RESEARCH ARTICLE

Network-Level Safety Metrics for Overall Traffic Safety Assessment: A Case Study

XIWEN CHEN1, (Graduate Student Member, IEEE), HAO WANG1, (Graduate Student Member, IEEE), ABOLFAZL RAZI1, (Senior Member, IEEE), BRENDAN RUSSO2, JASON PACHECO3, JOHN ROBERTS4, JEFFREY WISHART5, LARRY HEAD6, AND ALONSO GRANADOS BACA3

1School of Computing, Clemson University, Clemson, SC 29631, USA
2Department of Civil Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA
3Department of Computer Science, The University of Arizona, Tucson, AZ 85721, USA
4Arizona Department of Transportation, Phoenix, AZ 85009, USA
5Exponent Company, Phoenix, AZ 85721, USA
6Department of Systems and Industrial Engineering, The University of Arizona, Tucson, AZ 85027, USA

Corresponding author: Abolfazl Razi (arazi@clemson.edu)

This work was supported in part by the National Science Foundation under Grant 2204721, and in part by the Arizona Commerce Authority under Institute of Automated Mobility (IAM) Project.

ABSTRACT
Driving safety analysis has recently experienced unprecedented improvements thanks to technological advances in precise positioning sensors, artificial intelligence (AI)-based safety features, autonomous driving systems, connected vehicles, high-throughput computing, and edge computing servers. Particularly, deep learning (DL) methods empowered volume video processing to extract safety-related features from massive videos captured by roadside units (RSU). Safety metrics are commonly used measures to investigate crashes and near-conflict events. However, these metrics provide limited insight into the overall network-level traffic management. On the other hand, some safety assessment efforts are devoted to processing crash reports and identifying spatial and temporal patterns of crashes that correlate with road geometry, traffic volume, and weather conditions. This approach relies merely on crash reports and ignores the rich information of traffic videos that can help identify the role of safety violations in crashes. To bridge these two perspectives, we define a new set of network-level safety metrics (NSM) to assess the overall safety profile of traffic flow by processing imagery taken by RSU cameras. Our analysis suggests that NSMs show significant statistical associations with crash rates. This approach is different than simply generalizing the results of individual crash analyses, since all vehicles contribute to calculating NSMs, not only the ones involved in crash incidents. This perspective considers the traffic flow as a complex dynamic system where actions of some nodes can propagate through the network and influence the crash risk for other nodes. The analysis is carried out using six video cameras in the state of Arizona along with a 5-year crash report obtained from the Arizona Department of Transportation (ADOT). The results confirm that NSMs modulate the baseline crash probability. Therefore, online monitoring of NSMs can be used by traffic management teams and AI-based traffic monitoring systems for risk analysis and traffic control.

INDEX TERMS
Deep learning, driving safety analysis, safety metrics, autonomous vehicles.

I. INTRODUCTION

Vehicular technology has witnessed key milestones in recent years. Most cars are heavily equipped with advanced visual and radio sensors, cameras, control units, and artificial-intelligence (AI)-platforms that make driving safer and more convenient than ever. Electric vehicles (EVs) equipped with automated driving systems (ADS) have achieved higher levels of autonomy and continue to expand their territory in the global car market [1]. Crowd-souring and connected
Vehicle (CV) services have been utilized to improve the overall operation of the vehicular networks through data and model sharing. For instance, Uber Advanced Technologies Group has recently proposed a unified deep learning (DL) framework that assists automated vehicles (AVs) to map, perceive, predict, and plan sequentially to enhance driving safety [2]. Another example is Tesla which employs a cluster with 5,760 A100 GPUs to conveniently train their multi-modality neural network with a 1.5 petabytes dataset [3]. Intel’s mobile eye [4], [5], Google’s Waymo one [6], [7], and Nvidia’s Drive [8], [9] are other examples of using DL-based AI platforms for autonomous and safe driving application.

The use of AI platforms is not limited to car manufacturing. It indeed made revolutionary changes to traffic monitoring and control systems, and roadside infrastructures. Particularly, web-based high-performance computing (HPC), and vehicular edge computing (VEC) servers with graphics/tensor processing units (GPU/TPUs) have made volume data aggregation and processing, more feasible than ever [10].

Despite these technological advances, driving safety still remains one of the key challenges of today’s society. Statistics show that the mortality of motor-vehicle related injuries has been almost constant in years 2015 to 2019 in the US [11], while the car crash fatalities even have been increased since the COVID-19 pandemic started [12], [13]. This is a global issue, and about 1.3 million people die by car accidents worldwide, and millions are injured every year according to the world health organization (WHO) [14]. These statistics reveal that modern technology has not yet been fully utilized to prevent avoidable accident casualties and fatalities.

Driving safety is pursued from different perspectives by research communities, as shown in Fig. 1. New safety and warning systems are under design and development. Examples are applying eye-tracking technology to assess drivers’ distraction and fatigue [15], [16], and onboard collision avoidance systems to offer safety alerts and active brake upon detecting a danger [17], [18], [19]. Another research direction is exploring the casualty of incidents based on crash surrogate events (as the measures of accident proximity) from long-period naturalistic driving data [20], [21]. These works show that the expected number of crashes to occur during a specific time period is related to the number of observed surrogate events and crash-to-surrogate factors [22]. A triggered event often is determined by a set of key parameters known as surrogate safety metrics (SSM) designed for human-driven cars [23] as well as AVs equipped with autonomous driving system (ADS) [24], [25]. We also provide a comprehensive review of SSM in the appendix (attached as a supplementary material).

The community also has taken advantage of the recent developments in DL-based image/video processing that yield superior performance far beyond the conventional methods [26]. DL methods have also enabled developing well-annotated volume datasets (e.g., highD dataset 2018) [27] that in turn led to developing even more powerful video processing methods for autonomous and safe driving applications, such as vehicle detection [28], [29], plate recognition [30], [31], traffic sign classification [28], [29], lane detection [32], [33], and abnormal driving detection from surveillance video [34], [35].

Nowadays, safety-related events such as traffic congestion, red light violation, over speeding, unauthorized-vehicle stops on the highway shoulder, etc., can be detected, interpreted, and predicted by learning-based video analysis frameworks, such as generative adversarial network (GAN)-based architectures [34], query-based approaches [36], 3D convolutional Networks [37], and YOLO-family detectors [38].

To the best of our knowledge, most current methods favor the investigation of individual and independent crashes based on the extracted incident-level safety metrics and disjoint safety events. Therefore, they are not well-positioned to make relations between the crash distributions and the dynamics of the entire traffic flow.

Fig. 2 demonstrates a merge scenario, where individual safety metrics may not be capable of capturing overall safety risks, hence can yield misleading results. Gray rectangles in this figure represent the normal traffic flow of the highway, and vehicle v1 intends to join the traffic. Figs. 2(a) and 2(b) present an aggressive merge before and after the joining epochs, while Figs. 2(c) and 2(d) show a safe merge. Let’s investigate this scenario using the time-to-collision (TTC) metrics, which is one of the most commonly used safety metrics for safety analysis, especially rear-end crashes [39], [40]. Calculating TTC for the leading and following cars (v1 and v2) will favor the aggressive join, since a faster merge leaves more reaction time for the following car v2, and hence appears safer from v2’s perspective. However, it causes more risk to the following vehicle on the highway after joining the traffic (v3 in Fig. 2), hence disrupting the stability of
The importance of macro-scale analysis of crash data has been recognized by some researchers [41], [42], [43], [44], [45]. However, most of these studies investigate the geo-spatial distribution of crashes and their consequences with limited insight to finding diverse causes of crashes. More specifically, they emphasize geographical mapping of traffic properties (e.g., volume, density, congestion condition, etc.) as well as the road network topology. The high-resolution micro-level driving behaviors, such as multi-agent trajectory prediction [46], motion planning [47] have also been studied. These analyses predict crash probabilities and safety factors but they didn’t try verifying their results with readily available crash reports.

A brief review of these methods is provided in Section II. It is noteworthy that our analysis nicely integrates with these studies, since these methods find the baseline crash probability for each road section, whereas our method quantifies the modulated crash probability (the amount of increase/decrease) based on the temporal safety profiling of the traffic flow.

In this work, we offer a generalizable framework for finding meaningful correlations between representative traffic safety indicators and crash probability geo-distribution through integrative analysis of crash reports and traffic video captured by the roadside infrastructure. Our contribution is two-fold. First, we define a set of network-level safety metrics (NSM) from an external observer’s perspective that captures the inherent relations between local traffic flows and gauge the overall safety profile of the traffic network (Fig. 3(b)). These metrics are not claimed to be comprehensive, and are presented only to show the utility of such metrics for traffic analysis as a proof-of-concept, and can be extended to a more comprehensive list. Secondly, we provide a formal association analysis to assess the contribution of each safety metric in mediating the crash frequency. This is done by investigating the spatial and temporal correlation between the crash data points and the traffic disruption represented by the proposed safety metrics. The results of our framework can be used for developing online traffic advisory systems, crash explanation, risk prediction, and traffic optimization. It is noteworthy that crash severity analysis is another legitimate problem which is out of the scope of this paper. We have to acknowledge that our metrics may be not directly used to analyze crash severity since crash severity is more heterogeneous by nature and involves more confound factors beyond the developed NSMs.

The overall pipeline of the proposed integrative framework is presented in Fig. 4, which includes two parallel processing modules for crash analysis and video-based safety metric extraction followed by an association analysis step.

![FIGURE 2. Individual analysis fails to interpret a merge scenario. Vehicle v1 on the entrance ramp intends to join the highway traffic. The top row (a,b) shows unsafe (aggressive) join before and after the merge. This is considered favorable by the following car (v2), since it provides a higher TTC, while disrupting the overall highway traffic stability (c). The bottom row shows the safe join by v1 after yielding the traffic flow before (d) and after the merge (e). Although this merge provides a lower TTC for the following vehicle on the ramp (v2), it is advantageous from the traffic stability point of view (f). This can be reflected in the TTC of the car following v1 after joining the highway (v3).](image)
FIGURE 3. Different levels of traffic safety along with their regions of interest for (a) individual safety analysis from the vehicle’s perspective using standard safety metrics, (b) global safety analysis from an external observer’s perspective using the proposed network-level metrics, and (c) local analysis of traffic clusters.

FIGURE 4. Proposed network-level traffic safety analysis framework includes multiple modules: (i) video processing that includes video preparation (e.g., video stabilization, noise reduction, and personal information masking), trajectory extraction, geo-mapping with perspective translation, and network-level safety metric extraction, (ii) crash analysis which includes crash temporal and spatial mapping, crash distribution (histogram), and crash type mapping (representation), and (iii) temporal and spatial association analysis.

II. RELATED WORK ON NETWORK-LEVEL ANALYSIS

Traffic analysis can be performed at different levels. In micro-level analysis, typically an individual incident is analyzed by a set of fine-resolution parameters such as the involving vehicles’ geometry and motion dynamics as well as the local road and traffic flow properties, based on the sensor readings and captured imagery. A comprehensive review of micro-level analysis of crashes, DL-based driving behavior modeling, AI-platforms for driving safety enhancement have been reviewed in [48], [46], and [47].

In macro-level analysis, the traffic flow is considered a dynamic network with inherent relations among network nodes. This view is adopted for network structural analysis, traffic forecasting, abnormal pattern detection, and global traffic safety analysis. For instance, [49], [50], [51], [52] consider traffic as an application of complex network theory, where the network dynamics can be represented by the collection of small-world networks [53] and random scale-free networks [54]. A small-world network is defined as a network constructed with a high clustering coefficient with
small average geodesics, namely the pairwise shortest path lengths. In other words, most network nodes are expected to be accessible with a low number of hops, (e.g., logarithmic in the number of nodes). This model best fits a collection of lattice-shaped local neighborhood roads linked with a few highways.

