ARTIFICIAL INTELLIGENCE

Parallel testing of vehicle intelligence via virtual-real interaction

Li Li[1]*, Xiao Wang[2,3]*, Kunfeng Wang[2,3]*, Yilun Lin[2,3,4]*, Jingmin Xin[5]*, Long Chen[6,7], Linhai Xu[5], Bin Tian[2,7], Yunfeng Ai[4,7], Jian Wang[7,8], Dongpu Cao[7,8,9], Yuehu Liu[5], Chenghong Wang[10,11], Nanning Zheng[5†], Fei-Yue Wang[2†]

A self-driven closed-loop parallel testing system implements more challenging tests to accelerate evaluation and development of autonomous vehicles.

Although researchers and automobile manufacturers have built several proving grounds (*1*) and testing datasets (*2*) dedicated to autonomous driving, tests for intelligent vehicles remain time-consuming, inefficient, and sometimes dangerous for people who use the same roads.

According to Turing (*3*), a system could be said to be intelligent enough for special kind of tasks if and only if it could finish all the possible tasks of its kind. Therefore, we can begin to achieve safe and reliable artificial intelligence (AI) systems if and only if the tests have clear definitions of tasks and efficient methods to generate abundant data for tests. As a result, appropriate AI testing methods should be task-driven and data-centric.

Many existing test systems for autonomous vehicles do not provide a systematic, standard, and practical way to describe the driving tasks so that we can equivalently translate, modify, and reuse the tasks of both field and simulation tests (*4*–*6*). We can neither sufficiently sample the driving scenarios that we may encounter in practice nor learn to generate the challenging testing tasks to promote the capability of autonomous vehicles.

Many autonomous vehicle companies resort to simulation-based tests to save time and money. The simulation-based testing system can handle thousands of quantitative judgments in a short time and be more objective than a human expert. However, such a system relies heavily on human knowledge to properly design the scenarios. Some scenarios tested by simulations should also be re-evaluated and verified in field tests to validate the effectiveness of the simulation systems and the reliability of the hardware of autonomous vehicles.

Therefore, a human-in-loop simulation system is useful to evaluate the performance of vehicles (*5*–*7*) efficiently. A human expert can first vaguely define the tasks and perform qualitative judgments, and then the simulation-based system can make more precise task definitions, generating more tests, and receive feedback from humans to validate the test results.

We built a closed-loop testing system that focuses on implementing more challenging tests to accelerate the building and testing of autonomous vehicles. As shown in the system overview flowchart of Fig. 1, there are three parts in this system.

The first part establishes a set of semantic definitions to characterize the tasks that should be finished by autonomous vehicles (*5*, *6*). Each semantic entity of the tested driving scenario will be retrieved and reproduced in the semantic task space. The semantic task atoms for each entity will be labeled, with a special focus on the spatiotemporal range of each task atom. The merit of semantic analysis lies in its ability to capture the abstract attributes of a special task and discard the unnecessary details. The complexity of the abstracted semantic tasks also provides harmonized classification levels that describe the capabilities of autonomous vehicles.

We can rearrange the spatiotemporal ranges of task rectangles to sample different driving scenarios that belong to the same category so as to ensure that autonomous vehicles could work for these driving scenarios. Adding more semantic task atoms over time permits us to increase the complexity of tests and expand the range of scenarios in which autonomous vehicles can be tested.

The second part implements the tests for the specified task instances. The field test and simulation test are tightly integrated to ensure test safety and accelerate testing speed. Given the tight integration, we call this a "parallel testing system." Unlike many systems in which the behaviors of all the agents are manually coded and rigid, the parallel testing system keeps collecting new field data to update the simulation system. Moreover, we can carry out field tests that exactly correspond to the simulation tests held in the simulation system and compare their outputs to update the simulation system from time to time. To this end, we built an integration system to automatically and accurately collect various measurements of vehicles in a real-time manner.

We had developed various methods to mix both real scenario data and virtual scenario data generated by simulation engines to provide diverse scenarios for testing autonomous vehicles. Especially, we used the parallel vision techniques to transfer the real-world sensing data collected in the normal daytime to virtual-world sensing data in less frequently encountered situations (e.g., adverse weather and emergency events) (*8*).

The third part evaluates both vehicle performances and task difficulties to seek the most challenging new tasks. We set up several quantitative performance indices associated with each particular semantic task to fairly and quickly evaluate the performance of an autonomous vehicle (*9*, *10*). We designed a statistical learning model to simultaneously

[1]Department of Automation, BNRist, Tsinghua University, Beijing 100084, China. [2]The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. [3]Qingdao Academy of Intelligent Industries, Qingdao 266000, China. [4]University of Chinese Academy of Sciences, Beijing 100049, China. [5]Institute of Artificial Intelligence and Robotics (IAIR), Xi'an Jiaotong University, Xi'an 710049, China. [6]Sun Yat-sen University, Guangzhou 510275, China. [7]VIPioneers (HuiTuo) Inc., Qingdao 266109, Shandong, China. [8]Jilin University, Changchun 130012, China. [9]Waterloo Cognitive Autonomous Driving (CogDrive) Lab, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [10]Chinese Association of Automation, Beijing 100190, China. [11]National Natural Science Foundation of China, Beijing 100085, China.
*These authors contributed equally to this work.
†Corresponding author. Email: nnzheng@mail.xjtu.edu.cn (N.Z.); feiyue.wang@ia.ac.cn (F.-Y.W.)

