Supplementary Material for KING: Generating Safety-Critical Driving Scenarios for Robust Imitation via Kinematics Gradients

Niklas Hanselmann[1,2,3], Katrin Renz[2,3], Kashyap Chitta[2,3], Apratim Bhattacharyya[2,3], and Andreas Geiger[2,3]

[1] Mercedes-Benz AG R&D, Stuttgart
[2] University of Tübingen
[3] Max Planck Institute for Intelligent Systems, Tübingen

Abstract. In this **supplementary document**, we first provide additional information regarding AIM-BEV (Sec. 1.1) and the rule-based expert (Sec. 1.2). Next, we provide additional information on the kinematics model (Sec. 2) and a detailed description of our cost functions (Sec. 3). Furthermore, we discuss details regarding the maps, routes and traffic initialization of the scenarios in Sec. 4. Finally, we present details on the hyperparameter choices for all optimization methods, further results, as well as additional qualitative examples in Sec. 5.

1 Driving Agents

In this work, we consider three driving agents: (1) Our AIM-VA [3] variant termed AIM-BEV, (2) TransFuser [10] and (3) a privileged rule-based expert, which provides supervision for AIM-BEV and TransFuser and is additionally used to determine solvability for the scenarios generated by KING. In the following we provide additional details on AIM-BEV and the rule-based expert. For TransFuser, we refer readers to the original paper. In addition, we provide further information on the driving metrics used to benchmark their performance in the CARLA simulator.

1.1 AIM-BEV

Model: Our architecture is similar to the AIM models introduced in previous work [3, 10]. For efficiency during simulation, we use a lightweight MobileNetV3-Small [8] instead of the ResNet34 used in AIM as a convolutional backbone. As described in the main paper, it takes a bird's-eye view (BEV) semantic occupancy grid observation $\mathbf{o}_t$ as input. From this, a 512-dimensional feature vector is extracted (same as in AIM). This is then further processed by three fully-connected layers (256, 128 and 64 units with ReLU activations) to yield a 64-dimensional representation of $\mathbf{o}_t$. To be able to train and evaluate the model in the CARLA simulator [5], we concatenate a binary variable indicating if the vehicle should stop for a red light to this representation. As we do not model traffic lights in KING scenarios, this variable is always set to zero

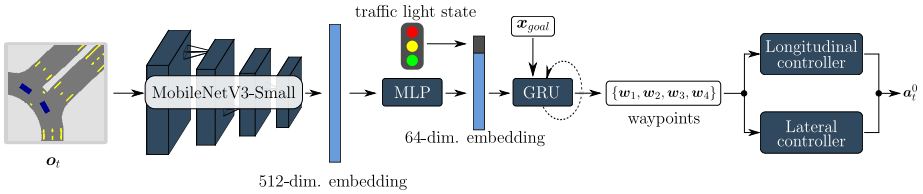


Fig. 1: **Architecture Overview.** Illustration of the AIM-BEV architecture. First, a low-dimensional embedding from the input BEV semantic grid map observation $\mathbf{o}_t$ is extracted. This is then used as the basis for autoregressive waypoint prediction, from which the final actions are obtained via lateral and longitudinal PID controllers.

in our proxy simulation.[4] Finally, the resulting feature is input to a single layer gated recurrent unit (GRU) with a linear decoder, which autoregressively estimates a set of four future waypoints $\{\mathbf{w}_i\}_{i=1}^{4}$ representing the desired trajectory. In each timestep, the GRU is additionally conditioned on a sparse goal location $\mathbf{x}_{goal}$ indicating the intended high-level route. To convert the waypoints into low-level actions we use a set of two PID controllers, which we describe next.

Controllers: We obtain actions for steering and acceleration from separate lateral and longitudinal controllers (LatPID and LonPID in Algorithm 1). These are similar to those used in [2, 3, 9], with slight modifications. In particular, we replace hard clipping and thresholding operations with soft versions to maintain differentiability. For these, we tune the shape parameters $\beta_{eb}, \beta_{decel}$ and clipping bounds $c_\delta, c_{throttle}$ to approximate the original non-differentiable implementation. For steering, we obtain a control target by computing the relative orientation Φ_d of the halfway point $\mathbf{w}^*$ between the first and second waypoint. This is then input to the lateral PID controller (with gains $K_p = 1.25$, $K_i = 0.75$, $K_d = 0.3$), which returns the corresponding steering action. Because this is unbounded, we constrain it to the $[-1, 1]$-interval by applying the *tanh* function. We have found a buffer of length 4 to be sufficient for computing the integral and derivative terms in both PID controllers. For throttle, we obtain the desired speed v_d by computing the norm of the vector connecting the first and second waypoint. From this we then compute a throttle value for the desired acceleration using the longitudinal PID controller ($K_p = 5$, $K_i = 0.5$, $K_d = 1$), as well as a weighting factor α indicating if the agent should decelerate (throttle = -1). The latter is high either when the current desired speed v_d is lower than an emergency brake threshold τ_{eb} or when the current speed is higher than v_d. In our kinematics model, which we detail in Section 2, the state update is then computed using the α-weighted sum of both the acceleration and deceleration throttle values. As this effectively allows to express the entire $[-1, 1]$ range of throttle values, we denote the action space of the driving policy as $\mathbf{a}_t^0 \in [-1, 1]^2$ in the main paper for notational convenience.

[4] Similarly, we use an additional channel in the BEV representation to encode pedestrians for deployment in CARLA, which is inactive when optimizing for and evaluating on KING scenarios.