A scale-free network is a network whose degree distribution obeys the Power-law, meaning that there exist only a few nodes with higher degrees (such as downtown or traffic hubs). Urban traffic can be modeled by such networks [49].

Some other works [51], [52], [55], [56], [57], [58], [59] analyze the network structure using complexity networks theory to evaluate, design and optimize traffic networks with sustainability and maintainability. Detecting network bottlenecks with poor connectivity and high congestion vulnerability is studied in [60], [61], [62], [63], [64], [65], and [66]. The authors of [67] provide a comprehensive review of continuum models that consider the traffic flow as a compressible fluid and employ some physical knowledge to explain the real-world phenomena in traffic. It is noteworthy that some recent works [68], [69], [70] address mixed traffic taking autonomous vehicles (AVs) into consideration in their analysis.

Traffic safety can be improved by more accurate traffic forecasting, namely by predicting the future properties of the traffic flow based on current/historical features. Traffic risk modeling and prediction can provide hints and guidelines on traffic management to minimize factors that elevate the crash risk. Specifically, [71], [72] transform the road network to 2D grids and then apply convolutional neural network (CNN) to predict the traffic crowd flow. The traffic flow is modeled as a diffusion process on a directed graph and deploys a diffusion convolutional recurrent neural network (DCRNN) to learn the spatio-temporal features of traffic based on the historical data and road structure [73]. A method called Deep Transport is proposed in [74] which combines CNN and recurrent neural network (RNN) architectures equipped with an attention mechanism to predict traffic volume. Recently, graph neural networks (GNNs), a class of DL networks performing inference over arbitrary graphs are proven to yield superior performance in predicting traffic-related parameters [75], [76], [77], [78], [79].

Crashes can also be viewed as severe disruptions in the network flow. Therefore, many papers have focused on abnormal network pattern detection, such as inferring abnormal patterns caused by unexpected events (e.g., natural disasters, serious car accidents, traffic control) [80], [81], [82]. For instance, interpreting crashes by analyzing the frequency of irregular patterns over traffic networks is considered in the following works. The role of road factors in characterizing the severity of incidents using logistic regression with chi-squared test is presented in [83]. Kernel density estimation (KDE) is used in [41] and al2021mapping to find the spatio-temporal patterns of traffic accidents, and rank them based on their statistical significance. Negative binomial and Poisson models are used in [43] to identify traffic factors that contribute to the crash frequency. Crash prediction based on random forest (RF), gradient boosting decision tree (GBDT), and Xgboost is adopted to find associations between the road links and incidents [45].

The above-mentioned works provide insightful results for different traffic problems by network-level analysis of traffic flow and crash statistics. However, they still suffer from a few drawbacks. Some works are devoted to explaining the traffic flow and crash distributions using deep learning models. Although successful from modeling and prediction perspectives, these methods lack the interpretability and generalizability features. Moreover, due to the difficulty of creating crash scenarios, the micro-level analysis is hardly verifiable, and most works settle with verifying the results with virtual simulators such as car learning to act (CARLA) [84] and simulation of urban mobility (SUMO) [85], ignoring the valuable information exploitable from a rich set of well-documented crash data. Furthermore, the utilized safety metrics are appropriate only for individual crash analysis, hence fail in modeling the inherent and complex relations among network nodes. In this work, we make a connection between the global analysis and deep analysis of individual incidents by introducing newly-defined network-level metrics. The integrative analysis of traffic video and crash data enables us to look for statistical relations between traffic properties and crash risks and drawing general conclusions on traffic safety enhancement. It is notable that there is a gap between the association analysis using data-driven methods and the real causality identification studied in [86].

Finally, we note that our method does not replace the micro-level and macro-level analyses, but complements them. It extends the notion of safety metrics to more insightful network-level metrics for global crash analysis while profiling the overall traffic flow safety that can fine-tune the baseline crash probability obtainable from macro-scale analysis of crash reports.

III. PROPOSED METRICS FOR NETWORK-LEVEL ANALYSIS

Conventional safety metrics are defined for scenarios where only two (or a few) vehicles are involved. A summary of the most commonly used safety metrics is provided in Appendix A, for the sake of completeness. In this appendix, we also introduce a taxonomy for safety metrics based on the level of access required to ADS data, when calculating these metrics for self-driving cars. In this paper, we introduce a set of NMS that can be used for the overall and long-term safety assessment of traffic flow. For instance, the composition of traffic (e.g., the ratio of trucks to all vehicles) can contribute to the frequency and severity of collisions. Likewise, the overall variations of car velocities on the road can reveal information about potential risk factors. A list of the proposed network-level safety metrics is provided in Table 2.

Time-to-Collision-Cluster Variation (TTC-CV) is defined to evaluate the relative velocity of car clusters that...
can pose safety risks. TTC is perhaps the most commonly used safety metric that evaluates the risk of rear-end crash by quantifying the time of the following car colliding with the leading car if they both retain their current speeds. We develop a new metric that extends this metric to car clusters since vehicle clustering naturally occurs on the road. We also conjecture that the instantaneous geo-distribution of the cars on the road can play essential roles on crash rates. For instance, car platooning is considered an important feature for autonomous and connected vehicles. For such scenarios, due to the coordination between platooning vehicles, inter-platoon crashes are rare and cluster-based TTCs can be useful.

Our approach is clustering vehicles based on a predefined threshold and assess the relative mobility of car clusters. More specifically, we use the down-sampled version of the traffic video, e.g., at rate 1 FPS) to cluster the cars based on their pairwise distances. If $S$ is the set of cars in the current video frame, then a cluster $C_i = \{n_i\} \subset S$ is defined so that for every $n_i \in C_i$, if $|C_i| > 1$, there exists at least one node $n_k \in C_i$, $j \neq k$ with $d(n_j, n_k) \leq d_c$ and likewise any car $n_l$ satisfying $d(n_k, n_l) \leq d_c$ for some $n_l \in C_i$ must be a member of $C_i$. Note that the converse is not true, and it is not necessary that any pair of the cluster members have a pairwise distance below the threshold. A special case is with only one member $|C_i| = 1$ when there is no other car within the certain distance (defined by $d_c$) of its only member vehicle $n_i$, meaning that for any $n_k \in S, j \neq k$, we have $d(n_j, n_k) > d_c$. Here, $d(n_j, n_k)$ is the Euclidean distance between vehicles $n_j$ and $n_k$, and $d_c$ is a predefined threshold, and, $|C|$ is the cardinality of set $C$ (i.e. the number of cars in cluster $C$). The clusters are non-overlapping and we have $C_i \cap C_j = {}$ for all $i \neq j$. Then, the cluster $C_i$ is considered as a virtual point object at the centroid of the cluster, i.e. $l(C_i) = \sum_{n \in C_i} l(n)/|C_i|$ with velocity $v(C_i) = \sum_{n \in C_i} v(n)/|C_i|$. Here, $l(n)$, and $v(n)$ represent the location and velocity of node $n$. The TTC of $C_i$ is calculated with respect to the potential collision point with the latest leading cluster $C_j$ moving at a lower speed $v(C_j) \leq v(C_i)$ as follows:

$$CTTC_i = \frac{l(C_i) - l(C_j)}{v(C_j) - v(C_i)},$$

For segments with crossing road segments (intersections and merging points), we use the stationary intersection point as the potential collision point when calculating cluster TTCs. Also, note that cluster TTCs are calculated using the original video with high FPS 30. The cluster TTC reduces to the regular inter-vehicle TTCs if the threshold $d_c$ is selected close to 0, so each car becomes a cluster. Next, the coefficient of variation of cluster-level TTCs for frame $j$, is calculated as:

$$TTC-CV(f_j) = \frac{\text{std}(CTTC)}{\text{mean}(CTTC)} \rho_v,$$

as an instantaneous network-level collision risk factor for the road segment covered by video frame $f_j$. Here, $N_j$ and $N_C$ are the numbers of vehicles and clusters in the frame, and $\rho_v = N_j/N_C$ is the average number of vehicles in each cluster used to emphasize higher risk for more crowded clusters. This metric is more robust against outliers and extremums (which often occur in computing TTC) compared to other statistics of cluster-level TTCs, including min$(CTTC)$, mean($CTTC$), or max$(CTTC)$.

**Individual Velocity Variation Rate (IVVR)** is defined as the variation of velocities for each vehicle in a specific zone or time interval. We define it as

$$IVVR = \frac{1}{N} \sum_{i=1}^{N} \frac{|v_i^\text{av} - v_i^\text{min}|}{v_i^\text{av}},$$

where $v_i^\text{max}$, $v_i^\text{min}$, and $v_i^\text{av}$ are the minimum, maximum, and average velocities of vehicle $i$. The higher values of this metric means that on average each vehicle changes its speed more often by accelerating and decelerating. This can be due to the road profile, density of intersections and exits, traffic volume, or road conditions.

**Overall Velocity Variation Rate (OVVR)** is defined as the variation of average velocities among vehicles in a specific zone or time interval. OVVR is defined as

$$OVVR = \frac{1}{N} \sum_{i=1}^{N} \frac{|v_i^\text{av} - v^\text{av}|}{v^\text{av}},$$

where $v_i^\text{av}$ is the average velocity of vehicle $i$, and $v^\text{av} = \sum_{i=1}^{N} v_i^\text{av}/N$ is the average velocity of all vehicles. Similar to IVVR, this metric can be associated with crash rate in specific highway sections.

**Over Speeding Rate (OSR)** is defined as the rate of over-speeding vehicles as follows:

$$OSR = \frac{1}{N} \sum_{i=1}^{N} I(v_i^\text{max} > v_L),$$

where $v_L$ is the speed limit, and $I(x > y) =$ the indicator function with $I(x > y) = 1$ for $x > y$ and $I(x > y) = 0$ otherwise. Speed limits are typically set based on a standardized set of national guidelines, taking into account road geometry (e.g., radii of curves, sight distance, weather conditions) and the location profile (e.g., residential versus rural areas). A high OSR, when associated with a high crash rate, may indicate the need for taking more warning and prevention measures to avoid over-speeding, noting that over-speeding can be an important contributor to crashes. On the other hand, high OSR, when it does not correlate with a high accident rate, can indicate that speed limits could be considered for potential increase without compromising traffic safety.

The aforementioned traffic metrics can be characterized by processing the roadside cameras or by crowd-sourcing and accumulating information provided by vehicles’ dash cameras.

