Tsinghua Science and Technology
Doi: s41095-0xx-xxxx-x

Research Article

Predictive estimation of tire-road friction for autonomous driving via vision-dynamics collaborative fusion

Jiahui Liu, Yang Liu, Chenglin Liu (✉), Liang Wang (✉), and Xiaohan Liu

Abstract The tire-road friction coefficient (TRFC) is a decisive factor governing vehicle safety, yet estimating it accurately in real time remains a challenge for autonomous driving. Existing dynamics-based methods rely on vehicle dynamics responses and require sufficient tire excitation, while vision-based methods can preview upcoming road conditions but typically yield only coarse friction ranges rather than precise values. To fill this gap, this paper proposes a vision-dynamics collaborative fusion framework for predictive TRFC estimation. A road surface condition classifier augmented with a confusion-aware training strategy is developed to distinguish ”wet” from ”water” surfaces. The classification results, together with the ego vehicle's kinematic state, are mapped into a continuously updatable preview friction field (PFF) that spatially anticipates the road friction ahead. The PFF is then adaptively fused with an adaptive cubature Kalman filter (A-CKF) estimator through a dual-confidence fusion mechanism. Real-vehicle experiments demonstrate that the proposed Fusion method consistently outperforms purely dynamics-based or vision-based baselines, achieving a mean absolute percentage error as low as 4.48% in low-excitation conditions, where the dynamics-only A-CKF degrades to 42.60%. Moreover, benefiting from the PFF's predictive capability, the Fusion method converges to the true TRFC markedly faster than the dynamics-only estimation. These results confirm the accuracy, robustness, and predictive capability of the proposed framework for autonomous driving applications.

Keywords Autonomous driving; Tire-road friction estimation; Information fusion; Vehicle dynamics; Road surface condition classification

• Jiahui Liu, Yang Liu, Chenglin Liu, and Liang Wang are with the School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China. E-mail: jh-liu22@mails.tsinghua.edu.cn, thu_ets_lab@tsinghua.edu.cn, liuchenglin@tsinghua.edu.cn, wangliang0117@mail.tsinghua.edu.cn.
• Yang Liu and Liang Wang are also with the State Key Laboratory of Intelligent Green Vehicle and Mobility, Tsinghua University, Beijing 100084, China.
• Xiaohan Liu is with the Department of Architecture and Civil Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. E-mail: xiaohanl@chalmers.se
Manuscript received: 20xx-xx-xx; revised: 20xx-xx-xx; accepted: 20xx-xx-xx

1 Introduction

1.1 Background

Onboard computing and sensing techniques have advanced rapidly, enabling intelligent vehicles to perceive their surroundings with increasing accuracy in real time. While a substantial body of autonomous driving research has focused on detecting traffic participants and environmental obstacles [1, 2], the tire-road friction coefficient (TRFC), a factor that directly affects vehicle safety, is still predominantly estimated using conventional techniques in real-world deployments. The TRFC indicates the achievable limits of longitudinal braking and lateral steering, making it a critical input for decision-making and trajectory planning [3, 4]. When the road surface is wet or otherwise contaminated, failure to anticipate and accurately track the reduction of TRFC in time would easily lead to traffic accidents. In China, statistics indicate that over 80% of traffic accidents stem from inadequate road anti-skid performance, with sideslip commonly occurring under strong control inputs such as braking [5, 6]. An examination of more than 500 crash sites in Texas, USA, revealed a pronounced link between the incidence of traffic accidents and diminished road anti-skid performance [7, 8].

TRFC is usually denoted as μ, and is widely recognized in autonomous driving research as a pivotal determinant of motion behavior and dynamic performance [9, 10]. The TRFC greatly effects how faithfully a vehicle can follow its intended trajectory. For instance, on low-μ surfaces, a path-tracking controller may fail to anticipate the true trajectory, producing errors in both heading and lateral distance. This is an effect that becomes especially pronounced when negotiating curves [11–13]. Accurate estimation of the TRFC at all four wheels enables the vehicle to properly adjust boundary conditions such as cornering speed and safe following distance, thereby generating motion trajectories that exploit the available friction as efficiently as possible while preserving obstacle-avoidance safety and handling stability[]. Therefore, reliable TRFC estimation is central to both the control performance of autonomous driving and the broader effectiveness of intelligent transportation systems.

Just Accepted

1.2 Related Work

TRFC estimation methods can generally be divided into effect-based and cause-based methods [14]. Effect-based methods estimate the TRFC by monitoring the responses of tires or vehicles under external excitation. The required signals, such as vehicle speed, tire forces, accelerations, and yaw rate, can usually be obtained from onboard sensors. Some studies have also measured tire noise, deformation, pressure, and vibration to characterize tire-road interactions [15, 16]. For example, acoustic signals collected near the tires have been used to distinguish different road surfaces [17]. These methods can provide accurate TRFC estimates when sufficient longitudinal or lateral excitation is available. However, their performance depends strongly on the current driving state, and they cannot obtain the friction condition of the road ahead.

In contrast, cause-based methods identify factors that directly determine the TRFC, including the road material and contaminants such as water, snow, and ice [18]. Cameras, LiDAR, and other optical or acoustic sensors are commonly used to acquire road-surface characteristics [19]. After the road type is identified, an empirical TRFC is assigned according to the classification result. This approach does not require sufficient tire excitation and can provide information about the road ahead. However, the estimated value generally represents a friction range or a typical value of each road category rather than the actual tire-road interaction.

A classical effect-based solution is to construct vehicle and tire dynamics models and subsequently estimate the TRFC using a Kalman filter (KF). The KF recursively combines model predictions with sensor measurements, making it suitable for estimating parameters that cannot be measured directly. The original KF is designed for linear systems, while the extended Kalman filter (EKF), unscented Kalman filter (UKF), and cubature Kalman filter (CKF) have been developed to address the nonlinear characteristics of vehicle and tire dynamics [20–22].

Since TRFC estimation involves measurement noise, uncertain tire parameters, and strongly nonlinear tire forces, current studies mainly improve the filter structure, tire model, and noise adaptation mechanism. Recent methods have enhanced estimation accuracy and convergence by optimizing the underlying tire model [23, 24] or adaptively adjusting the process and measurement noise [25]. For instance, a strong-tracking CKF has been combined with a Dugoff tire model [26], while central-difference and square-root CKFs have been employed to improve numerical stability and estimation robustness [27, 28]. These KF variants can provide continuous and physically consistent TRFC estimates and have relatively low computational costs. However, the observability of TRFC remains dependent on tire excitation. During straight or steady driving, the measured dynamic responses contain insufficient information about the friction limit.

