It is the reciprocal of the dynamic map used in the pedestrian behavior model §3.4, and contains a separate channel for cars and pedestrians. Further details are given in the supplementary. Finally $\mu$ observes the product of the prior and the occlusion map $PO$. The prior indicates to $\mu$ which car should be challenged by the pedestrian initialization.

We use a small architecture for $\mu$, see the supplement for details. It consists of two convolutional layers, where the output of each convolutional layer is interpolated back to the original image size and added to each-other. The sum of the up-scaled convolutional layers is passed through softmax giving rise to the output. An entropy term is subtracted from the loss $J^\mu$ to encourage enough exploration of the scene space before convergence.

The reward function of $\mu$ is defined as the cumulative sum of rewards $r^\mu(x_0) = \sum_t \gamma^t r^\mu(x_t, a^\tau_t, y_t, a^\rho_t)$ attained at each timestep $t$ in the episode. The reward $r^\mu$ encourages initial positions $x_0$ that lead the pedestrian to pedestrian-like goal reaching motion while avoiding collisions with all external objects, cars and pedestrians but encouraging collisions with the AV. The reward $r^\mu(x_0)$ is a slightly adapted multiplicative reward consisting of the reward components used in the pedestrian behavior model §3.4 and an additional term $R^\rho$ that is an indicator function that is 1 when $\pi$ and $\rho$ collide.
3.3 Autonomous Car Model

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

The car $\rho$ is assumed to have a constant direction and the policy gradient model learns to select the speed of the car. In a number of cases the speed control can be enough to avoid collisions. Allowing direction changes complicates the learning objective of the AV somewhat as the car receives two conflicting objectives: to move to a goal location further ahead and to avoid collisions. There is vast literature on AV’s and on balancing such conflicting objectives, and in the future we aim to replace the minimal car model with a state of the art AV model. This change in the car model would also require additional constraints to the learning problem in the form of realistic noise modelling on pedestrian detection. Without introducing additional constraints a state of the art car model may outperform the pedestrian initial distribution model $\mu$ in early training leading to 0 gradients.

The reward $r_\rho$ function provides a negative reward for collisions with cars, people and static objects. The car also receives a negative reward proportional to fraction of overlap with the sidewalk. To motivate the car to move, a positive reward is given at the end of the episode for travelled distance $\|y_0 - y_T\|$. The exact reward structure is provided in the supplementary.

3.4 Pedestrian Model

The collision avoiding pedestrian behaviour policy $\pi$ is the CARLA SPL-goal model [31]. The $\pi$ is a policy gradient agent that is trained by alternatively optimizing $\pi$ for the maximum likelihood objective for pedestrian trajectory forecasting and for the policy gradient objective for collision avoidance. The reward function $r_\pi$ of $\pi$ encourages motion in pedestrian dense areas $R_{\text{ped}}$ and penalizes collisions with cars (including $\rho$), other pedestrians and static objects in the reward terms $R_{\text{coll}}$. The reward component $R_g$ encourages movement towards the goal location $g''$, and $R_\phi$ 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 $s''$ contains a history of past movements and poses taken by the pedestrian in the past $N = 12$ timesteps, the displacement to the closest car and the displacement to the goal $g''$. The policy gradient model takes a step $a''_t$ that is described by a discretized direction and a speed, that is then articulated by the Human Locomotion Network. In the majority of the experiments the pedestrian models weights are kept constant to not deviate from the learnt pedestrian motion.

It should be noted that the CARLA SPL-goal model is trained to avoid collisions on a dataset where the cars have much lower average speed than what can be exhibited by $\rho$. This provides sufficient constraints on the collision avoidance abilities of $\pi$ to ensure that the initial pedestrian distribution can learn in early stages. The pedestrian model’s incorrect internal estimate of the car’s motion acts as a noise on the pedestrian’s prediction of the car’s motion, and could be instead be replaced by a noise model. Alternatively with a more detailed pedestrian model the test case generator could be given more control to induce collisions by controlling $\pi$’s latent behaviour and thus affecting the pedestrian’s distractedness, urgency or possibly mood.

4 Experiments

We experiment on a dataset gathered from CARLA[37]. 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 car’s perspective. We also introduce a smaller more object-dense dataset consisting of 4 different simulations of 5 scenes, these scenes are gathered from a drone’s perspective. The latter dataset is only used for training and will referred to as the denser CARLA dataset. In all experiments the scenes $51m \times 25.6m$
The number of collisions of the pedestrian initial distribution models are similar for the collision avoiding SPL-goal, the distracted SPL+\( \epsilon \) and the adversarial pedestrian policy Adv. SPL. The pedestrian initializer \( \mu \) does not learn to utilize the distractedness of SPL+\( \epsilon \), as the added noise is without an occlusion map.