**Traffic Composition Indicator (TCI)** is defined to gauge the diversity of vehicle types in specific road sections. For instance, it is known that a higher density of trucks on the roads can correlate with the frequency and severity of road
TABLE 2. Set of proposed network-level safety metrics.

| Metric | Definition | Features |
|--------|------------|----------|
| TTC-CV | $TTC-CV(f_j) = \frac{\sum_{i} f_{i}}{\sum_{j} \sum_{i} f_{i}}$ | Extend the conventional TTC to cluster level; Evaluate the global risk of the traffic flow at the network-level. |
| IVVR  | $IVVR = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{n=1}^{N} v_{n} - \frac{v}{2}}{v_{n}}$ | Reflects the instability of traffic flow by car speed variations and crash rate; Can partially offer network re-design suggestions. |
| OVVR  | $OVVR = \frac{1}{N} \sum_{i=1}^{N} I(v_{i} > \text{Threshold})$ | Overall speed variation of vehicles; Defined similar to IVVR for overall traffic speed variation; Not easily affected by outliers. |
| OSR   | $OSR_i = \frac{1}{N_i} \sum_{n=1}^{N_i} I(v_{n} < v_{L})$ | Indication of speed limit violations; Can offer suggestions for network design. |
| TCI   | $TCI = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{v_{i}}{v_{L}} \right) \left( 1 - \frac{v_{i}}{v_{L}} \right)$ | Flow composition indicator; Can be associated with elevated crash risks. |
| NTC   | $NTC = \frac{1}{N} \sum_{i=1}^{N} l_{i}$ | Consider vehicle shape; Reflects traffic density; Can be associated with elevated crash risks. |
| TRT   | $TRT = \frac{1}{N_e} \sum_{i=1}^{N_e} |t_{e}(i) - t_{r}(i)|$ | A simple and reasonable way to evaluate severity of the accidents; Can also take traffic stability into consideration. |

accidents; hence trucks are prohibited in some road sections in highly-populated areas [87]. This can be due to trucks’ larger deceleration inertia, more frequent brake failures, and lower maneuverability levels. In general, if vehicles classified into classes $c = 1, 2, 3, \ldots, C$, then the traffic composition can be defined as

$$TCI = \frac{\left( \sum_{c=1}^{C} N_c \right)^2}{C \sum_{c=1}^{C} N_c^2}$$

(6)

where $N_c$ is the number of vehicles in class $c$, and the metric is defined similarly to the Jain fairness index. This metric ranges from $1/C$ for the most unbalanced composition to 1 for an equal number of vehicles of each type. If one is interested in evaluating the rate of a specific class like trucks to all vehicles, the following metric can be used:

$$f_c = \frac{N_c}{\sum_{i=1}^{N} N_i}$$

(7)

When classifying cars into two classes $c = 1$ for trucks and $c = 2$ for non-trucks, these two metrics are related as

$$TCI = \frac{1}{2(1 - 2f_1f_2)}$$

(8)

Lower $TCI$ values mean that most cars are of the same type with lower risks. This metric can be estimated by processing roadside videos but requires high-complexity learning methods for vehicle detection and classification.

**Normalized Traffic Density (NTC)** is defined as the density of vehicles on road sections as

$$NTC = \frac{\sum_{i=1}^{N} l_{i}}{N_i \times L}$$

(9)

where $l_{i}$ is the length of vehicle $n_{i}$, $N$ is the number of all observed vehicles, $N_i$ is the number of lanes, and $L$ is the length of the road section. This parameter simply represents what portion of the road is occupied by the moving or standing vehicles. For an empty road, this factor is zero and approaches one in fully-packed heavy traffic behind intersections.

This parameter can be easily extracted from roadside videos by video processing and vehicle detection. Higher NTC values are expected to be associated with higher crash rates, and may raise the request for traffic load balance strategies.

**Traffic Recovery Time (TRT)** is defined as the time required for traffic flow recovery after incidents. The system of vehicular flow can be considered as a non-equilibrium system of interacting particles [88], and the instability of a flow-free state is induced by the collective effects of the increase of fluctuations.

It is known that any traffic event can suddenly lead to a jamming state, and TRT reflects the time span from an event epoch to the time point the status changes to free flow. This time is expected to be much shorter than the interval between the consecutive events for flow stability. TRT can be defined as:

$$TRT = \frac{1}{N_e} \sum_{i=1}^{N_e} |t_{e}(i) - t_{r}(i)|$$

(10)

where $N_e$ is the number of events in the monitoring interval, and $t_{e}(i)$ and $t_{r}(i)$, respectively, denote the event start epoch, and the flow recovery epoch for event $i$. This parameter can be easily obtained from roadside videos or by crowd-sourcing the position information obtained from dash cameras.

The summary of the proposed network-level safety metrics is presented in Table 2. We believe that further efforts are needed to develop a more complete list of network-level safety metrics.

**IV. METHODS**

**A. DATA ACQUISITION**

In this work, we use six camera feeds collected by the Arizona Department of Transportation (ADOT) roadside infrastructure. Each camera covers one segment of highway I-10 and records five 2-hour MP4 videos with resolution 1280 × 720 and FPS 30 (each video file is about 8 Gigabytes). Six
exemplary covered highway segments along with the approximate camera locations are shown in Fig. 5.

**B. VIDEO PREPROCESSING**

In order to calculate safety metrics, we need to extract vehicles’ motion trajectories from the video files. To this end, we integrate a tracking algorithm called DeepSORT [89] with the commonly used detector algorithm called you only look once (YOLO)-v5 [90] to obtain trajectories of labeled objects. DeepSORT is a real-time multi-object tracking algorithm based on Kalman filtering and Hungarian algorithm, which can consider both bounding box parameters and appearance simultaneously. We use YOLOv5 to obtain the bounding box information and fed them into the Deepsort algorithm to enable precise tracking and obtain fine trajectories. YOLOv5 is an advanced proposal-free detector, commonly used for object detection for its speed and accuracy. To achieve high performance, YOLOv5 employs a series of well-designed components to improve the accuracy and efficiency of the detection. The network processes the input images by mosaic data augmentation (when training) to improve the generalization of the algorithm. It uses adaptive image filling to accelerate the inference. It also applies adaptive anchor calculation to input so that it can automatically assign the size of initial anchors based on the size of the input. The backbone is essentially based on a cross-stage partial network (CSP) [91] and spatial pyramid pooling (SPP). The early version of YOLOv5 used a Focus module before CSP but then replaced by a $6 \times 6$ Conv layer. CSP decreases the computation and memory cost by suppressing the repeated gradient information, while SPP extracts the rich features in different granularities. Then these feature maps are fed into the neck network to be aggregated and fused. The neck network consists of a feature pyramid network (FPN) [92] and a path aggregation network (PAN) [93]. They allow the representation to flow top-down and bottom-up feature maps at different levels, respectively, which enhances the capacity of the network to learn abundant semantic features and localization features. Finally, the head performs the object prediction task, and outputs the coordinate of the bounding box, along with its class and confidence score. Non-max suppression (NMS) is used to post-process by filtering the redundant bounding boxes of the same object. In our framework, YOLOv5 is pre-trained on a large-scale image dataset named COCO [94]. Various vehicle types (car, bus, truck, etc.) are already included in the dataset, meaning that it is not required to annotate data further to train our framework. The centroid of detected bounding boxes is used as the position of the objects. This combination offers real-time tracking ($\sim 40$ FPS with a GeForce RTX 2070 GPU) with acceptable mostly tracked (MT) accuracy ($>90\%$). MT accuracy is defined as the ratio of MT objects over all objects. An object is considered MT [95] when it is successfully tracked for at least 70% of the time points during its real trajectory. This setup is sufficient to perform experiments on certain road segments. To the best of our knowledge, as the computer science and industry thrive, more lightening and accurate algorithms with hardware acceleration would arise that could allow widespread deployment of our framework on RSU of large-scale road networks. Another advantage of this approach is tracking objects even with long occlusion periods, a frequent issue in multi-vehicle tracking.

The extracted trajectories are in the pixel domain from the camera’s perspective, hence the exploited distances and velocities are not proportional to real values. In order to extract safety metrics from trajectories, we translate the position information ($u, v$) from the 2D pixel domain into 3D GPS positions ($x, y, z$) using perspective projection, as shown in Fig. 5 and Fig. 6. To this end, we solve the following projection equations for a set of key points with known GPS positions. Considering a flat surface with no elevation change, we can skip $z$ in our calculations.

$$
\begin{pmatrix}
    x \\
    y \\
    z
\end{pmatrix}
= 
\begin{pmatrix}
   a_{11} & a_{12} & a_{13} & a_{14} \\
   a_{21} & a_{22} & a_{23} & a_{24} \\
   a_{31} & a_{32} & a_{33} & a_{34} \\
   a_{41} & a_{42} & a_{43} & a_{44}
\end{pmatrix}
\begin{pmatrix}
   u \\
   v \\
   1
\end{pmatrix},
\quad (11)
$$
where, \((x/\lambda, y/\lambda, z/\lambda)\) denotes the GPS positions of the pixel after transformation from the pixel index values \((u, v)\) with \(\lambda\) being the scale factor. The optimal transformation coefficients \(a_{11}, \ldots, a_{43}\) are obtained by applying the least squares estimation (LSE) to the set of selected keypoint pairs with known GPS positions. Under linear transformation (no camera edge distortion) four keypoints are sufficient to recover the projection matrix, but a higher accuracy can be achieved using more key points.

The obtained trajectories represent noise-like fluctuations mainly due to the drift in the bounding boxes position around the object. We utilize a Savitzky–Golay (SG) filter [96] to smooth out the trajectories before performing the subsequent association analysis (Fig. 7).

**C. CRASH DATA**

The crash data is also provided by ADOT, which includes crash incident details in terms of date, time, location, collision type, etc. for 5-years, from 2015 to 2019. We extract the data for covered segments and use the most dominant crash types, including the rear-end, side-swipe, and all-types, since other types like head-on, angle, and rear-to-side crashes are sparse with not enough samples for association analysis. Figs. 8, and 9 represent the geo-distribution of crashes and the crash statistics of the six segments, respectively.

The temporal analysis aims to verify the utility of the proposed metrics in predicting crash count in each road segment for a given time interval. The spatial analysis investigates the generalization of the identified relations to other road segments with similar conditions. To this end, we split time into 1-hour intervals. For each interval, we calculate the average of safety metrics extracted from the traffic video for a specific road segment. Likewise, we take the average of crash counts for the same interval and road segment over the 5-year period. The validity of association analysis relies on two assumptions (i) different time intervals (e.g., [8:00 am-9:00 am] and [12:00 pm-1:00 pm]) are statistically different, and (ii) the crash counts over the same time intervals are statistically identical across different dates.

The results in Table 3 present the contingency Chi-square test, after applying Yates’s correction, to determine the statistical difference between the entire population and the sampled subset. For this test, we randomly sample the crash dataset and select 10% of the dataset, then compare the crash count per 1-hour time intervals between the entire and the sub-sampled dataset. Examples of this count for six 1-hour intervals and three crash types are presented in this Table. The achieved p-values are higher than 0.6, meaning that there is no significant difference between the subset and the entire dataset (usually, p-values larger than 0.05 are sufficient to accept the null hypothesis). Therefore, the 5-year average of crash reports can be used for temporal analysis with a 1-day traffic video. To avoid sampling bias, we repeat this test for 1,000 different sub-samples and present the average results in Table 4, which shows the same trend. We also performed similar tests over different weekdays, months, and years and obtained similar results suggesting that crash counts over time intervals are consistent among weekdays, months, and years. We observed the same consistency among different road segments as shown in Fig. 9.

To investigate the statistical difference between hours, we apply a one-way Chi-square test over different intervals averaged over the entire dataset. We repeat the test independently for all-types, rear-end, and sideswipe crashes. The P-values are extremely small for all cases suggesting that...
where \( x \) is the position, i.e., instantaneous velocity at time \( t \) of the corresponding bounding box at time point \( t \). More particularly, the position of vehicle are solely based on the position and velocity of the vehicle objects, are used to compute the proposed metrics.

metrics including IVVR, OVVR, and OSR metrics which are a scale factor into the desired metric unit and a reasonable level.

For some other metrics, such as TCI and NTC, we also avoid in this work to keep the computational complexity at a reasonable level.