Algorithm 1: Computing low-level actions from waypoints.

parameters: $\tau_{eb} \leftarrow 0.6$, $\beta_{eb} \leftarrow 10$, $\beta_{decel} \leftarrow 1$, $c_\delta \leftarrow 0.25$, $c_{throttle} \leftarrow 0.75$
input : v, $\{\mathbf{w}_i\}_{i=1}^4$
output : steer $\in [-1, 1]$, throttle $\in (0, 1)$, $\alpha \in (0, 1)$

$\mathbf{w}^* \leftarrow \frac{\mathbf{w}_1 + \mathbf{w}_2}{2}$;
$\Phi_d \leftarrow \tan^{-1}\left(\frac{\mathbf{w}^*[1]}{\mathbf{w}^*[0]}\right)$;
steer $\leftarrow \tanh\left(\text{LatPID}\left(\Phi_d\right)\right)$;
$v_d \leftarrow \|\mathbf{w}_2 - \mathbf{w}_1\| \times 2$;
$\alpha_{eb} \leftarrow \text{sigmoid}\left(-\beta_{eb} \times (v_d - \tau_{eb})\right)$;
$\alpha_{decel} \leftarrow \tanh\left(\beta_{decel} \times \text{ReLU}\left(v - v_d\right)\right)$;
if $v_d < \tau_{eb}$ **then**
  $\alpha \leftarrow \alpha_{eb}$
else
  $\alpha \leftarrow \alpha_{decel}$
end
$\delta \leftarrow c_\delta \times \tanh\left(\beta_\delta \times \text{ReLU}\left(v_d - v\right)\right)$;
throttle $\leftarrow c_{throttle} \times \tanh\left(\text{ReLU}\left(\text{LonPID}\left(\delta\right)\right)\right)$;

Training on regular data: For initial training of AIM-BEV, we collect a dataset $\mathcal{D}_{reg}$ of demonstrations by the rule-based expert described in Section 1.2 in regular traffic. We sample a dense set of routes for the expert to complete, covering all possible junctions in Town01-Town06 available in the CARLA simulator as well as routes outside of the junctions (e.g., highways). We include the hand-crafted CARLA scenarios during data collection to cover basic critical situations. In total, the dataset contains 91000 demonstrations, the equivalent of $12.6\,\text{h}$ of driving. We obtain ground truth supervision on the waypoints predicted by the driving policy from the future trajectory driven by the expert during data collection. The goal location $\mathbf{x}_{goal}$ is obtained by the route planner of [2], which is based on an A* planner. Using this, the sparse waypoint at least $7.5\,\text{m}$ ahead of the current position provided by CARLA is queried to obtain $\mathbf{x}_{goal}$. To train the driving policy using this dataset, we further employ data augmentation. For this, we rotate the BEV, waypoints and goal location around the vehicle center to simulate situations where the policy has to recover to stable lane keeping. We augment 50% of the samples with rotation angles sampled uniformly between ± 20 degrees. The policy is then trained using an L1-Loss on the waypoints for 350 epochs. We use the AdamW optimizer with a learning rate of 0.001 and a batch size of 128.

Fine-tuning with safety-critical scenarios: Similar to $\mathcal{D}_{reg}$, we build $\mathcal{D}_{crit}$ by collecting demonstrations from the rule-based expert. The overall approach remains similar. However now, instead of driving in CARLA under regular traffic, the expert solves scenarios generated by KING in our closed-loop proxy simulation. As described in the main paper, we first collect a larger set of KING scenarios that are solvable by the expert, containing around 8000 training samples. When training on a union of both datasets $\mathcal{D}_{crit} \cup \mathcal{D}_{reg}$, we ensure a certain proportion of critical to regular data in each minibatch. We find a mix of 60% regular and 40% critical data to work well in

practice. Aside from an increased batch size of 512 the training hyperparameters and augmentations remain the same.

1.2 Rule-based Expert

We use a rule-based expert policy based on privileged information to generate training data and determine solvability for the scenarios generated by KING. For the initialization of non-critical scenarios (see Section 4) we further adapt this expert to also work in our proxy simulation. We build on the code provided by [2, 3], which is based on the structure of the CARLA traffic manager[5]. It consists of three main components, (1) a collision stage, (2) a traffic light stage and (3) a motion planner stage, which we detail next.

Collision stage: In order to predict possible future collisions we use a kinematic bicycle model to extrapolate the position and orientation of all other agents within a radius of $30\,\mathrm{m}$. For the extrapolation we use an action repeat assumption (i.e., the exact same throttle and steer value is repeated for all extrapolation timesteps) since we do not have any information about the future ground truth actions. For the ego agent, we extrapolate assuming actions that maintain its current target speed. This allows reasoning about the possibility of a collision and necessary braking maneuvers. Outside of junctions, we extrapolate for a horizon of $1\,\mathrm{s}$. In junctions we extrapolate for a longer horizon of $4\,\mathrm{s}$ due to a higher probability of collision. We express each future timestep through a bounding box with the size of the respective vehicle and the extrapolated future position and orientation. With these bounding boxes we then check for overlaps between the ego and the other vehicles in each timestep to detect possible future collisions. In case of a bounding box intersection we set a brake hazard flag. As an additional security measure we extend the ego bounding box of the current timesteps to the front and check for intersections in this timestep. The extent is computed depending on the braking distance for the current ego speed.

Traffic light stage: In the traffic light stage we detect traffic lights and stop signs. For this we use privileged information provided by CARLA through trigger boxes. If there is an intersection of the boxes and the internal CARLA state of the traffic light is red or yellow we set the brake hazard flag.

Motion planner stage: For the main logic we refer to Section 2.2 of the supplementary material of [3]. However, we observed a higher driving score when modifying the set of target speeds. We use a default target speed of $4\,\mathrm{m/s}$. When entering an intersection this is increased to $5\,\mathrm{m/s}$. We observed that moving swiftly in intersections lowers the probability of collision. When the brake hazard flag is set, the target speed is always $0\,\mathrm{m/s}$. To adapt to the lower frame rate in our simulator ($4\,\mathrm{fps}$ vs $20\,\mathrm{fps}$), we use adjusted gains (K_p = 1.75, K_i = 0.75, K_d = 3.5) for the lateral controller (only in our simulator). We leave the longitudinal controller unchanged.

