2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

A Graded Offline Evaluation Framework for Intelligent Vehicle's Cognitive Ability*

Chi Zhang¶, Yuehu Liu†, Qilin Zhang‡, Le Wang†

***Abstract*— Cognitive ability evaluation in intelligent vehicles is conventionally evaluated by classical autonomous driving dataset, which lacks comprehensive annotations of driving difficulty. Realistically, different driving conditions require vast different level of cognitive ability, e.g., driving in highly congested traffic is much more challenging than driving on limited access highway; driving in a blizzard/hurricane requires much more robust environmental cognition abilities than driving under ordinary conditions. Different datasets contain different proportions of various driving conditions, rendering intelligent vehicle evaluation susceptible to dataset variations. To overcome such limitations, we propose to first benchmark the driving difficulty with the proposed "Cascaded Tanks Model" and obtain a fine-grained per-segment difficulty rating based on our proposed *Semantic Descriptor*. With the proposed Graded Offline Evaluation (GOE) framework, it is demonstrated that offline validation of the cognitive abilities in Intelligent Vehicles (IV) is more consistent regardless of dataset choice.**

I. INTRODUCTION

Cognitive abilities are the fundamentals of autonomous driving systems. Given some multi-source heterogeneous sensor data, intelligent vehicles are expected to answer the questions such as "*What are the current traffic conditions like? Where are other vehicles and pedestrians?*" The need of testing and validating the cognitive module (e.g., environmental recognition algorithms of the advanced driving assistant systems (ADAS) and the autonomous driving systems) has never been so urgent in the new era of wide applications of artificial intelligence in IV. To minimize potential safety risk, it is essential to assess how well a cognitive module works during its development procedure and before it is deployed on commercial systems. To achieve this goal, offline evaluation techniques and schemes [16] are widely utilized in the life cycle of such developments of cognitive modules.

For a typical task in an offline test, there are primarily two design problems. The first one is how to prepare the test data [10], [24]; and the other one is how to evaluate the intelligent vehicle's performance given the outputs from such vehicle. From 2012, various benchmarks [4], [14] on cognitive algorithms from different datasets [6], [7], [11] have been proposed. They provide numerous heterogeneous multi-modal time-synchronized sensory data acquired from the real traffic scenarios with video cameras, laser scanners (especially LiDAR for remote sensing applications [1], [17], [18]) and GPS localization systems. Besides, they provide baseline benchmarks on cognitive tasks, such as 2D/3D object detection, object tracking, lane detection.

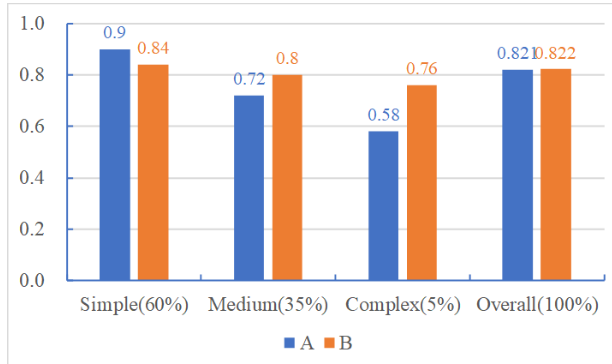

Fig. 1. F1 scores of two intelligent vehicle contest participators A (shown in blue) and B (shown in orange), in the proposed Graded offline Evaluation (GOE) framework, with individual scores for roadway segments with three scenario complexities. 60%, 35% and 5% of the roadway segments are rated as simple, medium, and complex, respectively.

However, existing benchmark dataset often ignores the correlation between the cognitive performance and the intrinsic complexity of the driving scenarios in the sensory data. For example, driving in highly congested traffic is much more challenging than driving on limited access highway; driving in a blizzard/hurricane requires much more robust autonomous driving system than driving under ordinary conditions. The sensory data included in different datasets contains different proportions of various driving conditions, but in their evaluation methods, autonomous driving systems are evaluated merely by a collection of overall quantitative indexes such as Precision, Recall and F1-Measure without proper indication of the intrinsic difficulty.

For instance, given two intelligent vehicles (denoted as participators A and B in Fig. 1 and their overall F1 scores (0.821 and 0.822, respectively) in the same cognitive functionality offline test, it is not immediately obvious which participator is comparably better. However, if a graded test results as well as the distribution of scenario complexity are obtained (both the blue and orange histograms in Fig. 1), it can be found out that the participator A achieves better performance on sensory data from simple scenarios while the participator B performs better in the more complex scenarios.

