

Driving Risk Recognition from Dashcam Video via License-Plate Bounding-Box Scale Variation

Zhaohui Jing

Department of Information and Intelligent Engineering, Shanghai Publishing and Printing College, Shanghai 200093, China

Email: jzhui@sina.com

How to cite this paper: Jing, Z. (2026). Driving Risk Recognition from Dashcam Video via License-Plate Bounding-Box Scale Variation. Journal of Computer Science and Frontier Technologies, 4(3), 1–13. ISSN Print: 3104-4204, ISSN Online: 3104-4212.

<https://doi.org/10.63313/JCSFT.9103>

Published: 2026-10-08

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

Dashcams are widely deployed, yet fleet management and insurance risk control still depend on manual frame-by-frame replay for batch retrospective review, while retrofitted recorders offer neither calibration conditions nor real-time links. This paper proposes a calibration-free method that recognizes driving risk from recorded dashcam video through license-plate bounding-box scale variation. Frames are sampled at 5 fps and processed by HyperLPR3 detection and recognition with a confidence threshold of 0.65. Taking the statutory GA 36 small-vehicle plate width of 440 mm as a prior, distance is estimated as $D=f \times 0.44/w$, and the scale change rate drives four types of risk rules gated by phase correlation-based stationary suppression. Per-character confidence-weighted voting, three-tier confidence grading, and a manual review loop secure plate identities, and cross-video same-vehicle merging operates within a $-30\text{ s} \sim +150\text{ s}$ temporal window. On 35 midday real-world videos totaling 2098.2 s, the pipeline screened out 41 high-risk events (21 lateral cut-in, 18 rapid approach, 1 too-close following, 1 adjacent-vehicle proximity) with a median minimum distance of 5.21 m; 165 trajectories were merged into 118 vehicles, and 38 of 86 medium- and low-confidence plates were corrected in review. The results suggest that the pipeline supports batch, calibration-free, relative-tier screening. The detected events are heuristic screening warnings, provided solely as a reference for post-event retrospective review, and do not constitute a basis for accident liability determination.

Keywords

Dashcam; License Plate Recognition; Monocular Distance Estimation; Scale Change Rate; Driving Risk Screening; Trajectory Merging

1. Introduction

1.1. Research Background

Fleet management and insurance risk control increasingly require batch retrospective screening of dashcam video, but manual frame-by-frame replay is costly. Factory-installed forward collision warning (FCW) and automatic emergency

braking (AEB) reduce rear-end crashes by approximately 49% according to data from 13 U.S. states [1], yet both depend on factory-installed real-time links, and retrofitted recorders can only be replayed after the event. We therefore target "post-event screening plus a manual review loop" under calibration-free conditions. Existing risk recognition studies mostly follow the SHRP2 video-plus-telematics paradigm [2] and do not exploit plate scale as a quantifiable physical cue.

1.2. Research Questions and Contributions

Three research questions are formulated. RQ1: under calibration-free conditions without distance ground truth, can distance estimation based on the 0.44 m plate-width prior and the focal length recovered from the nominal FOV support relative-tier risk screening? RQ2: can four heuristic rules built on license-plate bounding-box scale and its change rate screen out risk-representative events? RQ3: under real-world image quality, how can the reliability of plate recognition and vehicle counting be ensured?

The contributions are fourfold. (1) A "calibration-free distance estimation $\rightarrow$ scale change rate $\rightarrow$ four-type rule tiering" pipeline for Chinese GA36 plates with the 440 mm width as prior [3]; it requires neither calibration nor character segmentation, unlike the closest T-MDE series (preprints, controlled-condition results), which relies on U.S. FHWA-font character segmentation and three-point weak calibration [4][5]. (2) A reliability mechanism combining per-character confidence-weighted voting, three-level confidence tiering, and a manual review loop: among 165 trajectories, 79/52/34 are of high/medium/low confidence, and 38 of the 86 reviewed plates were corrected (44.2%; definition in §4.2). (3) Cross-video same-vehicle merging that combines plate symbol-level identity with a $-30\text{ s} \sim +150\text{ s}$ temporal proximity window, reducing 165 trajectories to 118 vehicles (18 across videos); unlike the appearance-free spatio-temporal linking of StrongSORT [6], our merging exploits plate identity. (4) Empirical evaluation on 35 midday fleet videos (2098.2 s): 41 high-risk events were screened out, with lateral cut-in and rapid approach together accounting for approximately 95.1% (39/41) and a median $d_{\min}$ of 5.21 m. All proportions are computed from counts.

