第 13 卷 第 2 期
2024 年 3 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 13 No. 2
Mar. 2024

引文格式：
桑明, 蒋拯民, 李慧云. 自动驾驶汽车的高效对抗性场景测试方法研究 [J]. 集成技术, 2024, 13(2): 15-28.
Sang M, Jiang ZM, Li HY. Efficient adversarial scenario test for autonomous vehicles [J]. Journal of Integration Technology, 2024, 13(2): 15-28.

自动驾驶汽车的高效对抗性场景测试方法研究

桑 明[1,2] 蒋拯民[1,2] 李慧云[1*]

[1]（中国科学院深圳先进技术研究院 深圳 518055）
[2]（中国科学院大学 北京 100049）

摘 要 在自动驾驶安全性的研究和应用中，测试里程长、暴露危险场景单一的问题使自动驾驶安全性能的提升受到限制。使用对抗性场景进行测试被认为是解决上述问题的重要手段，然而，现有研究采用通用的优化算法作为框架，将大量计算资源浪费在对参数空间的探索过程中，效率低下。在计算成本的约束下，这些算法甚至无法在更复杂的环境中测试出足够多、足够丰富的失效样本。复杂环境中的对抗性场景测试面临三大挑战：信息匮乏；对抗性样本在庞大的参数空间中稀疏分布；搜索过程中探索与利用难以平衡。该文从这三大挑战出发，提出一种高效的对抗性场景测试框架，通过代理模型来获取更多关于参数空间的信息，精选小样本，以打破庞大空间中稀疏事件的制约，对未知区域和对抗性样本附近的目标进行有针对性的搜索和更新，以实现探索和利用的平衡。实验证明，该文提出方法的搜索效率是随机采样的 4 倍，与通用遗传算法相比，效率提升一倍以上，在有限的仿真测试次数下，生成了更多容易使被测自动驾驶系统失效的对抗性测试用例。特别地，该文提出的方法能够找出许多离群的对抗性样本，揭示出现有算法无法识别的失效模式。此外，该文提出的方法能够快速、全面地定位出被测算法的脆弱场景，为自动驾驶算法的测试验证、迭代升级提供支持。

关键词 自动驾驶；安全验证；场景测试；代理模型；智能优化算法；Kriging 模型
中图分类号 TP 391.9；U 463.6 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20230726001

Efficient Adversarial Scenario Test for Autonomous Vehicles

SANG Ming[1,2] JIANG Zhengmin[1,2] LI Huiyun[1*]

[1](*Shenzhen Institute of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[2](*University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)
[*]*Corresponding Author*: *hy.li@siat.ac.cn*

Abstract In the field of autonomous driving safety research and application, the limitations of limited testing

收稿日期：2023-07-26 **修回日期**：2023-10-17
基金项目：深圳市基础研究重点项目（JCYJ20200109115414354, JCYJ20200109115403807）；广东省基金项目（2020B515130004, 2023A1515011813）
作者简介：桑明，硕士研究生，研究方向为自动驾驶安全性；蒋拯民，博士研究生，研究方向为自动驾驶仿真测试；李慧云（通讯作者），研究员，研究方向为自动驾驶、智能网联汽车等，E-mail：hy.li@siat.ac.cn。

mileage and exposure to only a single hazardous scenario hinder the improvement of autonomous driving safety performance. To address these issues, testing with adversarial scenarios is considered crucial. However, existing studies utilize generic optimization algorithms as frameworks, resulting in a wastage of computational resources in exploring the parameter space, thereby leading to low efficiency. Moreover, under the constraint of computational cost, these algorithms may not be able to test a sufficient number of diverse failure samples, especially in complex environments. Adversarial scenario testing in complex environments faces three major challenges: information scarcity, sparse distribution of adversarial samples in a vast parameter space, and the difficulty in balancing exploration and exploitation during the search process. To tackle these challenges, this paper proposes an efficient framework for adversarial scenario testing. This framework employs a surrogate model to gather more information about the parameter space, selects small samples to overcome the sparse event constraints in the vast space, and focuses on the unknown regions and adversarial samples for targeted search and update, thereby achieving a balance between exploration and exploitation. Experimental results demonstrate that the proposed method in this paper exhibits a search efficiency four times higher than random sampling and more than double the efficiency compared to general genetic algorithms. Additionally, with a limited number of simulation test runs, it generates a greater number of adversarial test cases that are likely to cause the tested autonomous driving system to fail. Notably, the proposed method can identify many outlier adversarial samples, unveiling failure modes that existing algorithms fail to recognize. Furthermore, the proposed method can swiftly and comprehensively identify the vulnerable scenarios of the tested algorithm, providing support for the testing, validation, and iterative upgrade of autonomous driving algorithms.

Keywords autonomous driving; safety test and validation; scenario-based test; meta model; intelligent optimization algorithms; Kriging model

Funding This work is supported by Shenzhen Basic Key Research Project (JCYJ20200109115414354, JCYJ20200109115403807) and Foundation of Guangdong Province of China (2020B515130004, 2023A1515011813)

1 引 言

自动驾驶的相关研究和技术蓬勃发展，已在一些国家运用部署。但是，如何保证自动驾驶汽车在各类场景下均能安全运行，仍然是制约自动驾驶大规模落地应用的主要因素[1]。在这一背景下，对自动驾驶系统充分地进行测试，并确定其安全边界，是一种重要的安全保障手段。基于场景的自动驾驶仿真安全测试方法以测试场景为核心，通过演绎被测系统可能经历的环境和事件，评估被测系统在各个场景下的表现，进而定位出被测自动驾驶系统的薄弱环节。当前，在安全性制约自动驾驶系统大规模部署的背景下，安全冗余备份系统、功能降级保障方案、自动驾驶系统准入认证及自动驾驶算法本身的优化迭代都必须以高效、丰富的场景测试为基础[2-4]。

根据使用场景的不同，自动驾驶的测试方法分为一般性测试和强化测试两大类[5]。前者使用的场景是随机生成的，或从自然驾驶数据中随机采样得到；而后者则使用容易使被测系统发生危险的场景进行针对性测试。一般性测试简便快速，常被用于测试自动驾驶系统的基本功能能否实现。但其测试完备性差、效率低，难以发掘被测自动驾驶系统的安全边界，对提升车辆安全性

的作用有限。强化测试因其高效性和针对性，被认为是探索自动驾驶系统安全边界的重要手段。但是，如何构建高价值的测试场景仍然是当前亟待解决的问题。国际标准化组织[6]从预期功能安全角度将自动驾驶汽车的测试场景分为已知安全场景、已知危险场景、未知安全场景和未知危险场景。其中，未知危险场景对自动驾驶系统的威胁最大，在相关文献[7]中也被称为对抗性场景。对抗性场景的构建和测试方法是目前研究的重点[8]。