Among cause-based methods, vision-based approaches are widely used to identify road conditions ahead. Since front-facing cameras are widely deployed for autonomous driving, these methods generally require lower additional hardware costs than approaches based on specialized road-characteristic sensors. Traditional machine-learning methods extract color, texture, or statistical features from road images and employ classifiers such as support vector machines and naive Bayes classifiers [29, 30]. Deep-learning methods learn road surface features using convolutional neural networks and end-to-end architectures [31, 32]. For example, color moments and an SVM have been used to distinguish dry and wet asphalt [33]. Segmentation and lightweight classification networks have also been developed to recognize dry asphalt, wet asphalt, and snow- or ice-covered roads [34, 35]. Multi-source datasets and spatiotemporal image information have subsequently been introduced to improve recognition robustness [36, 37]. Nevertheless, vision-based TRFC estimation essentially converts a continuous estimation problem into a road-type classification problem [38]. Different surfaces within the same category may have substantially different TRFC values, while visually similar surfaces, such as thin-film wet roads and water-covered roads, may correspond to different friction levels.

A recent research trend is to combine dynamics-based estimation with vision-based road perception [39, 40]. The two sources provide complementary information. Dynamics-based estimators yield accurate and continuous estimates when sufficient excitation is present, whereas vision-based methods provide a preview of upcoming road conditions without relying on tire excitation. Existing fusion frameworks use visual results to initialize a dynamics estimator, constrain its estimation range, or update its TRFC estimate at regular intervals [41]. Guo *et al.* combined a lightweight road-classification network with a UKF and fused their results after spatio-temporal synchronization according to confidence levels [42]. In addition to result-level fusion, visual and dynamic features have been integrated at the feature and decision levels using neural networks [43]. However, vision-based branches can introduce considerable computational costs [44, 45], and their outputs are commonly discrete or updated at a lower frequency than dynamics estimates. Existing methods therefore remain limited by spatiotemporal inconsistency, visual misclassification, and unreliable dynamics estimation under low excitation. Furthermore, the reliability of each source varies

Just Accepted

continuously with the driving maneuver and visual environment, but this variation is not always adequately represented during fusion [46].

In summary, dynamics-based methods can provide accurate and continuous TRFC estimates, but they cannot reliably identify friction changes when the available vehicle excitation is insufficient. Vision-based methods can perceive the road ahead, but their estimates are limited by classification errors and the use of empirical TRFC values. Existing fusion methods demonstrate the potential of combining these complementary sources, while their performance still depends on reliable spatiotemporal association and an appropriate evaluation of source confidence. Therefore, a predictive TRFC estimation method is required to preserve the physical consistency of dynamics-based estimation, exploit visual information about the road ahead, and adaptively adjust the contribution of each branch according to its current reliability.

1.3 Contributions

To address the challenges in existing research, this paper couples an adaptive CKF estimator with a vision-based friction perception module and develops a dual-confidence fusion mechanism. The resulting framework provides continuous tire-level TRFC estimates and responds to upcoming TRFC changes under both low- and high-excitation driving conditions. The main contributions of our study can be concluded from the following three aspects:

(1) **A road friction condition classifier that further differentiates wet and water surfaces is proposed.** Wet and water surfaces are distinguished as separate classes, given their difference in peak friction coefficient under different vehicle speeds.

(2) **A dynamic grid mapping method is proposed to provide friction estimation with finer spatial resolution.** The road ahead is divided into grids aligned with the spatial resolution of the planning module.

(3) **A dual-confidence strategy is proposed to adaptively combine vision- and dynamics-based estimation.** The visual confidence reflects the reliability of the preview friction field, while the dynamics confidence reflects the lateral excitation level.

The rest of this paper is organized as follows. Section 2 formulates the estimation problem and outlines the overall vision-dynamics fusion architecture. Section 3 develops a dynamics-based estimator used to track the peak TRFC in real time. Section 4 presents the generation of the vision-based estimation by constructing a preview friction field. Section 5 introduces a dual-confidence fusion mechanism that reconciles the dynamics- and vision-based estimates. Section 6 reports real-vehicle experimental results. Finally, Section 7 concludes the paper.

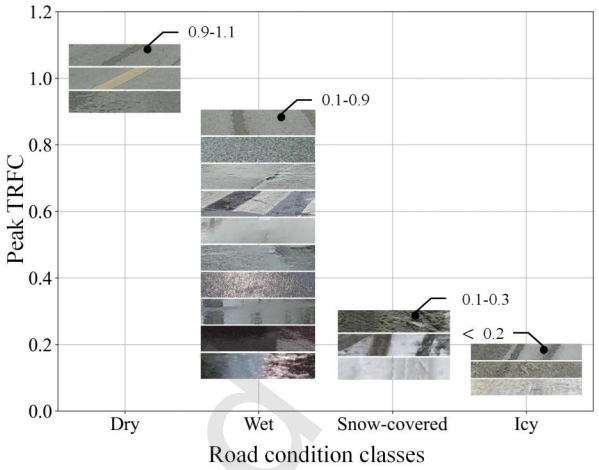


Fig. 1 TRFC value ranges of different road condition classes.

2 Problem Formulation and System Overview

In this study, snowy and icy road surface conditions are excluded from the research scope. As illustrated in Fig. 1, snow and ice provide an extremely low TRFC of approximately 0.1 to 0.3, which leaves little margin for intelligent trajectory optimization and reduces the control objective to safe deceleration and stability [47]. Thus, our research focuses on dry and wet surfaces, where the accurate tracking of TRFC values contributes to the planning module's performance.

To rigorously define these surface conditions, we further divide the "wet" category into "wet" and "water" subclasses, depending on their physical and visual properties. According to research on pavement engineering [48], the relationship between peak TRFC and vehicle speed shifts significantly once the water film height exceeds 1 mm. We therefore define the "wet" class as a water film height of 0–1 mm and the "water" class as exceeding 1 mm. During image acquisition, it is confirmed that the two categories carry distinct visual characteristics. Specifically, "water" surfaces show visible puddles with pronounced specular reflections and shadows, whereas "wet" surfaces resemble dry pavement but appear darker. The mapping between TRFC and the three road friction types is detailed in Section 4.

Visual classification alone offers a basis for previewing road friction but cannot guarantee reliable TRFC estimation. Dynamics-based estimation is more stable but reacts only to friction changes already encountered, offering no preview of upcoming conditions. To overcome the limitations of each method used in isolation, this paper proposes a vision-dynamics fusion framework that achieves stable estimation

Just Accepted

while constructing a predictive representation of the friction distribution ahead of the vehicle. To reflect the independent dynamic state of each wheel, the estimation target is defined as the four individual peak TRFC values:

$$\boldsymbol{\mu} = [\mu_{fl,k}, \mu_{fr,k}, \mu_{rl,k}, \mu_{rr,k}]^\top \tag{1}$$

where $\mu_{fl,k}$, $\mu_{fr,k}$, $\mu_{rl,k}$, and $\mu_{rr,k}$ denote the peak TRFC at time k for the front left, front right, rear left, and rear right wheels, respectively.

The overall framework, shown in Fig. 2, consists of a dynamics-based estimation module, a visual perception module, and a dual-confidence fusion module. The dynamics branch takes measurable vehicle states and estimates the four-wheel peak TRFC in real time through vehicle and tire dynamical modeling. The vision branch takes real-world road images from the front-facing camera, recognizes the road surface condition, and then constructs a spatiotemporal field for storing friction information to further perceive the TRFC changes ahead. The fusion module takes the reliability of estimates from the two branches into account and then outputs the weighted fusion estimation, combining real-time responsiveness, stability, and preview capability.