In this paper, we propose an efficient Graded Offline Evaluation (GOE) framework (Fig. 2) to assess the cognitive

*This work is supported by National Natural Science Foundation of China (Grant No: 91520301).

¶Chi Zhang is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China. Chi Zhang is the corresponding author. colorzc@stu.xjtu.edu.cn

†Yuehu Liu and Le Wang are with the Faculty of the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China.

‡Qilin Zhang is with HERE Technologies, Chicago, Illinois, 60606, USA.

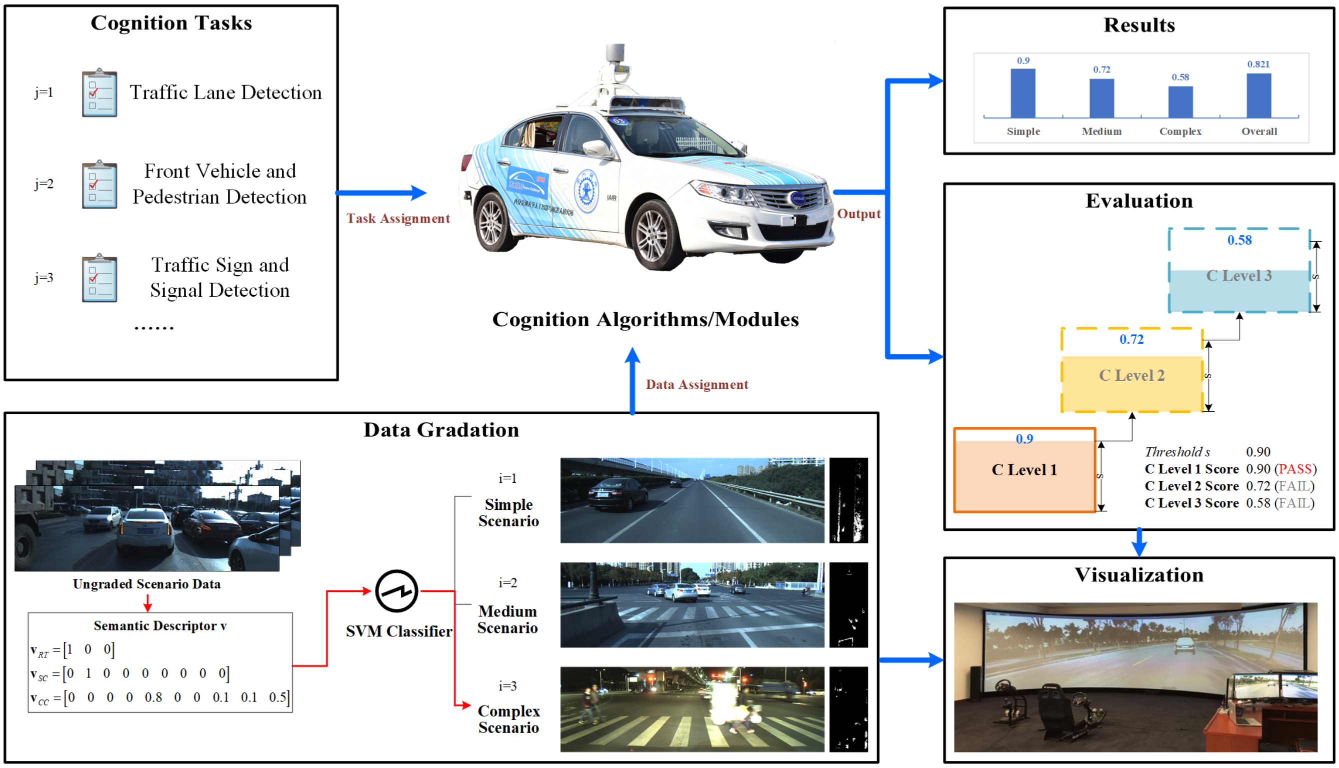

Fig. 2. The proposed Graded Offline Evaluation (GOE) framework for IV Cognition. As demonstrated on the bottom left, sensory data are quantified by the proposed *Semantic Descriptor* and classified into 3 levels of scenario complexity. With the graded and annotated offline testing data and the predefined cognition tasks (shown on the top left), the cognition abilities are evaluated with the *Cascaded Tanks Model* (shown on the right), with Level 1, 2 and 3 tests containing data rated as simple, medium and complex, respectively. The obtained results can be analyzed statistically or visualized for easier human interpretation.