现有的对抗性场景生成方法包括专家经验法、数据驱动法和优化搜索法等[1]。专家经验法通过对造成危险的因素进行分析和归纳，总结出造成危险的因素，进而手动设置场景条件和参数以构造潜在的危险场景[9]。受限于对场景的理解，专家经验法在不同的功能场景下表现不均衡，在复杂环境下难以保证关键危险场景的完备性。数据驱动法以大型交通数据集为基础，从大规模数据中发掘造成危险的特征，通过生成对抗网络等方法泛化生成新的对抗性场景[10]。该方法虽然克服了对专家知识的依赖，但其性能仍然受限于交通场景数据的丰富度。在现阶段，建立足够丰富的交通场景数据集仍面临许多难以解决的困难。此外，随着深度强化学习的发展，有学者提出通过强化学习的方法生成对抗性场景[11-13]。例如，Chen 等[14]提出将对抗强化学习方法用于对抗性场景的生成，该研究结合非零和博弈理论设计新的奖励函数，并通过聚类和组合的技巧得到了分布在多个局部最优区域内的对抗性场景。然而，由于对抗性场景在测试空间中是稀疏的，因此，在强化学习方法的早期，需要进行大量探索，在实际运用中，效率低下。在相对复杂的场景下，该方法需要进行的评估测试次数在工程上几乎是不可实现的。基于优化搜索的场景生成方法[15-17]将危险场景的构建问题转化为优化问题：以危险程度为目标函数，在功能场景或逻辑场景定义的参数空间中进行搜索，目标函数最优解集对应的参数子空间即为所需的危险场景。基于优化搜索的场景生成方法几乎不依赖专家知识，也无须事先建立场景数据集，并且其生成的危险场景具有较高的测试性和覆盖度，因而得到了广泛应用[4,13]。

与随机采样相比，现有基于优化搜索生成关键危险场景的方法虽然取得了一定的效率提升，但这些基于优化搜索的方法并不完全契合关键危险场景生成的特征，将大部分算力都耗费在了“探索”过程中。Yasasa 等[18]提出通过贝叶斯优化方法来平衡算法的探索和利用过程，并在一个简单场景下进行了验证，生成了一定数量的高质量对抗性场景。该方法在一定程度上解决了对抗性场景分布稀疏的问题，但当扩展到更接近真实交通环境的复杂功能场景时，运算成本会呈指数级增加，因而无法在可接受的成本下生成数量足够多、失效模式足够丰富的对抗性场景。

在复杂功能场景下，通过优化方法进行对抗性场景测试面临如下挑战。

(1)对抗性场景分布信息匮乏：目标函数不可导，搜索算法没有梯度信息可以使用，这使得基于梯度的优化算法不适用[11]；而对于以遗传、进化算法为代表的智能搜索算法来说，虽然不需要梯度信息，但仍然需要计算大量样本点的目标函数值，这一操作需要对每个样本点对应的具体场景进行测试，成本十分高昂。

(2)对抗性样本分布稀疏：参数空间包括自主行驶车辆、环境背景车辆、交通道路设施及自然环境状况等多种要素，这些要素组合后高达数十个维度[19]；在该庞大的状态空间中，占比不高的对抗性样本是稀疏分布的。

(3)探索与利用难以平衡：在优化搜索的过程中，若倾向于利用已知的对抗性样本，在已知对抗性样本附近采样进行测试，虽然能够获得大量失效用例，但其失效模式是趋同的，得到的用

例价值有限；反之，若倾向于探索未知空间，则无法尽快得到足够多的对抗性样本，难以确定参数空间的安全边界。

针对上述挑战，本文提出一种用于自动驾驶的高效对抗性场景测试(efficiency adversarial scenarios test，EAST)方法，通过以下 3 种手段的组合，大幅提高算法在复杂、庞大的特征空间下的效率。

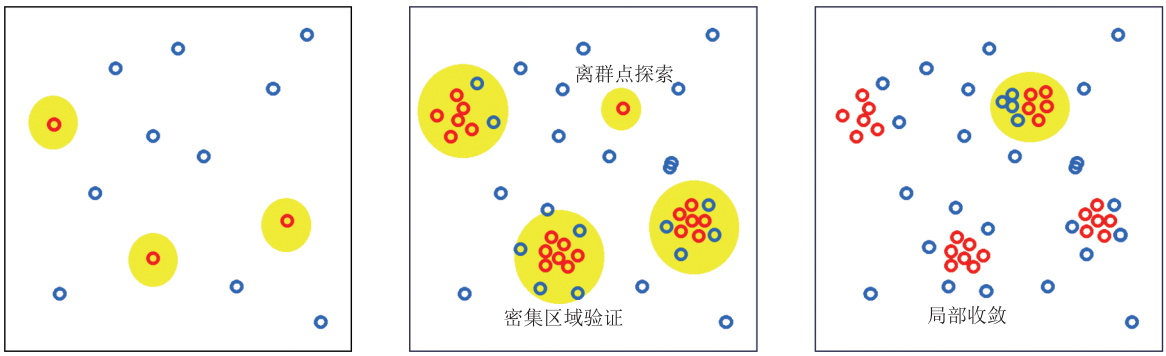
(1)以代理模型为枢纽，捕获并建模搜索过程中的样本分布信息，从而为搜索算法提供信息引导；

(2)通过拉丁超方采样精选少量样本，构建参数空间中稀疏事件的大致分布，以此作为搜索算法的先验，从而避免算法在庞大空间中搜索稀疏事件的盲目性[20]；

(3)提出一种搜索更新策略，使用代理模型更加充分地对未知区域和对抗性样本附近的目标进行“探索”，而后对代理模型认为价值较高的目标点进行验证，以实现“利用”。

在上述 3 种手段的结合下，本文提出的方法达成了一种高效的搜索过程，如图 1 所示。首先精选少量样本进行仿真测试，并将对抗性样本附近的区域作为感兴趣区域，其结果反映了参数空间中对抗性样本和非对抗性样本的大致分布情况(图 1(a))。其次，在这一先验信息的基础上，一方面，在已知对抗性样本周边区域密集地进行验证，以扩大各个感兴趣区域；另一方面，充分探索未知区域，将新出现的离群样本点周边一定范围标记为感兴趣区域(图 1(b))。最后，各个感兴趣区域附近不再存在未验证的对抗性样本，表明该区域局部收敛。当各个子区域均收敛后，整个参数空间下的搜索进程终止。

在一组涉及跟驰、换道、超车和避障等行为的综合性场景下，本文使用 EAST 方法对一个决策规划基线算法进行测试。结果表明，本文提出的 EAST 方法的失效检出率是随机采样方法的 4 倍左右，与现有的优化搜索方法相比，效率提高了一倍；同时，EAST 得到的失效危险场景具有较高的覆盖度。特别地，本文提出的 EAST 方法在被测环境中测出了现有方法无法检测的离群样本点，表明 EAST 能够发掘易被忽略的脆弱场景。本文提出的方法在不牺牲检出场景丰富性的前提下，大幅提高了对抗性场景测试的效率。在算力有限的客观约束下，本文提出的方法能够帮助相关设计和研发人员更充分地对自动驾驶系统进行测试，明确其安全边界，继而为算法的迭代升级及自动驾驶系统的安全保障提供支持。