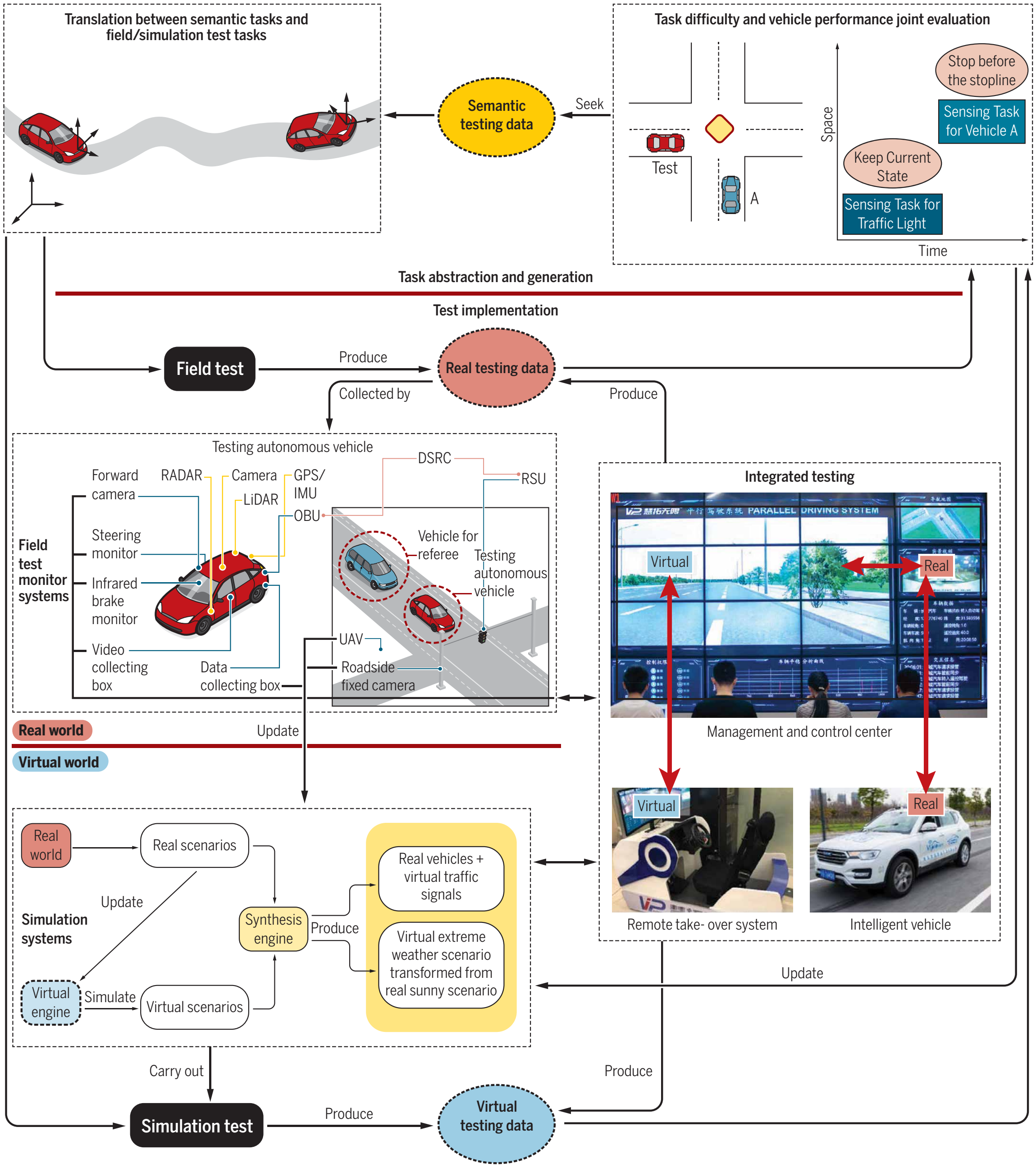

Fig. 1. System overview flowchart.

CREDIT: A. KITTERMAN/SCIENCE ROBOTICS

determine the relative difficulties of different tasks and the relative capabilities (ranks) of different autonomous driving systems under tests. Such ranking models help us not only find the challenging tasks, but also understand the real capability levels of the tested autonomous vehicles (*10*).

We also designed an adversarial learning model to automatically generate new task instances that may be harder than existing task instances based on the past testing results, aiming to push the autonomous vehicles to improve its capability. From this viewpoint, the test in our systems is a self-upgrading process occasionally guided by human experts. Such designs make the test of vehicle intelligence more quantifiable and automatic. We believe that this design philosophy may also be useful for building and testing other intelligent systems.

