Understanding Drivers' Visual and Comprehension Loads in Traffic Violation Hotspots Leveraging Crowd-Based Driving Simulation

Zhihan Jiang, Xin He, Chenhui Lu, Binbin Zhou, Xiaoliang Fan, *Senior Member, IEEE*, Cheng Wang, *Senior Member, IEEE*, Xiaojuan Ma, *Senior Member, IEEE*, Edith C.H. Ngai, *Senior Member, IEEE*, and Longbiao Chen, *Member, IEEE*

Abstract—**Traffic violations have become one of the major threats to urban transportation systems, undermining road safety and causing economic losses. Although various methods have been proposed by road authorities and researchers to find out the possible causes of traffic violations, existing methods often fail to diagnose traffic violations from drivers' perspectives and contexts or consider their visual and comprehension loads while driving. In this work, we propose a driver-centered simulation platform to inspect drivers' loads in traffic violation hotspots. Specifically, we first build a driving simulator based on the 3D point clouds of real-world traffic violation hotspots. We then recruit drivers to simulate driving in designated traffic scenes. Indicators for drivers' visual and comprehension loads are derived based on drivers' feedback. Upon this basis, we build an explainable model to automatically indicate drivers' visual and comprehension loads under various crowd-sensed traffic scenes. Experiments using real-world data from a Chinese City (Xiamen) and case studies show that our approach successfully derives a set of prominent indicators to effectively diagnose drivers' visual and comprehension loads in real-world traffic violation hotspots.**

Index Terms—**Traffic violation, crowdsensing, data analytics, driving simulation.**

I. Introduction

TRAFFIC violations, such as marked lanes violations and illegal turns, are one of the leading causes of traffic accidents [1], undermining human safety and causing economic

losses [2]. Therefore, it is important for urban authorities to find out the possible causes of traffic violations in urban road networks and implement corresponding strategies to alleviate the traffic violation and accident problem.

There are various causes of traffic violations. Some violations are results of drivers' intentional behaviors [3], such as drunk driving and anger driving. Others are unintentional and are usually caused by unfriendly traffic environments (e.g., worn-out road markings and appalling weather conditions) [4]. Unfriendly environmental factors significantly increase drivers' visual and comprehension loads [5], [6] and thereby lead to accidental breaking of traffic rules or regulations [7]. These high load-induced traffic violations are usually observed in specific road intersections or segments (i.e., traffic violation hotspots) during specific periods (i.e., traffic violation peaks). By understanding how environmental factors add to drivers' visual and comprehension loads in these traffic violation hotspots, urban authorities can improve traffic infrastructure accordingly to reduce the loads and ultimately mitigate the unintentional traffic violation issue.

Traditionally, the possible visual and comprehension loads of traffic violation hotspots are diagnosed by field studies of inspectors [8]. However, field trips are usually labor-intensive and time-consuming, and it is challenging to capture all traffic and external environments relating to traffic violations peaks [9]. As an alternative, researchers have proposed to diagnose possible factors relevant to drivers' loads from street view imagery [10], dashboard traffic cameras [11], Automated Enforcement System (AES) [12], etc. However, these methods are incapable of restoring the complex traffic environments at the time of traffic violation peaks (e.g., weather, illumination, and vehicle density) and often fail to identify issues from drivers' perspectives and trajectories. Therefore, a comprehensive and driver-centered approach to diagnose load-inducing factors at traffic violation hotspots is in great need.

With the development of autonomous driving and augmented reality, more and more vehicles and mobile devices are equipped with Mobile Laser Scanning (MLS) systems [13], making it possible to conduct widespread 3D scanning and collect large-scale city-wide 3D LiDAR point clouds [14]. By analyzing the point clouds, stereo views of the scanned location can be restored [15], providing researchers with new opportunities to understand the visual

Fig. 1. The structures of the driving simulator. (a) An illustration of the driving simulator implement and a traffic scene with overcrowded signs. (b) The real-world deployment of the driving simulator.

and comprehension loads in given traffic environments. For example, Zhang *et al.* [16] presented a quantitative visual recognizability evaluation method for traffic signs based on traffic recognition theory and 3D LiDAR point clouds. However, point clouds only contain the spatial structure information of the scanned locations while lacking contexts such as weather conditions and vehicle densities. To reconstruct specific drivers' contexts, we need to augment the point clouds with contextual features.

Therefore, we first locate traffic violation hotspots and restore violation-prone scenes based on crowd-sensed 3D point clouds and contexts. Second, we build a driving simulator in Virtual Reality (VR) (Fig. 1) for drivers to experience driving in various traffic scenes modeled from real-world traffic violation hotspots. Based on drivers' feedback, we derived a set of indicators of drivers' visual and comprehension loads related to traffic violations, namely *Blurred Signs&Markings*, *Worn-out Markings*, *Occluded Signs*, *Overcrowded Signs*, *Incompatible Signs*, and *Contradictory Signs&Markings*. For example, the traffic signs in Fig. 1(a) are overcrowded with redundant information, increasing drivers' comprehension loads (detailed in Section V). Furthermore, we build a diagnosing model based on the derived indicators to evaluate their effectiveness and apply the model to real-world traffic scenes to inspect the possible visual and comprehension loads experienced by drivers. To this end, the following issues need to be addressed.

1) *Restoration of Traffic violation-prone scenes.* Traffic violations are distributed across the road network. To restore traffic violation-prone scenes, it is essential to extract violation hotspots and the peaks of violation incidents at each hotspot. We first perform hotspot extraction on traffic violation data using a density-based spatiotemporal data clustering method. Second, we select the typical traffic violation hotspots based on their spatiotemporal characteristics and construct their structures with point clouds. Third, we augment the typical traffic violation hotspots with contexts derived from crowd-sensed data and model the traffic violation-prone scenes.

2) *Derivation of indicators for drivers' visual and comprehension loads in traffic violation-prone scenes.* To understand drivers' visual and comprehension loads in violation-prone scenes, we build a driving simulator for

drivers to experience driving in various traffic scenes and provide feedback on the perceived visual and comprehension demand. A hybrid approach of inductive and deductive analysis [17] (Fig. 5) is used to derive indicators for drivers' visual and comprehension loads related to traffic violations from drivers' input and literature of unfriendly environmental factors.

3) *Diagnosing visual and comprehension loads in a new traffic scene.* It is labor-intensive and time-consuming to recruit drivers to diagnose all traffic violation hotspots. We build a diagnosing model based on the derived indicators to automatically estimate drivers' visual and comprehension loads in traffic scenes. Specifically, road marking and traffic sign detectors based on Computer Vision are trained to identify these traffic infrastructural elements in image-based traffic scenes. Scores of the derived load indicators are calculated by a series of evaluators, and a classifier is trained to classify traffic scenes into violation-prone and non-violation-prone scenes.

To summarize, the key contributions of this paper include:

- To the best of our knowledge, this is the first work on understanding drivers' visual and comprehension loads leveraging driving simulation and building a diagnosing model for traffic scenes accordingly. Such a driver-centered framework can help urban authorities better inspect risky factors in road networks and take corresponding measures to alleviate the traffic violation issue.

- We propose a three-phase framework to diagnose the drivers' visual and comprehension loads in traffic scenes. Our method can not only derive indicators of drivers' visual and comprehension loads related to traffic violations *effectively* but also estimate loads in various crowd-sensed traffic scenes *automatically* and *systematically*.

- We implement the driving simulator and evaluate our approach using real-world datasets from a Chinese City (Xiamen). The results show that our approach successfully derives a set of prominent indicators to effectively diagnose drivers' visual and comprehension loads in real-world traffic violation hotspots.

TABLE I
DRIVER AND ENVIRONMENTAL FACTORS RELATED TO TRAFFIC VIOLATIONS

Driver	Intrinsic (e.g., age groups [18], education level [19], occupation [19], personality [20], distractibility [21]) Extrinsic (e.g., fatigue symptom [22], emotional states [23], alcohol [24], drugs [25], mobile phone interaction [26])
Environ-mental	Weather (e.g., foggy, rainy, snowy, dust) [27]–[29], Illumination (e.g., nighttime driving) [4], [30], Vehicle Density (e.g., traffic congestion) [31] Traffic Facility (e.g., compatibility, familiarity, standardization, simplicity, semantic distance, meaningfulness, concreteness) [32]

II. RELATED WORK

This section introduces the traffic violations and then reviews the works on factors leading to traffic violations. We divide these factors into driver factors and environmental factors (TABLE I) and elaborate on environmental factors which tend to form traffic violation hotspots. We further discuss the existing works on the relationships between environmental factors and drivers' workload and the measurement of drivers' workload. Finally, we discuss the existing works on mobile laser scanning and driving simulators.

A. Traffic Violations

Road traffic injuries have a significant impact on health and development [33]. There have been numerous works for promoting road safety in the literature. Some researchers worked on reducing in-vehicle risks, such as increasing the reliability of the controller area network in buses [34] and deploying driving assistance applications for drivers [35]. Others tried to improve traffic environments, such as countermeasures to reduce yellow-light running in dilemma zone [36]. As one of the major results of risky in-vehicle behaviors and unfriendly traffic environments, traffic violations have received extensive attention in both industry and academia.

The widespread traffic violation monitoring technologies, such as surveillance cameras and speed enforcement cameras, make it possible to collect a large amount of traffic violation data [37]. Also, many traffic violation crowdsensing platforms have been developed, making traffic violation reporting easier. These large-scale city-wide traffic violation data provide us with unprecedented opportunities to look into their causes.

B. Factors Related to Traffic Violations

Many studies have been conducted to identify the factors leading to traffic violations. Based on the literature review, we divide these factors into driver factors and environmental factors, as shown in TABLE I. Driver factors are drivers' characteristics related to traffic violation behaviors. Some are drivers' intrinsic characteristics, such as drivers' age, personalities [20], distractibility [21], education levels, and occupations [19]. Others are extrinsic characteristics, such as fatigue symptom [22], emotional states [23], alcohol and drug [24], [25], and drivers' interaction with mobile phones [26].