abilities of IV, by exploiting the dependence of cognitive ability level on fine-grained per-segment scenario complexity. Firstly, sensory data for offline test are described quantitatively by the proposed *Semantic Descriptor*. Secondly, with the pre-defined three levels (simple, medium and complex) of scenario complexity, the previously annotated offline testing data are graded into three classes by a Support-Vector Machine (SVM) [3] multi-class classifier. To evaluate the graded test results efficiently, we incorporate the *Cascaded Tanks Model* to assess different IV with a list of cognition tasks (e.g., traffic lane detection, front vehicle and pedestrian detection and traffic signal detection). In this *Cascaded Tanks Model*, roadway segments with scenario complexity levels simple, medium and complex are included in the Level 1, Level 2 and Level 3, respectively. A detailed case study is conducted on two sample autonomous driving platforms to validate the efficacy of the proposed GOE assessment framework. The major contributions of this paper are as follows.

1) A quantitative *Semantic Descriptor* is proposed to summarize realistic driving difficulties.
2) An SVM-based data gradation system is proposed to automatically predict per-segment roadway scenario complexity ratings.
3) Based on scenario complexity ratings, the *Cascaded Tanks Model* is incorporated in the proposed GOE framework, offering efficient and interpretable assessments for IV cognition.

II. Proposed GOE Framework for IV Cognition

Multiple testing frameworks [22], [23] have been proposed in recent years for the development and study on autonomous driving algorithms, functionalities and systems. Although the cognition ability have been recognized as an essential module in advanced driving assistant systems (ADAS) and self-driving cars [9], the existing test and evaluation frameworks are essentially identical to general purpose testing system for generic algorithms, which are susceptible to dataset variations.

As illustrated in Fig. 2, the GOE framework is proposed in this paper to offer efficient, consistent and easily interpretable evaluation of IV cognition. This GOE framework consists of the multi-sensor data gradation with the *semantic descriptor*, the cognition task list, the *Cascaded Tanks Model* based offline evaluation and results analysis and visualization.

A. Semantic Descriptor

Classical autonomous driving datasets, like KITTI [6], [7], RobotCar [11] categorize implicitly semantic contents of scenarios with coarse labels, i.e., labels are assigned to entire set of scenario data without precise locations. For instance, KITTI divides all scenarios into five scene types, “Road”, “City”, “Residential”, “Campus” and “Person”. Such annotations have overlaps and are ambiguous when only a small portion of images in a given set contain the scenes. On the contrary, annotations in the RobotCar dataset consists of an incomplete set of weather conditions and road users, such as "pedestrian", "bicycle" and "vehicle traffic", "light rain", "heavy rain", "direct sunlight", etc. Such annotations could not cover the full set of all possible weather scenarios and road users, thus it is not directly applicable for the gradation of scenario data.

In this paper, we propose the *Semantic Descriptor* $\mathbf{v}$ to encode the scenario data with perceptive data annotations from three different perspectives: road type, scenario content and challenging conditions.

1) **Road Type** (RT). Road types reflect the fundamental layout of a traffic scenario. The characteristics and quantities of challenging conditions also vary according to specific road types, too. For instance, road intersections and pedestrians are more prevalent in the "Urban" category; while in the "Highway" class, vehicles might need to drive across an overpass or pass toll booths. We use an n-dimension binary-valued vector $\mathbf{v}_{RT}$ with one-hot encoding to describe the road type information per roadway segment. The n road types are defined to cover all types present in a dataset, such as "Urban", "Highway", "Country". Unlike previous road type encodings, the proposed method assigns different $\mathbf{v}_{RT}$ values per roadway segments, where each segment contain only one category of Road Type and Scenario Content to eliminate ambiguity. Specifically, $\mathbf{v}_{RT}$ will have a single 1 bit and all the others are 0 bits, due to the one-hot encoding.
2) **Scenario Content** (SC). Scenario content reflects the semantic content of the corresponding roadway segment. We use an m-dimension binary-valued vector $\mathbf{v}_{sc}$ with one-hot encoding to describe such scenarios. The m scenario types cover nearly all the possible scenario contents encountered by an IV, including "Normal Driving", "Intersections", "Elevated Road", "Tollbooth", "Tunnel", "Roundabout", "Slope", "Bridge", "Railway" and etc. Due to the fine-grained per-segment labeling, different scenario content labels are mutually exclusive, i.e., $\mathbf{v}_{sc}$ will have a single 1 bit and all the others are 0 bits.
3) **Challenging Conditions** (CC). Challenging conditions $\mathbf{v}_{cc}$ are defined to reflect the per-segment driving difficulty due to a predefined o types of challenges. $\mathbf{v}_{cc}$ is an o-dimensional floating-point vector with values between 0 and 1, where 0 represents no challenge at all and 1 represents the most challenging conditions possible. The o types of CC are generalized from almost all possible challenges an IV might encounter, including "Curve", "Overtaking", "Pedestrians", "Road Construction", "Heavy Traffic", "Fog & Haze", "Night", "Marked Road", "Fuzzy Markers" and "Special Illumination".