[5] https://carla.readthedocs.io/en/latest/adv_traffic_manager/

1.3 Driving Metrics for CARLA Benchmarking

In the main paper we benchmark the driving agents wrt. to their driving performance in CARLA. This is measured by the *Driving Score (DS)*, which is composed of two terms, the Route Completion and the Infraction Score, which we detail in the following. The *Route Completion (RC)* is given by

$$\text{RC} = \frac{1}{N_R} \sum_{i}^{N_R} R_i \tag{1}$$

where R_i is the percentage of the total distance to be driven on route i that the agent has completed, and N_r is the total number of routes. This is complemented by the *Infraction Score (IS)*, which measures the agent's compliance with traffic rules. It is computed as a product of infraction penalties p_i^j for each instance i of an infraction of type j. An overview of the type-specific penalties is shown in Table 1. In addition to these, there is one further penalty for deviations from the drivable area. This is computed as $(1-R_{offroad})$, where $R_{offroad}$ denotes the percentage of the route completed outside of the drivable area. Overall, the agent starts with a base IS of one, which is reduced for each infraction. The DS is then computed as the product of the RC and IS, averaged over all routes.

Infraction Type	Penalty
Collisions with pedestrians	0.50
Collisions with other vehicles	0.60
Collisions with static elements	0.65
Running a red light	0.70
Running a stop sign	0.80

Table 1: **Infraction Score (IS) penalties for different infraction types.**

2 Kinematics Model

For the kinematic bicycle model we follow previous work [1] and use the author provided code and parameters, which were fitted to the CARLA dynamics and default vehicle model. Their implementation constrains the speed to be positive via a hard threshold to prevent backwards motion. This zeroes out gradients through the indirect path in timesteps where the current throttle action would result in negative speeds. We hence replace the hard threshold with a softplus activation. We set its shape parameter β to 7, which we found to result in sufficiently low error.

3 Cost Functions

Here, we provide more details of the costs from Section 3 Eq. 2 of the main paper. We begin with the ego-collision cost $\phi_{col}^{ego}(\mathcal{S})$ followed by the two regularization terms $\phi_{col}^{adv}(\mathcal{S})$ and $\phi_{dev}^{adv}(\mathcal{S})$.

Ego Collision Cost: The ego collision cost $\phi_{col}^{ego}(\mathcal{S})$ is an attractive potential encouraging close encounters between the driving policy and the adversarial agents. This attractive potential is proportional to the distance between the ego agent and the adversarial agents, for which the geometry of the ego agent and the adversarial agents is taken into account. Let $d_{bb}(\mathbf{s}_t^i, \mathbf{s}_t^j)$ denote the Euclidean distance in $\mathbb{R}^2$ between closest points on the bounding polygons of the i^{th} and j^{th} agents at time t. We minimize the Euclidean distance between the ego agent and closest adversarial agent. We choose only the closest adversarial agent in order to discourage situations where multiple adversaries deviate from their trajectory to collide with the ego agent:

$$\phi_{col}^{ego}(\mathcal{S}) = \min_{i \in \{1,\dots,N\}} \frac{1}{T} \sum_{t=0}^{T} d_{bb}(\mathbf{s}_t^0, \mathbf{s}_t^i). \tag{2}$$

The distance $d_{bb}(\mathbf{s}_t^i, \mathbf{s}_t^j)$ can be efficiently computed in parallel as the minimum of the distances between every vertex-edge pair of the two agents' bounding boxes.

Adversarial Collision Regularization: To improve the physical plausibility of the generated scenarios, we discourage collisions between adversarial agents through a repulsive potential $\phi_{col}^{adv}(\mathcal{S})$. This potential is inversely proportional to the distance between the closest pair of adversarial agents:

$$\phi_{col}^{adv}(\mathcal{S}) = -\min\Big(\min_{i,j \in \{1,\dots,N\},\ t \in \{0,\dots,T\}} d_{bb}(\mathbf{s}_t^i, \mathbf{s}_t^j), \tau\Big). \tag{3}$$

We find it sufficient to apply the repulsive potential ϕ_{col}^{adv} to the closest pair of adversarial agents. Finally, we apply a threshold τ to the distance, acting as a safety margin. This ensures that the repulsive potential is active only in a local neighbourhood around each adversarial agent j, preventing penalization of agents which are further apart than this threshold. We find $\tau = 1.25\,\mathrm{m}$ to work well in practice.

Out-of-bounds Regularization: Finally, we regularize adversarial agents to stay within drivable parts of the map. For this, we use a repulsive potential $\phi_{dev}^{adv}(\mathcal{S})$ between the adversarial agents and off-road areas. To ensure that this is compliant with the geometry of the vehicle, the overall cost is composed of multiple terms defined as Gaussian potentials $g(\mathrm{B}_t^i(c_k))$ at each corner c_k of the bounding rectangle B_t^i corresponding to the adversarial agent i. The potential at each corner is obtained through a convolution operation between the Gaussian potential $g(\mathrm{B}_t^i(c_k))$ and the map $\mathcal{M}_{oob}$, which is $g(\mathrm{B}_t^i(c_k)) \circledast \mathcal{M}_{oob}$. We consider a binary map $\mathcal{M}_{oob}$ which marks the non-drivable areas. The potential $g(\mathrm{B}_t^i(c_k)) \circledast \mathcal{M}_{oob}$ is thus large when the corner c_k is near (or on) a non-drivable area. The full "out-of-bounds" repulsive potential is defined as sum over all corners of the polygon B_t^i across agents and over all timesteps:

$$\phi_{rddev}^{adv} = \frac{1}{T} \sum_{i=0}^{N} \sum_{t=0}^{T} \sum_{k=0}^{C} g(\mathrm{B}_t^i(c_k)) \circledast \mathcal{M} \tag{4}$$

In practice, a variance of $1\,\mathrm{m}$ works well for the Gaussians. This ensures that the potential is large when any corner is $\sim 0.5\,\mathrm{m}$ from any non-drivable area.