Environmental factors are characteristics of traffic environments. Unlike traffic violations induced by driver factors that are usually isolated and incidental, those induced by unfriendly environmental factors tend to form traffic violation hotspots and can be prevented by improving traffic environments.

The environmental factors can influence drivers' visibility and comprehension loads and thereby influence their driving behaviors [5], [6]. We divide the potential environmental factors into four categories (TABLE I). Weather conditions show a significant influence on drivers' visual loads [27], [29]. Nighttime driving increases drivers' visual loads [4]. As for the vehicle density, researchers have different conclusions. For example, Shinar [31] claimed that traffic congestion increases driver aggression and thus increases the risks of traffic violations and accidents, while Lajunen *et al.* [38] suggested that congestion does not increase driver aggression directly. The conditions of traffic facilities also play an essential role. Many works studied the relationships between the design features of traffic facilities (e.g., familiarity, concreteness, complexity, meaningfulness, and semantic distance [32]) and their visibility and comprehensibility. Carlson [39] investigated the effects of the brightness of traffic signs on their visibility. Kersavage *et al.* [40] evaluated the effectiveness of using different wording, text, background colors, and types of information on the visibility and comprehensibility of signs.

C. Relationships Between Environmental Factors and Drivers' Workload

Unfriendly traffic environmental factors increase drivers' workload [5], [6], making drivers prone to breaking traffic rules or regulations accidentally there [7] and leading to *traffic violation hotspots*. The modeling of drivers' workload has been widely studied. Some works simulated the mechanism of human information processors and proposed computational models (e.g., queuing network model [41] and neural network [42]) to estimate drivers' workload and performance. However, these works lack the analysis of relationships between specific environmental factors and traffic violations.

Besides, some works estimate drivers' workload directly based on drivers' physiological data and vehicle states influenced by driving environments. For example, Kim *et al.* [43] analyze the electroencephalogram data collected through an urban road driving test to explain the cognitive workload characteristics of different driving sections. Xing *et al.* [44] estimated workload using a support vector regression model based on the physiological data collected from drivers and vehicle speed and position data. Based on drivers' physiological data, vehicle signals, and traffic contexts, in [45] and [46], deep learning architectures were developed to assess drivers' workload. Noh *et al.* [47] proposed a framework to generate personalized driver workload profiles using physiological and operational data. However, these works mainly emphasize individuals. They are good at measuring workloads with physiological data collected from different drivers rather than directly assessing workloads for traffic scenes without drivers.

On the other hand, some works studied the relationship between traffic environment factors and drivers' workload. For example, in [48], researchers found that the mental workload of drivers increased with an increase in other heavy goods

vehicles. Precht *et al.* [27] reviewed the effects of various weather conditions on drivers' workload. Yared *et al.* [49] found that nighttime driving creates a high perceived workload. However, these works mainly focus on several specific conditions. Although they quantified the environmental factors and established the relationship between these factors and drivers' workload, they are not able to directly estimate traffic violation risks for traffic scenes quantitatively.

D. Drivers' Visual and Comprehension Loads

Drivers' workload (e.g., cognitive and visual loads) is usually measured by physiological data and rating scales. For example, various strategies, such as empirical techniques, rating scales, and psycho-physiological, are used to determine the cognitive load. The visual load is usually quantified using eye movement measures, and higher visual demands lead to reduced speed and increased lane-keeping variation [50]. Based on the literature review, we can find that the increasing cognitive load usually leads to decreasing comprehensibility [51], and reducing the visual load can improve the visibility [50].

However, cognitive load is usually measured from the perspective of drivers' factors. To build a diagnosing model which can automatically figure out unfriendly environmental factors in traffic scenes, it is critical to reduce the influence of drivers' factors and measure the drivers' workload from the perspective of environmental factors.

Therefore, in this paper, inspired by the cognitive load theory, we define the difficulty of understanding the meaning of traffic facilities as the comprehension load. We use the comprehension load to describe the cognitive load added to drivers while comprehending traffic facilities. In this way, we infer drivers' comprehension and visual loads based on the comprehensibility and visibility of traffic facilities in traffic scenes, and thus directly assess violation risks of traffic scenes. Also, this kind of measurement is much easier to implement, minimizing the possible interference during driving.

E. Mobile Laser Scanning

Mobile Laser Scanning (MLS) collects geospatial data from vehicles and mobile devices fitted with LiDAR, cameras, and other remote sensors [52]. It can rapidly collect accurate and reliable 3D point clouds to represent 3D shapes or objects [53] and has been widely used in Autonomous Driving [14], Augmented Reality [54], etc. In recent years, more and more vehicles and mobile devices (e.g., iPad Pro and iPhone Pro [55]) are equipped with MLS systems, facilitating the collection of large-scale 3D point clouds [14].

By processing point clouds, stereo views of the urban traffic environment can be restored [16], providing researchers with new opportunities to understand drivers' visual and comprehension loads. For example, Zhang *et al.* [16] presented a quantitative visual recognizability evaluation method for traffic signs based on traffic recognition theory and 3D LiDAR point clouds. Takeuchi *et al.* [56] proposed a fast lane visibility check method using a high-density point cloud map. However, point clouds only contain the spatial structure information of

the scanned locations [57] while lacking contextual features. Therefore, we use point clouds to reconstruct traffic violation hotspots, and augment their contexts with vehicle density, weather, and illumination features.

F. Driving Simulation Environments

Driving simulation has been used in various scenarios, such as driving training [58], commercial games [59], and driving performance studies [60]. There are many existing driving simulators. Some simulators are based on virtual labs. For example, Dosovitskiy *et al.* [61] provided open digital assets to support autonomous driving. STISIM[1] provides professional driving simulation for occupational therapy, driving research, and driver training. OpenDS[2] is an open-source driving simulator with configurable vehicles and an informative GUI. The virtual lab-based simulators are more flexible and can be customized. However, although some driving simulators, such as NADS [62], can provide high-fidelity driving environments, they require specialized equipment and are inaccessible to the general public. Besides, they usually have limited scenarios. The contents are usually created from scratch by digital artists, unable to directly generate scenes based on real-world data and fail to simulate the complex natural urban traffic environments.

Other simulators work on roads. For example, Baltodano *et al.* [63] introduced a real road autonomous driving simulator. Wang *et al.* [64] studied on-road partial and fully autonomous driving interaction with a driving simulator in commercial passenger vehicles. Goedicke *et al.* [65] developed tools to enable VR driving simulation in a vehicle as it travels on the road. However, although these simulators have natural driving environments, they require much more labor and time than the virtual lab-based and have inherent risks.

Therefore, we build the driving simulator based on the highly accurate point clouds collected from real-world traffic scenes directly so that it can simulate complex urban traffic environments with high *fidelity* and *flexibility*. The VR driving simulator can provide users with immersive driving experiences and significantly reduce costs.

III. PRELIMINARIES AND FRAMEWORK

Definition 1 (Traffic Violation Hotspot and Peak): A traffic violation hotspot refers to a location with more traffic violations than others. Traffic violation peaks are the peaks of the time distribution of traffic violations in traffic violation hotspots.

Definition 2 (Traffic Scenes): Traffic scenes are vehicles' surroundings from the perspective of vehicles [66]. Real-world crowd-sensed traffic scenes are traffic scenes taken by the crowd in the real world (e.g., photos of road environments taken by dashcams or passengers).

Definition 3 (Traffic Facilities): Traffic facilities include traffic signs, road markings (e.g., lane markings, arrow markings, etc.), and other pieces of equipment that are used to control, regulate, and guide traffic. This paper mainly focuses on traffic signs, lane markings, and arrow markings.

[1] https://stisimdrive.com/
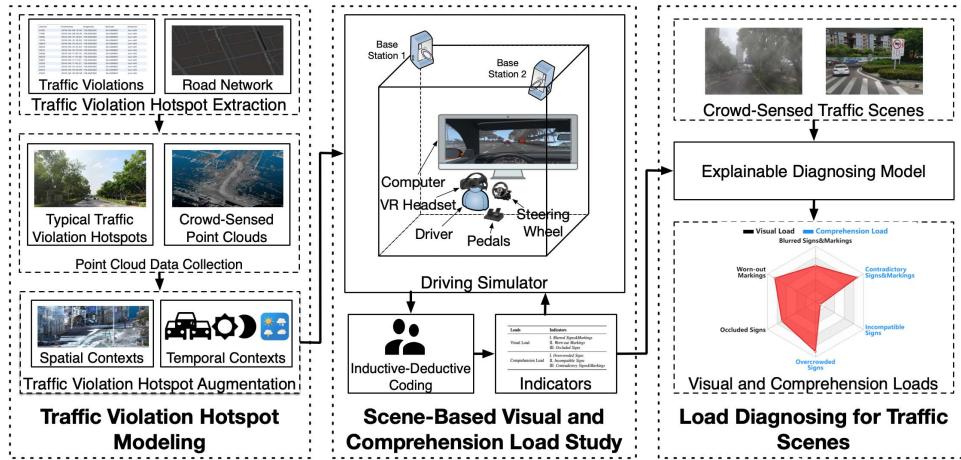[2] https://opends.dfki.de/

Fig. 2. Framework overview.

Definition 4 (Point Clouds and Mobile Laser Scanning): Point clouds are a collection of data points representing a 3D object and usually generated by 3D laser scanning or photogrammetric approaches [53]. Mobile Laser Scanning (MLS) allows the collection of accurate 3D point clouds using laser systems mounted on vehicles and mobile devices fitted with LiDAR, cameras, and other remote sensors [53].

Definition 5 (Visual and Comprehension Loads): Visual load refers to the difficulty of perceiving objects [50]. We use the visual load to describe the load added to drivers while trying to see traffic facilities clearly in traffic scenes, reflecting the visibility of traffic facilities. Inspired by the cognitive load theory [67], we define comprehension load as the difficulty of understanding the meaning of objects. We use the comprehension load to describe the cognitive load added to drivers while comprehending traffic facilities, which reflects the comprehensibility of traffic facilities.