After obtaining the individual components, the *Semantic Descriptor* $\mathbf{v}$ is obtained by concatenating them, $\mathbf{v} = \left[\mathbf{v}_{RT}^T, \mathbf{v}_{sc}^T, \mathbf{v}_{cc}^T\right]^T$. The obtained $\mathbf{v}$ serves as the basis for the subsequent data gradation without relying on other computationally expensive computer vision-based recognition algorithms (e.g., [8], [12], [13], [15]).

B. Data Gradation

Data gradation is a critical component for the proposed GOE framework. To thoroughly test IV cognition under different difficulties, it requires massive amount of annotated scenario data. Therefore, manual labeling of all scenario data in the format of the proposed $\mathbf{v}$ is prohibitively time-consuming. Instead, a data-driven gradation method is proposed to classify the scenario data automatically. First, the gradation is recast into a supervised multi-class classification problem, which can be solved via a combination of binary SVM classifiers [3]. In the proposed GOE framework, we utilize the Error-Correcting Output Codes [5] to code this combination and decode the predicted output for classification. The coding strategy is the one-versus-all [2], where each class is discriminated against the rest of classes. Given a set of quantitative descriptions $\{\mathbf{v}\}$ of segmented scenario data and the predefined 3 levels of scenario complexity $C_p \in \{1, 2, 3\}$, the prediction is carried out by

$$C_p(\mathbf{v}) = \arg \max_{i=1,2,3} f_i(\mathbf{v}), \tag{1}$$

where f_i is the output score of the ith-vs-rest SVM classifier.

The training procedure of such 3-class classifier is divided into two steps. Firstly, we manually annotate a small amount (approximately 1000) of scenario data segments as Simple (1), Medium (2) and Complex (3), according to their RT, SC and the quantity/difficulty of CC. Afterwards, the relationship between Semantic Descriptors $\mathbf{v}$ and manually labeled complexity C_p are learned by the classifier. The average training time is 1.18 second per 100 samples on an Intel Core i5 7200U laptop, and the accuracy of the predicted complexity on the training set is 93.23%.

C. Cascaded Tanks Model

According to empirical experiences, higher scenario complexity often leads to lower cognition performance. Based on this observation, we proposed the "Cascaded Tanks" evaluation model as illustrated in Fig. 3, with stepwise increasingly challenging IV cognition test. In this Cascaded Tanks model, IV cognition are evaluated with scenario data rated as simple, medium and complex at Level 1, 2 and 3, respectively. At each level, overall quantitative indexes such as precision, recall and F1 scores are calculated based on the average of per-frame performance. At the ith level for the jth cognition task ($j = 1, \cdots, N$),

$$\text{Precision}_i(j) = \frac{TP_i(j)}{TP_i(j) + FP_i(j)}, \quad i = 1, 2, 3. \tag{2}$$

$$\text{Recall}_i(j) = \frac{TP_i(j)}{TP_i(j) + FN_i(j)}, \quad i = 1, 2, 3. \tag{3}$$

$$F_{1,i}(j) = \frac{2\text{Precision}_i(j) \times \text{Recall}_i(j)}{\text{Precision}_i(j) + \text{Recall}_i(j)}, \quad i = 1, 2, 3. \tag{4}$$