(a) 通过少量精选的样本构建参数空间中对抗性样本和一般样本的大致分布

(b) 密集地验证对抗性样本周边区域，同时探索未知区域的离群样本点

(c) 局部收敛的区域停止搜索，其他区域继续执行

注：红、蓝圆圈分别表示对抗性样本和非对抗性样本

图 1 对抗性场景搜索过程

Fig. 1 Procedure of searching adversarial scenarios

2　高效场景测试框架

EAST 的算法流程如图 2 所示，它分为 3 个模块。先验获取模块通过拉丁超方采样的试验设计方法挑选少量样本进行仿真，构建初始代理模型。该初始代理模型反映了对抗性场景在参数空间中的大致分布，成为后续搜索模块的先验信息。在随后的搜索模块中，使用代理模型代替直接调用仿真测试来计算目标函数。因此，算法可以更加充分地进行探索，不会受到调用仿真带来的运算代价限制。为实现这一目标，本文提出了一种改进的搜索算法，以适应危险场景搜索问题维度高、解集稀疏的特性。最后，本文提出一种“优中选优”的策略，以对代理模型进行更新：在每次迭代中，先由代理模型得到一个粗糙解集，再根据该粗糙解集选择少量样本进行仿真测试，最后以少量精选样本的仿真测试结果更新代理模型。该策略限制了搜索过程中调用仿真的样本范围：只对代理模型认为可能是危险场景的样本调用仿真进程。

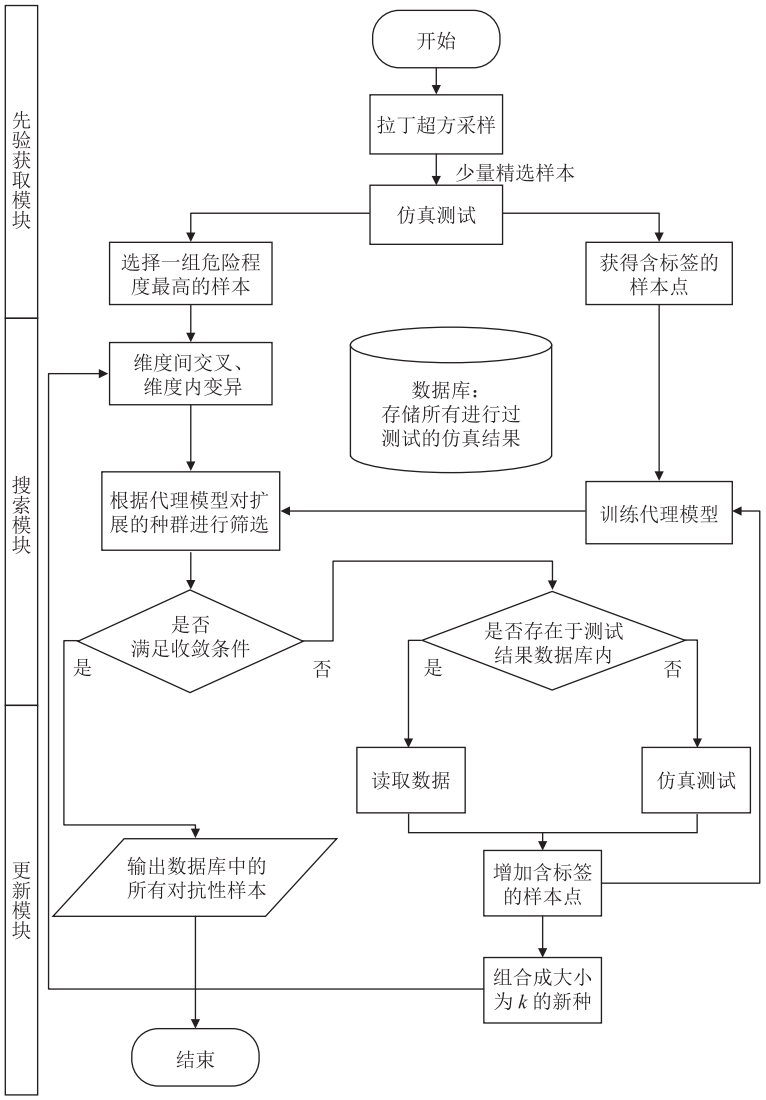


图 2　EAST 算法流程图

Fig. 2　Flowchart of the EAST algorithm

2.1 代理模型

代理模型是本文提出方法的枢纽，在先验获取模块、搜索模块和更新模块均有所运用。在本研究中，代理模型被用来拟合测试空间中对抗性样本的分布情况。在一个高维、庞大的测试空间中，对抗性样本是稀疏分布的。不仅如此，对于每一个样本，均需要通过仿真测试才能判定其是否为对抗性样本。如果直接使用智能优化算法进行搜索，则每一次迭代都需要进行大量的仿真测试，这造成了计算资源的浪费。本文将代理模型机制用于对抗性场景的搜索，使得搜索的早期阶段能够通过少量稀疏样本拟合出测试空间中对抗性样本的分布情况。进一步地，本文提出方法在后续迭代中对代理模型的局部进行更新，使得关键区段的拟合精度不断提高。

具体来说，由精选小样本得到的先验信息通过代理模型传递给后续搜索算法，搜索算法使用代理模型计算目标函数，以进行搜索，而本文提出的“优中选优”策略也借助代理模型实现。因此，设置合适的代理模型对本文提出的 EAST 方法至关重要。

在本文的研究中，代理模型用于拟合特征空间中样本危险程度的分布：

$$\boldsymbol{Y}_{\text{estimate}}=F_{\text{metamodel}}\left(\boldsymbol{X}\right) \tag{1}$$

其中，$\boldsymbol{X}=\left[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\right]$为整个特征空间，空间中任意一个样本 $\boldsymbol{x}_i=\left[d_1,d_2,\cdots,d_t\right]^{\mathrm{T}}$ 为一个能够进行单次测试的具体场景，它具有 t 个维度，每个维度表示一个仿真参数，例如：路径曲率、初始车道、车辆限速、规划步长等；$\boldsymbol{Y}_{\text{estimate}}=\left[y_1,y_2,\cdots,y_n\right]$为特征空间中各个样本的危险程度。

与代理模型整体的精度相比，EAST 更希望代理模型在包含关键危险场景对应样本的子空间中具有较高的精度。定义 $\boldsymbol{X}$ 的子集 $\boldsymbol{X}_{\text{pop}}$ 为搜索算法的种群，其大小为固定值 k，则本文构建代理模型的目标误差 E_{pop} 应表示为：

$$E_{\text{pop}}=\sum_{\boldsymbol{x}_p\in\hat{\boldsymbol{X}}_{\text{pop}}}\frac{\left(F_{\text{metamodel}}\left(\boldsymbol{x}_p\right)-G\left(\boldsymbol{x}_p\right)\right)^2}{k} \tag{2}$$

其中，$G\left(\boldsymbol{x}_p\right)$为样本 $\boldsymbol{x}_p$ 经过仿真得到的危险程度测试值；$\hat{\boldsymbol{X}}_{\text{pop}}$ 为目标种群，即当前代理模型认为危险程度最高的 k 个样本组成的种群，可表示为：

$$\hat{\boldsymbol{X}}_{\text{pop}}=\underset{\boldsymbol{X}_{\text{pop}}}{\operatorname{argmax}}\left(\sum_{\boldsymbol{x}_p\in\boldsymbol{X}_{\text{pop}}}F_{\text{metamodel}}\left(\boldsymbol{x}_p\right)\right) \tag{3}$$