1.3. Scope and Organization

This paper is positioned as a post-event offline batch screening tool, not a real-time or factory-installed application; the experiments are limited to midday scenarios. Section 2 reviews related work; Sections 3–5 present the methodology, experiments, and limitations; Section 6 concludes with point-by-point answers to RQ1–RQ3.

2. Related Work

2.1. License Plate Recognition and Multi-Frame Voting

Per-character voting over frames raised the recognition rate from 85.45% to 93.53%

[7], and temporal voting outperforms single-frame recognition by 15.5 percentage points [8]. The CCPD baseline [9] reaches an AP of only 43.42%, mainly due to blur and small scale; reliability scoring for degraded plates supports confidence gating [10]. GA36-2018 [3] specifies a plate width of 440 mm (480 mm for new-energy-vehicle (NEV) plates), the plate-width prior used for ranging.

2.2. Monocular Distance Estimation and Scale Change

Lee's tau theory [11] approximates time-to-collision (TTC) by the reciprocal of the retinal expansion rate; Stein et al. [12] derived accuracy bounds for scale-based methods; Kilicarslan and Zheng [13] warn directly via "TTC = size ÷ size change rate," the closest purely scale-based precedent. T-MDE Enhanced (preprint, controlled-condition results) [4] reaches an MAE of 2.3% at 10 m using U.S. plate character height; the initial version (preprint) [5] warns that perspective distortion of the whole-plate box can inflate scale by over 50%. Yu and Zhang's vanishing-point weak calibration [14] reports single-digit-percentage errors within a pixel-resolution-constrained range.

2.3. Multi-Object Tracking

ByteTrack [15] reduces trajectory fragmentation through two-stage matching of high- and low-score detections, an idea echoed by our retention of low-confidence plate detections; StrongSORT [6] performs appearance-free fragment linking and interpolation via AFLink and GSI; our cross-video merging instead relies on plate identity.

2.4. Risk Metrics and Warning Standards

Time headway (THW) and TTC are independent and complementary [16]; ISO 15623 [17] defines TTC, cut-in, and two-level warning; GB/T 33577-2017 [18] specifies warning TTC thresholds of 2.1/2.4/2.0 s. In the Shanghai naturalistic driving study [19], nearly half of 5608 cut-ins occurred without turn-signal activation. FCW alone reduces rear-end crashes by 16–19% and FCW+AEB by approximately 49% [1], so warning alone has a limited benefit ceiling. Plate anonymization is mandated by the Several Provisions on Automobile Data Security Management [20]. The risk events in this paper are heuristic screening warnings for post-event retrospective review, not a basis for accident liability determination.

2.5. Positioning Comparison

Table 1. Positioning comparison with the closest works

Dimension	T-MDE Enhanced [4]	Kilicarslan & Zheng [13]	This paper
Baseline object	U.S. plate character height (FHWA font)	Image motion/scale	Chinese GA36 whole-plate width of 440 mm
Calibration	Three-point weak	Calibration-free	Calibration-free (FOV +

	calibration		plate-width prior)
Character segmentation	Required	Not required	Not required
Risk output	TTC two-level warning	Approach alarm	Four-type rule tiering (high/medium)
Data validation	Controlled scenarios	Driving videos	35 fleet midday trips (approx. 35 min)
Identity management	None	None	Clustering + voting + merging + review
Review status	Preprint	Peer-reviewed	This paper

Against the T-MDE series [4][5], this work removes character segmentation and calibration via the GA36 plate-width and FOV priors and is validated on 35 real-world fleet trips; against Kilcarslan and Zheng [13], it adds meter-scale tiered rules with identity attribution—an adaptation, not a replacement.