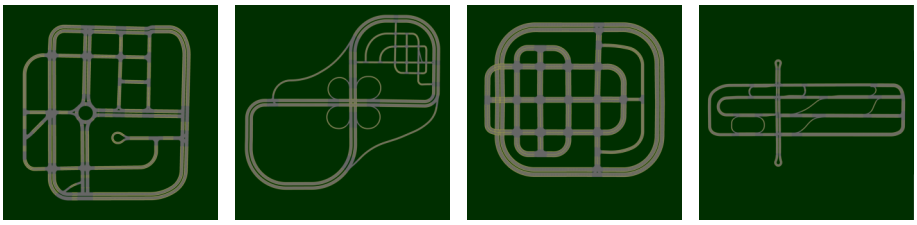
Cost Weighting: As the number of actors in the scene is always known ahead of optimization, we use separate cost weightings for each traffic density. This accommodates for the fact that for different densities, a slightly different emphasis of the regularization terms can be optimal. The used weightings can be found in Tab. 2.

# agents	γ (for ϕ_{dev}^{adv})	λ (for ϕ_{col}^{adv})
1	20	0
2	23	5
4	20	3

Table 2: **Cost weightings.** Weighting of the costs for different traffic densities.

4 Scenarios

In this section we provide additional information on the maps and routes as well as the traffic initialization scheme, which provide the basis for the dataset of initial non-critical scenarios used throughout the main paper.



(a) Town03 (b) Town04 (c) Town05 (d) Town06

Fig. 2: **Town layouts.** Each town has unique features, such as roundabouts (a) or highway junctions (b, d).

Maps and Routes: We use routes from towns 3-6 provided by the CARLA simulator for optimization. The layouts of these towns are illustrated in Fig. 2. Each town has unique features in terms of road topology, such as multi-lane highways, highway exits or roundabouts. From each of these towns we densely sample routes through junctions and intersections to obtain a set of 714 routes in total, each of which is a potential mission for the ego agent. This forms a diverse basis for possible scenarios. For evaluation on the Town10 intersections, we sample additional routes using the same sampling mechanism from CARLA's Town10, which we found to be the most challenging for the driving agents. This set consists of 82 different routes. We use the maximum traffic density for Town10 (156 background vehicles) and spawn one hand-crafted CARLA

scenario[6] along each route. Specifically, we consider Scenarios 7, 8, 9, and 10, which involve multiple vehicles simultaneously entering the intersection.

Initializating Traffic: As a starting point for safety-critical scenario generation, we require initial non-critical scenarios. Furthermore, to be able to study the effects of different traffic densities, the number of adversarial agents present in these initializations should be controllable. To this end, we mimic CARLA traffic by controlling the adversarial agents with the rule-based expert described in Section 1.2 to obtain an initialization. Specifically, we first sample a route that passes by the ego agent's route from the dense set of 714 routes for every adversarial agent. For this we consider routes which start and end at least $4\,\mathrm{m}$ and at most $100\,\mathrm{m}$ away from the ego agent's route. If possible, each agent is spawned on an individual route at the corresponding start location. If the set of candidate routes is smaller than the number of adversarial agents, additional spawn locations further along already occupied routes are used for the remaining agents while making sure a distance of $6\,\mathrm{m}$ between spawn locations is kept. Finally, using the expert, we roll out initial trajectories for the adversarial agents that follow these routes while avoiding collisions with each other and the ego agent. While this strategy lends itself well to study the effects of different traffic densities, KING is agnostic to the source of initial scenarios and other choices, for example a dataset of recorded traffic logs, may be more suitable in practice.

Pre-filtering: For our experiments we sample a subset of the available routes. Instead of directly sampling from the total set of 714 routes, we first apply a pre-filtering step. For this, we initialize a scenario on every route as described above. Then, we remove all routes for which the scenario terminates due to an adversarial agent collision within the first 10 timesteps. This is done to handle edge cases not covered by the spawning logic. After pre-filtering, we sample 20 routes and corresponding initial scenarios per town from the remaining set to obtain the benchmark routes used to compare KING against black-box optimization (BBO) baselines in Tab. 2 of the main paper. To build the larger set used in the fine-tuning experiments for improving robustness, we use the entire set of routes remaining after pre-filtering.

5 Additional Results and Details

In this section we provide additional results supplementing the findings in the main paper. First, we detail the optimization hyperparameter choices for all methods used to generate safety-critical scenarios, investigate the efficacy of using second-order optimizers for gradient-based generation and perform an ablation study on the regularization terms in the generation objective. Next, we assess the impact of their higher computational burden for KING with both the direct and indirect path and for Bandit-TD. Then, we provide experiments on an additional simple baseline. Furthermore, we evaluate AIM-BEV on a benchmark of longer routes involving rural and highway driving. Finally, we provide additional qualitative examples of safety-critical scenarios generated by KING.