为了使代理模型在目标种群内的误差 E_{pop} 更小，每次迭代都选择当前 $\hat{\boldsymbol{X}}_{\text{pop}}$ 内的样本来拟合代理模型，其更新是局部的。本文定义的误差并不追求全局的精确性，而是期望在目标函数值较高的局部较为精确。因此，本研究应选用适应高维空间且具有较强的局部拟合能力的模型作为代理模型。相关研究表明，Kriging 模型在拟合样本点附近具有极高的精度[21]，这与本文的目标完全契合。而在理论上，神经网络模型能够拟合任何函数，在高维、复杂空间也能表现出极强的拟合能力[22]。在后续的实验中，本文分别将这两种代理模型集成到了本文提出的方法中，并对它们的性能进行了对比。

2.2 先验获取

通过预实验，本文发现，在参数空间中，危险场景的对应样本虽然是稀疏的，但仍然呈现出局部集中的趋势。基于这一现象，本文提出获取先验的策略：先精选少量样本构建代理模型，然后将拟合参数空间中场景危险程度的大致分布作为先验，该先验信息为后续的搜索算法提供引导。

为使初始代理模型能够尽可能全面地表示参数空间中关键危险场景的分布，选取的少量样本必须在各个维度内均表现出空间均布性，并且各维度间的投影不应重合。由于朴素的随机抽样方法在高维参数空间上难以保证每次随机抽样得到的样本都具备空间均布性和投影均布性，因

而难以保证初始代理模型的性能表现。针对这一现象，本文使用拉丁超方设计[20](Latin Hyper Design，LHD)方法来解决这一问题。在 EAST 处理的高维空间对象中，LHD 分为两个步骤：(1)在单个维度内先根据样本数量进行分组，而后在每个组内分别进行抽样，从而保证单个维度内采样的均匀性；(2)将不同维度的抽样序列打乱后再进行组合，并剔除可能重合的组合方式，从而保证样本在不同维度之间的投影均布性。

进一步地，LHD 中存在随机环节，可能导致不同实验间选出的样本均匀性程度不一致。为缓解这一问题，本文定义在 t 维空间中使用 LHD 得到 n 个样本的最小距离 $d_{\min}$ 如下：

$$d_{\min}=\min_{\substack{1\leqslant i,j\leqslant n\\ i\neq j}}\sqrt{\sum_{d=1}^{d=t}\left(\boldsymbol{x}_{id}-\boldsymbol{x}_{jd}\right)^2} \tag{4}$$

本文提出的 EAST 方法多次进行 LHD 操作，选取最小距离 $d_{\min}$ 大于设定阈值的 k 个样本作为精选小样本群体 $\boldsymbol{X}_{\text{selected}}$。随后，对 $\boldsymbol{X}_{\text{selected}}$ 内的所有样本进行仿真测试，得到其对应的真实危险程度度量 $G\left(\boldsymbol{X}_{\text{selected}}\right)$，并将其用于训练初始代理模型 $F^1_{\text{proxymodel}}\left(\boldsymbol{X}_{\text{selected}}\right)$。这一代理模型和精选小样本群体 $\boldsymbol{X}_{\text{selected}}$ 都将被传递给随后的搜索模型，作为先验信息指引算法进行搜索。

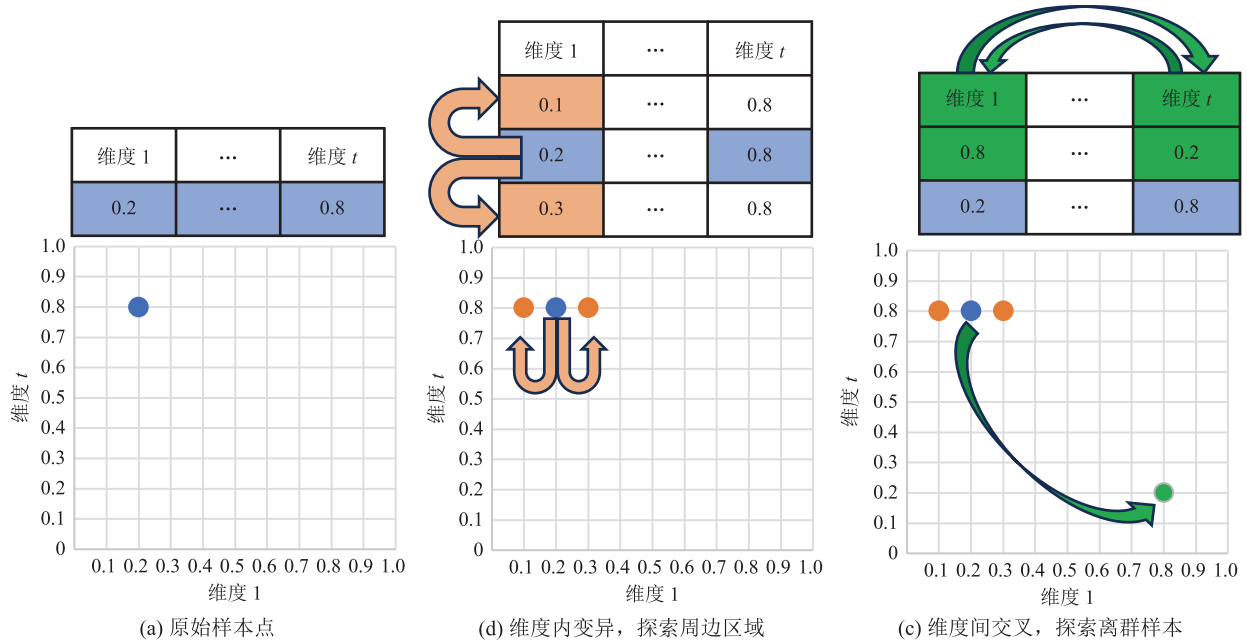
2.3　搜索算法和更新策略

在本文所指的对抗性场景搜索任务中，“探索”指对不确定区域进行测试，在较大范围尺度上找出新的潜在对抗性样本点；而“利用”指对模型认为价值较高的区段进行测试，从而快速得到多个经过验证的真实对抗性样本点。当算法倾向于“探索”时，有利于找出分布在不同子区域的对抗性样本，使得测试的覆盖度更高，但会导致效率低下；反之，当算法倾向于“利用”时，它将在较少的测试次数内找出更多失效样本，使得测试效率更高，但覆盖度会降低。现有方法通过调整超参数来权衡探索和利用的倾向性，但在实际运用中往往难以取舍。本文针对对抗性场景搜索问题的特性，使用新的搜索和更新策略，实现探索和利用的兼顾。