3 Dynamics-based Real-time Friction Estimation

In this section, a dynamics-based estimator is established to estimate the four-wheel TRFC in real time. Dynamic signals are obtained by on-board vehicle sensors, including longitudinal speed v_x, lateral speed v_y, longitudinal acceleration a_x, lateral acceleration a_y, steering wheel angle δ_f, yaw rate $\dot{\psi}$, and wheel angular speeds ω_i ($i \in \{fl, fr, rl, rr\}$), collectively denoted as the sensor signal vector $\boldsymbol{u}_k$. These signals are mapped to tire forces via a 3-DOF vehicle model and the nonlinear Dugoff tire model. Then the four-wheel TRFC μ_i is estimated by deploying a CKF estimator with an adaptive noise mechanism, which is referred to as the A-CKF in the following content.

3.1 Vehicle and Tire Modeling

Considering the longitudinal, lateral, and yaw motion of the vehicle, a planar 3-DOF rigid-body model is adopted, neglecting roll and pitch (see Fig. 3). Each wheel i carries longitudinal and lateral forces $F_{x,i}, F_{y,i}$ and steering angle δ_i. In this study, we take front-steering vehicles into account, leading to $\delta_i = \delta_f$ for front wheels and $\delta_i = 0$ for rear wheels. The equations of motion and yaw moment are formulated as:

$$m(\dot{v}_x - v_y\dot{\psi}) = \sum_i (F_{x,i}\cos\delta_i - F_{y,i}\sin\delta_i) \tag{2a}$$

$$m(\dot{v}_y + v_x\dot{\psi}) = \sum_i (F_{x,i}\sin\delta_i + F_{y,i}\cos\delta_i) \tag{2b}$$

$$I_z\ddot{\psi} = \sum_i \Big[(F_{x,i}\sin\delta_i + F_{y,i}\cos\delta_i)\, l_i - (F_{x,i}\cos\delta_i - F_{y,i}\sin\delta_i)\, \frac{t_i}{2}\mathrm{sgn}_i \Big] \tag{2c}$$

where m is the vehicle mass, I_z is the yaw moment of inertia, l_i is the longitudinal distance from wheel i to the vehicle's center of gravity (CG). For wheels on the front axle, $l_i = l_f$, and for wheels on the rear axle, $l_i = -l_r$. t_i indicates the corresponding track width, and $\mathrm{sgn}_i = +1$ for left wheels and -1 for right wheels.

The vertical loads $F_{z,i}$ on tires are dynamically reconstructed using a quasi-static load transfer model, which corrects the static axle loads with the measured longitudinal and lateral accelerations. Let $L = l_f + l_r$ denote the wheelbase and h_g indicate the height of the CG. For axle $a \in \{f, r\}$, $l_{\bar{a}}$ is the distance from the CG to the opposite axle, which means $l_{\bar{f}} = l_r$, $l_{\bar{r}} = l_f$. Assuming the lateral load transfer is distributed between axles in proportion to their static load, the tire vertical loads are defined as:

$$F_{z,a} = \frac{mg\, l_{\bar{a}}}{2L} + s_a \frac{ma_x h_g}{2L} \tag{3a}$$

$$F_{z,al} = F_{z,a} - \frac{ma_y h_g}{t_a} \cdot \frac{l_{\bar{a}}}{L} \tag{3b}$$

$$F_{z,ar} = F_{z,a} + \frac{ma_y h_g}{t_a} \cdot \frac{l_{\bar{a}}}{L} \tag{3c}$$

where $s_f = -1$, $s_r = +1$, and $F_{z,al}, F_{z,ar}$ denote the vertical load of the left and right wheels on axle a.

The tire longitudinal slip ratio σ_i and slip angle α_i are two core variables in tire force calculation, which cannot be measured directly. The longitudinal and lateral speed $v_{x,i}$, $v_{y,i}$ at each wheel center are obtained from the vehicle's speeds, yaw rate, and wheel position:

$$v_{x,al} = v_x - \dot{\psi}\,\frac{t_a}{2}, \quad v_{x,ar} = v_x + \dot{\psi}\,\frac{t_a}{2} \tag{4}$$

$$v_{y,al} = v_y + \dot{\psi}\, l_a, \quad v_{y,ar} = v_y - \dot{\psi}\, l_a \tag{5}$$

where $a \in \{f, r\}$ denotes the axle, l and r denote the left and right wheels, respectively, and $l_a = l_f$ for the front axle and $l_a = -l_r$ for the rear axle. Based on these wheel-center speeds, the slip ratio and slip angle of wheel i are defined as

$$\sigma_i = \frac{\omega_i R_e - v_{x,i}}{\max(|v_{x,i}|,\, |\omega_i R_e|,\, \epsilon)} \tag{6a}$$

$$\alpha_i = \arctan\left(\frac{v_{y,i}}{v_{x,i}}\right) - \delta_i \tag{6b}$$

where R_e is the effective rolling radius, and ϵ prevents singularity at zero velocity.

The Dugoff tire model is then adopted, deploying Eqs. (3)–

Just Accepted

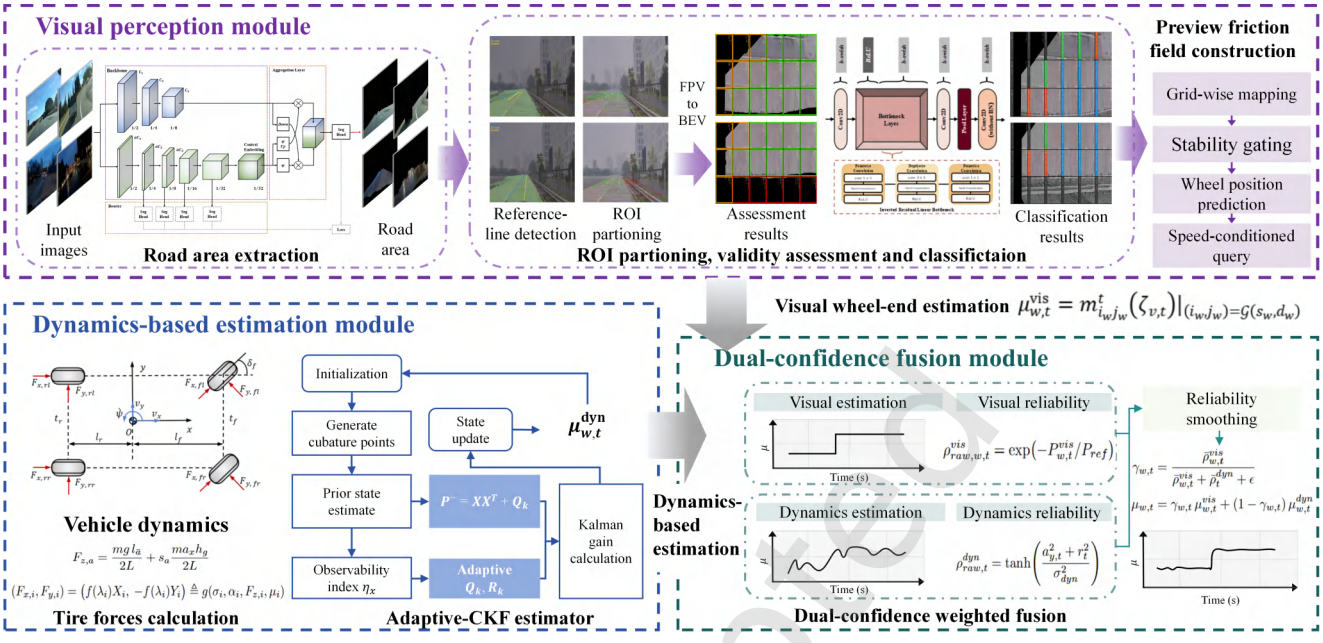


Fig. 2 Overall framework of the proposed fusion estimation method. Note that the architectures of BiSeNet-V2 and MobileNet-V3 are standard backbones adopted without structural changes, and their complete layer parameters and configurations are detailed in Table 1, Table 3, as well as in the original literature [49, 50].