[6] https://leaderboard.carla.org/scenarios/

Method	Hyperparameter	Sweep Range	$N = 1$	$N = 2$	$N = 4$
Random Search	Perturbation Bounds	$[0.01, 0.3]$	0.2	0.2	0.2
Bayesian Optimization	Exploration Kappa κ	$[0.1, 10]$	5	10	0.1
	Initial Random Points	$[1, 50]$	20	20	20
SimBA [6]	Perturbation Bounds	$[0.001, 0.2]$	0.05	0.05	0.05
CMA-ES [7]	Initial Standard Deviation σ	$[0.1, 0.4]$	0.2	0.1	0.4

Table 3: **Hyperparameter choices for BBO baselines.**

5.1 Optimization Hyperparameters

For gradient-based generation via KING, we use the Adam optimizer with a learning rate of $5e-3$. We use a decay rate of $\beta_1 = 0.8$ for all traffic densities, and $\beta_2 = 0.999$ for a traffic density of $N = 1$ and $\beta_2 = 0.99$ for $N \in \{2, 4\}$. We also adopt this to optimize the scenario parameters via the numerical gradients obtained from Bandit-TD. For the experiment involving both the direct and the indirect path (cf. Tab 2 of the main paper), we observed exploding gradients. To combat this, we clip the norm of the gradients through the rasterized BEV, where we found a maximum norm of $5e-2$ to work well. An overview of the remaining hyperparameters can be found in Tab. 3. For Random Search and SimBA [6], we tune the bounds of the interval from which perturbations can be sampled. For Bayesian Optimization, we search for both the optimal exploration value κ and the number of initial random points. Finally, for CMA-ES [7], we select an optimal initial standard deviation σ.

5.2 Second-Order Optimizers

Relative to the standards of modern deep learning, the parameter space is low-dimensional and the computational graph is simple for KING. This suggests that utilizing second-order gradients might be feasible for this problem, which we investigate in this experiment. On the example of Newton's method, we find that computing the exact Hessian with PyTorch's autograd engine, even for our low-dimensional search space, leads to 1-2 orders of magnitude slower runtime compared to Adam (see Table 4). This is a result of the additional backward passes needed to compute second-order derivatives. This overhead does not permit a single optimization step within our computational budget for the 4 agents setting. We also investigated quasi-Newton methods (L-BFGS), but did not observe any benefits over Adam (Table 5). This is due to two factors: the runtime penalty of requiring multiple function evaluations, each of which is a complete unroll of the simulation (we found a maximum of five to work best), and a susceptibility to local minima.

5.3 Cost Ablation Study

We study the influence of the regularization terms in our objective in Tab 6. While the out-of-bounds term $\phi_{dev}^{adv}(\mathcal{S})$ leads to a drop in performance when considering the regularization terms in isolation due to its computational expense, it is evident that

Method	1 Agent	2 Agents	4 Agents
KING, 1st-order (cf. Tab. 2 of main paper)	1.78	1.88	2.03
KING, 2nd-order	81.31	156.71	316.89

Table 4: **Timings for second-order gradients in s/it.**

Optimizer	CR	$t_{50\%}$	s/it
Adam	78.75	6.40	2.03
L-BFGS	65.00	16.79	9.31

Table 5: **Results on KING using L-BGFS in the 4 agent setting.**

ϕ_{col}^{ego}	ϕ_{dev}^{adv}	ϕ_{col}^{adv}	CR ↑	$t_{50\%}$ ↓	s/it ↓
✓	-	✓	66.25	6.94	1.59
✓	✓	-	57.50	10.21	2.03
✓	-	-	60.00	7.08	1.59
✓	✓	✓	**78.75**	**6.40**	2.03

Table 6: **Cost ablation.** We show the performance for different regularization combinations for the 4 agent setting. Both regularization terms in combination with our main objective results in the best performance.

combining both terms is crucial to achieve good results. Note that the losses for ego-adversarial and adversarial-adversarial collisions are computed in parallel as a batched operation, resulting in the same runtime when deactivating one of them.

5.4 Convergence

Bandit-TD and KING with both the direct and indirect gradient path are disadvantaged by their higher computational expense when imposing a fixed wall clock budget. We are interested in assessing their performance in the absence of this. For this, we show both the overall CR over GPU seconds for a less strict budget and the overall CR over optimization iterations (Fig. 3). As can be seen, given a high enough budget, Bandit-TD can eventually partially close the gap to other BBO methods. For KING with both paths, we observe that given the same number of optimization iterations, it achieves similar results compared to the indirect path only. Note that the results for KING with both paths stem from a preliminary experiment and thus perform slightly worse than the final version reported in Tab. 2 of the main paper.

5.5 Additional Baseline: Pose Jittering

As both AIM-BEV and TransFuser do not take temporal inputs, simply jittering agent poses for the single input frame is a reasonable simple baseline. However, even though they do not require temporally consistent inputs, temporally consistent rollouts from the jittered frame need to be possible in order to obtain ground-truth future waypoints from

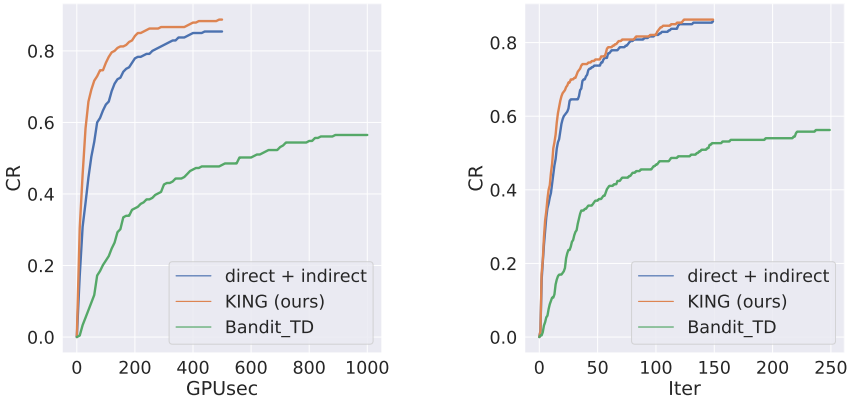


Fig. 3: **Results at higher compute budget.** We compare KING to KING with both the direct and indirect path, and to Bandit-TD for a larger compute budget (left) and in terms of optimization iterations (right).

the expert. To gauge the effectiveness of jittering the pose to find safety-critical perturbations, while maintaining temporally consistent rollouts, we construct the following experiment: We first find the time instant and adversarial agent closest to causing a collision in the initial scenario. We then jitter the pose of that agent in a Δ_t seconds earlier timestep by sampling perturbations uniformly from a pre-defined interval, while making sure the perturbations do not trigger our termination conditions. We then continue rolling out the scenario from the jittered state for 2 seconds to be able to collect ground-truth waypoints. The results for different Δ_t and perturbation magnitudes are shown in Table 7. In addition to being applicable to only single frame agents, we observe this approach to be much less effective than KING at generating safety-critical data. It struggles to find plausible perturbations, resulting in a significantly reduced CR. Furthermore, we observe that among the obtained perturbations a much smaller percentage are solvable by the rule-based expert. In conjunction, this results in an overall significantly smaller volume of usable safety-critical data.