首先，EAST 的探索过程类似于遗传算法中的交叉、变异，但本文对这一过程进行了改进，使其更适用于高维、庞大的目标空间。在算法开始前，参数空间的每个维度都进行了归一化处理。每个样本被严格编码为一个固定长度的向量，向量中的每个元素都是一个介于 0～1 之间的标量值，表示样本的一个维度在值域中取值的相对位置。这保证了不同尺度的维度能够进行交换。在 EAST 中，交叉和变异分别承担着不同的作用。图 3 为本文提出的 EAST 方法通过交叉和变异来探索未知区域的过程。图 3(a)为一个拟进行交叉和变异操作的样本。交叉不改变样本中各个单一维度的取值，而直接将不同样本的部分维度进行交换。这一操作专门用于搜索算法在不同维度间的探索，能够高效地探索高维空间中的未知区域(图 3(c))。变异操作只针对样本向量内的单个元素，从而实现了在样本附近进行探索的目的(图 3(b))。

其次，EAST 使用代理模型在探索过程中计算目标函数的粗糙值，计算成本十分低，因而无须顾虑计算成本的限制。在这个过程中，交叉和变异会在多个维度上进行多次，且交叉、变异前后的种群都会被暂时保留，因而种群规模会扩张 10 倍以上，这保证了探索的充分性。随后，EAST 根据代理模型给出的目标函数值对种群进行筛选，使得种群恢复到预先定义的大小。筛选后的种群是代理模型认为可能会比较危险的种群，如图 2 所示。

最后，EAST 的更新分为两部分：其一，经过代理模型筛选的种群 $\hat{\boldsymbol{X}}_{\text{pop}}$ 会成为下一次更新的初始种群；其二，对 $\hat{\boldsymbol{X}}_{\text{pop}}$ 中的新出现样本进行仿真测试，将其结果加入到代理模型的训练集中，继续训练代理模型。



(a) 原始样本点　(d) 维度内变异，探索周边区域　(c) 维度间交叉，探索离群样本

图 3　搜索算法的探索过程

Fig. 3　Procedure of searching algorithm's exploration

随着每次迭代的进行，代理模型的精度不断提高。当代理模型粗选得到的种群中绝大部分样本都是已经进行过测试的对抗性样本时，算法收敛。

3　实验平台

为了对本文提出的框架进行验证，本节选取一个典型的综合性仿真场景，对 EAST 的性能进行测试。

Frenet 坐标系下的多项式路径规划算法计算效率高，生成的路径平滑度好，是目前有车道线作为参考时的主流局部路径规划算法[23-24]，广泛运用于高级驾驶辅助(ADAS)系统中。本文选取 MATLAB 软件中 Mathworks 公司官方提供的路径规划基线算法作为被测对象，该算法在 Frenet 坐标系下使用 5 次多项式进行路径生成，并综合运动学特性和周边交通环境进行动作决策，实现高速公路环境下的自主行驶。

本文选取的功能场景(参见图 4 中的仿真测试截图)包含多辆环境背景车辆(BVs)，这些背景车辆分别以不同的规则行驶，它们可能匀速行驶、加速超车、换道，甚至意外地停在车道内。由被测算法控制的目标车辆(EV)不仅需要根据车道参考线规划出最优行驶路径，还需要根据周边环境背景车辆的行为和环境特征来进行行为决策。

EAST 与被测算法的接入方式如图 4 所示。首先，指定一个参数空间，EAST 在该参数空间内进行测试。实验过程中，EAST 不断将代理模型认为比较危险的样本传递给仿真测试平台，经过仿真得到样本的危险程度测试值（图 4 橙色部分）。具有仿真测试结果的样本不仅会被存储在数据栈，还会回传给 EAST，用于算法迭代。

根据被测算法和仿真平台环境的特性，本文选取 7 个维度来对 EAST 的性能进行评估，如表 1 所示。

EAST 的最终输出是一组参数空间中的样本点及其危险性度量值，即 $\hat{\boldsymbol{X}}_{\text{pop}}$ 和 $G\left(\hat{\boldsymbol{X}}_{\text{pop}}\right)$。其中，任意一个样本点 $\boldsymbol{x}_i \in \hat{\boldsymbol{X}}_{\text{pop}}$ 都代表在当前测试平台下使得被测算法发生危险的一组参数设定，能够还原出一个对抗性场景。

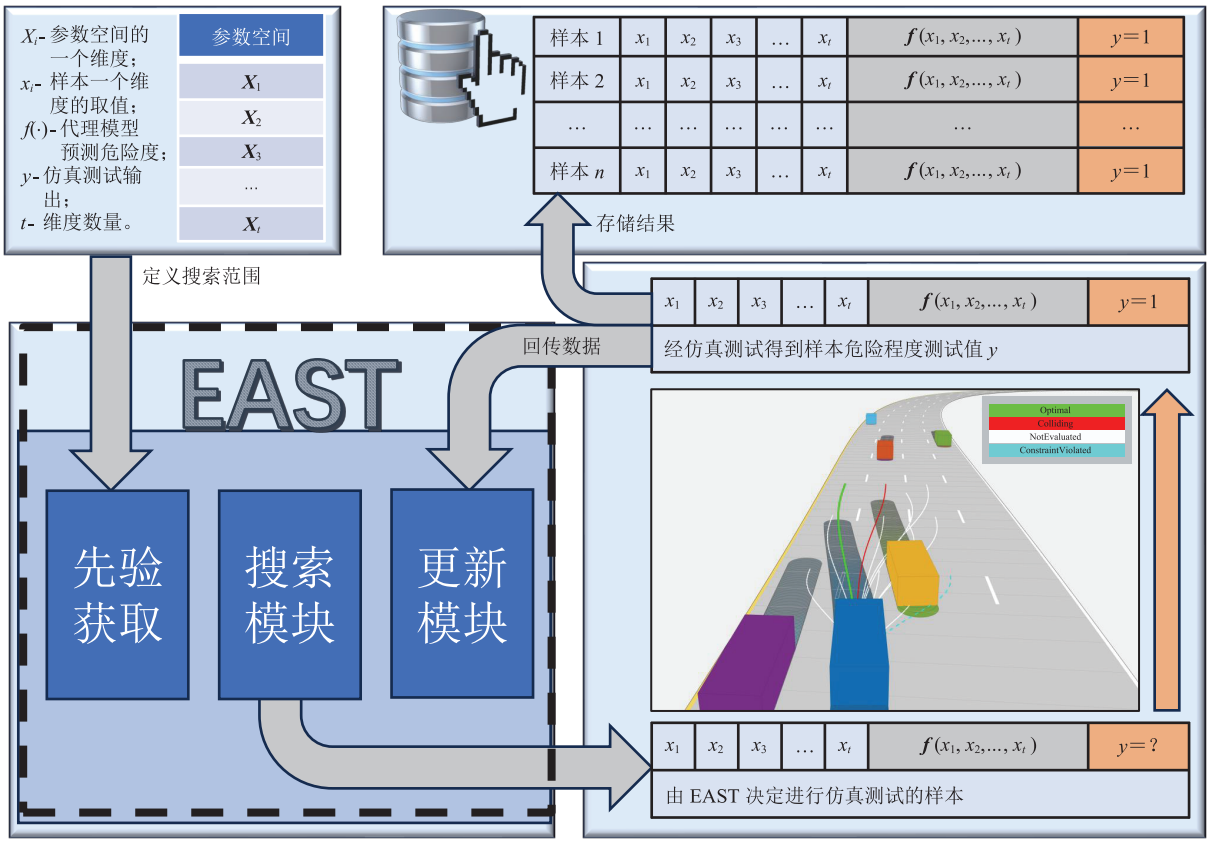


图 4　实验平台

Fig. 4　The experiment platform

表 1　参数空间定义

Table 1　Definition of the parameters space

参数名称	初始车道	最大车速 (m/s)	最大加速度 (m/s^2)	路径最大曲率 (m^{-1})	安全车距 (m)	规划步长	规划次数比率 (%)
取值范围	[1,4]，整数	[8,13.5]	[11,16]	[0.5,1.4]	[4,10]	[1,10]，整数	[10,100]