We propose a three-phase framework to diagnose the drivers' visual and comprehension loads related to traffic violations (Fig. 2). First, we extract hotspots from traffic violations and select typical hotspots to collect their point clouds and augment their contexts. Then, we build a driving simulator, recruit drivers to experience driving in typical traffic violation-prone scenes using the simulator, and derive indicators for drivers' visual and comprehension loads based on analyzing the drivers' feedback. Based on the indicators, we build a model to automatically estimate drivers' visual and comprehension loads in traffic scenes.

IV. CROWD-BASED TRAFFIC VIOLATION HOTSPOT MODELING

In this phase, our goal is to model traffic violation hotspots leveraging heterogeneous crowd-sensed data. First, we extract traffic violation hotspots from traffic violation data leveraging a density-based spatial-temporal data clustering method. Then, we select typical traffic violation hotspots and reconstruct their spatial structures by their 3D point clouds. After that, we augment the hotspots using contextual features. We elaborate on the details of our approach as follows.

A. Traffic Violation Hotspot Extraction

We first extract traffic violation peaks from scattered traffic violations. For each traffic violation p, its location is denoted

by γ (a road or road intersection). We aggregate traffic violations with the same locations (i.e., the same roads or road intersections) and thereby get a set of locations Γ, i.e., $\Gamma = I \cup R = \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$, $\gamma_i = (i, P_i), 1 \leq i \leq N$, where I, R, i, P_i, and N are the set of road intersections with traffic violations, the set of roads with traffic violations, the ID of road intersections or roads, the set of traffic violations in the road or road intersection i, and the number of road intersections or roads, respectively. Specifically, for some long roads, we need to extract road segments with frequent traffic violations from them. Besides, since the time distribution of traffic violations in the same road intersection or segment is uneven, we also need to extract periods with frequent traffic violations from the continuous time. To this end, for each location $\gamma_i \in \Gamma$ $(1 \leq i \leq N)$, we cluster its traffic violations into different clusters with a density-based spatiotemporal data clustering method, *ST-DBSCAN* [68]. *ST-DBSCAN* can discover clusters according to non-spatial and spatiotemporal values of objects, which is very suitable for traffic violation extraction. After clustering, the scattered traffic violations are clustered into traffic violation peaks.

After getting the set of traffic violation peaks Λ, we first initialize the set of traffic violation hotspots H with Λ. Then, we get the closest distance of sets in H, denoted as d_{min} and if d_{min} is smaller than δ_r, the closest pair of sets are combined as one set. The closes pair of sets are iteratively combined until the d_{min} is no smaller than δ_r. Note that the distance between two sets are determined by the geographic distance of the average longitude/latitude in each set. Finally, we remove the sets whose size (the number of peaks) is smaller than δ_k from H, and get a set of traffic violation hotspots H, and for each hotspot $h_i \in H$, it consists of traffic violation peaks in this location, i.e., $H = \{h_1, h_2, \ldots, h_K\}$, where K is the number of traffic violation hotspots extracted, $\forall i, j \in [1, K], i \neq j, h_i \cap h_j = \emptyset$, and $(h_1 \cup h_2 \cup \ldots \cup h_K) \subseteq \Lambda$.

B. Hotspot Structure Construction

After extracting the traffic violation hotspot, our next step is to construct their spatial structures using their point clouds. We build a Point Cloud Crowdsensing Platform to collect their point clouds. As shown in Fig. 3, MLS providers (detailed in Appendix III) upload the point clouds of traffic environments

(a) MLS Providers   (b) Point Cloud Crowdsensing Platform   (c) Point Clouds

Fig. 3. The point cloud Crowdsensing platform.

to the crowdsensing platform. We can view the collected point clouds on the platform.

Since investigating all traffic violation hotspots requires rather high costs, we classify traffic violation hotspots based on their spatiotemporal contexts and select typical traffic violation hotspots from them. More specifically, in terms of time, we divide the time into T hours, and for each hotspot $h_i \in H$ ($i \in [1, K]$), we represent its traffic violation peak distribution using a T-dimensional 0-1 vector v_i, where the jth entry is set to one if the jth hour is in a traffic violation peak. We form the traffic violation hotspots as a weighted graph whose nodes are hotspots and weights of edges are the Euclidean Distance Similarity between two nodes (hotspots). A hierarchical clustering algorithm, Louvain algorithm [69], is used to divide traffic violation hotspots into different communities.

In terms of space, we assign traffic violation hotspots to different urban functional areas based on their Points of Interest (POIs) features [70]. We divide POIs into five categories, i.e., business, accommodation, entertainment, infrastructure, and tourism. We count the number of POIs for each category and build the POI features for each hotspot, i.e., $\Phi(h_i) = [o_1, o_2, o_3, o_4, o_5]'$, where $0 \le o_1, o_2, o_3, o_4, o_5 \le 1$ are the ratios of business, accommodation, entertainment, infrastructure, and tourism POIs to all POIs for hotspot h_i, $\sum_{i=1}^{5} o_i = 1$. Then we build five core POI features corresponding to business, accommodation, entertainment, infrastructure, and tourism, respectively. For core feature $\Phi_k(1 \le k \le 5)$, $o_j = 1(j = k), o_j = 0(j \ne k)$. We assign each traffic violation hotspot to one of the five spatial categories whose core feature has the closest Euclidean distance to them.

Finally, for each pair of temporal category and spatial category, if there exist traffic violation hotspots, we look up the point cloud crowdsensing platform to get, if any, point clouds of one traffic violation hotspot, and process the point clouds by removing the outliers and filtering out interferences such as pedestrians and vehicles to get the spatial structures.

C. Hotspot Context Augmentation

Our next step is to augment the typical traffic violation hotspots. Based on the literature review (TABLE I), we augment the following contextual features for traffic violation peaks in each typical traffic violation hotspot.

- **Vehicle Density** is estimated by the Baidu Traffic Congestion Index (TCI), which is open-access and has been used in many Chinese cities for years [71]. The vehicle density of each traffic violation peak is divided into

four categories, i.e., *smooth*, *slight*, *moderate*, and *severe* (detailed in Appendix I).
- **Weather**. We retrieve the weather conditions using the Weather Underground API[3] and Chinese Meteorological Big Data Platform,[4] and divide weather conditions into six categories, {cloudy, rainy, snow, fog, dust, sunny}. Furthermore, for category *rainy, snow, fog, dust*, and *cloudy*, they are further divided into four categories {Light, Moderate, Heavy, Violent} according to China Meteorological Bureau (detailed in Appendix II).
- **Illumination**. Since weather conditions influence the illumination and we have considered weather conditions, we classify the illumination conditions into daytime driving and nighttime driving.

D. Evaluation

1) Datasets: The traffic violation data used in this paper are provided by Xiamen Transportation Bureau, collected from March 2017 to February 2018 by traffic enforcement systems, traffic police, and crowd reporting, including 651 traffic sign violations, 74,921 road marking violations, 1,473 illegal parking, 698 speeding violations, and 1,283 other violations such as overtaking and not yielding to pedestrians, all committed by car users. The road network includes Xiamen City. As shown in Fig. 4(g), part of Xiamen City is connected to the mainland, and the other part is an island. We can find that traffic violations are unevenly distributed across the time and space (Fig. 4(a)(b)(c)(g)).

2) Results: Based on the traffic violation data, we extracted 9,539 traffic violation peaks, and they are clustered into 68 traffic violation hotspots. The POI distribution around the hotspots is shown in Fig. 4(d). These hotspots were classified into five spatial categories based on their POI distribution, and classified into three temporal categories based on their traffic violation peak distributions. For each pair of temporal category and spatial category, we select one traffic violation hotspot whose point clouds have been collected. If there exists no available hotspot, we skip this pair of categories. If there exist more than one available hotspot, we randomly select one of them. Finally, we obtained eight typical hotspots. The weather distribution of all data collection periods and with traffic violation peaks in typical hotspots is shown in Fig. 4(e). The distribution of vehicle density of traffic violation peaks in typical hotspots are shown in Fig. 4(f).

---

[3]https://www.wunderground.com/weather/api/
[4]http://www.weatherdt.com

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

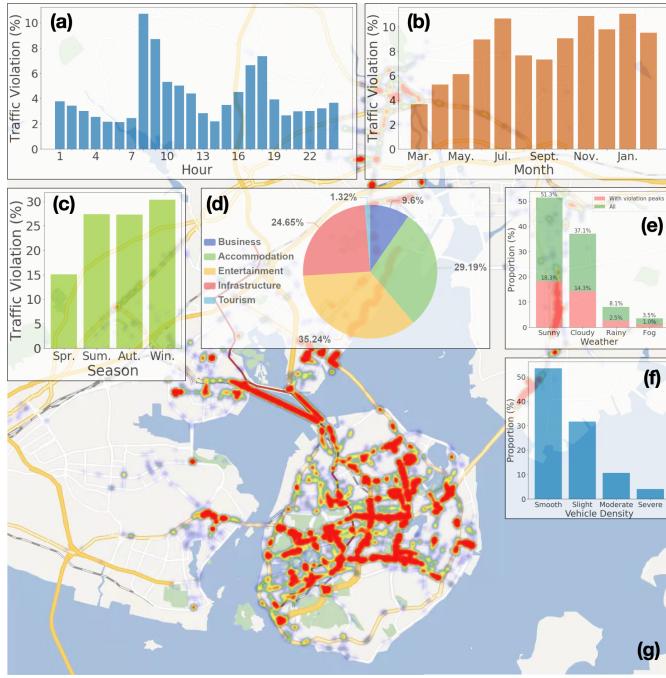JIANG *et al.*: UNDERSTANDING DRIVERS' VISUAL AND COMPREHENSION LOADS

7

Fig. 4. The overview of datasets and results. (a) The average traffic violations in a day. (b) The average traffic violations across different months. (c) The average traffic violations across different seasons. (d) The POI distribution of traffic violation hotspots. (e) The weather distribution of all data collection period and with traffic violation peaks in typical hotspots. (f) The distribution of vehicle density of traffic violation peaks in typical hotspots. (g) The heat map of traffic violations.