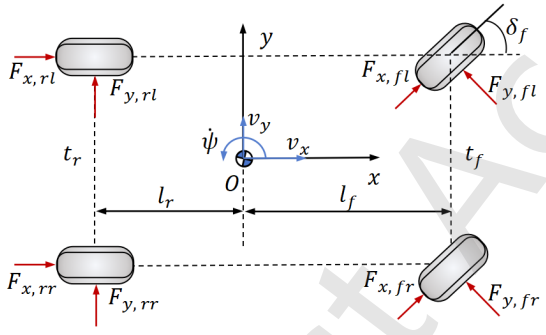


Fig. 3 Vehicle dynamics modeling.

(6) to calculate longitudinal and lateral tire forces. It is a classic nonlinear tire model with clear physical interpretation and analytical invertibility. Compared with empirical models like the Pacejka Magic Formula, its explicit parameterization of μ avoids prohibitive computational overhead and numerical instability during online calculation. For wheel i, the nominal longitudinal and lateral forces are first defined as:

$$X_i = C_\sigma F_{z,i}\,\sigma_i, \quad Y_i = C_\alpha F_{z,i}\,\tan\alpha_i \tag{7}$$

where C_σ, C_α are pre-calibrated normalized longitudinal and lateral stiffness coefficients, which remain constant during real-time estimation. A combined saturation indicator S_i and the friction utilization factor λ_i are then defined as:

$$S_i = \sqrt{X_i^2 + Y_i^2}, \quad \lambda_i = \frac{\mu_i F_{z,i}(1+\sigma_i)}{2S_i} \tag{8}$$

where μ_i is the TRFC of wheel i. The Dugoff nonlinear saturation function is formulated as:

$$f(\lambda_i) = \begin{cases} 1, & \lambda_i \geqslant 1 \\ (2-\lambda_i)\lambda_i, & \lambda_i < 1 \end{cases} \tag{9}$$

thus the resulting tire forces are expressed as:

$$(F_{x,i}, F_{y,i}) = \big(f(\lambda_i)X_i,\ -f(\lambda_i)Y_i\big) \triangleq g(\sigma_i, \alpha_i, F_{z,i}, \mu_i) \tag{10}$$

An important implication of Eqs. (7)–(10) is that the observability of μ_i depends on vehicle operating conditions. In the regime where slip ratio and slip angle are small, the corresponding S_i is small and consequently $\lambda_i \geqslant 1$. According to Eq. (9), $f(\lambda_i)$ equals one in this regime, which leads to $(F_{x,i}, F_{y,i}) = (X_i, -Y_i)$. Since X_i and Y_i in Eq. (7) do not depend on μ_i, the tire forces are insensitive to μ_i in this regime. Therefore, μ_i cannot be identified from the vehicle-response measurements under such operating conditions. As the tires enter the operating regime where $\lambda_i < 1$, the saturation function becomes dependent on λ_i. Then the resulting tire forces become sensitive to μ_i, providing informative measurements for estimation.

3.2 Adaptive-CKF for Peak TRFC Estimation

To establish the dynamics-based estimator, the TRFC values of the four tires are selected as the state vector, with the observation vector consisting of measurable vehicle parameters:

$$\boldsymbol{x}_k = [\mu_{fl,k}, \mu_{fr,k}, \mu_{rl,k}, \mu_{rr,k}]^\mathsf{T} \in \mathbb{R}^4 \tag{11}$$

Just Accepted

$$\boldsymbol{z}_k = \left[a_{x,k}, a_{y,k}, \ddot{\psi}_k\right]^{\mathsf{T}} \tag{12}$$

where $\boldsymbol{z}_k$ is obtained from sensor signal vector $\boldsymbol{u}_k$. Moreover, a first-order random-walk process model is adopted, together with a nonlinear observation model:

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}, \quad \boldsymbol{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, Q_{k-1}) \tag{13}$$

$$\boldsymbol{z}_k = h(\boldsymbol{x}_k) + \boldsymbol{v}_k, \quad \boldsymbol{v}_k \sim \mathcal{N}(\mathbf{0}, R_k) \tag{14}$$

where $h(\cdot)$ is constructed by combining the tire model and the vehicle dynamics equations in Section 3.1. Specifically, given the state $\boldsymbol{x}_k$, the slip ratio σ_i, the slip angle α_i, and vertical loads $F_{z,i}$, the tire forces are calculated using Eqs. (7)–(10). These tire forces are then substituted into Eqs. (2a)–(2c) to obtain the predicted observation vector:

$$h(\boldsymbol{x}_k) = \left[a_{x,k}^{\text{pred}}, a_{y,k}^{\text{pred}}, \ddot{\psi}_k^{\text{pred}}\right]^{\mathsf{T}} \tag{15}$$

Then a third-degree spherical-radial CKF with $2n_x = 8$ equally weighted cubature points ($n_x = 4$) is adopted. This is because $h(\cdot)$ contains the piecewise nonlinear saturation function in Eq. (9), whose Jacobian is discontinuous at the saturation boundary. The CKF follows the standard time update, cubature-propagation, and measurement update recursion.

The normalized resultant force of wheel i is defined as an observability indicator, and is further averaged as a composite observability index:

$$\eta_i = \frac{2S_i}{F_{z,i}}, \quad \eta_k = \frac{1}{4}\sum_{i=1}^{4}\eta_{i,k} \tag{16}$$

According to Eq. (8), η_i and λ_i are related by:

$$\lambda_i = \frac{\mu_i(1+\sigma_i)}{\eta_i} \tag{17}$$

where for a given μ_i and σ_i, a larger η_i corresponds to a smaller λ_i. This means the tire operates closer to the friction limit, leading to a stronger observability of μ_i.

To adaptively adjust the core parameters in the estimator depending on different vehicle operating conditions, a monotonically increasing sigmoid mapping function is then constructed, formulated as $g(\eta_k) = [1 + \exp(-\kappa(\eta_k - \eta_0))]^{-1} \in (0,1)$, which is used to adjust the process and observation noise covariances:

$$Q_k = Q_0 \cdot g(\eta_k) \tag{18a}$$

$$R_k = \frac{R_0}{g(\eta_k)} \tag{18b}$$

where Q_0, R_0 are baseline noise covariance matrices. When the tire approaches the friction limit and observability strengthens, the process noise is increased, and the observation noise is decreased, so the filter trusts the current observation more and converges faster to the true TRFC. When the tire remains in the linear operating regime, the filter instead retains trust in the historical estimate, avoiding meaningless state drift driven by uninformative observations.