3. Methodology

3.1. Pipeline Overview

Figure 1 shows the offline post-event screening pipeline: frame sampling at 5 fps, license-plate detection and recognition with confidence gating, trajectory clustering, per-character confidence-weighted voting, plate-width prior ranging and scale change rate estimation, four risk-rule decisions with stationary suppression, and cross-video same-vehicle merging, ending in three Excel sheets for manual review. The pipeline is calibration-free and relies on only two priors: the nominal field of view and the statutory plate width.

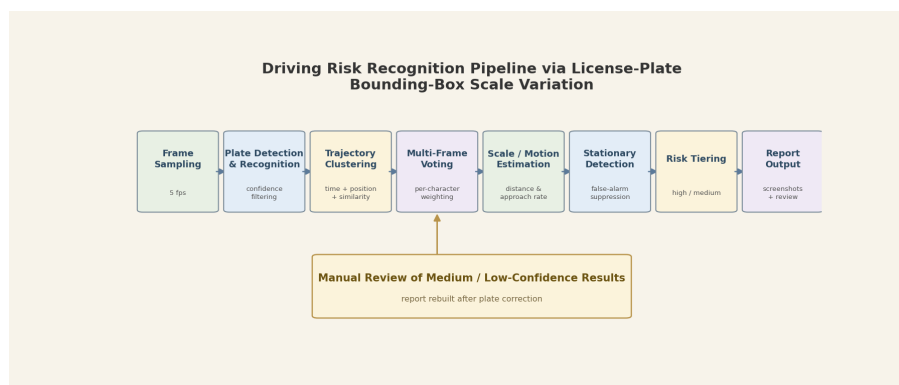


Figure 1. Overview of the proposed pipeline

3.2. License-Plate Detection and Trajectory Clustering

3.2.1. Detection, Recognition, and Confidence Gating

We use the open-source HyperLPR3, whose recognizer follows a segmentation-free design that regresses the character sequence directly from the whole plate; the detection-box width w serves as the scale observation. Only detections with a whole-plate confidence of at least 0.65 that pass the regular-expression check $^{\wedge}[\text{一} \text{ 𠃍} \text{ 𠃊}][\text{A-Z}][\text{A-Z0-9}]\{5,6\}\$$ are retained (seven characters for ordinary blue plates and

eight for new-energy green plates; 一-顛 denotes the CJK Unicode range U+4E00–U+9FA5, i.e., a leading provincial character). This threshold is an empirical value tuned on earlier data batches, balancing recall and false-detection suppression. The videos have a frame width of 1920 px and a source frame rate of about 25 fps; frames are sampled at 5 fps, and the 0.2 s interval suffices to capture the visible growth of plate width during approach events.

3.2.2. Trajectory Clustering

The same vehicle is detected repeatedly and must be aggregated into a trajectory before identity determination: adjacent detections separated by no more than 3 s are first clustered into candidate trajectory segments, which are then subdivided by recognition similarity (character-level similarity ≥ 0.55) or positional proximity (a compensation of 0.6 is added to the similarity score when positions are close). This echoes ByteTrack's idea of preserving trajectories through a second association of low-score boxes [15]: plate confidence fluctuates with scale and blur, so discarding low-score detections easily breaks trajectories. We take plate recognition as the primary evidence and positional continuity as the auxiliary cue.

3.3. Per-Character Weighted Voting and Confidence Tiering

3.3.1. Per-Character Confidence-Weighted Voting

Multi-frame recognitions of the same vehicle contain OCR variants (misread characters, one missing character on NEV plates, etc.) that must be fused into a single plate identity. A confidence-weighted vote is performed per character position j over the valid frames of trajectory τ ,

$$\hat{c}_j = \arg \max (c \in A) \sum (i \in \tau \wedge \alpha_i [c_{i,j} = c]), \quad \alpha_i = \text{conf}_i \quad (1)$$

where A is the character set, $c_{i,j}$ is the character at position j in frame i , the weight α_i is the whole-plate confidence conf_i of that frame, and $\hat{c}_j$ is the decided character at position j . High-confidence frames dominate the result and jitter from distant blurry frames is suppressed; the superiority of temporal aggregation over single-frame selection has been demonstrated empirically by majority-voting studies [7][8].