V. SCENE-BASED VISUAL AND COMPREHENSION LOAD STUDY

In this section, a hybrid approach of inductive and deductive analysis is used to derive load indicators related to traffic violation from drivers' input. As shown in Fig. 5, we first model the traffic violation-prone scenes and build a driving simulator to conduct driving simulation experiments (I). We derive a set of reasons for traffic violations and unfriendly elements through interviews. We further model the experimental traffic scenes with unfriendly elements improved and recruit 70 drivers to conduct driving simulation experiments (II) and score their visual and comprehension loads with questionnaires. The experiments are within-subject designed to decrease the influence of individual differences and increase the probability of finding the differences between the control scene and experimental scene with fewer participants [72]. Finally, we conduct statistical analysis to derive load indicators.

A. Driving Simulator Development

We develop a VR driving simulator to conduct driving simulation experiments, as shown in Fig. 1(a).

1) Traffic Violation-Prone Scene Modeling: To model the traffic violation-prone scenes, we first extract contextual features significantly related to violation peaks in typical hotspots selected. Two T-dimensional 0-1 vectors v and f represent the traffic violation peak distribution and contextual features for each typical hotspot. For example, for sunny weather, if the jth hour is sunny, the jth entry of f is set to one, and otherwise set to zero. For other features, f is set in

TABLE II
THE SIGNIFICANT CONTEXTUAL FEATURES (SIG. FEATURES) AND IDS OF NINE TRAFFIC VIOLATION-PRONE SCENES MODELED FROM EIGHT TRAFFIC VIOLATION HOTSPOTS IN XIAMEN

Scene ID (Hotspot ID)	Sig. Features
S1 (H1)	Heavy Rain
S2 (H1)	Light Fog
S3 (H2)	Moderate Congestion
S4 (H3)	Nighttime
S5 (H4), S6 (H5), S7 (H6),S8 (H7), S9 (H8)	None

a similar way. Then, we conduct the Chi-Square tests [73] based on the contingency table displaying the joint frequencies of values to evaluate whether the variables are associated or independent. For the 8 typical hotspots, we modeled 9 traffic violation-prone scenes based on their significant contextual features (TABLE II). Specifically, if the violation peaks in a traffic violation hotspot are significantly related to more than one feature, we would model more than one scene accordingly, such as S1 and S2 for H1.

In each hotspot, if some types of contextual features are not significantly related to its violation peaks, such as *Vehicle Density* and *Illumination* in H1, these types of features in the modeled scenes are set to the most friendly conditions based on the literature review. For example, in S1 and S2, the vehicle density is set to *smooth* and the illumination condition is set to *daytime driving*. For hotspots without significant contextual features, we set all conditions in their traffic violation-prone scenes as the most friendly ones. All traffic scenes modeled in this paper can be found in Appendix VII. The software part of the driving simulator is developed using the Unity 3D Engine.[5]

2) Hardware Components:

- **VR Headset and Base Stations**. The VR Headset and base stations are from HTC Vive Virtual Reality System,[6] with which drivers can experience driving using controllers and headset tracking with realistic graphics, directional audio, and HD haptic feedback [74]. To make the driving experience more realistic, we replaced the controllers by the steering wheel and car pedals.
- **Computer**. The computer is with Intel Core i5-6500 CPU, 16GB RAM, and Nvidia GeForce GTX1080Ti 11G. Three screen monitors are used to help observe the drivers' behaviors.
- **Steering Wheel and Car Pedals**. The steering wheel and car pedals are PXN-V3II.[7] The wheel has 180-degree rotation and a realistic wheel design, and the pedals are with gas and braking control.
- **Environments**. We deploy the simulator in a quiet room, as shown in Fig. 1(b).

B. Driving Simulation Experiments I

1) Participants: In driving simulation experiments I, we recruit 12 drivers (6 males and 6 females) Their ages range from 23 to 24 ($M = 23.7, SD = 0.471$). Their years of driving experiences range from 1 to 5 years ($M = 2.58, SD = 1.26$).

[5]https://unity.com/
[6]https://www.vive.com/eu/
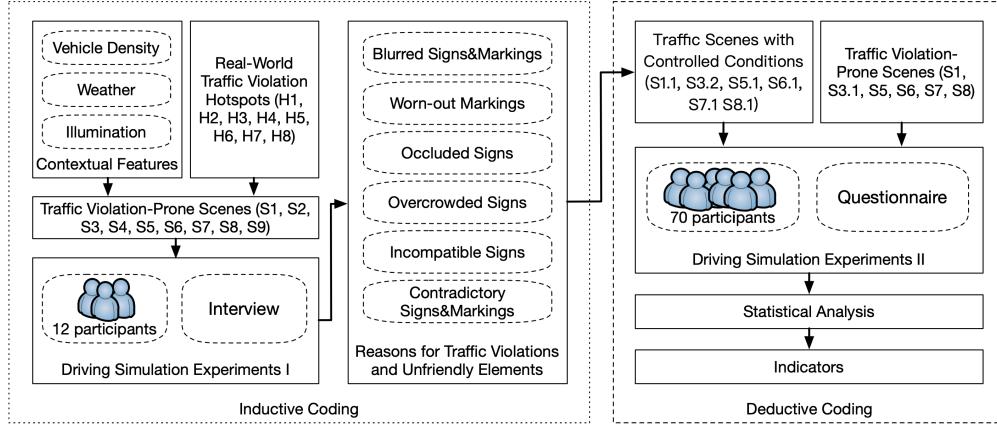[7]http://www.e-pxn.com/

Fig. 5. The diagram of the hybrid approach of inductive and deductive analysis. Inductive coding: a bottom up approach where you start with no codes and develops codes as you analyze the data. Deductive coding: a top down approach where you start with a set of predetermined codes and then find excerpts that fit those codes.

Fig. 6. View of participants sitting in the simulated car. Note that the enlarged traffic sign is not in the view.

Half of them are local residents relatively familiar with the typical traffic violation hotspots, and the other half are non-local. Their detailed information can be found in Appendix IV.

2) Experiment Settings: In the experimental room, a participant sits in the seat and wears a VR headset through which the participant sees himself/herself sitting in the seat of a virtual car (Fig. 6). Before driving, the participant would practice controlling the virtual car using the steering wheel and pedals with our instructions. In each traffic violation-prone scene, the start points for participants are fixed, while the endpoints depend on their driving behaviors. The domain of the scene is limited to guarantee that the participant would drive across the traffic violation hotspot and finish driving in 10 seconds.

Upon finishing driving in a scene, if the participant violated traffic rules, we would ask about the possible reasons for the violation. Otherwise, we would inquire about what factors in this scene are unfriendly to driving. Then, the participant would proceed to the next scene until all scenes have been experienced. The order of scenes is random. The participant would take a break for 15 minutes after completing five scenes and then resume to finish the remaining scenes.

3) Experiment Results: Three authors conducted reflexive thematic analysis on the text of interviews with drivers following the guidelines of [75], detailed in Appendix VI. In general, three authors first read through the text independently to get familiar with the data. Second, the authors coded the data by extracting key phrases or sentences in an inductive manner, during which authors discussed and compared their codes to refine the codes. Then, the authors reviewed the codes together and discussed the code importance and relationships among codes to generate key themes. After reviewing the themes by all authors, we named and defined the following six key factors related to drivers' traffic violation behaviors.

a) Blurred signs&markings: Many participants violated traffic rules (5/12) because they failed to see traffic signs and Markings clearly. In S1, all participants claimed that the heavy rain added to their burden when trying to recognize the traffic signs. In S2, fog weather influenced the visibility of traffic signs, making drivers more likely to violate traffic rules (7/12 participants complained about the fog). In S4 where traffic violation peaks usually happened during nighttime, most of the participants (9/12) claimed that the illumination there is insufficient and the no-left-turn sign is not very obvious. Two participants turned left illegally there.

b) Worn-out markings: Some road lanes are seriously worn-out, leading to marked lanes violations. For example, in S7, participants (5/12) complained that the solid road line had faded, making it look like a skip line. Drivers are thereby very likely to make a marked lanes violation there.

c) Occluded signs: Some traffic signs are occluded by trees or other objects, making it difficult for drivers to notice traffic signs quickly or see them clearly. In S5, drivers can only make a U-turn on the far left lane, and there is a U-turn sign on the left for a reminder. However, some participants (3/12) complained that the roadside trees covered the U-turn sign and they did not notice it when they drove through fast. Although some participants (6/12) did not violate the traffic rule there, they still complained about it.

d) Overcrowded signs: It is difficult for drivers to timely recognize too many traffic signs in a traffic scene. In S3, four non-local participants complained that there are too many traffic signs, and some have similar meanings. Different from previous studies [31] but consistent with [38], although 8 participants expressed annoyance about the high vehicle density in S3, none of them thought it would lead to traffic violations. No traffic violation was committed in the experiments.

e) Incompatible signs: Signs are incompatible means that their physical arrangement in space is unrelated to or even opposite the information or directions conveyed. Although the traffic sign and road markings in S6 are in good condition and visually apparent, the design of the traffic sign is spatially incompatible. Three non-local participants pointed out the confusing arrangement of elements in the traffic sign. The text "Island Ring South Road" is on the left of the sign, and the text "Island Ring Main Road" is on the right of the sign (see the enlarged traffic sign in Fig. 6). Drivers may intuitively think the left road is "Island Ring South Road" and the right road is "Island Ring Main Road". However, according to the arrows, the "Island Ring South Road" is the road on the right, and the "Island Ring Main Road" is on the left. Drivers new to this area are prone to make mistakes, especially when driving at high speed. After realizing they are driving on the wrong side, they are likely to cross the solid lines to change lanes at the last minute. Even two local participants said they were unsatisfied with this sign and had made mistakes there before.

f) Contradictory signs&markings: Unlike the Incompatible Signs, signs&markings are contradictory means that the information conveyed by these signs or markings is contradictory. For example, in S8, there is a no-left-turn sign, while the road is painted with a dashed line on the left side of the lane, meaning that drivers can turn left there. Four out of 12 participants pointed out that they were unsure whether they could turn left or not there. Similarly, eleven participants reported that the new road marking coexisted with the old one in S9, making people prone to drive in the wrong lane.