Finally, a physical-consistency correction is deployed to enforce $\boldsymbol{x}_{k|k} \leftarrow \text{clip}\left(\boldsymbol{x}_{k|k},\ \mu_{\min},\ \mu_{\max}\right)$. The CKF with the noise covariance adjustment mechanism proposed above is completely established, of which the workflow is described in Algorithm 1.

Algorithm 1 A-CKF for TRFC Estimation

Input: $\boldsymbol{x}_{k-1|k-1}$, $P_{k-1|k-1}$, sensor signals $\boldsymbol{u}_k$, Q_0, R_0
Output: $\boldsymbol{x}_{k|k} = [\mu_{fl,k}, \mu_{fr,k}, \mu_{rl,k}, \mu_{rr,k}]^{\mathsf{T}}$, $P_{k|k}$
Initialize $\boldsymbol{x}_{0|0}$, $P_{0|0}$;
for $k = 1$ **to** N **do**
  Reconstruct $\sigma_i, \alpha_i, F_{z,i}$ from $\boldsymbol{u}_k$ via Eqs. (3)–(6);
  Compute observability index η_k; adapt Q_k, R_k via Eqs. (16)–(18b);
  Standard CKF time update: predict $\boldsymbol{x}_{k|k-1}$, $P_{k|k-1}$;
  Cholesky-decompose $P_{k|k-1}$, generate $2n_x$ cubature points $\mathcal{X}_{j,k|k-1}$;
  for $j = 1$ **to** $2n_x$ **do**
    Calculate tire forces using Eqs. (7)–(10), and substitute them into Eqs. (2a)–(2c) to obtain $\mathcal{Z}_{j,k|k-1}$;
  end for
  Compute predicted $\boldsymbol{z}_{k|k-1}$, $P_{zz,k}$, $P_{xz,k}$; calculate gain $K_k = P_{xz,k}P_{zz,k}^{-1}$;
  Standard CKF measurement update: $\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + K_k(\boldsymbol{z}_k - \boldsymbol{z}_{k|k-1})$, $P_{k|k} = P_{k|k-1} - K_k P_{zz,k} K_k^{\mathsf{T}}$;
  Clip $\boldsymbol{x}_{k|k}$ to $[\mu_{\min}, \mu_{\max}]$;
end for
return $\{\boldsymbol{x}_{k|k}\}_{k=1}^{N}$;

4 Vision-based Predictive Friction Perception

Dynamics-based estimators, including the proposed A-CKF, can only estimate the friction of the road patch currently in contact with four tires, and therefore lack predictive capability. This section proposes a method for constructing a Preview Friction Field (PFF). By fusing road condition classification with the ego vehicle's kinematic state, it provides a continuous, updatable spatial prediction of the road friction condition ahead of the vehicle.

4.1 Drivable road area extraction

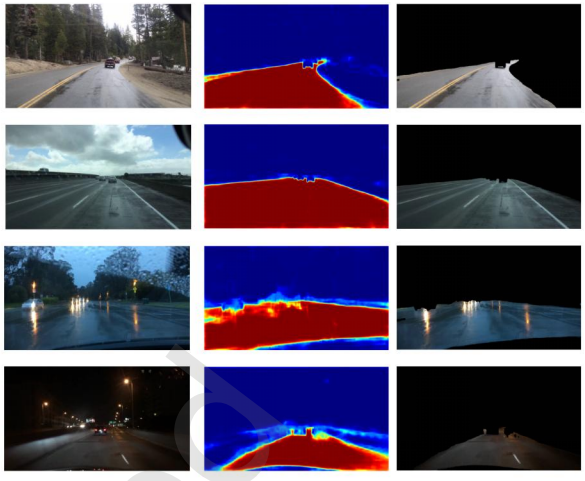
To establish a friction field ahead of the vehicle, a BiSeNet-V2 semantic segmentation model [49] is first used to extract the

Just Accepted

Table 1 BiSeNet-V2 structure for road area segmentation.

Module	Sub-module	Out. Ch.	Stride/Rep.	Output
Detail Branch	Stem Block	64	2/1	$H/2\times W/2$
	Conv Block	64	2/1	$H/4\times W/4$
	Conv Block	128	2/1	$H/8\times W/8$
Semantic Branch	Stem Block	16	2/1	$H/2\times W/2$
	GE Layer	32	2/1	$H/4\times W/4$
	GE Layer	64	2/2	$H/8\times W/8$
	GE Layer	128	2/2	$H/16\times W/16$
	GE Layer	128	1/4	$H/16\times W/16$
Semantic Branch	CE Block	128	1/1	$H/16\times W/16$
Fusion	BGA Layer	128	—/1	$H/8\times W/8$
Seg. Head	Conv+Upsample	2	—/1	$H\times W\times 2$

Note: GE refers to Gather-and-Expansion Layer and CE refers to Context Embedding Block.



(a) Original images (b) Road mask (c) Road area

Fig. 4 Segmentation results by BiSeNet-V2.

drivable road area from forward-facing camera images. It adopts a two-branch architecture, with a Detail Branch that captures fine-grained spatial detail and a Semantic Branch that captures contextual semantic information. The two branches are aggregated via a Bilateral Guided Aggregation (BGA) layer to balance segmentation accuracy and inference efficiency, and the network structure is summarized in Table 1. To reduce task complexity, the segmentation task is formulated as a binary classification problem, distinguishing only "road" from "non-road", and the model is trained on the BDD100K dataset[51], whose extensive samples cover varied illumination, weather, and road surface conditions consistent with the highway driving scenarios targeted in this study.

The main segmentation head is trained with a hybrid loss combining Online Hard Example Mining (OHEM) and the Lovász loss, targeting background pixels like sky and guardrails, which are hard to classify, and irregular segmentation boundaries, respectively. Four auxiliary segmentation heads are trained with standard cross-entropy loss and jointly weighted with the main loss. Optimization uses momentum-SGD, with momentum $= 0.9$ and weight decay $= 5 \times 10^{-4}$, and deploys polynomial learning-rate decay with power $= 0.9$, for a total of 190 epochs. The segmentation results on typical image samples are illustrated in Fig. 4.

4.2 ROI partitioning and validity assessment

The friction field constructed in this study must satisfy spatiotemporal continuity and sufficient resolution to serve as a continuous constraint for trajectory planning defined in the Frenet frame. Therefore, the extracted road area is partitioned according to the vehicle's Frenet coordinate system, where the ego lane's centerline is defined as $C_t(s)$, with s and d respectively indicating arc length along and lateral offset from $C_t(s)$. In this study, the ego-lane markings and the resulting centerline $C_t(s)$ are extracted using a pre-trained Ultra Fast Lane Detection (UFLD) network [52] due to its computational efficiency. After the region of interest (ROI) is partitioned within the drivable road area, the ROI patches are transitioned from the FPV (first-person-view) image plane to a bird's-eye-view (BEV) plane via the Inverse Perspective Mapping (IPM). This transformation is essential because perspective distortion causes identical road patches to exhibit drastically varying scales and geometries at different longitudinal distances.