For some other metrics, such as TCI and NTC, we also need the vehicle counts and types. In this work, we use only two classes of vehicles: small vehicles (e.g., cars, SUVs) and large commercial vehicles (e.g., trailer trucks, busses). We do not consider motorcycles here since they rarely appear in our videos. Since the dimensions of each object are already provided by the object detection stage (after the perspective translation and scaling), we use the object dimensions for object classification to incur minimal additional computational cost to the system. Moreover, we use two-class classification since it yields more robust results compared to mult-level classification. For example, a sedan may have 50% to be classified as an SUV but <5% to be classified as a bus by the detector. If the misclassification happens by the detector, it still has a big chance of being finally classified as a small vehicle. Noting the fact that the average length of personal cars and trucks is respectively about 4.5 m and 22 m [97], [98], the classification results are fairly accurate, except for the overlapping and temporarily occluded objects. For most of these objects taking the average over consecutive frames solves the transitional issues. There exist some prior work on fine-resolution vehicle classification into multiple subtypes (sedans, SUVs, truck, minivans, etc.) [99], [100], [101], [102], which we skip here in the advantage of low computational complexity for real-time monitoring systems. A summary of safety metrics is shown in Table 6.

E. ASSOCIATION ANALYSIS

To investigate the correlation between each metric and the crash rate, we use three correlation coefficients,

\[
\rho_{ij} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

where \( \rho_{ij} \) is the correlation coefficient between metrics \( i \) and \( j \), \( x_i \) and \( y_i \) are the values of metric \( i \) and \( j \) at time \( t_i \), and \( \bar{x} \) and \( \bar{y} \) are the means of \( x \) and \( y \) respectively.

D. SAFETY METRICS EXTRACTION

The utilized trajectory extraction algorithm provides vehicle IDs, and the objects’ locations (in terms of bounding box corner points per video frame), and the category of each object. We model vehicles as point objects located at the center of the bounding box. The extracted trajectories, after proper handling such as perspective transformation, denoising and smoothing, and filling the missing values by linear interpolation and excluding transitional and stationary non-vehicle objects, are used to compute the proposed metrics.

This information is sufficient to calculate most of the metrics including IVVR, OVVR, and OSR metrics which are solely based on the position and velocity of the vehicles. More particularly, the position of vehicle \( n_i \) at time \( t \) is \( x_i(t) = (x_{i1}(t) + x_{i2}(t))/2, y_i(t) = (y_{i1}(t) + y_{i2}(t))/2, \) where \( (x_{i1}(t), x_{i2}(t), y_{i1}(t), y_{i2}(t)) \) represent the corner points of the corresponding bounding box at time point \( t \). The instantaneous velocity at time \( t \) is simply the derivative of the position, i.e. \( v_x(t) = \sqrt{v_{x1}^2(t) + v_{x2}^2(t)}, \) \( v_y(t) = \alpha(x(t) - x(t - dt))/dt, \) \( v_y(t) = \alpha(y(t) - y(t - dt))/dt \) with \( \alpha \) being a scale factor into the desired metric unit and \( dt = 1/FPS \) is the time step. Higher order derivatives, and joint smoothing of the positions and velocities using methods such as Kalman filtering can also be used for higher accuracy, which we avoid in this work to keep the computational complexity at a reasonable level.

For some other metrics, such as TCI and NTC, we also need the vehicle counts and types. In this work, we use only two classes of vehicles: small vehicles (e.g., cars, SUVs) and large commercial vehicles (e.g., trailer trucks, busses). We do not consider motorcycles here since they rarely appear in our videos. Since the dimensions of each object are already provided by the object detection stage (after the perspective translation and scaling), we use the object dimensions for object classification to incur minimal additional computational cost to the system. Moreover, we use two-class classification since it yields more robust results compared to mult-level classification. For example, a sedan may have 50% to be classified as an SUV but <5% to be classified as a bus by the detector. If the misclassification happens by the detector, it still has a big chance of being finally classified as a small vehicle. Noting the fact that the average length of personal cars and trucks is respectively about 4.5 m and 22 m [97], [98], the classification results are fairly accurate, except for the overlapping and temporarily occluded objects. For most of these objects taking the average over consecutive frames solves the transitional issues. There exist some prior work on fine-resolution vehicle classification into multiple subtypes (sedans, SUVs, truck, minivans, etc.) [99], [100], [101], [102], which we skip here in the advantage of low computational complexity for real-time monitoring systems. A summary of safety metrics is shown in Table 6.

| Interval | Entire | Subset | Entire | Subset | Entire | Subset |
|----------|--------|--------|--------|--------|--------|--------|
| 0        | 23     | 2      | 7      | 0      | 3      | 1      |
| 1        | 16     | 0      | 1      | 0      | 5      | 0      |
| 2        | 18     | 1      | 6      | 1      | 2      | 0      |
| 10       | 130    | 9      | 86     | 6      | 36     | 3      |
| 11       | 168    | 16     | 92     | 8      | 62     | 6      |
| 12       | 311    | 19     | 225    | 16     | 66     | 3      |

Statistics 19.334 20.359 15.184
P-value 0.682 0.620 0.888

| All Type | Rear-end | Sideswipe |
|----------|----------|-----------|
| Statistics | 18.692 | 18.073 | 18.265 |
| P-value | 0.688 | 0.716 | 0.707 |
defined as
\[ r_p = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \] (Pearson),
\[ r_s = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}} \] (Spearman),
\[ \tau = \frac{2}{n(n-1)} \sum_{i<j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j) \] (Kendall),

(12)
where \( \bar{x}, \text{cov}(x), \sigma(x), R(x), \) and \( \text{sgn}(x) \), are the expected value, covariance, standard deviation, rank, and sign of \( x \), respectively. Pearson’s coefficient quantifies the strength of linear correlations and is most appropriate for normally distributed variables, whereas Spearman’s correlation is a rank-based method that does not assume linearity or normality of the variables. Kendall offers a rank correlation based on the concordance of the pair of observations, which is less informative but more robust than the other two methods.

We used correlation methods to find pairwise relations between the individual NSMs and crash rates. We can also use regression models to evaluate the collective prediction power of NSMs in predicting crash rate (not crash incidents).

It is noteworthy that crash counts can be used as approximate surrogate for risk probability, therefore the discrete values can be viewed as the quantized (and noisy) versions of the continuous-valued crash probabilities. This is more reasonable when crash counts are larger numbers (for 5-year crash data). Continuously-valued crash probabilities. This is more reasonable when crash counts are larger numbers (for 5-year crash data).

More specifically, we have \( y_i = x_i^T \beta + \epsilon_i \), where \( x_i = [x_{i1}, x_{i2}, \ldots, x_{iM}] \) is the set of \( M \) extracted safety metrics for a given time interval, and \( y_i \) is the crash count in the same interval. \( \beta = [\beta_1, \beta_2, \ldots, \beta_M] \) is the vector of model parameters, and \( \epsilon_i \sim N(0, \sigma^2) \) is the zero mean model noise with variance \( \sigma^2 \). This model is appropriate for continuous-valued outputs. To cover the count data, we also used Poisson regression model, where the crash counts \( y_i \) are Poisson distributed with mean \( \lambda_i \) linearly related to respective NMSs in the same interval, \( x_{i1}, x_{i2}, \ldots, x_{iM} \).

Note that the solution of linear regression model \( y_i = x_i^T \hat{\beta} + \hat{\epsilon}_i \) (or its compact format \( y = X \beta + \epsilon \)) using the ordinary least square (OLS) is \( \hat{\beta} = (X^T X)^{-1} X^T y \). From the frequentist’s perspective, the true \( \beta \) is deterministic but unknown. Based on these assumptions, \( \hat{\beta} \) is normally distributed as \( \mathcal{N}(\beta, \sigma^2) \) [103]:
\[ \hat{\beta} \sim \mathcal{N}(\beta, \sigma^2) \]
\[ \hat{\beta} \sim \mathcal{N}(\beta, (X^T X)^{-1} \epsilon^2) \]
This justifies the validity of using the subsequent statistical analysis for the relevance of normally distributed model parameters.

- **Adjusted R-squared**: This test is commonly used to validate the goodness of fit for linear regression models. R-squared statistics is defined as
\[ R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}, \quad 0 \leq R^2 \leq 1, \] (13)
where \( \hat{y}_i \) and \( \bar{y} \) are the estimated value and the average of the outputs \( y_i \). Obviously, more predictors can result in stronger models in the presence of sufficient data samples. In order to account for the number of predictors and the trivial gain for using more predictors, we use Adjusted R-squared which increases when the new predictor improves the model performance more than would be expected by chance. It is defined as:
\[ R^2_a = 1 - (1 - R^2) \frac{n-1}{n-p}, \] (14)
where \( n \) is the number of samples and \( p \) is the number of predictors.

- **F-test**: This test evaluates the significance of the model by evaluating all observed variables simultaneously. More specifically, we have the following null hypothesis and alternative hypothesis:
\[ H_0: \beta_1 = \ldots = \beta_k = 0, \] (15)
\[ H_a: \text{Not all } \beta_j \text{ are zero}. \]

The F-test statistic can be computed by:
\[ F = \frac{(\text{TSS} - \text{RSS})/(p-1)}{\text{RSS}/(n-p)} \sim F(p-1, n-p). \] (16)
If the achieved p-value is less than the given significance level \( \alpha \), we reject the hypothesis of all model parameters being zero (irrelevant).

- **Shapley Value**: This concept is borrowed from Game Theory and can be used to assess the obtained value of a player (predictor) \( x_i \) when combined with all other permutation of preceding players \( S \) in Coalition games [104]. In our case, NSM predictors are the players, and the value of coalition \( S \) is the prediction power of a linear model constructed using these features. More specifically, we have
\[ \phi(x_i) = \sum_{S \subseteq \mathcal{X} \setminus x_i} \frac{|S| - 1}{|\mathcal{X}|} \frac{S !}{\left| v(S \cup x_i) - v(S) \right|}, \] (17)
where \( \mathcal{X} = \{x_1, x_2, \ldots, x_M\} \) is the ordered set of NSMs, and \( v(S) \) is the accuracy of linear regression model made by features \( x_i \in S \). Here, \( \phi(x_i) \) quantifies the added accuracy in terms of adjusted R-squared score of the model made by a set of preceding features \( S \) after \( x_i \) joins the coalition, \( v(S \cup x_i) - v(S) \) averaged over all computations of \( S \) with preceding predictors.
TABLE 6. Summary of proposed safety metrics along with the utilized hyperparameters. *: Vehicle’s Length is calculated as 4,500mm for cars and 16,000mm for trucks.

| Metrics     | Data Requirement       | Hyperparameter       |
|-------------|------------------------|----------------------|
| TTC-CV      | Cluster Distance and Velocity | Cluster min distance \(d_C\) |
| IVVR        | Velocity, #of Vehicles | None                 |
| OVVR        | Velocity, #of Vehicles | None                 |
| OSR         | Velocity, #of Vehicles | Speed Limit          |
| TCI         | #of Vehicles in Each Class | None                 |
| NTC         | #of Vehicles          | Vehicle’s Length*    |
| TRT         | Timestamp             | None                 |

TABLE 7. Average performance of all-predictor model relating the crash rate to all predictors (i.e. safety metrics) by 5-fold cross-validation. Normalized MSE (N-MSE) is calculated as \(\frac{\sum (y_i - \hat{y}_i)^2}{(y_i + \hat{y}_i)^2/2}\).

|        | P-value | \(R^2\) | \(\text{adj. } R^2\) | N-MSE (Linear) | N-MSE (Poisson) |
|--------|---------|---------|-------------------|---------------|-----------------|
| All Type | 6.52E-05 | 0.578   | 0.519             | 0.120         | 0.122           |
| Rear-end | 1.77E-04 | 0.565   | 0.505             | 0.179         | 0.173           |
| Sideswipe | 2.76E-01 | 0.269   | 0.159             | 0.536         | 0.531           |

![FIGURE 10. True and estimated rear-end crash counts for two segments using the full-predictor model.](image)

V. RESULTS

The association analysis is performed for six cameras covering six disjoint road segments as shown in Fig. 5. Video is collected for 10 hours, 8:00 am - 6:00 pm, on different days. In this analysis, each data point is a pair \((x_i, y_i)\) with two components: \(x_i\) the average of extracted safety metrics during a 10-min interval, and \(y_i\) the crash count during the same interval averaged over 5 years of crash reports. Some data points are excluded due to camera issues (camera off, covered, mid-oriented).