Method	Interval	Δ_t	CR	$t_{50\%}$	s/it	Solvable
Jitter	$\pm 1m$ / $\pm 15°$	-1s	42.50	-	1.10	64.71%
	$\pm 1m$ / $\pm 15°$	-2s	43.75	-	1.10	65.72%
	$\pm 2m$ / $\pm 15°$	-1s	61.25	9.04	1.10	53.06%
	$\pm 2m$ / $\pm 15°$	-2s	55.00	7.84	1.10	47.73%
KING (Ours)	–	–	78.75	6.40	2.03	76.25%

Table 7: **Pose jittering in the 4 agent setting for AIM-BEV.**

5.6 Evaluation on the Longest6 Benchmark

We also evaluate AIM-BEV before and after fine-tuning on $\mathcal{D}_{crit} \cup \mathcal{D}_{reg}$ on the Longest6 benchmark [4]. It contains long routes, which in addition to urban settings,

also involve rural and highway driving similar to the NEAT validation benchmark [3], but at the maximum traffic density allowed by the CARLA simulator. Here, both the regular and fine-tuned version of AIM-BEV achieve driving scores of $35.38_{\pm 1.55}$ and $38.84_{\pm 2.96}$ respectively. Because this benchmark places additional emphasis on other aspects of the driving task not specifically targeted by KING, such as safe lane-changing and -keeping on multilane roads over extended distances, we instead choose the dense urban intersection setting in Town10 for evaluation in the main paper.

5.7 Additional Qualitative Examples and Failure Cases

Here, we show additional qualitative examples of safety-critical scenarios for both AIM-BEV (Fig. 4-9) and TransFuser (Fig. 10). As can be seen, they cover diverse situations and highlight the brittleness of current imitation learning (IL)-based driving agents.

Finally, we discuss failure cases of KING and possible strategies for mitigation. As we aim to cover the long tail of the distribution of traffic scenarios, we impose few constraints on the trajectories of other traffic participants. The formulation of the objective $\mathcal{C}(\mathcal{S})$ in Eq. (2) of the main paper only penalizes deviations from drivable areas, but does not enforce any further restrictions within those areas. While this allows flexibility in the discovery of diverse safety-critical scenarios, in some cases, the trajectories of background agents that do not interact with the ego agent may be perturbed, causing multiple vehicles to leave the trajectory they followed at initialization. This is illustrated in Fig. 11. Nonetheless, as can be seen from the experiments in Sec. 4.4 of the main paper, these scenarios provide useful training data that lead to improved robustness.

These non-interactive perturbations could be mitigated by adding additional regularizing terms to the overall objective, e.g. a penalty for deviation from an assigned lane center. However, using a more restrictive objective may lead to less safety-critical data overall and reduced coverage of out-of-distribution situations.

Fig. 4: **AIM-BEV - Cluster (a).** The driving agent collides with a slow moving vehicle while taking a turn (left). Additionally, it often struggles to handle lane-change scenarios safely (right). This is also evident in cluster (c).

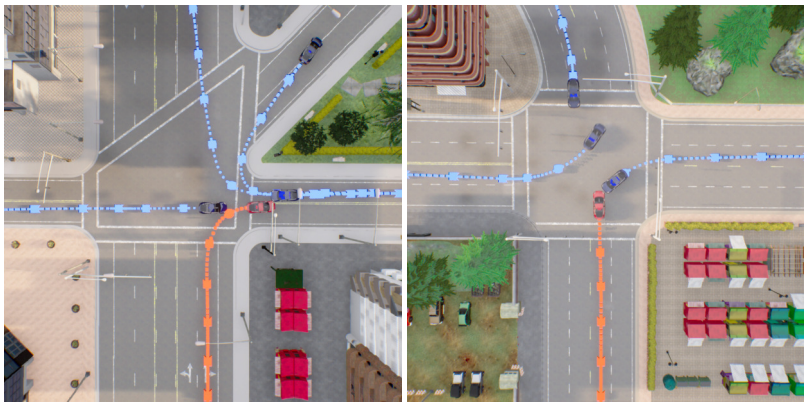


Fig. 5: **AIM-BEV - Cluster (b).** The driving agent understeers when taking a turn, colliding with another vehicle that has been slightly displaced from its lane center by the safety-critical perturbation (left). Background traffic in CARLA usually closely adheres to the lane centers, making penalization of bad lane following less likely. On the right hand side, overconfident behavior in an unprotected intersection leads to collision.