Lateral partitioning rule divides highway lanes into n_d bins with a bin offset, specifically defined as:

$$d_j = \left(\frac{N_{lat}n_d}{2} - j + \frac{1}{2}\right)\frac{W_{lane}}{n_d}, \quad j = 1, \ldots, N_{lat}n_d \tag{19}$$

where N_{lat} is the total number of lanes, W_{lane} is the width of a lane. This rule enables detection of split-μ variation, i.e., cases where the two tires track within a single lane experience different peak TRFC values. As for the longitudinal partitioning rule, the preview road range $[s_{near}, s_{far}]$ is divided into N_{long} non-overlapping bins whose boundaries correspond to a set of preview-time nodes $t_1 < t_2 < \cdots < t_{N_{long}+1}$ at a reference speed v_{ref}:

$$s_n = v_{ref}\, t_n, \quad n = 1, \ldots, N_{long} + 1 \tag{20}$$

where the time nodes t_n are chosen with intervals $\Delta t_n = t_{n+1} - t_n$, so that the resulting bin widths $\Delta s_n = v_{ref}\Delta t_n$ increase monotonically with distance.

The validity of each ROI patch b is assessed using two independent criteria. First, its road-mask pixel ratio ρ_b^t must exceed a threshold τ_ρ. Then, its geometric usability G_b^t in the BEV view transition is checked. This is because patches

Just Accepted

Table 2 TRFC value ranges under different conditions.

Speed Range ζ_v	$[\underline{\mu}_c(\zeta_v), \overline{\mu}_c(\zeta_v)]$		
	Dry	Wet	Water
$v \leqslant 60$ km/h	$[0.75,\ 1.00]$	$[0.50,\ 0.75]$	$[0.35,\ 0.55]$
$60 < v \leqslant 100$ km/h		$[0.40,\ 0.65]$	$[0.25,\ 0.45]$
$v > 100$ km/h		$[0.35,\ 0.55]$	$[0.15,\ 0.30]$

near the image edges or bottom row may not map to a valid rectangle under IPM transition. Eventually, the validity indicator is concluded as:

$$V_b^t = \begin{cases} 1, & \rho_b^t \geqslant \tau_\rho \text{ and } G_b^t = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

ROI patches judged invalid are excluded from road friction condition classification, and are instead handled by the history maintenance rules in constructing the friction field.

4.3 Road friction condition classification

A three-class classifier considering "dry", "wet" and "water" is applied to valid ROI patches, each mapped to a baseline peak-TRFC interval $[\underline{\mu}_c^0, \overline{\mu}_c^0]$: $[0.75,\ 1.00]$ for "dry", $[0.35,\ 0.75]$ for "wet", and $[0.15,\ 0.55]$ for "water". Since wet and water surfaces exhibit markedly speed-dependent peak TRFC while dry surfaces do not [53], each baseline interval is further conditioned on a vehicle speed interval ζ_v to yield $[\underline{\mu}_c(\zeta_v), \overline{\mu}_c(\zeta_v)]$, giving the mapping relationship used to construct the conditional PFF, as summarized in Table 2.

4.3.1 Backbone and training configuration

The MobileNet-V3 model [50] is adopted as the backbone for onboard real-time road type classification, and the small variant is selected for its lower computational cost. The network employs depthwise separable convolutions, reducing computation to approximately $1/k^2$ of a standard convolution with kernel size k. It also incorporates a Squeeze-and-Excitation (SE) module that reweights channel responses to strengthen discriminative texture and reflection cues, aiding the distinction between wet and standing-water surfaces. The network further applies the Hard-Swish (H-Swish) activation in its latter stages, while earlier layers retain the ReLU activation. The full network structure, adapted with the three-class output head, is given in Table 3.

4.3.2 Confusion-aware training strategy

Since "wet" and "water" surfaces are visually similar, while "dry" surfaces often differ sharply from both, the "wet–water" confusion is explicitly penalized. The classifier outputs a softmax vector $\boldsymbol{p} = [p_1, p_2, p_3]$ over the three classes, and the true label is denoted $y \in \{1, 2, 3\}$. A confusion indicator matrix M is defined with $M_{23} = M_{32} = 1$ and all other

Table 3 MobileNet-V3 structure for road classification.

Layer	Op.	Exp. Ch.	Out. Ch.	SE / Act.
Conv2d	3×3 Conv	—	16	–/H-Swish
Bottleneck	3×3 DW	16	16	Y/ReLU
Bottleneck	3×3 DW	72	24	N/ReLU
Bottleneck	3×3 DW	88	24	N/ReLU
Bottleneck	5×5 DW	96	40	Y/H-Swish
Bottleneck (×2)	5×5 DW	240	40	Y/H-Swish
Bottleneck	5×5 DW	120	48	Y/H-Swish
Bottleneck	5×5 DW	144	48	Y/H-Swish
Bottleneck	5×5 DW	288	96	Y/H-Swish
Bottleneck (×2)	5×5 DW	576	96	Y/H-Swish
Conv2d	1×1 Conv	—	576	Y/H-Swish
Pool	GAP	—	576	—
Conv2d	1×1 Conv	—	1024	–/H-Swish
Classifier	FC/Softmax	—	3	—

entries equal to 0, marking the "wet-water" pair as the only penalized confusion. A Confusion-Aware Training Strategy (CATS) loss is then defined as:

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{conf} \tag{22}$$

where the first term is the standard cross-entropy loss $\mathcal{L}_{CE} = -\log p_y$, and the second term is a confusion penalty $\mathcal{L}_{conf} = \sum_{j \neq y} M_{yj}\, p_j$ that directly penalizes probability mass assigned to the confusable class only when y and j form a "wet–water" pair. The weight $\lambda_c > 0$ is determined via grid search on the validation set, jointly evaluating overall classification accuracy and the "wet–water" confusion rate. This strategy improves robustness on the most easily confused road types without any additional inference computation.

4.4 Construction of the preview friction field

4.4.1 Grid-wise field generation and update

For each ROI b with friction condition label c_b, classification confidence $p_b \in [0, 1]$, and observation quality $q_b \in [0, 1]$, a baseline friction range $[\underline{\mu}_b^0, \overline{\mu}_b^0]$ is assigned. The observation mean and uncertainty of each ROI b are formulated as:

$$m_b^0 = \tfrac{1}{2}(\underline{\mu}_b^0 + \overline{\mu}_b^0) \tag{23a}$$

$$P_b^0 = \tfrac{1}{12}(\overline{\mu}_b^0 - \underline{\mu}_b^0)^2 + \alpha_p(1 - p_b) + \alpha_q(1 - q_b) \tag{23b}$$

where the uniform-interval variance term captures the intrinsic dispersion of TRFC within a road-condition class, and $\alpha_p, \alpha_q > 0$ introduce visual reliability into the uncertainty, so that ROIs with low confidence or quality are downweighted in further grid-wise mapping.