3.3.2. Confidence Tiering and the Manual Review Loop

A trajectory with at least 3 frames, a maximum confidence of at least 0.90, and a voted-character agreement rate of at least 0.60 is classified as high confidence; one with at least 2 frames and an agreement rate of at least 0.55 is medium confidence; all others are low confidence. All medium- and low-confidence trajectories enter the manual review loop, with conclusions written back to the detail sheet. This triage follows studies on reliability scoring for degraded license plates: confidence can serve as an acceptance criterion, while low-quality samples still require human confirmation [10].

3.4. Plate-Width Prior Ranging and Scale Change Rate

3.4.1. Focal-Length Back-Solving and Ranging Model

The license plate is a statutorily standardized planar object and can serve as a passive ranging reference: GA 36 stipulates that small-vehicle plates are 440 mm wide, so we adopt the prior $W=0.44$ m [3]. Let w be the detection-box width in pixels and f the equivalent focal length (px); the estimated distance D (m) follows from the similar triangles of pinhole imaging:

$$D = f \cdot W / w = f \cdot 0.44 / w \quad (2)$$

The focal length is back-solved from the dashcam's nominal horizontal field of view:

$$f = (H/2) / \tan(FOV/2) \quad (3)$$

where H is the frame width (px); with $FOV=90^\circ$ and $H=1920$ px, $f=H/2=960$ px. Deviations between the nominal FOV and the actual optics enter D systematically, so D is used only for relative tiering rather than absolute ranging (the error can reach about $\pm 30\%$). A preprint study (controlled-condition results) warns that perspective distortion of the whole-plate bounding box can inflate its width by more than 50% [5]; we apply median filtering and outlier removal to the w sequence, a smoothing that reduces variance without eliminating systematic bias [12]. The systematic error introduced by the 480 mm NEV plate is discussed in §5.1.

3.4.2. Scale Change Rate and Lane Offset

The approach situation is characterized by the relative growth rate of the plate-box width. Taking the plate widths w_1, w_2 at the two ends of a time window Δt (s) no shorter than 0.8 s within a trajectory (at least 3 frames), the scale change rate is defined as

$$r = (w_2 - w_1) / (w_1 \cdot \Delta t) \quad (4)$$

where r is multiplied by 100% and expressed in %/s; differencing over the two ends of a window rather than frame by frame suppresses single-frame detection-box jitter. The tau-theory coupling between optical expansion rate and time to collision [11] and the accuracy-bound analysis of the scale change rate [12] back the use of r without calibration; we use it only for approach tiering and do not compute an exact TTC. The lane relationship is determined by the lane-offset ratio $\rho = |\bar{x}_c - H/2| / H$, where $\bar{x}_c$ is the median horizontal coordinate of the plate-box centers within the trajectory; $\rho > 15\%$ is judged as an adjacent lane, otherwise as a preceding vehicle in the same lane.

3.5. Risk Rules and Stationary Suppression

3.5.1. Four Risk Rules

For each trajectory the minimum estimated distance is taken, and risk events are triggered according to the tiers in Table 2. In R4, Δx is the horizontal displacement (in pixels) of the plate-box center between adjacent valid detections; 150 px is about 8% of the 1920 px frame width, corresponding to the typical magnitude of lateral displacement in an adjacent-lane change.

Table 2. Thresholds of the four risk rules

Rule	Trigger condition (high)	Trigger condition (medium)	Output tier
R1 too-close following	same lane and $D < 8$ m	same lane and $8 \leq D < 15$ m	high/medium
R2 adjacent-vehicle proximity	adjacent lane and $D < 3$ m	—	high
R3 rapid approach	$r > 30\%/s$	$15\% < r \leq 30\%/s$	high/medium
R4 lateral cut-in	$\Delta x > 150$ px and $D < 10$ m	$\Delta x > 150$ px and $10 \leq D < 20$ m	high/medium

The dual thresholds on distance and change rate reflect the empirical finding that headway and TTC are mutually independent and complementary [16]; the cut-in rule aligns with the ISO 15623 definition of the cut-in vehicle [17]; the threshold magnitudes benchmark against the GB/T 33577-2017 forward-collision warning TTC alarm interval of 2.0–2.4 s [18] and agree with naturalistic-driving evidence that cut-in events have shorter TTCs [19]. Medium-risk events are only cached for reference and are not included in the final risk-event table; the risk events are heuristic screening warnings intended only for post-event retrospective review and do not constitute a basis for accident liability determination.