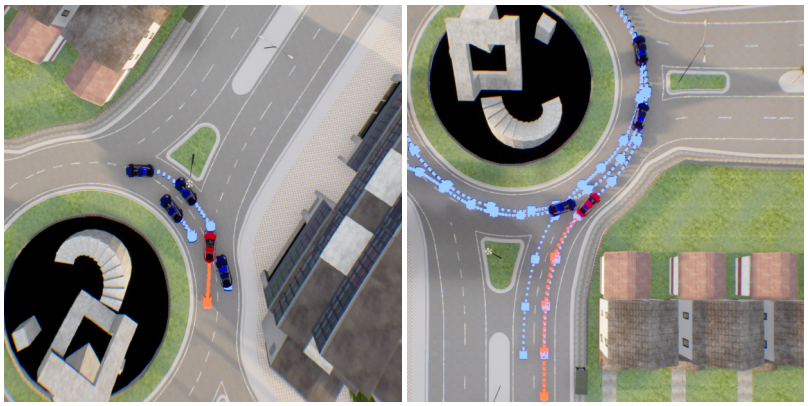


Fig. 6: **AIM-BEV - Cluster (c).** The driving agent fails to handle merges and lane-changes in a roundabout safely. These maneuvers are especially problematic due to AIM-BEV's forward facing field of view. This shortcoming is not as apparent in regular CARLA traffic, highlighting KING's ability to draw attention to possible failure modes.

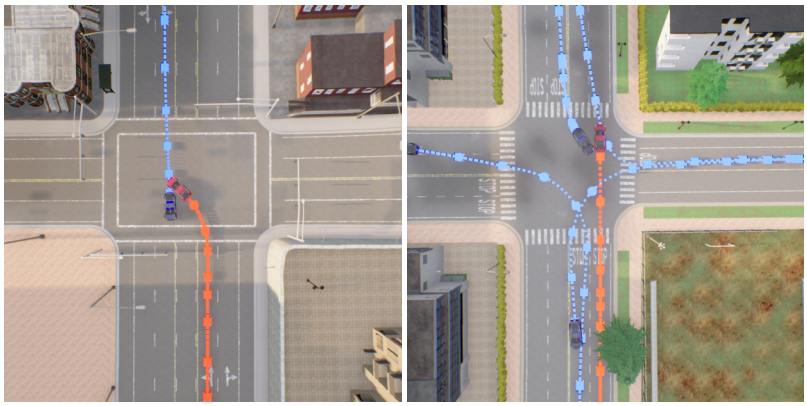


Fig. 7: **AIM-BEV - Cluster (d).** This cluster primarily features collisions at the side of the driving agent. As depicted, this type of collision can arise from insufficient safety distance while turning (left) or failure to yield for another vehicle taking an aggressive turn (right).

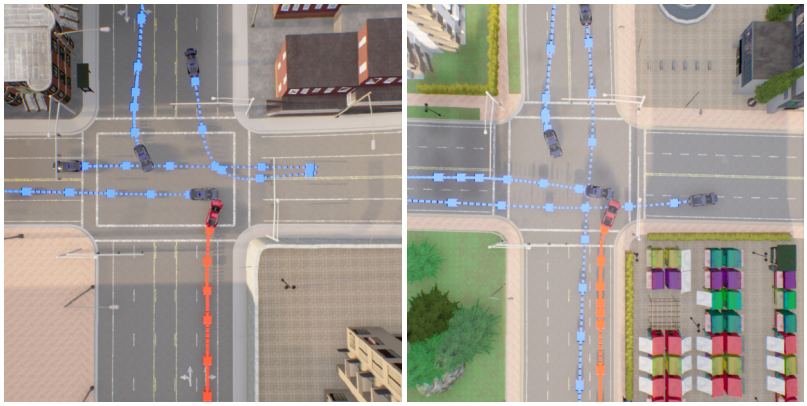


Fig. 8: **AIM-BEV - Cluster (e).** The driving agent fails to brake when traffic in the intersection does not yield. As a result the agent collides into the side of another vehicle.

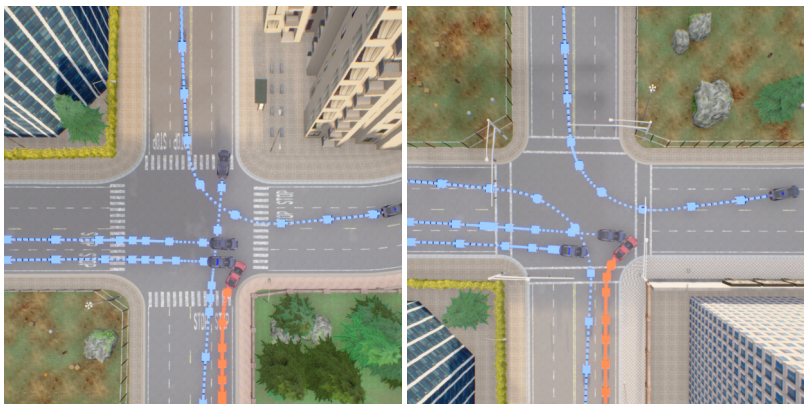


Fig. 9: **AIM-BEV - Cluster (f).** The examples shown here depict similar situations as in cluster (e). The ego agent fails to correctly assess if it is safe to continue on, resulting in other traffic colliding into its side.