The ROI region is partitioned into a set of grid cells, over which the PFF is defined, with spatial extent $\Omega = \{(s, d) : s \in [s_{\min}, s_{\max}], d \in [d_{\min}, d_{\max}]\}$ and discretization resolution $(\Delta s, \Delta d)$ matched to the sampling interval of the vehicle

Just Accepted

trajectory planner. Because a grid cell (i,j) may spatially overlap with several ROIs, the ROI-level observations are fused into a single grid-level observation. Let $\mathcal{B}_t^{ij}$ denote the set of ROIs overlapping cell (i,j) at frame t, and $\omega_{ij}^{b,t}$ indicate the corresponding coverage weight. Then the instantaneous grid observation for frame t is obtained by inverse-variance weighting when $\mathcal{B}_t^{ij} \neq \varnothing$:

$$\bar{m}_{0,ij}^t = \frac{\sum_{b\in\mathcal{B}_t^{ij}} \omega_{ij}^{b,t}(P_b^0)^{-1} m_b^0}{\sum_{b\in\mathcal{B}_t^{ij}} \omega_{ij}^{b,t}(P_b^0)^{-1}} \tag{24}$$

$$\bar{P}_{0,ij}^t = \Big(\sum_{b\in\mathcal{B}_t^{ij}} \omega_{ij}^{b,t}(P_b^0)^{-1}\Big)^{-1} \tag{25}$$

and the dominant class of the grid $c_{ij}^{t,\mathrm{obs}}$ is determined by a weighted vote among the overlapping ROIs, accounting for coverage and confidence.

Each grid cell then maintains a persistent state $\mathcal{G}_{ij}^t = (m_{0,ij}^t, P_{0,ij}^t, c_{ij}^t, N_{ij}^{\mathrm{miss},t})$, where $m_{0,ij}^t$ and $P_{0,ij}^t$ are the current best estimate of the TRFC mean and its uncertainty, c_{ij}^t is the current label, and $N_{ij}^{\mathrm{miss},t}$ counts the number of consecutive frames up to t in which the cell has received no valid ROI observation. In conclusion, $\mathcal{G}_{ij}^t$ is a filtered quantity that carries the cell's belief forward through time and is the information actually used for autonomous driving.

Three rules are designed to govern the PFF's update. First, if the grid cell can be observed ($\mathcal{B}_t^{ij} \neq \varnothing$), the historical uncertainty is inflated as $P_{0,ij}^{t|t-1} = \min\left(P_{0,ij}^{t-1} + q_0,\, P_0^{\mathrm{init}}\right)$, where $q_0 > 0$. Then it is fused with the instantaneous observation using a Kalman-style update method:

$$P_{0,ij}^t = (1 - K_{ij}^t)\, P_{0,ij}^{t|t-1} \tag{26a}$$

$$m_{0,ij}^t = m_{0,ij}^{t-1} + K_{ij}^t\left(\bar{m}_{0,ij}^t - m_{0,ij}^{t-1}\right) \tag{26b}$$

where $K_{ij}^t = P_{0,ij}^{t|t-1}/(P_{0,ij}^{t|t-1} + \bar{P}_{0,ij}^t)$, with $c_{ij}^t = c_{ij}^{t,\mathrm{obs}}$ and $N_{ij}^{\mathrm{miss},t}$ reset to 0. Second, if the cell is unobserved but still holds a valid history ($\mathcal{B}_t^{ij} = \varnothing$, $P_{0,ij}^{t-1} < P_0^{\mathrm{init}}$), the mean and label are carried over while the uncertainty inflates at a rate accelerated by $q_m > 0$ with the consecutive miss count, driving a stale belief toward P_0^{init}:

$$m_{0,ij}^t = m_{0,ij}^{t-1}, \quad c_{ij}^t = c_{ij}^{t-1} \tag{27a}$$

$$N_{ij}^{\mathrm{miss},t} = N_{ij}^{\mathrm{miss},t-1} + 1 \tag{27b}$$

$$P_{0,ij}^t = \min\left(P_{ij}^{t-1} + q_0 + q_m N_{ij}^{\mathrm{miss},t-1},\, P_0^{\mathrm{init}}\right) \tag{27c}$$

Last, if the cell has never been observed or its uncertainty has saturated at P_0^{init}, it is reset to a conservative initial state:

$$m_{0,ij}^t = m_0^{\mathrm{init}}, \quad P_{0,ij}^t = P_0^{\mathrm{init}} \tag{28a}$$

$$c_{ij}^t = \text{unknown}, \quad N_{ij}^{\mathrm{miss},t} = 0 \tag{28b}$$

4.4.2 *Field stability gating*

The grid-wise PFF is constructed anchored to the reference-line detection. Thus, transient reference-line distortion during lane changes or obstacle-avoidance maneuvers would contaminate the field with constant non-observations, leading to storing wrong information in the PFF. A stability flag $\Lambda_t \in \{0,1\}$ is therefore introduced to gate all grid cell updates, based on two indicators. The ratio of valid detection points detected on lane boundaries is formulated as:

$$r_{anchor,t} = \frac{1}{2}\left(\frac{N_{L,t}}{N_{\max}} + \frac{N_{R,t}}{N_{\max}}\right) \in [0,1], \tag{29}$$

where $N_{L,t}$, $N_{R,t}$ are the numbers of valid detected points on the left and right lane boundaries, and $N_{\max}$ is the theoretical maximum per side. The other indicator is the lateral jump of the reference line between frames:

$$\Delta d_{ref,t} = d_{C_t} - d_{C_{t-1}}. \tag{30}$$

The stability flag is then defined as:

$$\Lambda_t = \begin{cases} 1, & r_{anchor,t} \geqslant r_{th} \text{ and } |\Delta d_{ref,t}| \leqslant \Delta d_{th} \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

where r_{th} and d_{th} are respectively the thresholds of the two indicators. N_{stable} consecutive qualifying frames are required before Λ_t is permitted to switch from 0 to 1.

When $\Lambda_t = 0$, the field degrades to a pure maintenance process, where the historical mean and label are retained while the uncertainty grows monotonically with the duration of the missed-observation period, $P_{0,ij}^t = \min(P_{0,ij}^{t-1} + q_0 + q_m N_{ij}^{miss,t-1},\, P_0^{init})$. This preserves historically correct field states rather than contaminating them, while continuously signaling to the planner that the current friction information is becoming less trustworthy.