C. Experimental Traffic Scene Modeling

For each factor, we model the experimental scenes (Appendix VII) for comparison, detailed as follows.

- **Blurred Signs&Markings**. Scene S1 has poor clarity. We change its weather condition into sunny to model S1.1 (experimental scene) with good clarity.
- **Worn-out Markings**. The lane marking in scene S7 has been worn out, and in the corresponding experimental scene, we repaint the worn-out solid line to model S7.1.
- **Occluded Signs**. In S5, the traffic sign is covered by roadside trees, and in the corresponding experimental scene, we remove the trees before the traffic signs to model scene S5.1.
- **Overcrowded Signs**. In S3, Since many participants complained about the high vehicle density but did not think it was related to traffic violations, we change the vehicle density to smooth to eliminate its influence (S3.1). We remove signs with redundant information in the corresponding experimental scene (S3.2).
- **Incompatible Signs**. Corresponding to S6 with incompatible traffic signs, we change the traffic signs to make them more spatially compatible to model S6.1.
- **Contradictory Signs&Markings**. Scene S8 has contrary traffic facilities, and we change the dashed line into a solid line to model S8.1.

D. Driving Simulation Experiments II

We conduct comparative experiments to explore the relationships between the six unfriendly environmental factors concluded above and drivers' visual and comprehension loads.

1) Participants: We recruit 70 drivers (29 males, 41 females) with diverse education backgrounds (21 have education levels below university, and 49 are with university education or above). Participants' ages ranged from 23 to 62 years old ($M = 36.9$, $SD = 12.1$). Their years of driving experience ranged from 1 to 38 years ($M = 7.89$, $SD = 6.52$). Among the participants, 43 are local people familiar with the experimental traffic violation hotspots, and 27 are non-local with little knowledge about the hotspots. The detailed information can be found in Appendix V.

2) Experiment Settings: Every time participants finish driving in a scene, traffic violations would be recorded. Participants would rate their *Visual Load* and *Comprehension Load* on a 5-point Likert scale (1 strongly disagree - 5 strongly agree) questionnaire as follows. Other settings and procedures are the same as the driving simulation experiments I.

- It is difficult for me to see traffic signs/road markings in the scene clearly while driving.
- It is difficult for me to understand the meaning of traffic signs/road markings in the scene while driving.

3) Experiment Results: The detailed experiment results can be found in Appendix V. We conduct statistical analysis on the experiment results, as shown in TABLE III.

Based on the results, we find that participants commit fewer traffic violations in experimental scenes (with unfriendly environmental factors improved) than in control scenes (with unfriendly environmental factors). The *Blurred Signs&Markings* caused by environmental conditions (e.g., bad weather or low illumination) indicate an increase in drivers' visual load for detecting traffic signs. The *Worn-out Markings* indicate a higher visual load for road markings. The *Occluded Signs* add to drivers' visual load for traffic signs, depending on the extent to which the signs have been blocked by other objects, such as roadside trees, telegraph poles, and other traffic signs. The *Overcrowded Signs* increase drivers' comprehension load for differentiating relevant traffic signs, which is influenced by the number of traffic signs in a traffic scene, especially for drivers unfamiliar with the road. The *Incompatible Signs* indicate a higher drivers' comprehension load for traffic signs. The *Contradictory Signs&Markings* indicate a greater demand for drivers' comprehension resources as they may confuse drivers.

We further conduct Wilcoxon Signed Ranks Test [76] for statistical significance. The results are shown in TABLE III, we find the *Blurred Signs&Markings*, *Worn-out Markings*, *Occluded Signs* are significantly related to the visual and comprehension loads. If a scene had poor visibility, its comprehensibility would be impaired since people often have trouble making sense of things that they could hardly see, leading to both high visual and comprehension loads. The *Contradictory Signs&Markings*, *Incompatible Signs*, and *Overcrowded Signs* are significantly related to comprehension load but not significantly related to visual load. Based on the above discussion, we derive six indicators for drivers' visual and comprehension loads related to traffic violations, as shown in TABLE IV.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10   IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS

TABLE III
THE STATISTICAL ANALYSIS RESULTS FOR DRIVING SIMULATION EXPERIMENTS II (70 PARTICIPANTS)

Scene (Violations)	Clear S1.1 (12)	Blurred S1 (31)	$\mathcal{Z}$ (p)	Normal S7.1 (3)	Worn-out S7 (25)	$\mathcal{Z}$ (p)
Visual Load	2.97±0.80	3.67±0.99	-4.489 (0.000)	3.37±1.04	3.83±0.99	-4.703 (0.000)
Comprehension Load	3.23±0.92	3.84±0.86	-4.481 (0.000)	3.51±1.09	3.89±0.99	-4.326 (0.000)
Scene (Violations)	Uncontradictory S8.1 (2)	Contradictory S8 (9)	$\mathcal{Z}$ (p)	Without Occlusion S5.1 (8)	With Occlusion S5 (28)	$\mathcal{Z}$ (p)
Visual Load	3.93±0.87	4.06±0.83	-1.330 (0.183)	2.76±1.36	4.20±0.86	-5.933 (0.000)
Comprehension Load	3.49±1.06	3.87±1.02	-4.204 (0.000)	2.87±1.34	4.16±0.77	-5.560 (0.000)
Scene (Violations)	Compatible S6.1 (4)	Incompatible S6 (7)	$\mathcal{Z}$ (p)	Normal S3.2 (9)	Overcrowded S3.1 (16)	$\mathcal{Z}$ (p)
Visual Load	3.91±0.86	4.04±0.82	-1.435 (0.151)	4.13±0.76	4.34±0.63	-1.719 (0.086)
Comprehension Load	3.67±0.97	4.09±0.76	-3.335 (0.001)	3.57±0.94	4.09±0.74	-3.429 (0.001)

The visual and comprehension loads are reported in the format of mean±standard deviation. $\mathcal{Z}$ and p are the $\mathcal{Z}$ statistic and p value.

TABLE IV
INDICATORS FOR VISUAL AND COMPREHENSION LOADS IN TRAFFIC VIOLATION HOTSPOTS

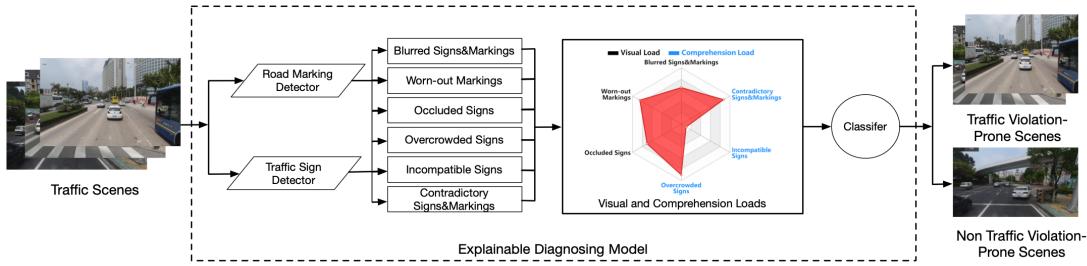| Loads	Indicators	Description
Visual Load	I. *Blurred Signs&Markings* II. *Worn-out Markings* III. *Occluded Signs*	The Blurred Signs&Markings increase the drivers visual load. The Worn-out Markings increase the drivers' visual load. The Occluded Signs increase the drivers' visual load.
Comprehension Load	I. *Overcrowded Signs* II. *Incompatible Signs* III. *Contradictory Signs&Markings*	The Overcrowded Signs increase the drivers' comprehension load. The Incompatible Signs increase the drivers' comprehension load. The Contradictory Signs&Markings increase the drivers' comprehension load.

Fig. 7.   The explainable model for visual and comprehension diagnosing in traffic scenes.

VI. LOAD DIAGNOSING FOR TRAFFIC SCENES

In this section, we build an explainable diagnosing model based on the indicators to automatically estimate drivers' visual and comprehension loads in traffic scenes.

A. Explainable Diagnosing Model

Image-based traffic scenes can be collected by mobile devices or from crowdsensing platforms, such as Baidu Panoramas.[8] As shown in Fig. 7, for traffic scenes, a road marking detector based on Mask R-CNN [77] is trained to detect road markings and worn-out road markings. A traffic sign detector based on YOLOv4 [78] is used to detect traffic signs, arrows in indication signs, and traffic signs with occlusion.

The detection results are input into 6 evaluators to get 6 scores for 6 indicators correspondingly. Suppose the traffic facilities detected in the scene are $F_1, F_2, \ldots, F_a$ (traffic signs and markings in normal condition), $W_1, W_2, \ldots, W_b$ (worn-out road markings), $O_1, O_2, \ldots, O_c$ (occluded traffic signs), and $A_1, A_2, \ldots, A_d$ (arrows in indication signs). We use $\{C_{F1}, C_{F2}, \ldots, C_{Fa}\}$, $\{C_{W1}, C_{W2}, \ldots, C_{Wb}\}$, $\{C_{O1}, C_{O2}, \ldots, C_{Oc}\}$, and $\{C_{A1}, C_{A2}, \ldots, C_{Ad}\}$ to denote the sets of confidence values of traffic signs and markings in

[8]https://quanjing.baidu.com/

normal condition, worn-out road markings, traffic signs with occlusion, and arrows in indication signs, respectively. The evaluators are detailed as follows.