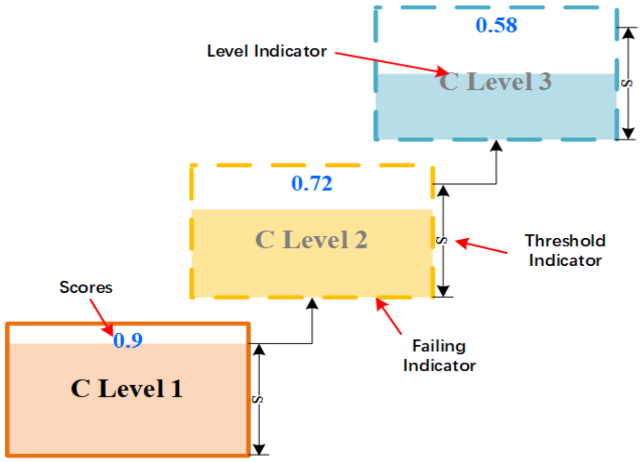

Fig. 3. Proposed Cascaded Tanks Model. Level 1, 2 and 3 evaluations are visualized in red, yellow and blue color, respectively. The $s = 0.90$ is denoted by threshold indicators, and the per-level all-task average score S_i is visualized by the "water level" in the respective "tanks". The PASS and FAIL outputs of C_i in Eq. (6) is represented the solid and dash borders of square "tanks", respectively.

TP_i, FP_i and FN_i represent the True Positive, False Positive and False Negative at the i-th level. A per-level all-task average score S_i is obtained by

$$S_i = \sum_{j=1}^{N} w_j F_{1,i}(j), \tag{5}$$

where w_j, $j = 1, \cdots, N$ denotes the predefined reference weights. Additionally, let $s \in [0, 1]$ denote a given global threshold, and an indicator function $C_i(S_i, s)$ for the ith level ($i = 1, 2, 3$) is constructed to provide easily interpretable results as

$$C_i(S_i, s) = \begin{cases} \text{PASS} & \text{if } S_i \geq s \\ \text{FAIL} & \text{otherwise} \end{cases}. \tag{6}$$

As in Fig. 3, the proposed cascaded tank model offers intuitive visualization and easy interpretation. Level 1, 2 and 3 evaluations are visualized in red, yellow and blue color, respectively. The $s = 0.90$ is denoted by threshold indicators, and the per-level all-task average score S_i is visualized by the "water level" in the respective "tanks". The PASS and FAIL outputs of C_i in Eq. (6) is represented the solid and dash borders of square "tanks", respectively.

III. Case Study: IV Cognition Evaluation at iVPC

In November 2017 at the Intelligent Vehicles Proving Center of China (iPVC), Changshu, China, the proposed GOE framework was applied in an invited offline cognition test on two autonomous driving platforms A and B (both anonymized), with the following tasks.

1) Lane Detection. Lane marking localization and corresponding line type classification based on video inputs.
2) Front Vehicle-and-Pedestrian (V&P) detection based on synchronized point cloud-image sequences.
3) Traffic Sign and Signal detection based on video inputs. All signs and signals are compliant with the national standards GB5768-2009 and GB14886-2006.

TABLE I

Statistics of the Chosen Segments Utilized in the Case Study.

Tasks	Simple	Medium	Complex	Total
Lane Detection	18	19	7	44
V&P Detection	19	7	3	29
S&S Detection	5	14	13	32

Data Acquisition and Selection. To ensure the objectiveness and effectiveness of the conducted test, instead of utilizing the existing traffic dataset (e.g., KITTI, RobotCar), we collected the wide-angle image sequences in conjunction with point cloud data from the autonomous driving platform shown in Fig. 2. The equipped Velodyne LiDAR and PointGrey camera was pre-calibrated following [7]. As commonly assumed in [19], [20] [21], multi-sensory data offers more discriminative information thus both video sequence and LiDAR point cloud are recorded while driving in the city of Changshu and on nearby highways, in regular daytime, rush hours and at night. Post-processing was conducted to ensure the time synchronization between video frames and point cloud frames.

As the raw dataset was of approximately 18TB in size, appropriate subsets need to be selected. Based on the proposed *Challenging Conditions*, we manually chose 105 segments of scenario data, each of which contain 80-200 video and point clouds frames, as summarized in Table I. For example, 44 segments were selected for the lane detection task with varying difficulty ratings, including curvy roads, roads with fuzzy markings, roads under strong illumination, roads in the night. All the selected data is encoded by the proposed *Semantic Descriptor* and subsequently graded by the classifier presented in Section II-B. The distribution of the obtained scenario difficulty grades are summarized in Table I.