3.5.2. Phase-Correlation Stationary Suppression

When the ego vehicle is stationary (red lights, congestion stops), distance-based rules inevitably generate continuous false alarms. We estimate the global displacement between adjacent sampled frames using phase correlation: frames are first resized to 320 px width, and a frame is judged stationary if the displacement is below 1.0 px at that scale; a phase-correlation response below 0.03 is treated as a textureless scene and no judgment is made; if the proportion of stationary frames over the period covered by a trajectory is no less than 0.6, the ego vehicle is judged essentially stationary, in which case R1 and R2 are suppressed while R3 and R4 are retained. Only the frame-level displacement is estimated for gating; no camera-motion compensation is applied.

3.6. Cross-Video Same-Vehicle Merging and Output

3.6.1. Cross-Video Same-Vehicle Merging Decision

Two trajectories are merged if and only if their plates agree and the gap between the end time of one and the start time of the other falls within the -30 s to +150 s window (the video filename contains the recording start time): -30 s absorbs timestamp-alignment errors, and +150 s is the upper bound for the same vehicle to reappear within a filming gap. Merging imposes a temporal mutual-exclusion constraint, so overlapping trajectory pairs are prohibited; when plate comparison is uncertain, positional continuity serves as a fallback cue [6].

3.6.2. Result Output

After merging, vehicle-level risk events and trip-level trajectories are formed and output as three Excel sheets: a summary sheet (duration, vehicle count, high-confidence count, high-risk count), a plate detail sheet (confidence tier,

timestamps, and review remarks), and a risk-event sheet (listing only high-risk events, with annotated full frames and enlarged plate snapshots), supporting manual review and post-event retrospective review.

4. Experiments and Analysis

4.1. Data and Implementation

4.1.1. Overview of the Experimental Data

The experimental data consist of 35 forward-facing midday dashcam video clips (Table 3), with parameters identical to those in §3. The data paradigm follows the SHRP2 event-classification study [2], which performs hazard/safe binary classification on 30 s, 15 fps trimmed clips combined with telematics (accuracy above 0.87); we instead target untrimmed trip videos for post-event screening.

Table 3. Overview of the experimental data

Item	Value
Number of video clips	35 clips (forward-facing, midday lighting)
Total duration	2098.2 s (approximately 35.0 min)
Frame-sampling rate	5 fps
Recognition confidence threshold	≥ 0.65
Vehicle trajectories (before merging)	165
Vehicles after merging	118 (18 spanning multiple videos)
High-risk events	41
Medium-risk events	20 (cached for reference)
Median $d_{\min}$ of high-risk events	5.21 m

4.1.2. Statement of the Evaluation Protocol

This batch of data contains no LiDAR, GPS, or calibrated reference objects, nor event-level annotations; mAP, precision, and recall are therefore not reported, and only counts, proportions, and medians are presented. The pipeline is a deterministic algorithm whose results are fully reproducible under fixed parameters, so the results of a single batch retrospective review are reported.

4.2. Recognition Reliability

Table 4 gives the confidence distribution and review statistics of the 165 trajectories, answering RQ3.

Table 4. Trajectory confidence distribution and review statistics

Metric	Value	Proportion and scope
High-confidence trajectories	79	47.9%
Medium-confidence trajectories	52	31.5%
Low-confidence trajectories	34	20.6%
8-character plate (NEV) trajectories	85	51.5% (85/165, computed from counts)
7-character plate trajectories	80	48.5%
Entered manual review	86	Full coverage of the medium/low-confidence subset (52+34)
Corrected after review	38	38/86 $\approx$ 44.2% (medium/low-confidence subset)

		only)
Machine-read result retained with annotation	48	Machine-read result retained with an annotation added