- **Blurred Signs&Markings**. The blurriness of traffic signs&markings influences their probabilities of being detected. We use the confidence values to estimate the scores for *Blurred Signs&Markings*. The score is determined by the blurriest traffic signs&markings in normal conditions, i.e., the minimum values among $C_{F1}, C_{F2}, \ldots,$ C_{Fn}. A lower blurring score means a higher blurring degree of the traffic signs&markings.

- **Worn-out Markings**. a higher the confidence value of the detected worn-out marking means a higher probability of an actual worn-out marking. The *Worn-out Markings* scores are determined by the most worn-out road markings in scenes, i.e., the minimum values among $1 - C_{W1}, 1 - C_{W2}, \ldots, 1 - C_{Wb}$. A lower score means a higher wearing degree of the road markings in the scene. If there is no worn-out road marking, the score is one.

- **Occluded Signs**. Similar to the *Worn-out Markings*, the *Occluded Signs* score is determined by the most occluded traffic Signs in the scene, i.e., the minimum values among $1 - C_{O1}, 1 - C_{O2}, \ldots, 1 - C_{Oc}$. A lower score means a higher occlusion degree of the traffic signs in the scene. If there is no traffic sign occluded, the score is one.

- **Overcrowded Signs**. The score for *Overcrowded Signs* is determined by the number of normal condition traffic signs detected weighted by their confidence values and normalized into $[0, 1]$, i.e., $1 - (\sum_{i=1}^{a} C_{Fi})/10$. If $(\sum_{i=1}^{a} C_{Fi})/10 > 1$, the score is set to 0. A lower score means more crowded signs.
- **Incompatible Signs**. The score for *Incompatible Signs* is determined by the most incompatible arrow arrangement of indicator signs in the scene. For every pair of arrows in the same indicator sign, we first check whether they are compatible according to their locations and types. If they (A_i, A_j) are incompatibility (e.g., left and right arrows with up and down arrangement), their incompatibility score is $1 - C_{Ai} \times C_{Aj}$. The incompatibility score of the scene is the minimal value among the incompatibility pairs. A lower score means a more incompatible scene. If there is no incompatible pair, the score is one.
- **Contradictory Signs&Markings**. The score for *Contradictory Signs&Markings* is determined by the most contradictory traffic signs&markings in the scene. For every pair of contradictory traffic signs&markings (F_i, F_j) in the scene (e.g., no left turn signs and left turn road markings, no u-turn signs and u-turn signs), their score is $1 - C_{Fi} \times C_{Fj}$. The score of the scene is the minimal value among the contradiction scores of the contradictory pairs. A lower score means a more contradictory scene. If there is no contradictory pair, the score is one.

A classifier is trained based on the scores for load indicators to classify traffic scenes into traffic violation-prone and non-traffic violation-prone scenes. Here we use XGBoost [79] since it is efficient and can give the importance of features.

B. Evaluation

1) Datasets: The datasets for training the traffic sign and marking detectors and model evaluation include:

Traffic Sign Dataset. The Traffic Sign Dataset is from [80], including 100,000 images containing 30,000 traffic-sign instances. Furthermore, we annotate 400 occluded traffic signs and 587 arrows in indication signs (including 259 left arrows, 329 right arrows, 144 up arrows, and 88 down arrows).

Road Marking Dataset. The Road Marking Dataset is from [81], including lanes (25,354 single white, 74,733 dashed white, 206 double white, 28,054 single yellow, 5,734 dashed yellow, and 8,998 double yellow) and arrows (1,186 left, 537 right, 6,968 straight, and 127 u-turn). Furthermore, we annotate 400 worn-out road markings.

Traffic Scenes. Fifty traffic violation-prone scenes and 50 non-traffic violation-prone scenes are used to evaluate the performance of the load diagnosing model. The traffic scenes are street view images collected from Baidu Panoramas whose data are uploaded by the company or the crowd, and we augment their contexts using Adobe Photoshop.

2) Baseline Methods: We compare our method with various baseline methods. We separate the 100 traffic scenes into a training set (70%), validation set (10%), and test set (20%). In order to achieve a fair comparison, we make sure that the three sets and parameters used in each method are the same.
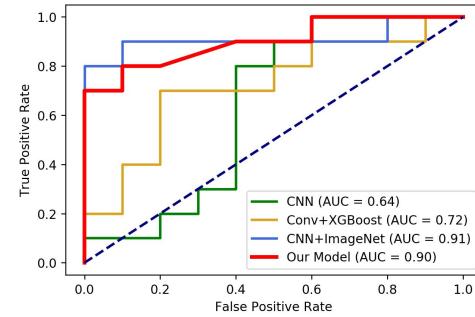

Fig. 8. The ROC curves of the classification performance of the proposed diagnosing model and baseline methods.

CNN is an end-to-end classifiers with convolutional neural networks [82]. It can directly extract features from images.

Conv + XGBoost extracts features from images using convolutional layers, but different from CNN which does classification using the fully connected layer, the extracted features are used to train a XGBoost classifier.

CNN + ImageNet is an end-to-end classifiers with convolutional neural networks fine-tuned from pre-trained weights on ImageNet [83].

3) Results: We use the ROC curve [84] to evaluate the performance of the proposed model and baseline methods on the test set, as shown in Fig. 8. **CNN** did not perform well, and **Conv + XGBoost** performed better than **CNN**, which may be resulted from insufficient training data. The deep-learning-based method usually requires a large amount of training data to achieve high performance, and thus transfer learning [85] is proposed to improve the performance on small datasets. **CNN + ImageNet** leveraged the weights pre-trained on ImageNet and achieved the best performance, which indicates the feasibility of identifying traffic violation-prone scenes based on traffic scene images. The proposed diagnosing model not only achieves comparable good performance with an AUC of 0.90 but also can present the load profiles to explain the potential causality of traffic violation-prone scenes.

C. Case Studies

1) A Marked Lane Violation Hotspot: As shown in Fig. 9, it is a traffic violation hotspot with frequent marked lane violations. The traffic scene has a poor traffic lane condition, increasing drivers' visual load and achieving a low score in *Worn-out Markings*. The worn-out lane detected used to be a solid white line that drivers are not permitted to press this line and change the driving lane. However, since the traffic lane is seriously worn-out, drivers are very likely to regard it as a dashed line or overlook it, and thus commit marked lane violations. Besides, the street view image has low resolution, and there is a no-right-turn sign in the distance, leading to some loss in the *Blurred Signs&Markings* score.

2) An Illegal Turning Violation Hotspot: Fig. 10 shows a traffic violation hotspot with frequent illegal turning violations. The traffic scene gets low scores in *Occluded Signs* and *Overcrowded Signs*. There is a no-right-turn traffic sign occluded by the streetlight pole, and a no-access sign occluded by trees, increasing the visual load. Moreover, there are many traffic signs, increasing the comprehension load in terms of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12

IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS

Fig. 9. A traffic violation hotspot caused by worn-out road markings. (a) The street view image of the scene with detection results. (b) The visual comprehension load profile.

Fig. 10. A traffic violation hotspot caused by occluded signs and overcrowded signs. (a) The street view image of the scene with detection results. (b) The visual comprehension load profile.

Overcrowded Signs. Specifically, the street view image is downloaded from Baidu Panorama, and the first traffic sign on the second row of the green signboard is automatically masked by Baidu Panorama. Similar to the first case, the low resolution of the picture influences the score of *Blurred Signs&Markings*.

VII. DISCUSSION

A. Generalizability

Compared with many existing driving simulators, our driving simulator is cheaper, enhancing the generalizability of our work. Point clouds are getting more accessible to the general public since more and more mobile devices are equipped with LiDAR, such as iPad Pro and iPhone [55]. The software part of the driving simulator is developed using the open-source Unity 3D Engine, and the hardware part consists of devices on sale to the public and popular in VR Games.

Besides, given more indicators, the diagnosing model can be easily extended since the detectors and evaluators in the model are independent of each other. For example, besides the road marking and traffic sign detectors, new detectors, such as a road condition detector, can be added to the model. New evaluators can also be easily added to get scores for new indicators. Then, the load profile and classifier can be updated with new features based on the new indicators.

Furthermore, our work also provides implications for other areas, such as transportation facility maintenance and driving assistance systems. For example, if a traffic scene gets a low score in *Overcrowded Signs*, the authority can remove the redundant signs in this scene to improve the score. The diagnosing model can be embedded into driving assistance systems to raise awareness about the potential risk factors in various traffic scenes.

B. Limitations and Future Work

In traffic violation hotspot modeling, we model the typical traffic violation hotspots based on real-world traffic violation data. However, since the traffic violation data are only from Xiamen City which has a tropical oceanic monsoon climate and no snowy or dusty days, the weather contexts we can explore are limited. Besides, although the typical hotspots selected are representative, there may still be some missing types due to the limited data. In the future, we can collect data from other cities with different climatic conditions, validate the current selected typical hotspots and obtain more typical hotspots by crowdsensing.

In the scene-based visual and comprehension load study, to explore the environmental indicators for the general public, we recruit drivers from various age groups and binary gender, and the experiments are conducted in a quiet room without interference from individuals' emotional status (e.g., anger) or contexts (e.g., rushing to work). However, this sample may still lack demographic diversity in terms of race and personality. The derived six indicators may not adequately reflect all unfriendly environmental factors in a traffic scene. To address this issue, we can recruit more drivers with different backgrounds to experience driving using the simulator and extend indicators based on their opinions.

Besides, we directly use the comprehensibility and visibility of traffic facilities in traffic scenes to represent the comprehension and visual load added to drivers. Although it is more cost-effective and intuitive, it may introduce bias. In the future, we may also use specialized equipment or standardized rating scales to measure drivers' workload to validate the effectiveness of the designed questionnaire.