A. CRASH RISK ASSESSMENT

We perform this analysis to show that monitoring RSU videos and extracting NSMs can be used for risk assessment by predicting expected crash counts. To this end, we build a full-predictor model which is a linear regression model to predict the crash count based on all NSMs using 5-fold cross validation during a given interval. The prediction results are presented in Fig. 10, Fig. 11, and Table 7.

Fig. 10 demonstrates a high alignment between the true and predicted rear-end crash counts using the full-predictor model. The same results are shown for all-type crash counts for all six segments as a scatter plot in Fig. 11, where most of the data samples concentrate around the unit-slope line (true value = predicted value). The results verify the usefulness of the proposed approach of using safety metrics to predict crash rates.

A more formal statistical analysis is provided in Table 7 using the metrics discussed in section III to validate the relevance of the constructed linear regression model. The results are based on the average of the 5-fold cross-validation (instead of the best results), which further confirms the validity of the developed models. It can be seen that the P-value is less than 0.05 for all crash types (using three different tests), which shows that the combination of all predictors (NSMs) makes a non-zero (significant) contribution to predicting crash rates, with any reasonable significance value. Similarly, the normalized MSE for both linear regression model as well as Poisson regression model are in an acceptable range for all-type and rear-end crashes. The results suggest that predicting crashes based on NSMs is more relevant.
TABLE 8. Absolute value of Pearson correlation, Spearman’s correlation, and Kendall correlation for the metrics across all segments.

|                | Proposed Metrics | Baseline |          |          |          |
|----------------|------------------|----------|----------|----------|----------|
|                | TTC-CV | IVVR | OVVR | OSR | TCI | NTC | Volume | E(TTC) |
| Pearson        |        |     |      |     |     |     |        |        |
| All Type       | 0.308  | 0.308 | 0.531 | 0.559 | 0.328 | 0.469 | 0.041  | 0.014  |
| Rear-end       | 0.310  | 0.305 | 0.521 | 0.555 | 0.315 | 0.462 | 0.015  | 0.004  |
| Sideswipe      | 0.090  | 0.213 | 0.335 | 0.314 | 0.228 | 0.270 | 0.142  | 0.040  |
| Spearman’s     |        |     |      |     |     |     |        |        |
| All Type       | 0.384  | 0.453 | 0.536 | 0.538 | -0.329 | 0.512 | 0.053  | 0.014  |
| Rear-end       | 0.373  | 0.429 | 0.515 | 0.535 | 0.313 | 0.503 | 0.060  | 0.004  |
| Sideswipe      | 0.104  | 0.330 | 0.350 | 0.283 | 0.223 | 0.247 | 0.071  | 0.040  |
| Kendall        |        |     |      |     |     |     |        |        |
| All Type       | 0.273  | 0.330 | 0.377 | 0.390 | -0.246 | 0.363 | 0.045  | 0.086  |
| Rear-end       | 0.265  | 0.305 | 0.358 | 0.386 | 0.235 | 0.354 | 0.050  | 0.082  |
| Sideswipe      | 0.091  | 0.255 | 0.267 | 0.212 | 0.165 | 0.188 | -0.044 | 0.009  |

FIGURE 12. Individual correlations between different metrics and all-types crash counts.

for rear-end crashes than the sideswipes. This is justifiable since most metrics (such as TTC, TTC-CV, IVVR, OVVR, OSR) consider longitudinal motions while sideswipe crashes highly depend on latitudinal motions. This reveals the need for developing a richer set of network-level safety metrics that are capable of predicting sideswipe metrics. Such metrics that indicate zigzag driving can be helpful. Also, sideswipe crashes seem to be more complicated in nature and potentially depend on other factors such as human mistakes.

B. TEMPORAL CORRELATION BETWEEN NMS AND CRASH COUNT

Now that the collective power of NSMs in predicting crash rates is established, we develop further analysis to estimate the relevance of each metric individually. To this end, we compute the individual correlation of each safety metric during a given interval with any of the crash types for each segment during the same time interval, using Pearson, Spearman, and Kendall correlation measures. To further highlight the importance of the achieved correlations, we compare the results against two reference values, (i) the correlation between the crash types and the traffic volume, and (ii) the correlation between the crash count and individual TTCs averaged over the entire traffic, E(TTC) discussed in Section I. The results are presented in Table 8 and Fig. 12, showing that the proposed metrics exhibit a high correlation (0.25 ~ 0.5) for rear-end and all-type crashes, which is much higher than the baseline correlation between the traffic volume and crash counts (0.01 ~ 0.08). Also, it is clearly seen that the achieved correlation for TTC-CV is in range (0.09 ~ 0.38), much higher than that of E(TTC) in range (0.004 ~ 0.08). This observation supports the idea of developing new network-level safety metrics.

C. SPATIAL ANALYSIS

So far, we showed the high correlation between the NSMs and crash rates of different types, by investigating each segment individually. To investigate the generalization of the proposed method, we perform cross-segment analysis. Specifically, we perform the same correlation analysis to all combinations of 2 to 5 segments, as shown in Table 9. The results are consistent with Table 8 that shows a reasonable consistency across segments.

To further investigate that the spatial generalizability of the results, we perform a cross-segment analysis. We construct a linear model (to predict crash counts based on NSMs) for five segments, and evaluate the model for the remaining segment (out-of-segment validation). Then, we repeat the test for all other combinations, so that each segment is tested once. The results are shown in Table 10.

D. COALITIONS OF PREDICTORS

It is known that the predictive power of each feature can depend on the presence of other features, due to inter-feature linear and non-linear correlations [104]. To account for this fact, we calculate the Shapley value for each metric as discussed in section IV-E. Here, \( v(S) \) is the value of a coalition \( S \) calculated as the adjusted R-squared statistics of a model built using predictors \( x_i \in S \). The results are presented in Table 11 that show these metrics play a relatively balanced role in the cooperative operation. Therefore, it is advantageous to use the full-predictor model based on all proposed NSMs.

Overall, we made the following observations, some of which require further investigations using more data samples to draw stronger conclusions. 1) Monitoring RSU traffic video and extracting network-level safety metrics can be used for...
TABLE 9. Average absolute Pearson correlation between NSMs and crash count across segments. The numbers in parentheses denote the number of combinations.

| Scenario       | Type     | TTC-CV | IVVR  | Metrics | OVVR  | OSR  | TCI  | NTC  |
|----------------|----------|--------|-------|---------|-------|------|------|------|
| Individual     | All-Type | 0.308  | 0.308 | 0.531   | 0.599 | 0.328| 0.469|      |
| segments       | Rear-end | 0.310  | 0.307 | 0.521   | 0.591 | 0.315| 0.462|      |
| (#6)           | Sideswipe| 0.090  | 0.211 | 0.335   | 0.329 | 0.228| 0.270|      |
| Combine 2      | All-Type | 0.297  | 0.380 | 0.550   | 0.575 | 0.441| 0.412|      |
| segments       | Rear-end | 0.286  | 0.384 | 0.536   | 0.568 | 0.441| 0.399|      |
| (#15)          | Sideswipe| 0.176  | 0.281 | 0.430   | 0.403 | 0.380| 0.304|      |
| Combine 3      | All-Type | 0.268  | 0.377 | 0.536   | 0.553 | 0.435| 0.350|      |
| segments       | Rear-end | 0.253  | 0.387 | 0.522   | 0.546 | 0.404| 0.336|      |
| (#20)          | Sideswipe| 0.186  | 0.276 | 0.445   | 0.415 | 0.403| 0.291|      |
| Combine 4      | All-Type | 0.238  | 0.384 | 0.529   | 0.540 | 0.421| 0.309|      |
| segments       | Rear-end | 0.222  | 0.395 | 0.515   | 0.534 | 0.389| 0.294|      |
| (#15)          | Sideswipe| 0.184  | 0.276 | 0.451   | 0.420 | 0.407| 0.279|      |
| Combine 5      | All-Type | 0.212  | 0.396 | 0.526   | 0.534 | 0.408| 0.281|      |
| segments       | Rear-end | 0.196  | 0.408 | 0.513   | 0.527 | 0.376| 0.265|      |
| (#6)           | Sideswipe| 0.181  | 0.280 | 0.455   | 0.422 | 0.409| 0.270|      |
| Combine all    | All-Type | 0.191  | 0.412 | 0.526   | 0.530 | 0.397| 0.262|      |
| segments       | Rear-end | 0.174  | 0.424 | 0.513   | 0.524 | 0.366| 0.246|      |
| (#1)           | Sideswipe| 0.178  | 0.284 | 0.458   | 0.424 | 0.409| 0.265|      |

TABLE 10. Cross-segment analysis for full model. The model is trained using 5 segments and tested with the unseen segment.

| F Value | R² | adj. R² | N-MSE |
|---------|----|--------|-------|
| All Type| 7.20E-31| 0.510 | 0.499 | 0.219 |
| Rear-end| 1.35E-28| 0.507 | 0.496 | 0.287 |
| Sideswipe| 3.51E-14| 0.295 | 0.278 | 0.622 |

TABLE 11. Shapley value of the proposed metrics for different types of crashes.

| TTC-CV | IVVR  | OVVR  | OSR  | TCI  | NTC  |
|---------|-------|-------|------|------|------|
| All Type| 2.49E-03| 1.97E-03| 2.55E-03| 4.86E-03| 2.74E-03| 2.57E-03|
| Rear-end| 2.49E-03| 1.95E-03| 2.38E-03| 4.69E-03| 2.85E-03| 2.59E-03|
| Sideswipe| 2.26E-04| 6.37E-04| 1.31E-03| 1.23E-03| 1.02E-03| 5.29E-04|

crash risk analysis; ii) The models build for dome road segments are generalizable to roads with similar geometry; iii) OSR exhibits a negative correlation with crash count, which is counter intuitive and can reflect the fact that crashes are more likely in busy hours than light traffics. This requires further investigation with larger datasets; iv) traffic composition represented by TCI shows that a more unbalanced traffic flow (exremely different number of small cars and trucks) is more prone to making crashes; v) cluster level analysis of TTC presents credibility in analyzing the traffic safety since it participates in most of the restricted models, vi) sideswipe crashes are harder to predict with metrics driven from longitudinal motions, and perhaps human factors or road geometry play more significant roles in modulating crash rates.

E. EXEMPLARY SIMULATION

Noting that most ADOT videos did not include real crashes, we use traffic simulators to investigate the associations between NSMs and elevated crash risk in near-reality conditions. In this study, we use Simulation of Urban MOBility (SUMO) platform to assess the impact of OVVR on crash probability in an on-ramp scenario, shown in Fig. 13(a). We define two traffic flows: i) vehicles in the main flow randomly start from one of the three lanes, and their initial speed distribution approximately obeys a normal distribution $\mathcal{N}(\mu_v, \sigma^2_v)$ (here we set mean as $\mu_v = 16$ m/s and standard deviation as $\sigma_v = \text{SpeedDev} \times \mu_v$ with $0 \leq \text{SpeedDev} \leq 1$, ii) vehicles in on-ramp flow has a fixed initial speed. All vehicles use Krauss model [105] to handle car-following. To generate crashes, we set $\text{jmIgnoreJunctionFoeProb} = 0.8$ of both flows, meaning that there is an 80% chance of vehicles ignoring vehicles of the other flow that have already entered the junction. To ensure reliability, we set 10 discrete values $\text{SpeedDev} = 0, 1, \ldots, 9$, and execute ten rounds of simulation for each value of $\text{SpeedDev}$.