Roughly half of the trajectories did not reach high confidence, and NEV plates account for more than half—by no means a marginal case. All 86 medium/low-confidence records were reviewed using full frames together with magnified plate crops, yielding a correction rate of $38/86 \approx 44.2\%$ (computed from the above counts; the scope is the medium/low-confidence subset only and is not extrapolated to overall recognition accuracy). Voting (Eq. (4)) alone cannot guarantee correctness, and the manual review loop is necessary; the direction of the gain agrees with majority-voting studies [7][8]. The main causes of low confidence—insufficient pixels at long range, motion blur, reflections, and visually similar characters—mirror the failure modes of the CCPD baseline, whose AP drops to 25.83% on the blur subset and 45.24% on the near/far-scale subsets [9].

4.3. Effect of Cross-Video Same-Vehicle Merging

Trajectory clustering and same-vehicle merging consolidate the 165 trajectories into 118 vehicles ($165 - 118 = 47$ duplicate trajectories), with 18 vehicles spanning multiple video clips, answering RQ3 from the counting dimension: without merging, split vehicles would be counted repeatedly. Since no ground-truth vehicle identities are available, no merging-accuracy metric is defined; the absence of plate-recognition conflicts after merging serves as internal-consistency evidence.

4.4. Risk-Event Statistics

4.4.1. Distribution of Event Types

Table 5 lists the type distribution of the 41 high-risk events; this subsection, together with §4.4.2, answers RQ2 and RQ1.

Table 5. Type distribution of high-risk events

Event type	Count	Share of high-risk events
Lateral cut-in (sideswipe risk)	21	51.2%
Rapid approach (rear-end collision risk)	18	43.9%
Too-close following	1	2.4%
Adjacent-vehicle proximity	1	2.4%
Total	41	100%

Lateral cut-in and rapid approach together account for 39 events, or 95.1% ($39/41$, computed from counts); the dominance of cut-in corroborates naturalistic-driving benchmarks [19]. Too-close following occurs only once, which is related to the rarity of extremely close following at midday and to the stationary-suppression gating of R1/R2. The 20 medium-risk events are only cached for reference and are excluded from the high-risk statistics. All 41 high-risk events are accompanied by exported full-frame annotations of the triggering frame and magnified plate crops (§3.6), and

their risk characterization was confirmed manually.

4.4.2. Distance Statistics and Threshold Consistency

The median of the minimum estimated distance $d_{\min}$ of high-risk events is 5.21 m, which falls within the triggering bands ($R1 < 8$ m, $R4 < 10$ m/ < 20 m). Because events are by definition triggered by these thresholds, this verification is a sanity check of dimensional plausibility rather than independent evidence: it indicates only that the ranging of Eq. (1) is of the same order of magnitude as the tiering thresholds, on the basis of which RQ1 is answered. As an internal self-check, the distance sequences of the 18 vehicles merged across videos were manually inspected and proved continuous at the junctions without order-of-magnitude jumps. The conclusion is strictly limited to "relative tiered screening being usable" and does not claim absolute ranging accuracy. A distant plate spans only a dozen pixels, and the ± 1 px quantization error of w is amplified by Eq. (1) and dominates the distance fluctuation, so the farthest high-risk threshold band is set at only 20 m. The review additionally found that some events were triggered by enlarged plates on vehicle side bodies or by small targets, which are unsuitable for precise ranging.

4.5. Representative Cases

One lateral cut-in event and one rapid approach event are selected for temporal qualitative analysis. In the cut-in event, the center of the adjacent-lane plate bounding box moved persistently toward the ego lane, the lateral displacement between adjacent detections exceeded 150 px, and the frame-by-frame increase of w caused D to converge within 10 m, triggering R4. In the approach event, the preceding vehicle's w grew continuously over a window of no fewer than 3 frames and no shorter than 0.8 s, and the change rate r in Eq. (3) exceeded 30%/s, triggering R3. In accordance with the anonymization requirements for faces and license plates outside the vehicle stipulated in the Several Provisions on the Management of Automobile Data Security (Trial) [20], no screenshots are shown and no complete plate numbers are disclosed; only textual descriptions are provided. These cases are examples of heuristic screening warnings, intended solely as a reference for post-event retrospective review, and do not constitute a basis for accident liability determination. Separate ablations of the voting, stationary-suppression, and cross-video same-vehicle merging modules are left to future work.