Our integrated parallel testing system successfully supported the Intelligent Vehicle Future Challenge of China (IVFC), which is the longest-lasting autonomous driving competition (*6*). Along with IVFC held from 2009 to 2018, our testing system was upgraded to implement systematic, quantitative, automatic, and safe tests for industrial vehicles. Results show that our testing system significantly reduced the burden of competition organizers and test engineers. For more details of the applications, see the Supplementary Materials.

SUPPLEMENTARY MATERIALS

robotics.sciencemag.org/cgi/content/full/4/28/eaaw4106/DC1
Text
Fig. S1. The locations and periods of the past 10 years IVFCs.
Fig. S2. Host of IVFC: The Intelligent Vehicles Proving Center.
Fig. S3. The demonstration of application of parallel testing system in the 10th Intelligent Vehicle Future Challenge (IVFC 2018).
Fig. S4. The integration of field and simulation tests adopts a better opportunity to take over.

REFERENCES AND NOTES

1. Mcity; https://mcity.umich.edu/ [accessed 15 January 2019].
2. A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.* **32**, 1231–1237 (2013).
3. A. M. Turing, Computing machinery and intelligence. *Mind* **59**, 433–460 (1950).
4. A. Broggi, M. Buzzoni, S. Debattisti, P. Grisleri, M. C. Laghi, P. Medici, P. Versari, Extensive tests of autonomous driving technologies. *IEEE Trans. Intell. Transport. Syst.* **14**, 1403–1415 (2013).
5. L. Li, W.-L. Huang, Y. Liu, N.-N. Zheng, F.-Y. Wang, Intelligence testing for autonomous vehicles: A new approach. *IEEE Trans. Intell. Veh.* **1**, 158–166 (2016).
6. L. Li, Y.-L. Lin, N.-N. Zheng, F.-Y. Wang, Y. Liu, D. Cao, K. Wang, W.-L. Huang, Artificial intelligence test: A case study of intelligent vehicles. *Artif. Intell. Rev.* **50**, 441–465 (2018).
7. F.-Y. Wang, X. Wang, L. Li, P. Mirchandani, Creating a digital vehicle proving ground. *IEEE Intell. Syst.* **18**, 12–15 (2003).
8. X. Li, K. Wang, Y. Tian, L. Yan, F. Deng, F.-Y. Wang, The ParallelEye dataset: A large collection of virtual images for traffic vision research. *IEEE Trans. Intell. Trans. Syst.* **99**, 1–13 (2018).
9. F.-Y. Wang, X. Wang, L. Li, L. Li, Steps toward Parallel Intelligence. *IEEE/CAA J. Autom. Sinica* **3**, 345–348 (2016).
10. C. Zhang, Y. Liu, L. Li, N.-N. Zheng, F.-Y. Wang, Joint difficulty estimation and testees ranking for intelligent evaluation. *IEEE Trans. Comput. Social Syst.* **6**, 222–227 (2019).

Funding: The work was supported, in part, by National Natural Science Foundation of China (61533019, 91720000, 91520301, 61702519, and 61773414), Beijing Municipal Science and Technology Commission (Z181100008918007), Intel Collaborative Research Institute for Intelligent and Automated Connected Vehicles ("ICRI-IACV"). **Author contributions:** F.-Y.W. and L.L. conceived the parallel testing framework; N.Z. and F.-Y.W. conceived and led the research program and IVFC competition over the last decade (each equally contributing 10% of the whole research; N.Z. was the Director of the Steering Group of Experts of KP-CCVAI; F.-Y.W. was the Chief Judge of IVFC competition and Director of IVPC). L.L., X.W., K.W., Y.L., and J.X. jointly designed and developed the parallel testing and evaluation system for intelligent vehicles and conducted the main research and analyses (each equally contributing 10% of the whole research work). L.X. and Y.L. assisted in organizing the IVFC competition and participated in parallel simulation and testing (each equally contributing 4% of the whole research work). L.C., B.T., Y.A., J.W., and D.C. developed the parallel driving system and participated in parallel testing and construction of IVPC (each equally contributing 4% of the whole research work). C.W. participated in organizing the IVFC competition and served as the Director of the Arbitration Committee of IVFC competition (contributing 2% of the whole research work). L.L., X.W., K.W., Y.L., and F.-Y.W. prepared the manuscript, and all authors provided feedback during the manuscript revisions and results discussions.

10.1126/scirobotics.aaw4106

Citation: L. Li, X. Wang, K. Wang, Y. Lin, J. Xin, L. Chen, L. Xu, B. Tian, Y. Ai, J. Wang, D. Cao, Y. Liu, C. Wang, N. Zheng, F.-Y. Wang, Parallel testing of vehicle intelligence via virtual-real interaction. *Sci. Robot.* **4**, eaaw4106 (2019).

Science Robotics

Parallel testing of vehicle intelligence via virtual-real interaction

Li Li, Xiao Wang, Kunfeng Wang, Yilun Lin, Jingmin Xin, Long Chen, Linhai Xu, Bin Tian, Yunfeng Ai, Jian Wang, Dongpu Cao, Yuehu Liu, Chenghong Wang, Nanning Zheng and Fei-Yue Wang

Sci. Robotics **4**, eaaw4106.
DOI: 10.1126/scirobotics.aaw4106