The simulation will output the vehicles’ trajectories as well as the collision information. Fig. 13(b) presents a strong correlation between the OVVR and the number of collisions, as expected.

VI. CONCLUSION

In this paper, we offered a set of network-level safety metrics to assess the overall safety characteristics of traffic flow in a given driving zone. This concept extends the popular notion of safety metrics to network-level analysis. We showed that the proposed safety metrics are highly correlated with crash frequency (temporally and spatially). We conducted a case study in the state of Arizona by integrative analysis of collected video files from the I-10 highway and 5-year crash reports that verify the association between the network-level safety metrics and crash frequency in the same time intervals. More specifically, metrics that gauge the overall speed variation of vehicles, the traffic composition and diversity of vehicles, the density of traffic volume, and also relative mobility of car clusters are highly correlated with the crash frequency (with p-values much lower than 5% for most scenarios). We also observed that rear-end crashes are easier to predict than side-swipe crashes, perhaps due to the stronger role of road geometry and human mistakes in side-swipe accidents. Also, it shows the need for developing safety metrics that mimic latitudinal motions in addition to longitudinal motions.
The practical use of this analysis is identifying risk factors by constant monitoring of traffic flow using AI-based roadside infrastructures to broadcast warning messages and take more efficient traffic control decisions. Also, traffic control teams can take redesign and long-term decisions to keep the safety metrics in an acceptable range to enhance the overall driving safety on the road. Developing lightweight deep learning models to process traffic video and extract safety metrics in a real-time fashion can pave the road for developing online risk assessment systems.

**APPENDIX. TAXONOMY FOR OPERATIONAL SAFETY METRICS**

An essential objective of operational safety analysis from a scenario perspective, as opposed to the network-level perspective that is the focus of this paper, is to implement *operational safety assessment (OSA) metrics*, which are quantifiable measures extracted from traffic videos (or other data sources). These OSA metrics allow for a determination of the operational safety of a vehicle (AV or human-driven) to be assessed as a given scenario is navigated. Here, we review key OSA metrics that have been broadly used for operational safety analysis. It is notable that many algorithms developed for AVs utilize safety metrics for safe navigation decisions and avoiding crashes; however, we consider both self-driving and human-driven vehicles.

In this Appendix, for the sake of completeness, we review key OSA metrics that have been broadly used for operational safety analysis along with the level of access to ADS data to extract these metrics for AVs. A recent paper [23] summarizes safety metrics according to different basis (temporal, distance, and deceleration):

i. **Temporal-based indicators**: Time to Collision (TTC), Extended Time to Collision (Time Exposed Time-to-Collision(TET), Time Integrated Time-to-Collision(TIT) [106]), Modified TTC (MTTC), Crash Index (CI), Time-to-Accident (TA), Time Headway (THW), and Post-Encroachment Time (PET).

ii. **Distance-based indicators**: Potential Index for Collision with Urgent Deceleration (PICUD), Proportion of stopping Distance (PSD), Margin to Collision (MTC), Difference of Space Distance and Stopping Distance (DSS), Time Integrated DSS (TIDSS), and Unsafe Density (UD);

iii. **Deceleration-based indicators**: Deceleration Rate to Avoid a Crash (DRAC), Crash Potential Index (CPI), and Criticality Index Function (CIF).

A more recent paper [24] by the metric team of the Institute of Automated Mobility (IAM), provides a comprehensive set of OSA metrics following an extensive literature review. The objective was to develop a set of metrics for both human-driven and AVs that includes existing, adapted, and novel metrics. In a follow-up paper [107], the IAM proposed a taxonomy for operational safety metrics that is explored and expanded upon here. The IAM work is also a component of an Recommended Practice standards being developed by the SAE Verification and Validation (V&V) Task Force under the On-Road Automated Driving (ORAD) Committee [108]. The taxonomy will be introduced first, and then selected OSA metrics will be discussed (including the metrics from [23] listed above).

The expanded taxonomy introduced here is shown in Figure 14. The highest taxonomic rank in the proposed taxonomy hierarchy consists of three types that are based, essentially, on the data source, which includes the level of access required of ADS data. This access to proprietary data could be challenging, depending on the implementation of the OSA metric; it should be noted that “lighter” metrics require more cooperation with the AV developer. The three types are (example metrics are given for each, and are described in more detail later in the section):

1. **Black Box Metric**: A metric that allows measurement of data that can be obtained without requiring any access.
to ADS data. This could be from an on-board or off-board source. ADS data may enhance the accuracy and precision of the measurement(s). EXAMPLE: Collision incident (CI).

2) **Grey Box Metric**: A metric that allows measurement of data that can only be obtained with limited access to ADS data. EXAMPLE: ADS DDT Execution (ADE).

3) **White Box Metric**: A metric that allows measurement of data that can only be obtained with significant access to ADS data. EXAMPLE: Perception Precision (PP).

The second rank in the taxonomic hierarchy is the classification rank, which consists of the following (again, an example metric that is described later in the section is included):

1) **Safety Envelope Metric**: A metric that allows for measurement of the subject vehicle’s maintenance of a safe boundary around itself. This includes situations that may not be within the subject vehicle’s control. EXAMPLE: Minimum Safety Envelope (MSE).

2) **Behavioral Metric**: A metric that allows for measurement of an improper behavior of the subject vehicle. EXAMPLE: Aggressive Driving (AD).

3) **Component Metric**: A metric that allows for measurement of the proper function of ADS components. EXAMPLE: Event Data Recorder Compliance (EDRC).

4) **Sensing Metric**: A metric that allows for measurement of the ability of the ADS sensors to receive adequate information from the AV environment. EXAMPLE: Camera Resolution (CR).

5) **Perception Metric**: A metric that allows for measurement of the ability of the ADS to interpret information about the AV environment obtained by the ADS sensors. EXAMPLE: Human Traffic Control Detection Error Rate (HTCDER).

6) **Planning Metric**: A metric that allows for measurement of the ability of the ADS to plan an appropriate route through the AV environment. EXAMPLE: Object Avoidance Plan Error (OAPE).

7) **Control Metric**: A metric that allows for measurement of the ability of the AV to execute the planned route devised by the ADS. EXAMPLE: Actuation Error (AE).

It is important to note that not all classification rank types link to data source rank types in the current version. For example, the Component Metric is not linked to the Black Box Metric because data at the component level is deemed to be more access to (potentially proprietary) data than is allowed for the Black Box Metric.

The third rank in the taxonomy hierarchy is the Leading/Lagging rank, which relates (in binary fashion) to either the prediction (i.e., before) of a potential operational safety outcome or report (i.e., after) of an operational safety outcome after it has occurred. Operational safety outcomes include conflicts, collisions, an ADS disengagement, or a violation of a traffic law:

1) **Leading**: An OSA identification of a potential future operational safety outcome. EXAMPLE: Any safety envelope metric.

2) **Lagging**: An OSA identification of a report of an operational safety outcome. EXAMPLE: Any traffic law violation metric.

The metrics currently being considered for inclusion in the SAE J3237 Recommended Practice are included in Table 12 (although the list may change)). The IAM has focused on Black Box Metrics and Grey Box Metrics as part of the comprehensive set introduced in [23].

We also mention the approach an infrastructure-based (i.e., off-board the vehicle) observer system takes to monitor video and extract OSA metrics. The observer system also called system in the rest of this appendix for convenience, is a terrestrial or aerial monitoring system that collects traffic video for processing from an external observer’s point of view. A list of these metrics, along with their brief descriptions, is presented in Table 13.

Our main references include [23], [24], and [109].

### A. SUMMARY OF BASIC OPERATIONAL SAFETY METRICS

**Maximum Speed (MaxS)**, when associated with a collision, denotes the maximum speed of the involved vehicles before the crash starts until the full stop (e.g., between time points $t_1$ and $t_4$ in Fig. 15). MaxS is a simple but effective measure directly related to the severity of the collision.

**Differential Speed ($\Delta s$)** is defined as the relative speed between the involved vehicles. It occurs at $t_2$ in Fig. 15. The system needs to calculate the speed of the involved vehicles (using methods like DL-based object tracking with or without explicitly extracting the motion trajectories) to determine MaxS and Deltas.

**Time to Collision (TTC)** is a commonly used surrogate measure to define the time of an upcoming rear-end collision between two vehicles, if they continue their current speed ($t_2$-$t_3$ in Fig. 15). The system needs to observe the relative position and calculate the relative velocity of the two involved vehicles. TTC is computed as:

$$TTC = \frac{X_L - X_F}{v_F - v_L},$$  \hfill (18)

where $X_i$ denotes the position, $v_i$ denotes the velocity, and the indexes $L$ and $F$, respectively, denote the leading and following vehicles. The collision occurs only if $v_F \geq v_L$. Usually, we have $v_L \geq 0$ assuming that the cars move in the same direction. $v_L \leq 0$ represents a vehicle approaching from the front. The severity of an encounter can be determined by

---

1 This SAE standards document is currently in development and the operational safety metrics list could be modified, expanded, or contracted. It should be noted that the objective for the SAE J3237 document is to identify the minimum number of safety envelope metrics that will allow for an unsafe situation to be identified. Ideally, a single safety envelope metric would suffice; however, two or more safety envelope metrics may be necessary. The IAM is currently conducting research into this question of which safety envelope metrics to select, and the result may be adopted in the SAE J3237 Recommended Practice document.


| Metric Name                                      | Data Source Taxonomy | Classification Taxonomy |
|-------------------------------------------------|----------------------|-------------------------|
| Minimum Safe Envelope (MSE)                     | Black                | Safety Envelope         |
| Proper Response (PR)                            | Black                | Safety Envelope         |
| Minimum Safe Distance Factor (MSDF)             | Black                | Safety Envelope         |
| Time-to-Collision (TTC)                         | Black                | Safety Envelope         |
| Modified Time-to-Collision (MTTC)               | Black                | Safety Envelope         |
| Time Headway (TH)                               | Black                | Safety Envelope         |
| Instantaneous Safety Metric (ISM)               | Black                | Safety Envelope         |
| Collision Avoidance Capability (CAC)            | Black                | Safety Envelope         |
| Minimum Safe Distance Infringement (MSDI)       | Black                | Safety Envelope         |
| Post-Encroachment Time (PET)                    | Black                | Safety Envelope         |
| Not-at-Fault Collision Incident (NAFCl)         | Black                | Safety Envelope         |
| Rear-End Collision Incident (RECI)              | Black                | Safety Envelope         |
| Deceleration Rate to Avoid the Crash (DRAC)     | Black                | Safety Envelope         |
| Difference of Space Distance and Stopping Distance (DSS) | Black      | Safety Envelope         |
| Maximum Speed (MaxS)                            | Black                | Behavioral              |
| Delta Velocity (ΔV)                             | Black                | Behavioral              |
| Traffic Law Violation (TLV)                     | Black                | Behavioral              |
| Evasive Action (EA)                             | Black                | Behavioral              |
| Aggressive Driving (AD)                         | Black                | Behavioral              |
| Collision Incident (CI)                         | Black                | Behavioral              |
| ADS DDT Execution (ADE)                         | Grey                 | Control                 |
| Achieved Behavioral Competency (ABC)            | Grey                 | Planning                |
| Human Traffic Control Perception Error Rate (HTCDER) | Grey             | Perception               |
| Human Traffic Control Violation Rate (HTCVR)    | Grey                 | Behavioral              |
| Minimum Safe Distance Calculation Error (MSDCE)  | Grey                 | Perception               |
| Safety-Related Component Failure Perception (SRCFD) | Grey            | Component               |
| Event Data Recorder Compliance (EDRC)           | Grey                 | Component               |
| Perception Precision (PP)                       | White                | Perception               |
| Perception Rate (PDR)                           | White                | Perception               |
| Perception Weighted Harmonic Mean (PWHM)        | White                | Perception               |
| Perception False Positive Rate (PFPR)           | White                | Perception               |
| Perception False Negative Rate (PFNR)           | White                | Perception               |
| Anomaly Perception Behavior (ADB)               | White                | Perception               |
| Multiple Object Perception Error Rate (MODER)    | White                | Perception               |
| Localization Error (LE)                         | White                | Perception               |
| Multiple Object Tracking Precision (MOTP)       | White                | Perception               |
| Data Conflict Perception Rate (DCDR)            | White                | Perception               |
| Intersection over Union (IoU)                   | White                | Perception               |
| Actuation Error (AE)                            | White                | Control                  |
| Software Execution Error (SEE)                  | White                | Component                |
| Object Avoidance Plan Error (OAPE)              | White                | Planning                 |

the minimum TTC (TTC_{min}). Although very useful in investigating crashes, this metric has some limitations. First, TTC assumes that involved vehicles are moving at constant speeds, which ignores the potential dangers caused by acceleration or deceleration. Secondly, it assumes that the cars drive in the same direction and not appropriate for side crashes.