5. Discussion and Limitations

5.1. Ranging Validity and Error Sources

In the absence of distance ground truth, plausibility is corroborated with orders of magnitude reported in the literature: the character-height method achieves an MAE of 2.3% at 10 m (preprint, controlled-condition results) [4]. Together with the consistency between $d_{\min}$ and the thresholds and the cross-segment distance

continuity, relative tiering is usable, although no claim is made of reaching the accuracy reported in the literature. When a uniform 440 mm prior is applied, the 480 mm NEV plate [3] has a larger pixel width, so the distance is underestimated by about 8.3% ($440/480-1$, computed from the plate-width parameters). The measured data contain 85 trajectories with 8-character plates and 80 with 7-character plates (51.5%/48.5%), so this is not a marginal case. Underestimation is conservative and thus favorable for screening, and the plate-width prior can be switched according to the number of plate characters. Detection-box jitter propagates through the width sequence to the distance and the change rate, potentially causing misjudgments; outlier removal, smoothing, and confidence gating mitigate this effect.

5.2. False-Alarm Characteristics and Review Cost

The 86 medium/low-confidence trajectories require manual review, of which 38 were corrected (§4.2); the optimal operating point for review cost remains to be studied. Stationary suppression reduces too-close-following false alarms under low-speed congestion, while anomalous distance readings from plates on vehicle side bodies and from small targets must be removed through review.

5.3. Limitations and Boundary Statements

The data cover only midday lighting and do not generalize to nighttime or rain/snow conditions; offline batch processing provides no real-time capability and is not applicable to factory-installed online warning. It bears repeating that the risk events are heuristic screening warnings, intended solely as a reference for post-event retrospective review, and do not constitute a basis for accident liability determination. The benefit ceiling of warning alone is lower than that of warning combined with autonomous braking [1], so the method does not replace factory-installed FCW/AEB. Using only bounding-box scale features and desensitizing screenshots conforms to the data-minimization and anonymization requirements of [20].

6. Conclusion

RQ1: with the 0.44 m plate-width prior and the focal length back-solved from a 90° FOV, the method produces, in a calibration-free manner, distance sequences of the same order of magnitude as the risk thresholds (3/8/10/15/20 m), with a median $d_{\min}$ of 5.21 m; the results suggest that it is sufficiently usable for relative tiered screening. RQ2: four heuristic rules screened out 41 high-risk events from the 35 midday video clips (2098.2 s), with lateral cut-in and rapid approach dominating (21+18 events) and too-close following and adjacent-vehicle proximity occurring once each; phase-correlation stationary suppression controlled false alarms in stationary scenarios. RQ3: confidence tiering, per-character confidence-weighted voting, the manual review loop (86 records reviewed, 38 corrected), and cross-video

same-vehicle merging (165 trajectories consolidated into 118 vehicles, 18 spanning multiple videos) secured engineering reliability. The risk events remain heuristic screening warnings, intended solely as a reference for post-event retrospective review, and do not constitute a basis for accident liability determination. Future work will collect distance ground truth to quantify ranging error, extend the data to nighttime and rain/snow conditions, calibrate the optimal review-cost operating point, conduct module-level ablations, and fuse the pipeline with general-purpose trackers, drawing on the low-score secondary matching of ByteTrack [15] and the AFLink/GSI linking modules of StrongSORT [6].