Lane Detecting Evaluation. We jointly evaluated lane marking detection and line type classification with the benchmark dataset, which comprises 6464 annotated lane markings with a total length of 91229.51 meters. With a given image frame, an algorithm is required to localize and classify all visible markings $10 \sim 50$ meters ahead. Algorithm outputs with distance to the ground truth of less than 40 cm are accepted as True Positives (TP).

As illustrated in Fig. 4, Platform A outperformed B on the simple scenario data with the score of 0.93, which earned it a "Level 1 PASSED" rating. On the contrary, limited to detecting only the ego-lane, Platform B failed in the all levels of evaluations. Besides, both platforms suffered dramatic performance degradations under low-light conditions (night driving scenarios).

Front V&P Detection Evaluation. The task of front vehicle and pedestrian detection provides time-synchronized RGB image-point cloud sequences and calibrated parameters of LiDAR-to-camera, therefore, a high degree of freedom is offered. A candidate platform could fuse data from both modalities or select relevant data from one modality.

Unlike general purpose object detection tests which typi-

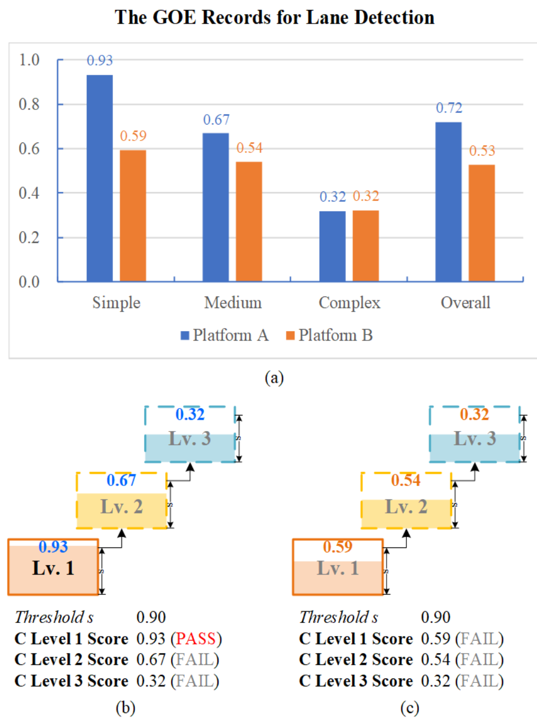

Fig. 4. Records of GOE in lane detection evaluation. (a) Histograms of GOE results. (b)-(c) Cascaded tank model based visualization and comparison of Platform A and B, respectively.

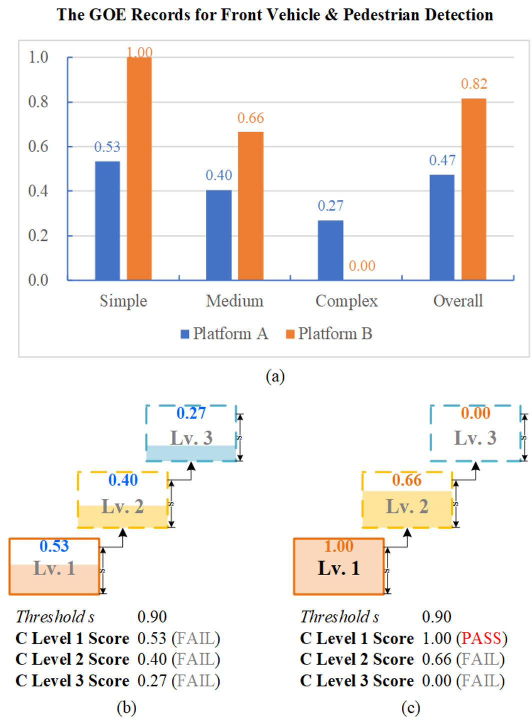

Fig. 5. Records of GOE in front vehicle and pedestrian detection. (a) Histograms of GOE results. (b)-(c) Cascaded tank model based visualization and comparison of Platform A and B, respectively.

cally require the detection of all vehicles within a predefined visible range, the objective of this evaluation is focused on the most relevant (i.e., nearest) frontal vehicle/pedestrian travelling in the ego-lane within 20 meters of range. Such setting is derived from the front collision mitigation feature in modern ADAS.