在评价指标方面，本文定义搜索效率和对抗性场景覆盖度两个定量指标来评价 EAST 的性能表现。在测试过程中，每一次迭代中运算成本最高的步骤是对目标场景进行仿真测试，以计算其危险度量值，这一步骤的运算成本是其他所有步骤的 100 倍以上。因此，本文忽略算法其他环节带来的计算代价，定义搜索效率为算法实测得到的对抗性场景数量与整个算法进程调用仿真次数之间的比值。测试覆盖度被定义为算法收敛后输出的真实对抗性场景数量与整个参数空间中对抗性场景数量的比值。

4　结果与讨论

在上述实验条件下，本文对 EAST 方法进行了验证，在同一实验条件下，还使用随机采样和普通遗传算法[16]进行了对比实验，以说明 EAST 的性能。

与现有方法相比，本文提出的 EAST 方法的测试效率取得了明显提升，在覆盖度方面也有所进步。当使用 Kriging 模型作为代理模型时，EAST 方法与现有方法在搜索效率和搜索覆盖度方面的对比如图 5 所示。当设定为效率优先时，

与普通遗传算法相比，本文提出的 EAST 方法的搜索效率达到 0.402，即平均每进行 10 次仿真测试，有 4 次以上属于对抗性测试；该效率是普通遗传算法的 2.03 倍。

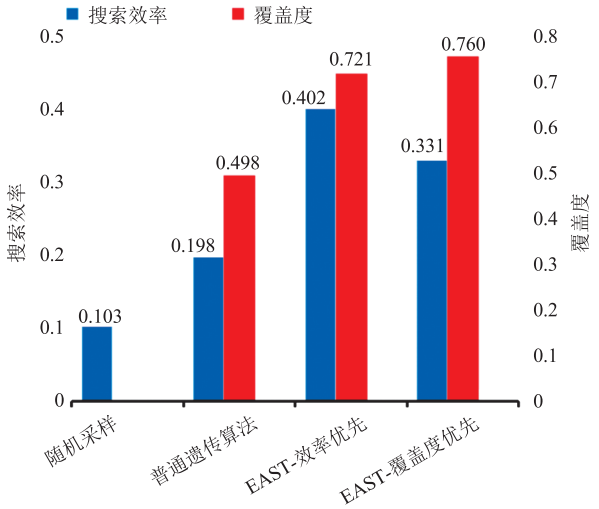


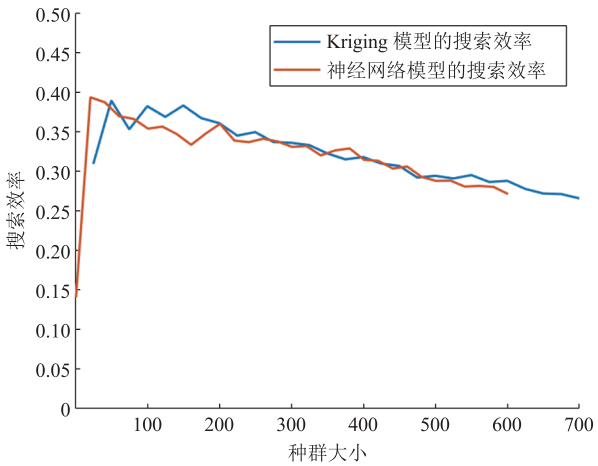
图 5　EAST 和现有方法的对比

Fig. 5　Comparation of EAST and existing methods

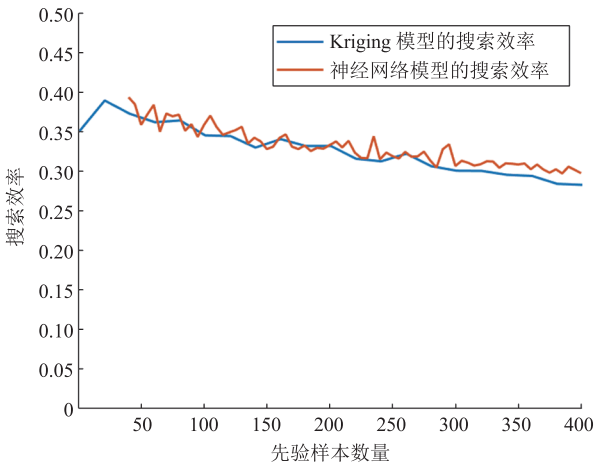
当使用不同代理模型时，本文提出的EAST方法的性能总体变化不大，但存在性能标线的细节差异。图 6(a) 显示了随着种群大小的变化，两种代理模型下危险场景搜索效率的变化趋势，图 6(b) 显示了不同先验样本数量对两种代理模型搜索效率的影响情况。

就趋势而言，与 EAST 的超参数(种群规模和先验样本数量)相比，选择何种代理模型对 EAST 的性能影响不大。这表明，两种模型在被测环境下均具有足够的拟合能力。具体来看，神经网络模型的鲁棒性比 Kriging 模型差，在不同的测试实验中，表现出明显的震荡。此外，当先验样本数量过少(50 以下)时，神经网络模型在初始阶段因训练样本过少而无法输出有效结果。这些现象表明，Kriging 模型更适合作为当前任务下 EAST 使用的代理模型。

图 7 展示了种群大小对 EAST 性能的影响，图 8 展示了先验样本数量对 EAST 性能的影响。其中，曲线的横轴表示当前讨论的超参数，而最大、最小和平均曲线分别表示在横轴取值下，在整个取值范围遍历另一超参数，得到测试指标的最大、最小和平均值。图 7 表明，种群规模对搜索效率和检出失效点数量的影响均十分显著。当种群大小较小时，搜索效率较低，随着种群大小的增加，搜索效率迅速上升；当种群大小超过 200 后，搜索效率开始下降(图 7(a))。随着种群数量上升，检出失效点数量持续上升(图 7(b))。针对这一现象，本文对算法运行过程中种群内样本的变化情况进行研究，结果表明，当种群数量