In load diagnosing for traffic scenes, we build a diagnosing model based on the derived six indicators to automatically estimate drivers' visual and comprehension loads in various

traffic scenes. Although the current diagnosing model achieves good performance, its explainability can be further improved by considering more indicators. Also, the model can only analyze a traffic scene with images, and the scores are estimated based on traffic sign and marking detectors, while the texts in traffic signs also play an important role. We can incorporate other literacies such as texts and audio to build a multimodal model [86] to further improve the explainability.

VIII. CONCLUSION

In this work, we propose a three-phase framework to derive indicators for drivers' visual and comprehension loads related to traffic violations and build an explainable diagnosing model. First, we extract traffic violation hotspots from traffic violations and select representative typical hotspots. Second, we model the typical traffic violation-prone scenes based on the point clouds and traffic environment data. Third, we build a driving simulator, recruit drivers with various backgrounds to experience driving in various traffic scenes, and derive drivers' visual and comprehension load indicators based on drivers' feedback. Finally, we build a diagnosing model based on the indicators to automatically estimate loads in traffic scenes and evaluate the model using real-world datasets.
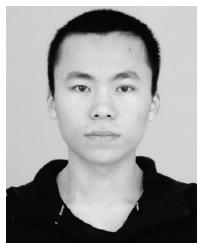
ACKNOWLEDGMENT

REFERENCES

[1] J. J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accident Anal. Prevention*, vol. 115, pp. 11–24, Jun. 2018.

[2] G. Zhang, K. K. W. Yau, and G. Chen, "Risk factors associated with traffic violations and accident severity in China," *Accident Anal. Prevention*, vol. 59, pp. 18–25, Oct. 2013.

[3] D. Parker, T. Lajunen, and S. Stradling, "Attitudinal predictors of interpersonally aggressive violations on the road," *Transp. Res., Traffic F, Psychol. Behav.*, vol. 1, no. 1, pp. 11–24, 1998.

[4] R. Madleňák, L. Madleňáková, D. Hošťáková, P. Drozdziel, and A. Török, "The analysis of the traffic signs visibility during night driving," *Adv. Sci. Technol. Res. J.*, vol. 12, no. 2, pp. 71–76, 2018.

[5] G. Murphy and C. M. Greene, "Load theory behind the wheel; perceptual and cognitive load effects," *Can. J. Exp. Psychol./Revue Canadienne de Psychologie Expérimentale*, vol. 71, no. 3, p. 191, 2017.

[6] K. Zeeb, A. Buchner, and M. Schrauf, "Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving," *Accident Anal. Prevention*, vol. 92, pp. 230–239, Jul. 2016.

[7] G. Zhang, Y. Tan, and R.-C. Jou, "Factors influencing traffic signal violations by car drivers, cyclists, and pedestrians: A case study from Guangdong, China," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 42, pp. 205–216, Oct. 2016.

[8] T. A. Dingus *et al.*, "The 100-car naturalistic driving study, phase II-results of the 100-car field experiment," Dept. Transp., Nat. Highway Traffic Saf., Washington, DC, USA, Tech. Rep. 37370, 2006.

[9] L. Larsen and P. Kines, "Multidisciplinary in-depth investigations of head-on and left-turn road collisions," *Accident Anal. Prevention*, vol. 34, no. 3, pp. 367–380, 2002.

[10] C. Monsere *et al.*, "Road user understanding of bicycle signal faces on traffic signals," Nat. Acad. Sci., Eng., Med., Washington, DC, USA, Final Rep. NCHRP Project 20-07/Task 420, 2019.

[11] J. Kim, S. Park, and U. Lee, "Dashcam witness: Video sharing motives and privacy concerns across different nations," *IEEE Access*, vol. 8, pp. 110425–110437, 2020.

[12] Y. Li, M. Abdel-Aty, J. Yuan, Z. Cheng, and J. Lu, "Analyzing traffic violation behavior at urban intersections: A spatio–temporal kernel density estimation approach using automated enforcement system data," *Accident Anal. Prevention*, vol. 141, Jun. 2020, Art. no. 105509.

[13] X. Yan, C. Zheng, Z. Li, S. Wang, and S. Cui, "PointASN: Robust point clouds processing using nonlocal neural networks with adaptive sampling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5589–5598.

[14] Y. Li *et al.*, "Deep learning for LiDAR point clouds in autonomous driving: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3412–3432, Aug. 2021.

[15] G. Vosselman, B. G. H. Gorte, G. Sithole, and T. Rabbani, "Recognising structure in laser scanner point clouds," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 46, no. 8, pp. 33–38, 2004.

[16] S. Zhang *et al.*, "Automated visual recognizability evaluation of traffic sign based on 3D LiDAR point clouds," *Remote Sens.*, vol. 11, no. 12, p. 1453, 2019.

[17] J. Fereday and E. Muir-Cochrane, "Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development," *Int. J. Qualitative Methods*, vol. 5, no. 1, pp. 80–92, 2006.

[18] T. Dukic and T. Broberg, "Older drivers' visual search behaviour at intersections," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 15, no. 4, pp. 462–470, 2012.

[19] Z. Feng, N. Ji, Y. Luo, N. N. Sze, J. Tian, and C. Zhao, "Exploring the influencing factors of public traffic safety awareness in China," *Cogn., Technol. Work*, vol. 23, no. 4, pp. 731–742, 2021.

[20] B. Shen, W. Qu, Y. Ge, X. Sun, and K. Zhang, "The relationship between personalities and self-report positive driving behavior in a Chinese sample," *PLoS ONE*, vol. 13, no. 1, 2018, Art. no. e0190746.

[21] S. J. Kass, K. E. Beede, and S. J. Vodanovich, "Self-report measures of distractibility as correlates of simulated driving performance," *Accident Anal. Prevention*, vol. 42, no. 3, pp. 874–880, 2010.

[22] F. Alonso, C. Esteban, S. A. Useche, and E. L. de Cózar, "Prevalence of physical and mental fatigue symptoms on Spanish drivers and its incidence on driving safety," *Adv. Psychol. Neurosci.*, vol. 1, no. 2, pp. 10–18, 2016.

[23] E. Roidl, B. Frehse, and R. Höger, "Emotional states of drivers and the impact on speed, acceleration and traffic violations—A simulator study," *Accident Anal. Prevention*, vol. 70, pp. 282–292, Sep. 2014.

[24] J. B. Jacobs, *Drunk Driving*. Chicago, IL, USA: Univ. of Chicago Press, 2013.

[25] J. M. Walsh, J. J. Gier, A. S. Christopherson, and A. G. Verstraete, "Drugs and driving," *Traffic Injury Prevention*, vol. 5, no. 3, pp. 241–253, 2004.

[26] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Demmel, "Driving behaviour while self-regulating mobile phone interactions: A human-machine system approach," *Accident Anal. Prevention*, vol. 118, pp. 253–262, Sep. 2018.

[27] L. Precht, A. Keinath, and J. F. Krems, "Identifying effects of driving and secondary task demands, passenger presence, and driver characteristics on driving errors and traffic violations–using naturalistic driving data segments preceding both safety critical events and matched baselines," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 51, pp. 103–144, Nov. 2017.

[28] Q. Wang *et al.*, "Common traffic violations of bus drivers in urban China: An observational study," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0137954.

[29] J. Hong, J. Park, G. Lee, and D. Park, "Endogenous commercial driver's traffic violations and freight truck-involved crashes on mainlines of expressway," *Accident Anal. Prevention*, vol. 131, pp. 327–335, Oct. 2019.

[30] X.-D. Pan, Y. Lin, X.-B. Guo, and S.-E. Fang, "Research on traffic sign discernible visual range under backlighting condition," *J. Highway Transp. Res. Develop.*, vol. 5, no. 5, pp. 118–120, 2006.

[31] D. Shinar, "Aggressive driving: The contribution of the drivers and the situation," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 1, no. 2, pp. 137–160, 1998.

[32] T. Ben-Bassat and D. Shinar, "Ergonomic guidelines for traffic sign design increase sign comprehension," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 48, no. 1, pp. 182–195, 2006.

[33] *Global Status Report on Road Safety*, World Health Org., Geneva, Switzerland, 2015.

[34] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and T. Reddy, "Canintelliids: Detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1456–1466, 2021.

[35] X. Fan, F. Wang, D. Song, Y. Lu, and J. Liu, "Gazmon: Eye gazing enabled driving behavior monitoring and prediction," *IEEE Trans. Mobile Comput.*, vol. 20, no. 4, pp. 1420–1433, Apr. 2019.

[36] X. Lin and L. Cheng, "Engineering countermeasures to reducing unintentional yellow-light running owing to dilemma zone," *Proc.-Social Behav. Sci.*, vol. 96, pp. 900–904, Nov. 2013.

[37] M. Bramberger, J. Brunner, B. Rinner, and H. Schwabach, "Real-time video analysis on an embedded smart camera for traffic surveillance," in *Proc. 10th IEEE Real-Time Embedded Technol. Appl. Symp. (RTAS)*, May 2004, pp. 174–181.

[38] T. Lajunen, D. Parker, and H. Summala, "Does traffic congestion increase driver aggression?" *Transp. Res. F, Traffic Psychol. Behav.*, vol. 2, no. 4, pp. 225–236, Dec. 1999.

[39] P. J. Carlson, "Can traffic signs be too bright on low-volume roads?" *Transp. Res. Rec.*, vol. 2472, no. 1, pp. 101–108, 2015.

[40] K. Kersavage, S. I. Guler, and M. Pietrucha, "Analysis of colored variable message signs for visibility and comprehensibility," *Transp. Res. Rec.*, vol. 2674, no. 1, pp. 125–134, 2020.

[41] Y. Zhang and C. Wu, "Modeling the effects of warning lead time, warning reliability and warning style on human performance under connected vehicle settings," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 62, no. 1. Los Angeles, CA, USA: SAGE, 2018, p. 701.