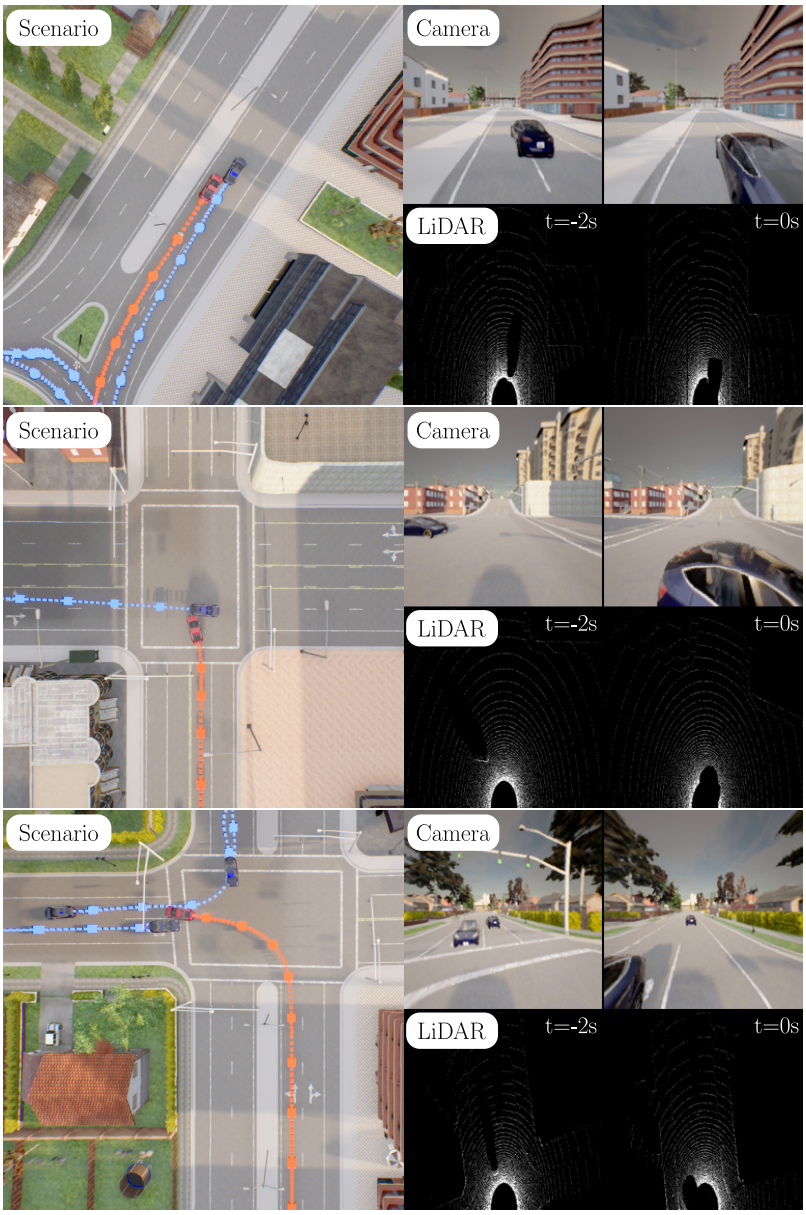


Fig. 10: **Additional Qualitative Examples for TransFuser.** We show additional examples of safety-critical perturbations for TransFuser. It shares many of the failure modes of AIM-BEV, such as difficulty handling merges (top), unsafe behaviour when moving through intersections (middle) or the reliance on other vehicles strictly adhering to their lane centers and stopping points at intersections (bottom). The common failure modes of AIM-BEV and TransFuser suggest that these problems indeed stem from insufficient representation in the regular traffic data obtained from CARLA.

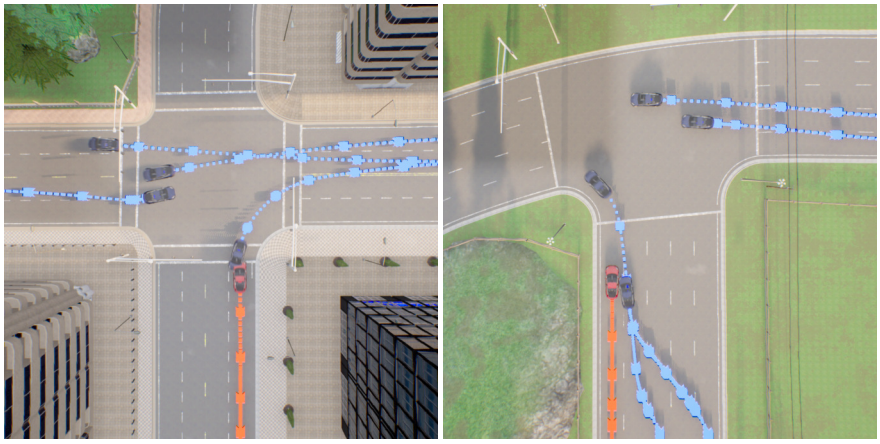


Fig. 11: **KING failure cases.** As we aim to cover the long tail of the distribution of traffic scenarios, we impose few constraints on the trajectories of other traffic participants. This can lead to unlikely behaviour in the background actors by definition, such as moving into opposing lanes (left) or performing seemingly unnecessary lane changes (right).

References

1. Chen, D., Koltun, V., Krähenbühl, P.: Learning to drive from a world on rails. In: Proc. of the IEEE International Conf. on Computer Vision (ICCV) (2021) 5
2. Chen, D., Zhou, B., Koltun, V., Krähenbühl, P.: Learning by cheating. In: Proc. Conf. on Robot Learning (CoRL) (2019) 2, 3, 4
3. Chitta, K., Prakash, A., Geiger, A.: Neat: Neural attention fields for end-to-end autonomous driving. In: Proc. of the IEEE International Conf. on Computer Vision (ICCV) (2021) 1, 2, 4, 12
4. Chitta, K., Prakash, A., Jaeger, B., Yu, Z., Renz, K., Geiger, A.: Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. In: arXiv.org (2022) 11
5. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An open urban driving simulator. In: Proc. Conf. on Robot Learning (CoRL) (2017) 1
6. Guo, C., Gardner, J.R., You, Y., Wilson, A.G., Weinberger, K.Q.: Simple black-box adversarial attacks. In: Proc. of the International Conf. on Machine learning (ICML) (2019) 9
7. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation **9**(2), 159–195 (2001) 9
8. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H.: Searching for mobilenetv3. In: Proc. of the IEEE International Conf. on Computer Vision (ICCV) (2019) 1
9. Prakash, A., Behl, A., Ohn-Bar, E., Chitta, K., Geiger, A.: Exploring data aggregation in policy learning for vision-based urban autonomous driving. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2020) 2
10. Prakash, A., Chitta, K., Geiger, A.: Multi-modal fusion transformer for end-to-end autonomous driving. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2021) 1