As in Fig. 5, with such evaluation setting, the algorithm in Platform B achieved higher scores than that in Platform A at Lv.1 and Lv.2. Nevertheless, neither algorithm can gracefully handle complex traffic scenarios at night. For example, the complex scenario depicted at the bottom of the Fig. 2 consists of blurry pedestrians and overexposed cyclists, which remained difficult for cognition algorithms in IV.

Sign & Signal Detection Evaluation. In the benchmark dataset for traffic signs and signals detection, a total of 1849 signs/signals of 25 types were manually annotated, as shown in Table II. The algorithm in a candidate platform is required to localize and classify the signs/signals from video inputs. Extremely small signs/signals with bounding boxes smaller than 16×9 pixels are excluded from the evaluation.

As in Fig. 6, both algorithms in the two platforms failed in all 3 levels. However, Platform B demonstrated its relative advantages in recognizing Warning Signs, Indication Signs and Prohibitory Signs. According to Table II, the ability to detect Guide Signs and Signals is still weak for both platforms. One possible explanation is that without zoomed-in image captured with a telescope lens, neither algorithm could distinguish traffic signal lights from other objects, as illustrated by the image at the bottom of Fig. 2.

A. *Experimental Analysis*

An overview of the results from the proposed GOE framework is provided in Table III. Although some of them passed Level 1 evaluations, both failed when encountering real, complex cognitive challenges, which in return validated the necessity of such graded evaluation frameworks. From the subfigure (a)s of Fig. 4–6, it empirically verifies our speculation that higher scenario complexity often leads to lower cognition performance.

TABLE II

THE RECOGNITION ACCURACIES IN DIFFERENT CATEGORIES.

Category	GT	TP	Precision
Warning Sign	88	0(A)/39(B)	0.0000(A)/**0.4432(B)**
Indication Sign	195	1(A)/99(B)	0.0051(A)/**0.5077(B)**
Prohibitory Sign	680	66(A)/319(B)	0.0971(A)/**0.4691(B)**
Guide Sign	382	65(A)/40(B)	**0.1702(A)**/0.1047(B)
Signals	504	72(A)/24(B)	**0.1429(A)**/0.0476(B)

TABLE III

OVERALL GOE COMBINED ACCURACIES AND PASSED LEVELS. N/A INDICATES AN ALGORITHM FAILS ALL 3 LEVELS OF EVALUATION.

Tasks	Platform A	Platform B
Lane Detection	**0.7208 (Lv.1)**	0.5276 (N/A)
V&P Detection	0.4747 (N/A)	**0.8157 (Lv.1)**
Signal Detection	0.1229 (N/A)	**0.3599 (N/A)**

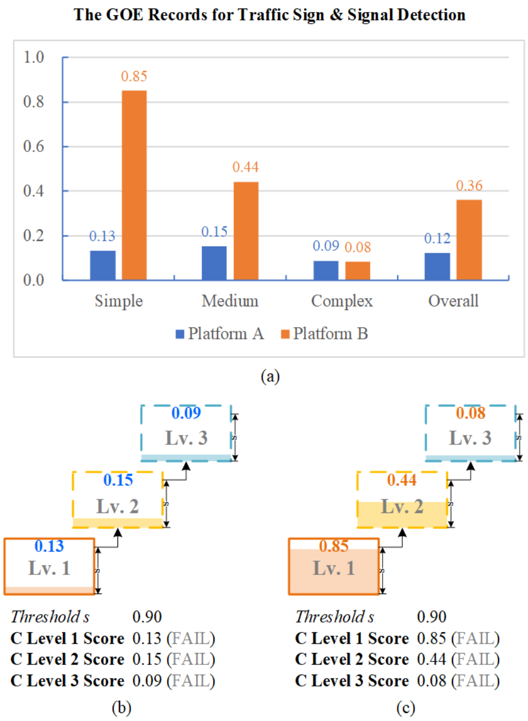

Fig. 6. Records of GOE in traffic sign & signal detection evaluation. (a) Histograms of GOE results. (b)-(c) Cascaded tank model based visualization and comparison of Platform A and B, respectively.