[42] Y. Lin, P. Tang, W. Zhang, and Q. Yu, "Artificial neural network modelling of driver handling behaviour in a driver-vehicle-environment system," *Int. J. Vehicle Des.*, vol. 37, no. 1, pp. 24–45, 2005.

[43] H. S. Kim, Y. S. Hwang, D. S. Yoon, W. G. Choi, and C. H. Park, "Driver workload characteristics analysis using EEG data from an urban road," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1844–1849, Aug. 2014.

[44] Y. Xing, C. Lv, D. Cao, H. Wang, and Y. Zhao, "Driver workload estimation using a novel hybrid method of error reduction ratio causality and support vector machine," *Measurement*, vol. 114, pp. 390–397, Jan. 2018.

[45] Y. Xie, Y. L. Murphey, and D. S. Kochhar, "Personalized driver workload estimation using deep neural network learning from physiological and vehicle signals," *IEEE Trans. Intell. Vehicles*, vol. 5, no. 3, pp. 439–448, Sep. 2020.

[46] U. E. Manawadu, T. Kawano, S. Murata, M. Kamezaki, J. Muramatsu, and S. Sugano, "Multiclass classification of driver perceived workload using long short-term memory based recurrent neural network," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1–6.

[47] Y. Noh, S. Kim, Y. J. Jang, and Y. Yoon, "Modeling individual differences in driver workload inference using physiological data," *Int. J. Automot. Technol.*, vol. 22, no. 1, pp. 201–212, 2021.

[48] D. de Waard, A. Kruizinga, and K. A. Brookhuis, "The consequences of an increase in heavy goods vehicles for passenger car drivers' mental workload and behaviour: A simulator study," *Accident Anal. Prevention*, vol. 40, no. 2, pp. 818–828, 2008.

[49] T. Yared and P. Patterson, "The impact of navigation system display size and environmental illumination on young driver mental workload," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 74, pp. 330–344, Oct. 2020.

[50] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 8, no. 2, pp. 97–120, 2005.

[51] K. Figl and R. Laue, "Influence factors for local comprehensibility of process models," *Int. J. Hum.-Comput. Stud.*, vol. 82, pp. 96–110, Oct. 2015.

[52] H. Guan, J. Li, Y. Yu, C. Wang, M. Chapman, and B. Yang, "Using mobile laser scanning data for automated extraction of road markings," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 93–107, Jan. 2014.

[53] *Stay Connected With the Latest in Industrial AI, Smart Engineering & IoT*. Accessed: Oct. 6, 2021. [Online]. Available: https://tech27.com/resources/point-clouds/

[54] S. Gupta and B. Lohani, "Augmented reality system using LiDAR point cloud data for displaying dimensional information of objects on mobile phones," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 5, p. 153, 2014.

[55] T. S. Perry, "Look out for apple's AR glasses: With head-up displays, cameras, inertial sensors, and LiDAR on board, apple's augmented-reality glasses could redefine wearables," *IEEE Spectr.*, vol. 58, no. 1, pp. 26–54, Dec. 2021.

[56] E. Takeuchi, Y. Ninomiya, and S. Kato, "Lane visibility check methods based on high precision maps and 3D LiDAR for traffic prediction," in *Proc. 3rd Int. Symp. Future Act. Saf. Technol. Toward Zero Traffic Accidents (FAST-Zero)*, 2015, pp. 61–66.

[57] T. T. Vu, M. Matsuoka, and F. Yamazaki, "LIDAR-based change detection of buildings in dense urban areas," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARS)*, vol. 5, Sep. 2004, pp. 3413–3416.

[58] G. Markkula, R. Romano, A. H. Jamson, L. Pariota, A. Bean, and E. R. Boer, "Using driver control models to understand and evaluate behavioral validity of driving simulators," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 6, pp. 592–603, Dec. 2018.

[59] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 102–118.

[60] P. Konstantopoulos, P. Chapman, and D. Crundall, "Driver's visual attention as a function of driving experience and visibility. using a driving simulator to explore drivers' eye movements in day, night and rain driving," *Accident Anal. Prevention*, vol. 42, no. 3, pp. 827–834, 2010.

[61] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. 1st Annu. Conf. Robot Learn.*, 2017, pp. 1–16.

[62] L. Chen, Y. Papelis, G. Waston, and D. Solis, "NADS at the university of IOWA: A tool for driving safety research," in *Proc. 1st Hum.-Centered Transp. Simulation Conf.*, 2001, pp. 1–14.

[63] S. Baltodano, S. Sibi, N. Martelaro, N. Gowda, and W. Ju, "The RRADS platform: A real road autonomous driving simulator," in *Proc. 7th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, 2015, pp. 281–288.

[64] P. Wang, S. Sibi, B. Mok, and W. Ju, "Marionette: Enabling on-road wizard-of-oz autonomous driving studies," in *Proc. ACM/IEEE Int. Conf. Robot Interact.*, Mar. 2017, pp. 234–243.

[65] D. Goedicke, J. Li, V. Evers, and W. Ju, "Vr-oom: Virtual reality on-road driving simulation," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–11.

[66] M. Oeljeklaus, F. Hoffmann, and T. Bertram, "A fast multi-task CNN for spatial understanding of traffic scenes," in *Proc.21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2825–2830.

[67] J. Sweller, "Cognitive load theory," in *Psychology of Learning and Motivation*, vol. 55. Amsterdam, The Netherlands: Elsevier, 2011, pp. 37–76.

[68] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial–temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, 2007.

[69] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, 2008, no. 10, Oct. 2008, Art. no. P10008.

[70] Z. Jiang, Y. Liu, X. Fan, C. Wang, J. Li, and L. Chen, "Understanding urban structures and crowd dynamics leveraging large-scale vehicle mobility data," *Frontiers Comput. Sci.*, vol. 14, no. 5, pp. 1–12, 2020.

[71] Y. Wen *et al.*, "Mapping dynamic road emissions for a megacity by using open-access traffic congestion index data," *Appl. Energy*, vol. 260, Feb. 2020, Art. no. 114357.

[72] G. Charness, U. Gneezy, and M. A. Kuhn, "Experimental methods: Between-subject and within-subject design," *J. Econ. Behav. Org.*, vol. 81, no. 1, pp. 1–8, 2012.

[73] M. L. McHugh, "The chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, pp. 143–149, 2013.

[74] Z. Lai, Y. C. Hu, Y. Cui, L. Sun, N. Dai, and H.-S. Lee, "Furion: Engineering high-quality immersive virtual reality on today's mobile devices," *IEEE Trans. Mobile Comput.*, vol. 19, no. 7, pp. 1586–1602, Oct. 2019.

[75] V. Braun and V. Clarke, *Thematic Analysis*. Washington, DC, USA: American Psychological Association, 2012.

[76] B. Rosner, R. J. Glynn, and M.-L.-T. Lee, "The Wilcoxon signed rank test for paired comparisons of clustered data," *Biometrics*, vol. 62, no. 1, pp. 185–192, 2006.

[77] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[78] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[79] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[80] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2110–2118.

[81] S. Lee *et al.*, "VPGNet: Vanishing point guided network for lane and road marking detection and recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1947–1955.

[82] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1–6.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIANG *et al.*: UNDERSTANDING DRIVERS' VISUAL AND COMPREHENSION LOADS

15

[83] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[84] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.

[85] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[86] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. ICML*, 2011, pp. 1–8.

Zhihan Jiang received the B.E. and M.E. degrees in computer science and technology from Xiamen University, China, in 2018 and 2021, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong. Her research interests include mobile computing, urban computing, crowdsensing, and data analysis.

Xin He received the B.E. degree in computer science and technology from Fujian Normal University, China, in 2019, and the M.E. degree in computer science and technology from Xiamen University, China, in 2022.

Chenhui Lu received the B.E. degree in electronic information engineering from Fujian Normal University, China, in 2018, and the M.E. degree in computer technology from Xiamen University, China, in 2022.

Binbin Zhou received the B.Math. degree from Hangzhou Normal University in 2008, the M.Phil. degree in computer science from Hongkong Polytechnic University in 2011, and the Ph.D. degree in computer science from Zhejiang University in 2021. She is currently an Associate Research Fellow of Zhejiang University City College. Her research interests are mainly in urban computing, big data applications, and ubiquitous computing.

Xiaoliang Fan (Senior Member, IEEE) received the Ph.D. degree from University Pierre and Marie Curie, France, in 2012. He is currently a Senior Research Specialist at Xiamen University. He has published more than 70 journals and top conferences papers. His research interests include trustworthy AI and federated learning, spatio-temporal data mining, and services computing. His works are funded by NSFC and many industry collaborators. He is a Senior Member of CCF.

Cheng Wang (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2002. He is currently a Professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, China. His research interests include remote sensing image processing, mobile LiDAR data analysis, and multisensor fusion.

Xiaojuan Ma (Senior Member, IEEE) received the Ph.D. degree in computer science from Princeton University. She is currently an Associate Professor in human–computer interaction at the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. She was a Post-Doctoral Researcher at the Human–Computer Interaction Institute, Carnegie Mellon University.

Edith C.H. Ngai (Senior Member, IEEE) is currently an Associate Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong. Before joining HKU in 2020, she was an Associate Professor with the Department of Information Technology, Uppsala University, Sweden. Her research interests include the Internet of Things, machine learning, data analytics, and smart cities.

Longbiao Chen (Member, IEEE) received the joint Ph.D. degree in computer science from Zhejiang University, China, in 2016, and Sorbonne University, France, in 2018. He is currently an Associate Professor with the Department of Computer Science, Xiamen University, China. Before joining Xiamen University, he worked as a Research Assistant with the Institut Mines-Télécom, France. His research interests include ubiquitous computing, mobile crowdsensing, urban computing, and big data analytics. He is a Senior Member of CCF, a Technical Committee Member of ACM SIGSPATIAL China Chapter, and a CCF Ubiquitous Computing Committee.