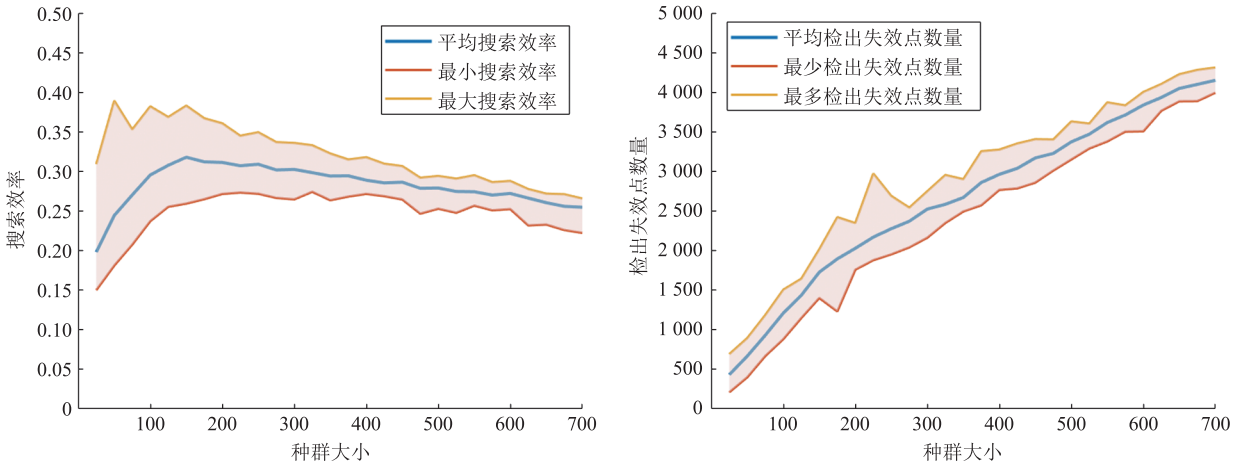
(a) 不同种群大小下两种代理模型的搜索效率



(b) 不同先验数量下两种代理模型的搜索效率

图 6　不同代理模型下 EAST 的性能

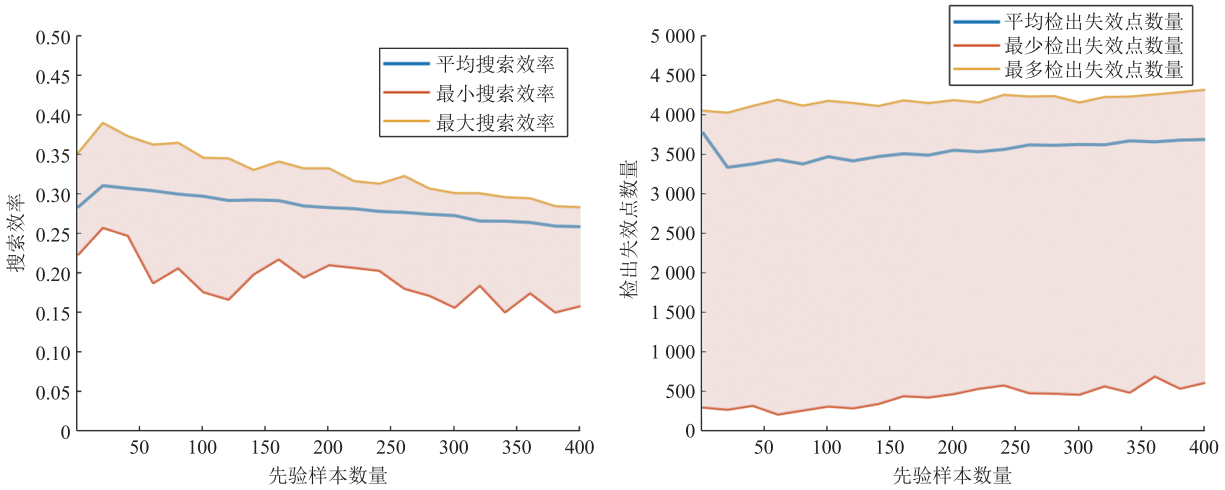
Fig. 6　EAST's performance with different proxy models



(a) 种群大小对搜索效率的影响　　(b) 种群大小对检出样本数量的影响

图 7　种群大小对 EAST 性能的影响

Fig. 7　Effect of EAST's performance on the population size



(a) 先验样本数量对搜索效率的影响　　(b) 先验样本数量对检出样本数量的影响

图 8　先验样本数量对 EAST 性能的影响

Fig. 8　Effect of EAST's performance on the quantity of prior samples

较小时，算法的每一代种群无法容纳足够多的对抗性样本，算法还未搜集到一定的参数空间的信息便已经停止，因而效率较低。当种群数量足够大之后，若继续增加种群规模，则会使种群难以填满，从而降低了 EAST 的测试效率。而种群数量越大，每次迭代中进行探索的样本点数量也越多，因而能够检出更多对抗性样本，即算法趋向于完备性优先而牺牲效率。

图 8 表明，与几乎没有先验信息时相比，选取少量样本作为先验(50 以下)能够提升 EAST 的搜索效率，之后若继续增加先验样本数量，则搜索效率会缓慢降低。这说明一定的先验信息能够启发算法更快地进行搜索，但进行太多先验信息的计算则将有限的仿真测试次数浪费在了非关键样本的测试上，进而导致效率降低。在覆盖度方面，随着先验数量的增加，检出失效样本点数量呈增加趋势，这表明先验信息确实能够启发搜索算法找出更丰富的对抗性场景。

最后，本文抽样对比了 EAST 方法和普通遗传算法检出的对抗性样本的具体分布情况。图 9 左侧部分显示某次实验下，将其余维度固定后，保留两个维度作为变量时，测出的对抗性样本的分布情况。其中，有两个对抗性样本点(＋符号表示)是本文提出的 EAST 方法所检出，而传统遗传算法未能检出。遗传算法能检出的(□符号表示)，EAST 方法都能检出。特别地，图中蓝色阴影标出的样本点不在密集发生失效的对抗性样本点集中区域内，它是一个离群样本，被本文提出的 EAST 方法成功检出。这表明本文提出的方法在效率提升的前提下，并未降低检出样本的丰富性。

5 结 论

本文针对复杂环境下自动驾驶场景测试效率低的问题，提出了一种高效的对抗性场景测试框架。针对信息匮乏、事件稀疏，以及探索与利用难以平衡的挑战，本文提出的方法通过构建代理模型、获取先验信息和改进搜索策略 3 种手段，实现了一种高效进行对抗性场景测试的方法。结果表明，与现有方法相比，本文提出的自动驾驶高效对抗性场景测试方法在搜索效率和测试覆盖度方面均显著提升。特别地，本文提出方法能够发掘现有方法缺失的离群对抗性样本，直接揭示自动驾驶系统的潜在安全隐患。

本文提出方法只适用于固定维度、固定大小的结构化参数空间。在非结构化的，甚至语义的场景定义方式下，如何高效进行对抗性场景测试将是下一步研究的重点。此外，基于本文提出方法揭示出的复杂场景下的新失效模式，如何改进相应的自动驾驶算法，提升其本身的安全性能也是具有重要价值的研究。

参 考 文 献

[1] 蒋拯民, 党少博, 李慧云, 等. 自动驾驶汽车场景测试研究进展综述 [J]. 汽车技术, 2022, 563(8): 10-22.
Jiang ZM, Dang SB, Li HY, et al. A survey on the research progress of scenario-based testing for autonomous vehicles [J]. Automobile Technology, 2022, 563(8): 10-22.