IV. CONCLUSIONS

In this paper, the GOE framework is proposed to provide time efficient, cost effective and repeatable offline evaluation of the cognitive ability in IVs. To achieve such goals, a quantitative *Semantic Descriptor* is proposed to summarize realistic driving difficulties, an SVM-based classifier is utilized in the per-segment data gradation subsystem (to classify data at the simple, medium and complex scenario difficulty ratings), and a cascaded tanks model is incorporated to provide efficient evaluation comparison and easily interpretable result visualization. With a thorough investigation in the section of case study, the efficacy of proposed GOE framework is demonstrated and verified.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (Grant No: 91520301). The authors would like to thank Guangzhong Liu, Song Liao, Jiajie Wang and Xudong Han's effort in case studies, and thank Dr. Dixiao Cui, Di Wang, Yang Zhong and Tao Li for constructive advices.

REFERENCES

[1] H. Abeida, Q. Zhang, J. Li, and N. Merabtine. Iterative sparse asymptotic minimum variance based approaches for array processing. *IEEE Trans. Signal Process.*, 61(4):933–944, 2013.

[2] M. B. Christopher. *PATTERN RECOGNITION AND MACHINE LEARNING*. Springer-Verlag New York, 2016.

[3] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[4] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, April 2012.

[5] S. Escalera, O. Pujol, and P. Radeva. On the decoding process in ternary error-correcting output codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):120–134, Jan 2010.

[6] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[7] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, June 2012.

[8] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li. Video-based sign language recognition without temporal segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA*, 2-7 February, 2018.

[9] L. Li, D. Wen, N. N. Zheng, and L. C. Shen. Cognitive cars: A new frontier for adas research. *IEEE Transactions on Intelligent Transportation Systems*, 13(1):395–407, March 2012.

[10] Y. Li, Y. Liu, C. Zhang, D. Zhao, and N. Zheng. The "floor-wall"traffic scenes construction for unmanned vehicle simulation evaluation. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1726–1731, Oct 2014.

[11] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017.

[12] L. Ran, Y. Zhang, W. Wei, and Q. Zhang. A hyperspectral image classification framework with spatial pixel pair features. *Sensors*, 17(10):2421, 2017.

[13] L. Ran, Y. Zhang, Q. Zhang, and T. Yang. Convolutional neural network-based robot navigation using uncalibrated spherical images. *Sensors*, 17(6):1341, 2017.

[14] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In *IEEE Winter Conference on Applications of Computer Vision*, pages 75–82, March 2014.

[15] J. Zang, L. Wang, Z. Liu, Q. Zhang, Z. Niu, G. Hua, and N. Zheng. Attention-based temporal weighted convolutional neural network for action recognition. In *Proceedings of the 14th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations, AIAI 2018, Rhodes,Greece*, 2018.

[16] C. Zhang, Y. Liu, D. Zhao, and Y. Su. Roadview: A traffic scene simulator for autonomous vehicle simulation testing. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1160–1165, Oct 2014.

[17] Q. Zhang, H. Abeida, M. Xue, W. Rowe, and J. Li. Fast implementation of sparse iterative covariance-based estimation for source localization. *J. Acoust. Soc*, 131(2):1249–1259, 2012.

[18] Q. Zhang, H. Abeida, M. Xue, W. Rowe, and J. Li. Fast implementation of sparse iterative covariance-based estimation for array processing. In *Proceedings of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA*, pages 2031–2035, 6-9 November, 2011.

[19] Q. Zhang and G. Hua. Multi-view visual recognition of imperfect testing data. In *Proceedings of the ACM International Conference on Multimedia, Brisbane, Australia*, pages 561–570, 26-30 October, 2015.

[20] Q. Zhang, G. Hua, W. Liu, Z. Liu, and Z. Zhang. Can visual recognition benefit from auxiliary information in training? In *Proceedings of the Asian Conference on Computer Vision, Singapore*, pages 65–80, 1-5 November, 2014.

[21] Q. Zhang, G. Hua, W. Liu, Z. Liu, and Z. Zhang. Auxiliary training information assisted visual recognition. *IPSJ Trans. Comput. Vis. and Appl.*, 7:138–150, 2015.

[22] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. LeBlanc. Accelerated evaluation of automated vehicles in car-following maneuvers. *IEEE Transactions on Intelligent Transportation Systems*, PP(99):1–12, 2017.

[23] D. Zhao, H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan. Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. *IEEE Transactions on Intelligent Transportation Systems*, 18(3):595–607, March 2017.

[24] D. Zhao, Y. Liu, C. Zhang, and Y. Li. Autonomous driving simulation for unmanned vehicles. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 185–190, Jan 2015.