The effects of the pedestrian behaviour model \( \pi \) on the pedestrian initial distribution model \( \mu \) can be seen in Table 2. The tested pedestrian behaviour policies are the following,

- **SPL-goal** - the goal reaching collision avoiding pedestrian model described in §3.4
- **SPL+\( \epsilon \)** - a distracted SPL-goal pedestrian. With a 0.3 probability at each timestep the pedestrian will not notice the car for \( m \sim Po(2) \) timesteps.
- **Adv. SPL** - an adversarial SPL-goal agent. A SPL-goal that is trained further simultaneously with \( \mu \) with the \( R^\mu \) reward.
- **CV** - constant velocity motion articulated by [32]. The movement direction is towards the goal and the speed is drawn from a Gaussian with \( \mu = 1.23 \) and \( \sigma = 0.3 \) (same as [38]).

The effects of the pedestrian behaviour model \( \pi \) on the pedestrian initial distribution model \( \mu \) can be seen in Table 2. The tested pedestrian behaviour policies are the following,

- **SPL-goal** - the goal reaching collision avoiding pedestrian model described in §3.4
- **SPL+\( \epsilon \)** - a distracted SPL-goal pedestrian. With a 0.3 probability at each timestep the pedestrian will not notice the car for \( m \sim Po(2) \) timesteps.
- **Adv. SPL** - an adversarial SPL-goal agent. A SPL-goal that is trained further simultaneously with \( \mu \) with the \( R^\mu \) reward.
- **CV** - constant velocity motion articulated by [32]. The movement direction is towards the goal and the speed is drawn from a Gaussian with \( \mu = 1.23 \) and \( \sigma = 0.3 \) (same as [38]).

The effects of the pedestrian behaviour model \( \pi \) on the pedestrian initial distribution model \( \mu \) can be seen in Table 2. The tested pedestrian behaviour policies are the following,

- **SPL-goal** - the goal reaching collision avoiding pedestrian model described in §3.4
- **SPL+\( \epsilon \)** - a distracted SPL-goal pedestrian. With a 0.3 probability at each timestep the pedestrian will not notice the car for \( m \sim Po(2) \) timesteps.
- **Adv. SPL** - an adversarial SPL-goal agent. A SPL-goal that is trained further simultaneously with \( \mu \) with the \( R^\mu \) reward.
- **CV** - constant velocity motion articulated by [32]. The movement direction is towards the goal and the speed is drawn from a Gaussian with \( \mu = 1.23 \) and \( \sigma = 0.3 \) (same as [38]).
Table 2: Left: Changing the collision avoiding pedestrian behaviour model SPL-goal for a distracted pedestrian SPL+ε or an adversarial pedestrian Adv. SPL does not change the number of collisions. A constant velocity pedestrian leads to a lower number of collisions and entropy. Right: Training the π and µ simultaneously Sim results in metrics similar to those of separately trained models. This is confirmed by testing the Alt-µ,Sim-µ against the baseline car, and the Alt-ρ,Sim-ρ against POµ.

<table>
<thead>
<tr>
<th></th>
<th>SPL</th>
<th>SPL+ε</th>
<th>A. SPL</th>
<th>CV</th>
<th>Alt-µ, ρ</th>
<th>Sim-µ, ρ</th>
<th>Alt-µ</th>
<th>Sim-µ</th>
<th>Alt-ρ</th>
<th>Sim-ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>#. collisions</td>
<td>0.21</td>
<td>0.21</td>
<td>0.20</td>
<td>0.12</td>
<td>0.20</td>
<td>0.19</td>
<td>0.20</td>
<td>0.18</td>
<td>0.40</td>
<td>0.22</td>
</tr>
<tr>
<td>Avg distance</td>
<td>6.9</td>
<td>7.1</td>
<td>7.1</td>
<td>8.7</td>
<td>7.1</td>
<td>7.1</td>
<td>7.2</td>
<td>7.3</td>
<td>5.1</td>
<td>6.9</td>
</tr>
<tr>
<td>π-entropy</td>
<td>0.22</td>
<td>0.21</td>
<td>0.11</td>
<td>0</td>
<td>0.11</td>
<td>0.24</td>
<td>0.24</td>
<td>0.24</td>
<td>0.18</td>
<td>0.22</td>
</tr>
</tbody>
</table>

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

5 Conclusion

The problem of adversarial pedestrian test case generation can advantageously be decomposed into the pedestrian initial distribution model learning and the pedestrian behaviour model learning. This problem decomposition allows for collision avoiding pedestrian models to be used to evaluate AVs. Collision avoiding pedestrian models mimic cooperative pedestrians and the experiments have shown that a collision inducing pedestrian initial distribution is learnable even for co-operative pedestrian behaviours. Further the problem decomposition also allows for a number of pedestrian behaviour models to be used in the testing of an AV.
References