[2] Zhu B, Zhang PX, Zhao J, et al. Hazardous scenario enhanced generation for automated vehicle testing based on optimization searching method [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 7321-7331.

[3] 朱冰, 张培兴, 刘斌, 等. 基于自然驾驶数据的自

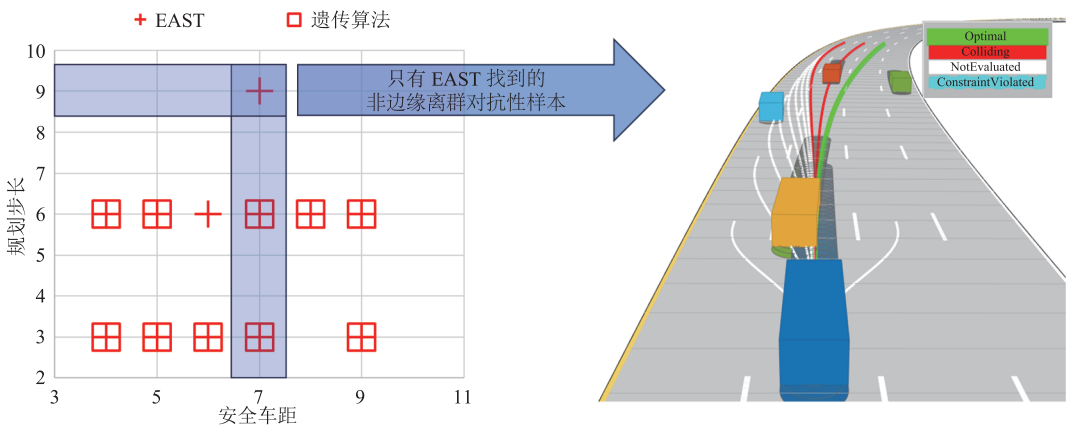


图 9 测试框架检出的对抗性样本

Fig. 9 The adversarial samples detected by the test framework

动驾驶汽车安全性评价方法 [J]. 中国公路学报, 2022, 35(7): 283-291.
Zhu B, Zhang PX, Liu B, et al. Safety evaluation method of automated vehicle based on naturalistic driving data [J]. China Journal of Highway and Transport, 2022, 35(7): 283-291.

[4] Li L, Wang X, Wang KF, et al. Parallel testing of vehicle intelligence via virtual-real interaction [J]. Science Robotics, 2019, 4(28): eaaw4106.

[5] 朱冰, 张培兴, 赵健, 等. 基于场景的自动驾驶汽车虚拟测试研究进展 [J]. 中国公路学报, 2019, 32(6): 1-19.
Zhu B, Zhang PX, Zhao J, et al. Review of scenario-based virtual validation methods for automated vehicles [J]. China Journal of Highway and Transport, 2019, 32(6): 1-19.

[6] Gelder ED, Elrofai H, Saberi AK, et al. Risk quantification for automated driving systems in real-world driving scenarios [J]. IEEE Access, 2021, 9: 168953-168970.

[7] 朱向雷, 杜志斌. 自动驾驶场景仿真与 ASAM OpenX 标准应用 [M]. 北京: 机械工业出版社, 2023.
Zhu XL, Du ZB. Scenario-based autonomous driving simulation tenology and application of ASAM OpenX standars [M]. Beijing: China Machine Press, 2023.

[8] Jiang ZM, Pan WB, Liu J, et al. Efficient and unbiased safety test for autonomous driving systems [J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(5): 3336-3348.

[9] Ahlstrom C, Victor T, Wege C, et al. Processing of eye/head-tracking data in large-scale naturalistic driving data sets [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(2): 553-564.

[10] Lee J, Shiotsuka D, Nishimori T, et al. GAN-based LiDAR translation between sunny and adverse weather for autonomous driving and driving simulation [J]. Sensors, 2022, 22(14): 5287.

[11] Feng G, Han ZD, Zhou JW, et al. Performance limit evaluation by evolution test with application to automatic parking system [J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(4): 3096-3105.

[12] Feng S, Sun HW, Yan XT, et al. Dense reinforcement learning for safety validation of autonomous vehicles [J]. Nature, 2023, 615(7953): 620-627.

[13] Feng TY, Liu LH, Xing XY, et al. Multimodal critical-scenarios search method for test of autonomous vehicles [J]. Journal of Intelligent and Connected Vehicles, 2022, 5(3): 167-176.

[14] Chen BM, Chen X, Wu Q, et al. Adversarial evaluation of autonomous vehicles in lane-change scenarios [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 10333-10342.

[15] 朱冰, 张培兴, 赵健. 面向多维度逻辑场景的自动驾驶安全性聚类评价方法 [J]. 汽车工程, 2020, 42(11): 1458-1463.
Zhu B, Zhang PX, Zhao J. Clustering evaluation method of autonomous driving safety for multi-dimensional logical scenario [J]. Automotive Engineering, 2020, 42(11): 1458-1463.

[16] 邢星宇, 吴旭阳, 刘力豪, 等. 基于目标优化的自动驾驶决策规划系统自动化测试方法 [J]. 同济大学学报(自然科学版), 2021, 49(8): 1162-1169.
Xing XY, Wu XY, Liu LH, et al. Automatic testing method based on optimization algorithms for the decision and planning system of autonomous vehicles [J]. Journal of Tongji University (Natural Science), 2021, 49(8): 1162-1169.

[17] 马依宁, 姜为, 吴靖宇, 等. 基于不同风格行驶模型的自动驾驶仿真测试自演绎场景研究 [J]. 中国公路学报, 2023, 36(2): 216-228.
Ma YN, Jiang W, Wu JY, et al. Self-evolution scenarios for simulation tests of autonomous vehicles based on different models of driving styles

[J]. China Journal of Highway and Transport, 2023, 36(2): 216-228.

[18] Yasasa A, Florian S, Gregory D. Generating adversarial driving scenarios in high-fidelity simulators [C] // Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), 2019: 8271-8277.

[19] Feng S, Yan XT, Sun HW, et al. Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment [J]. Nature Communications, 2021, 12(1): 748.

[20] Mckay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code [J]. Technometrics, 2000, 42(1): 55-61.

[21] 王家恒. 基于 Kriging 模型的多目标优化方法研究 [D]. 大连: 大连理工大学, 2019.
Wang JH. Research on multi-objective optimization method based on Kriging surrogate model [D]. Dalian: Dalian University of Technology, 2019.

[22] Abiodun OI, Jantan A, Omolara AE, et al. State-of-the-art in artificial neural network applications: a survey [J]. Heliyon, 2018, 4(11): e00938.

[23] Schwarting W, Alonso-Mora J, Rus D. Planning and decision-making for autonomous vehicles [J]. Annual Review of Control, Robotics, and Autonomous Systems, 2018, 1: 187-210.

[24] Gonzalez D, Perez J, Milanes V, et al. A review of motion planning techniques for automated vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1135-1145.