References

- [1] U.S. Department of Transportation, ITS Knowledge Resources (2023) Large-Scale Analysis Using Data from 13 States Found Vehicles Equipped with Forward Collision Warning and Automatic Emergency Braking Had 49 Percent Less Front-to-Rear Crashes. <https://www.itskrs.its.dot.gov/2023-b01753>
- [2] Taccari, L., Sambo, F., Bravi, L. et al. (2018) Classification of Crash and Near-Crash Events from Dashcam Videos and Telematics. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, 2460-2465.
- [3] Ministry of Public Security of the People's Republic of China (2018) License Plates of Motor Vehicles of the People's Republic of China: GA 36-2018. Standards Press of China, Beijing. (In Chinese)
- [4] Reddy, M.L. and Liu, Z. (2026) Physics-Grounded Monocular Vehicle Distance Estimation Using Standardized License Plate Typography. arXiv: 2604.12239. <https://arxiv.org/abs/2604.12239>
- [5] Reddy, M.L. and Liu, Z. (2026) Typography-Based Monocular Distance Estimation Framework for Vehicle Safety Systems. arXiv: 2603.22781. <https://arxiv.org/abs/2603.22781>
- [6] Du, Y., Zhao, Z., Song, Y. et al. (2023) StrongSORT: Make DeepSORT Great Again. IEEE Transactions on Multimedia, 25, 8725-8737. <https://doi.org/10.1109/TMM.2023.3240881>
- [7] Laroca, R., Severo, E., Zanlorensi, L.A. et al. (2018) A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector. 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 1-10. <https://doi.org/10.1109/IJCNN.2018.8489629>
- [8] Gonçalves, G.R., Menotti, D. and Schwartz, W.R. (2016) License Plate Recognition Based on Temporal Redundancy. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, 2577-2582. <https://doi.org/10.1109/ITSC.2016.7795970>
- [9] Xu, Z., Yang, W., Meng, A. et al. (2018) Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. In: Computer Vision – ECCV 2018, Lecture Notes in Computer Science, Vol. 11217, Springer, Cham, 261-277. https://doi.org/10.1007/978-3-030-01261-8_16
- [10] Maier, A., Moussa, D., Spruck, A. et al. (2022) Reliability Scoring for the Recognition of Degraded License Plates. 2022 18th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), Alicante, 1-8. <https://doi.org/10.1109/AVSS56176.2022.9959390>
- [11] Lee, D.N. (1976) A Theory of Visual Control of Braking Based on Information about Time-to-Collision. Perception, 5, 437-459.
- [12] Stein, G.P., Mano, O. and Shashua, A. (2003) Vision-Based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy. IEEE IV2003 Intelligent Vehicles

- Symposium, Columbus, 120-125. <https://doi.org/10.1109/IVS.2003.1212895>
- [13] Kilicarslan, M. and Zheng, J.Y. (2016) Bridge Motion to Collision Alarming Using Driving Video. 2016 23rd International Conference on Pattern Recognition (ICPR), Cancún, 1870-1875. <https://doi.org/10.1109/ICPR.2016.7899909>
 - [14] Yu, H. and Zhang, W. (2012) Method of Vehicle Distance Measurement for Following Car Based on Monocular Vision. Journal of Southeast University (Natural Science Edition), 42, 542-546. (In Chinese) <https://doi.org/10.3969/j.issn.1001-0505.2012.03.029>
 - [15] Zhang, Y., Sun, P., Jiang, Y. et al. (2022) ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In: Computer Vision – ECCV 2022, Lecture Notes in Computer Science, Vol. 13663, Springer, Cham, 1-21. https://doi.org/10.1007/978-3-031-20047-2_1
 - [16] Vogel, K. (2003) A Comparison of Headway and Time to Collision as Safety Indicators. Accident Analysis & Prevention, 35, 427-433. [https://doi.org/10.1016/S0001-4575\(02\)00022-2](https://doi.org/10.1016/S0001-4575(02)00022-2)
 - [17] International Organization for Standardization (2013) ISO 15623:2013 Intelligent Transport Systems—Forward Vehicle Collision Warning Systems—Performance Requirements and Test Procedures. ISO, Geneva.
 - [18] Standardization Administration of China (2017) GB/T 33577-2017 Intelligent Transportation Systems—Forward Vehicle Collision Warning Systems—Performance Requirements and Test Procedures. Standards Press of China, Beijing. (In Chinese)
 - [19] Wang, X., Yang, M. and Hurwitz, D. (2019) Analysis of Cut-In Behavior Based on Naturalistic Driving Data. Accident Analysis & Prevention, 124, 127-137.
 - [20] Cyberspace Administration of China et al. (2021) Several Provisions on the Management of Automobile Data Security (Trial). http://www.cac.gov.cn/2021-08/20/c_1631049984897667.htm (In Chinese)