**Post-Encroachment Time (PET)** is defined as the time span between the encroached vehicle leaving and the other vehicle with the right-of-way arriving at the conflict point. The system needs to observe the relative time of the involved vehicles and estimate the conflict point. PET in Fig.16 is computed as:

\[
PET = t_b - t_a, \tag{19}
\]

where \( t_a \) and \( t_b \) represent the arrival time of the two involved vehicles. PET can be easily captured and computed by the
TABLE 13. Summary of operational safety metrics along with their key properties. Noting that '*' denotes the metrics that are not selected by IAM but are basic metrics employed in other papers.

| Metric              | Definition                                                                 | Features                                                                 |
|---------------------|---------------------------------------------------------------------------|--------------------------------------------------------------------------|
| MaxS*               | \( \max v \)                                                             | A simple but effective measure of the severity of a collision.           |
| \( \Delta v^* \)    | \( \Delta v \)                                                            | An effective indicator of the severity of a collision.                   |
| TTC                 | \( TTC = \frac{X_L - X_F}{V_F} \)                                        | Provides more information than PET; assumes that involved vehicles are at a constant speed; cannot reflect the severity. |
| PET                 | \( PET = t_2 - t_1 \)                                                    | Can be easily captured and computed; suitable for assessing intersecting conflicts; only suitable for the cases of transversal trajectories; cannot reflect the severity. |
| DE           | \( DE = \frac{\sum_{i=1}^{n} P_i DRAC_i}{\sum_{i=1}^{n} MADR_i} \)      |                                                                                                                                   |
| CPI*               | \( CPI = \frac{\sum_{i=1}^{n} P_i DRAC_i}{\sum_{i=1}^{n} MADR_i} \)      | Considers more factors such as traffic and road conditions compared to DRAC;                                                |
| PSD*               | \( PSD = \frac{\sum_{i=1}^{n} P_i DRAC_i}{\sum_{i=1}^{n} MADR_i} \)      | Works for a conflict with a single vehicle involved; rarely used for specific safety problems.                                |
| UD                  | \( UD = \frac{\sum_{i=1}^{n} P_i DRAC_i}{\sum_{i=1}^{n} MADR_i} \)       |                                                                                                                                   |
| MSE                 | Indicate the subject vehicle violates safety boundary of another.         | Provides how to calculate the quantified risky distances; basis of many assessment methods.                                 |
| PR                  | An action to recover when \( d_{min} \) and the safe range of \( d_{saf} \) are violated | Adds more information beyond MSDV.                                         |
| MSDF                | \( MSDF = \frac{d_{min} - d_{saf}}{d_{saf}} \)                           | Can indicate the degree of defensive driving style; higher \( MSDF \) may not mean higher safety.                           |
| MSDCE               | \( MSDCE = \sqrt{\frac{d_{min}^2}{d_{saf}^2} + \frac{d_{saf}^2}{d_{min}^2}} \) | Indicate the ADS ability to determine the safety distances.                 |
| CI                  | Indicate the subject vehicle is in a collision.                           | Capture instances of collisions as a metric; severity of the collision can be defined by KABCO scale [110]: K(Fatal Injury), A(Incapacitating Injury), B(Non-Incapacitating Injury), C(Possible Injury), O(No Injury). |
| TLV                 | Indicate the subject vehicle violates traffic law                        | Emphasizes that an ADS vehicle must follow existing laws; exceptions are made, for example, when road closures require temporary violations of driving exclusively inside a traffic lane. |
| ABC                 | Indicate the subject vehicle can execute a specific behavior correctly.    | Indicate the safety of ADS; included in the preliminary list of metrics.                                                   |
| ADSA                | Indicate the ADS is active when executing behaviors.                     | Indicate the safety of ADS; included in the preliminary list of metrics; has some dispute in California [111].               |
| ITCDER              | \( ITCDER = \frac{CDF - CPI}{CDF} \)                                     | One of ITC measurements; manually instructions by officers should still be considered.                                       |
| ITCVR               | \( ITCVR = \frac{CDF - CPI}{CDF} \)                                     | One of ITC measurements; manually instructions by officers should still be considered.                                       |
| AD                  | Indicate the maneuvers (longitudinal/lateral accelerations) of a subject vehicle exceed specified thresholds. | The thresholds vary by jurisdiction and culture; AD is only involved the subject vehicle; implies the inherent and potential risks of natural driving behaviors; should be included when evaluating ADS. |

system. However, it is suitable for intersection conflicts with transversal trajectories, but it does not fully reflect the severity of the crash.

**Initial Deceleration Rate (Initial DR)** quantifies the avoidance behavior taken by a vehicle to avoid a collision. More specifically, Initial DR (the second derivative of Curve B at time point \( t_2 \) in Fig.15) is defined as the deceleration rate \( a_0 \) at the beginning of the decelerating state. Deceleration rate is an appreciated variable to assess the potential severity of a conflict. Other variants of this metric family include Deceleration Rate to Avoid a Crash (DRAC), Crash Potential Index (CPI), and Criticality Index Function (CIF).

**Deceleration Rate to Avoid a Crash (DRAC)** is defined as the minimum deceleration rate of the following vehicle to avoid a crash to the leading vehicle. To characterize this metric, the observer system should keep track of the relative positions and the relative velocities of the two vehicles. Note that DRAC fails to evaluate lateral movements and is applicable only to scenarios where both cars are on the same lane. Mathematically, DRAC is defined as

\[
DRAC = \frac{(V_F - V_L)^2}{X_L - X_F}.
\] (20)

**Crash Potential Index (CPI)** is defined as the probability of DRAC exceeding the Maximum Available Deceleration Rate (MADR) at any moment. MADR depends on the type of the vehicle as well as the environmental conditions. The system needs to calculate MADR of the target vehicles and instantaneous track time-variant deceleration to obtain the
Fig. 16. Schematic diagram of PET.

If MSD is considered under maximum deceleration, then PSD becomes:

$$PSD = \frac{v^2}{2MADR}.$$  \hspace{1cm} (23)

PSD is an easily observable indicator and is one of the few metrics that can be used for a conflict involving a single vehicle (e.g., crashing into a stationary obstacle like a tree). PSD is defined for evasive actions and usable only for specific safety problems. If PSD is large, one can say the situation is safe, but a small PSD doesn’t necessarily indicate a high crash risk [112].

Unsafe Density (UD) is defined as the severity of the potential crash when the leading vehicle is within the achievable maximum DR. This metric is introduced in [113]. When multiple cars involved, this metric considers all cars in a link, while the similar metric of UnSafety (UNS) considers only two of the cars from the link to calculate the $UN_S$ of car $v$ at time step $s$, denoted by $UN_{Sv, s}$.

According to [113], the severity of a read-end crash is proportional to $\Delta v$ and $v_F$. More specifically, UNS is defined as:

$$UNS = \Delta v \cdot v_F \cdot R_d,$$ \hspace{1cm} (24)

where $R_d$ denotes the ratio between the deceleration of the leading vehicle and its maximum deceleration capacity, namely

$$R_d = \begin{cases} 
\frac{b}{b_{\text{max}}} & b < 0 \\
0 & \text{else} 
\end{cases}.$$ \hspace{1cm} (25)

Then, UD can be written as:

$$UD = \frac{\sum_{S=1}^{S_t} \sum_{V=1}^{V_t} UNS_{V, S} \cdot d}{TL},$$ \hspace{1cm} (26)

where $S_t$ denotes the number of simulation steps within the aggregation period, $V_t$ denotes the number of vehicles in the link, $d$ denotes the time span between the two simulation steps, $T$ denotes the aggregation period, and $L$ denotes the section length for which the metrics are evaluated.

UD provides more accurate information than typical micro-simulation outputs and allows comparison between the simulation scenarios or links. In a transportation network, each intersection is considered as a node, and the traffic
flow between the nodes is represented by a link. The two limitations of the UD parameter include (i) the difficulty of quantifying its constituent parameters and (ii) its applicability only to identical trajectories, namely when two cars move in the same direction, and the type of the potential crash is rear-end collision.

B. OPERATIONAL SAFETY METRIC FOR ADS PROPOSED BY THE IAM

Here, we review part of the recently proposed metrics by the IAM [24]. These metrics are developed by a team whose leader is a co-author of this paper.

Minimum Safe Envelope (MSE) is defined to indicate the minimum longitudinal and lateral distances that the subject vehicle should maintain from other safety-relevant entity (often is another vehicle) for safety purposes. When the subject vehicle (subscript 1) is following behind another entity (subscript 2) and both of them are moving in the same direction. The longitudinal boundary can be defined as:

\[ d_{\text{long, same}}^{\text{long}} = v_1 \rho_1 + \frac{1}{2} a_{1,\text{max}, \text{acc}} \rho_1^2 + \frac{(v_1 + a_{1,\text{max}, \text{acc}} \rho_1)^2}{2 a_{2,\text{min}, \text{dec}}} - \frac{v_2 \rho_2}{2 a_{2,\text{max}, \text{dec}}}, \]

where \( v_{1/2}^{\text{long/lat}} \) denotes the current longitudinal/lateral velocity of vehicle \( i \), \( a_{i,\text{max/\text{min}, \text{acc/dec}}} \) denotes the longitudinal/lateral maximum/minimum acceleration/deceleration of vehicle \( i \), and \( \rho_i \) denotes the response time of vehicle \( i \). When two involved vehicles are moving in opposite directions towards each other, the longitudinal boundary can be defined as:

\[ d_{\text{long, opp}}^{\text{long}} = v_1 \rho_1 + \frac{1}{2} a_{1,\text{max}, \text{acc}} \rho_1^2 + \frac{(v_1 + a_{1,\text{max}, \text{acc}} \rho_1)^2}{2 a_{2,\text{min}, \text{dec}}} + \frac{|v_2| \rho_2 + a_{2,\text{max}, \text{acc}} \rho_2^2}{2 a_{2,\text{min}, \text{dec}}} - \frac{(v_2 + a_{2,\text{max}, \text{acc}} \rho_2)^2}{2 a_{2,\text{max}, \text{dec}}}, \]

The lateral boundary can be defined as:

\[ d_{\text{lat}}^{\text{lat}} = \mu + v_1 \rho_1 + \frac{1}{2} a_{1,\text{max}, \text{acc}} \rho_1^2 + \frac{(v_1 + a_{1,\text{max}, \text{acc}} \rho_1)^2}{2 a_{2,\text{min}, \text{dec}}} - \frac{|v_2| \rho_2 - a_{2,\text{max}, \text{acc}} \rho_2^2}{2 a_{2,\text{min}, \text{dec}}} - \frac{(v_2 - a_{2,\text{max}, \text{acc}} \rho_2)^2}{2 a_{2,\text{min}, \text{dec}}}, \]

where \( \mu \) is the lateral fluctuation margin. Noting that if the calculated result of \( d_{\text{min, same}}^{\text{long}} \) or \( d_{\text{lat}}^{\text{lat}} \) is negative, it should be rounded up to 0. While the two boundaries are violated by the subject vehicle, MSDV is active (\( MSDV = 1 \)), meaning that an avoidable accident may occur. This metric characterizes the quantified risky distances as the basis of many assessment methods.

Proper Response (PR) is defined to indicate a proper action taken by the subject vehicle to recover itself when the MSE’s safety boundaries (\( d_{\text{min, same}}^{\text{long}} \) and \( d_{\text{lat}}^{\text{lat}} \) and the safe range of \( d_{\text{long, dec}}^{\text{long}}, d_{\text{lat, dec}}^{\text{lat}} \)) are violated. It adds more information beyond MSE to evaluate the behavior of the subject vehicle.

Minimum Safe Distance Factor (MSDF) is defined as the ratio between the current distance[s] to the calculated safe boundaries from the surrounding entity. It can be found as:

\[ MSDF^{\text{lat}} = \frac{d_{\text{lat}}}{d_{\text{lat, min}}} \quad \text{and} \quad MSDF^{\text{long}} = \frac{d_{\text{long}}}{d_{\text{long, min}}} \]

where \( d_{\text{lat}} \) and \( d_{\text{long}} \) denote the measured distances. MSDF \( \geq 1 \) indicates the degree of defensive driving style. Note that a higher MSDF may not mean a higher safety, necessarily.

Minimum Safe Distance Calculation Error (MSDCE) is defined as the difference between the calculated results by ADS from the ground truth. It can be formulated as:

\[ MSDCE = \sqrt{\frac{d_{\text{long, dec}}^{\text{long}} - d_{\text{long, min}}^{\text{long}}}{d_{\text{long, dec}}^{\text{long}}} + \frac{d_{\text{lat, dec}}^{\text{lat}} - d_{\text{lat, min}}^{\text{lat}}}{d_{\text{lat, dec}}^{\text{lat}}}} \]

This metric is derived from the MSDV, which indicates the ADS ability to determine the safety distances.

Collision Incident (CI) is defined to indicate the subject vehicle is involved in a collision determined by the reasonably related data. CI is active when \( d_{\text{lat}} \) and \( d_{\text{long}} \) are equal to 0. Also, the severity of the collision can be defined by KABCO scale [110] with K(Fatal Injury), A(Incapacitating Injury), B (Non-Incapacitating Injury), C(Possible Injury), O(No Injury).

Traffic Law Violation (TLV) is defined to indicate the subject vehicle that violates a traffic law which would result in an infraction or citation. Violating these laws sets TLV active. This metric emphasizes that an ADS-vehicle must follow existing laws, and an active TLV may lead to severe safety issues. It should be noted that exceptions to a TLV are made, for example, when road closures require temporary violations of driving exclusively inside a traffic lane.

Achieved Behavioral Competency (ABC) is defined to indicate the subject vehicle can execute a specific behavior correctly. This metric indicates the safety of ADS and is included in the preliminary list of ADS vehicles’ metrics.

ADS Active (ADSA) is defined to indicate that the ADS is active when executing behaviors. This metric represents the safety of ADS and is included in the preliminary list of ADS vehicles’ metrics. The metric has some dispute in California [111].

Human Traffic Control Detection Error Rate (HTCDER) is defined as the capability to detect instructions
from a Human Traffic Control (HTC) actor correctly. It is calculated as:

\[ HTCDER = \frac{GTI - CDI}{CDI} \]  

(32)

where \( GTI \) is the number of ground truth instructions and \( CDI \) is the number of correctly detected instructions. This metric is one of the HTC measurements to evaluate the safety of the subject vehicle. Note that the manual instructions by officers should still be considered in this metric.

**Human Traffic Control Violation Rate (HTCVR)** is defined as the capability of a subject vehicle to follow the received instructions successfully. It is formulated as:

\[ HTCVR = \frac{CDI - CCI}{CDI} \]  

(33)

where \( CCI \) is the number of correctly compiled instructions.

**Aggressive Driving (AD)** is defined to indicate the maneuvers (longitudinal/lateral accelerations) of a subject vehicle exceeding specified thresholds. When exceeding \( (a \geq a_y) \), the metric is set active. The thresholds vary by jurisdiction and culture. This metric involves only the subject vehicle and implies the inherent and potential risks of natural driving behaviors. When evaluating the safety of ADS, this metric should be included.

**DECLARATION OF COMPETING INTEREST**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**ACKNOWLEDGMENT**

The authors would like to thank the NSF and the Institute of Automated Mobility (IAM) for supporting this work. They would also thank the Arizona Department of Transportation (ADOT) for sharing roadside infrastructure, crash reports, and other resources with them during the performance of this project. Also, the opinions, findings, and conclusions expressed in this manuscript are those of the author’s and not necessarily those of the IAM and ADOT.

**REFERENCES**

[1] IEA. (Apr. 2021). Global EV Outlook 2021—Analysis. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2021

[2] S. Casas, A. Sadat, and R. Urtasun, “MP3: A unified model to map, perceive, predict and plan,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jan. 2021, pp. 14403–14412.

[3] Workshop on Autonomous Driving. Accessed: Jan. 30, 2022. [Online]. Available: http://cvsr2021.wad.vision/

[4] (2022). Creating the Autonomous Future Takes Experience and Vision. Accessed: Feb. 1, 2022. [Online]. Available: https://www.mobilitye.com/

[5] (Jan 2020). CES 2020: Engines Powering L2+ To L4 (Mobileye). Accessed: Feb. 1, 2022. [Online]. Available: https://s21.qrcdn.com/60069269/files/doc_presentations/2020/1/Mobileye-CES-2020-presentation.pdf

[6] (2022). We’re Building the World’s Most Experienced DriverLin. [Online]. Available: https://waymo.com/

[7] Q. Xu, Y. Zhou, W. Wang, C. R. Qi, and D. Anguelov, “SPG: Unsupervised domain adaptation for 3D object detection via semantic point generation,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., Oct. 2021, pp. 15446–15456.

[8] Omniverse Platform for Virtual Collaboration. Accessed: Feb. 1, 2022. [Online]. Available: https://www.nvidia.com/en-us/omniverse/
D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.

A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

M. Batt, "Building a science of cities," *Cities*, vol. 29, pp. S9–S16, Mar. 2012.

M. Batt, *The New Science of Cities*, Cambridge, MA, USA: MIT Press, 2013.

R. Ding, N. Ujang, H. B. H. Hamid, M. S. A. Manan, Y. H. Ye, R. Li, and J. Wu, "Detecting the urban traffic network structure dynamics through the growth and analysis of multi-layer networks," *Phys. A, Stat. Mech. Appl.*, vol. 503, pp. 800–817, Aug. 2018.

S. Wang, L. Zheng, and D. Yu, "The improved degree of urban road traffic network: A case study of xiamen, China," *Phys. A, Stat. Mech. Appl.*, vol. 469, pp. 256–264, Mar. 2017.

Z. Ruan, C. Song, X.-H. Yang, G. Shen, and Z. Liu, "Empirical analysis of urban road traffic network: A case study in Hangzhou city, China," *Phys. A, Stat. Mech. Appl.*, vol. 527, Aug. 2019, Art. no. 121287.
ABOLFAZL RAZI (Senior Member, IEEE) received the B.S. degree in electrical engineering from Sharif University, in 1998, the M.S. degree in electrical engineering from Tehran Polytechnic, in 2001, and the Ph.D. degree in electrical engineering from The University of Maine, in 2013. He was a Postdoctoral Researcher of machine learning and predictive modeling at Duke University, Durham, NC, USA, from 2013 to 2014, and Case Western Reserve University, Cleveland, OH, USA, from 2014 to 2015. He is currently an Associate Professor of computer science with Clemson University, Clemson, SC, USA. His current research interests include smart connected vehicles, driving safety analysis, wireless networking, the Internet of Things, and predictive modeling. He was a recipient of several competitive awards, including the Best Research of Wireless Networking, the Internet of Things, and Predictive Modeling at Duke University.

BRENDAN RUSSO received the B.S. and M.S. degrees in civil engineering from Wayne State University, Detroit, MI, USA, and the Ph.D. degree in civil engineering with a focus in transportation from Iowa State University. He is currently an Associate Professor of civil engineering with Northern Arizona University, Flagstaff, AZ, USA, where he also serves as the Associate Director of the Arizona Laboratory for Applied Transportation Research (AZTrans). He is a Registered Professional Engineer with Arizona. He has more than 12 years of transportation-related research, teaching, and professional work experience. His research interest includes traffic safety. He is a member of the Institute of Transportation Engineers (ITE), the Transportation Research Board (TRB) ACS10 Transportation Safety Management Systems Committee (also serving as Committee Research Coordinator), and the TRB AKD20 Roadside Safety Design Committee.

JASON PACHECO received the Ph.D. degree in computer science from Brown University, Providence, RI, USA, with Erik Sudderth. He was a Postdoctoral Researcher at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), with John Fisher III. He is currently an Assistant Professor with the Department of Computer Science, The University of Arizona, Tucson, AZ, USA. His research interests include computer vision, signal processing, computational biology, and computational neuroscience. One of his research, statistical machine learning, aims to build intelligent systems for automated reasoning in stochastic systems.

JOHN ROBERTS received the B.Sc. degree in electrical engineering and the M.Sc. degree in information management. He has worked at the Arizona Department of Transportation (ADOT) for more than 13 years in IT and systems management. Before coming to ADOT, he held various engineering and management positions in the semiconductor and defense industries. He is currently a Professor of systems and industrial engineering with the Department of Transportation Systems Management and Operations (TSMO) a division of ADOT.

JEFFREY WISHART received the B.Sc. degree in engineering physics (minor in mechanical engineering) from The University of British Columbia, Vancouver, BC, Canada, the M.Sc. degree in engineering physics from the University of Saskatchewan, and the Ph.D. degree in mechanical engineering from the Institute for Integrated Systems, University of Victoria. He is currently a Managing Engineer with the Test and Engineering Center, Exponent Company, and an Adjunct Professor of automotive systems with the Ira A. Fulton Schools of Engineering, Arizona State University. His research interests include energy and advanced transportation, including advanced powertrains, connected and automated vehicles, electric vehicle supply equipment, energy storage systems, and micro-mobility applications. He is the Chair of the Verification and Validation (V&V) Task Force under the On-Road Automated Driving (ORAD) SAE Committee that is establishing standards for automated vehicles.

LARRY HEAD received the Ph.D. degree in systems and industrial engineering from The University of Arizona, Tucson, AZ, USA, in 1989. He is currently a Professor of systems and industrial engineering and the Director of the Craig M. Berge Engineering Design Program, The University of Arizona. He has more than 30 years of research and development experience in adaptive traffic signal control, signal priority, traffic management, and connected and autonomous vehicle systems.

ALONSO GRANADOS BACA received the bachelor’s degree in computer science from The University of Texas, El Paso, TX, USA. He is currently pursuing the Ph.D. degree with the Department of Computer Science, The University of Arizona, Tucson, AZ, USA. He works at the intersection of image processing and traffic modeling, with Dr. Jason Pacheco.

* * *