4.4.3 *Conditional field considering speed*

The baseline PFF is constructed independently of vehicle speed, driven solely by visual perception. But different vehicle speed significantly effects the peak TRFC especially on water-covered surfaces. Thus, a speed-related query operator $\mathcal{Q}(\cdot)$ is defined as:

$$\left(m_{ij}^t(\zeta_v),\, P_{ij}^t(\zeta_v)\right) = \mathcal{Q}\left(c_{ij}^t,\, m_{0,ij}^t,\, P_{0,ij}^t,\, \zeta_v\right) \tag{32}$$

which, for "dry" class that is insensitive to speed, simply passes through the baseline estimate,

$$m_{ij}^t(\zeta_v) = m_{0,ij}^t, \quad P_{ij}^t(\zeta_v) = P_{0,ij}^t \tag{33}$$

And for speed-sensitive "wet" and "water" classes, the operator looks up the conditional interval $[\underline{\mu}_c(\zeta_v), \overline{\mu}_c(\zeta_v)]$ from Table 2 and reconstructs the PFF:

$$m_{ij}^t(\zeta_v) = \frac{\underline{\mu}_c(\zeta_v) + \overline{\mu}_c(\zeta_v)}{2},\; P_{ij}^t(\zeta_v) = \frac{(\overline{\mu}_c(\zeta_v) - \underline{\mu}_c(\zeta_v))^2}{12} \tag{34}$$

Just Accepted

4.5 Vision-based wheel-end estimation

To obtain the wheel-end TRFC estimation from the constructed PFF, he positions of the four wheels should be predicted first. Considering highway driving scenarios, the heading deviation between the vehicle and the reference-line tangent is small and is neglected. The vehicle's longitudinal advance between two image frames is defined as $v_t \Delta t_{sys}$, where Δt_{sys} denotes the total system latency including sensor transmission, BiSeNet-V2 and MobileNet-V3 inference, and coordinate transformation.

Let the vehicle's Frenet longitudinal coordinate be s_{ego}, lateral offset be d_{ego}, wheelbase distance to wheel w be l_w (positive for front, negative for rear), and axle track width be B_w. The predicted position of each wheel $w \in \{fl, fr, rl, rr\}$ in the next frame is:

$$(s_w, d_w) \approx \Big(\max\big(s_{ego} + l_w + v_t \Delta t_{sys},\ s_{\min}\big), \\ d_{ego} + \mathrm{sgn}_w \tfrac{B_w}{2} \Big) \tag{35}$$

where $\mathrm{sgn}_w = +1$ for a left wheel and -1 for a right wheel, and $\max(\cdot, s_{\min})$ clamps the rear axle to the PFF's near boundary.

Each (s_w, d_w) is mapped to its grid cell by:

$$\mathcal{I}(s,d) = \left(\left\lfloor \frac{s - s_{\min}}{\Delta s} \right\rfloor, \left\lfloor \frac{d - d_{\min}}{\Delta d} \right\rfloor \right) \tag{36}$$

and, combined with the current vehicle-speed interval $\zeta_{v,t}$, the speed-conditioned query of Eq. (32) is executed to obtain the visual estimate for each wheel:

$$(\mu^{vis}_{w,t}, P^{vis}_{w,t}) = (m^t_{ij}(\zeta_{v,t}), P^t_{ij}(\zeta_{v,t}))\big|_{(i,j)=\mathcal{I}(s_w,d_w)} \tag{37}$$

5 Fusion of visual and dynamics estimation

The dynamics-based A-CKF estimator and the vision-based PFF are two TRFC information sources, each with its own applicability conditions and limitations, requiring dynamic coordination based on their respective current reliability. This section first derives the reliability measure of each estimate, and then establishes a dual-confidence fusion mechanism, which is referred to as the Fusion method. The spatiotemporal synchronization between the two heterogeneous estimation modules is mainly achieved via Eq. (35). When the vehicle moves forward, the visual estimate $\mu^{\text{vis}}_{w,t}$ at time t is obtained by querying the previously recorded friction value from the PFF based on the predicted wheel position (s_w, d_w), where the system latency Δt_{sys} is compensated. The previewed road friction and the real-time dynamic estimates are then spatially aligned at the four wheels and fused at around 10 Hz.

Two reliability measures are used to control the fusion. The A-CKF's identifiability of peak TRFC requires the tire to operate near force saturation, which is excited by lateral maneuvers. Thus, the dynamics reliability is defined from lateral acceleration a_y and yaw rate r as:

$$\rho^{dyn}_{raw,t} = \tanh\left(\frac{a^2_{y,t} + r^2_t}{\sigma^2_{dyn}} \right) \tag{38}$$

where σ_{dyn} is a calibrated scale constant. Meanwhile, the visual reliability reflects the freshness and consistency of the queried grid cell in the PFF, decaying with its uncertainty:

$$\rho^{vis}_{raw,w,t} = \exp\big(-P^{vis}_{w,t} / P_{ref}\big) \tag{39}$$

where P_{ref} is a reference uncertainty scale. This reliability indicator decays automatically when the field generation is locked, since $P^{vis}_{w,t}$ grows under this situation.

However, both reliabilities are prone to high-frequency oscillation. This arises from sensor noise on the dynamics side, and confidence fluctuation between frames on the vision side. Thus, substituting them directly into the fusion weight would cause the fused estimation to exhibit meaningless jumps. An exponential moving average is therefore applied to each reliability:

$$\bar{\rho}^{(\cdot)}_t = \alpha_{(\cdot)} \bar{\rho}^{(\cdot)}_{t-1} + (1 - \alpha_{(\cdot)}) \rho^{(\cdot)}_{raw,t} \tag{40}$$

where $(\cdot) \in \{dyn, vis\}$, with smoothing coefficients $\alpha_{dyn}, \alpha_{vis} \in [0, 1)$ controlling memory depth. The smoothed reliabilities are combined into the fusion weight, and the final fused estimate is formulated as:

$$\gamma_{w,t} = \frac{\bar{\rho}^{vis}_{w,t}}{\bar{\rho}^{vis}_{w,t} + \bar{\rho}^{dyn}_t + \epsilon} \tag{41a}$$

$$\mu_{w,t} = \gamma_{w,t}\, \mu^{vis}_{w,t} + (1 - \gamma_{w,t})\, \mu^{dyn}_{w,t} \tag{41b}$$

where $\epsilon > 0$ is a small constant preventing division by zero.

Under weak lateral excitation with a stable PFF, $\gamma_{w,t}$ is close to 1, and the estimate is visual-dominant. Under strong lateral excitation with a locked field, $\gamma_{w,t}$ is close to 0, and the estimate beconmes dynamics-dominant. When both sources are unreliable, $\gamma_{w,t} \approx 0.5$, thus yielding a conservative average. Because the weight varies continuously and smoothly with $\bar{\rho}^{dyn}_t$ and $\bar{\rho}^{vis}_{w,t}$, the fused estimate transitions gradually between the two regimes rather than switching discontinuously. For instance, when the vehicle starts a lane change from straight driving, $\rho^{dyn}_{raw,t}$ rises sharply, but the smoothed weight avoids an abrupt jump in estimation.

6 Experimental Results and Analysis

6.1 Validation of visual perception reliability

To validate the reliability of the proposed visual perception module, we conduct comparative experiments on the classifier proposed in Section 4.3, using a self-built dataset exclusively for the road surface friction condition classification.

Just Accepted

6.1.1 Dataset preparation

To build the dataset, original 1920×1080 road images captured by a passenger vehicle's forward-looking camera are processed through ROI partitioning and validity assessment and transformed into 224×224 pixel road ROI patches in the BEV format. The training and validation sets are collected at the same test field with various road segments, but at different times of the year. Specifically, the training set is collected during the spring rainy season, and the validation set is derived from data collected during the autumn rainy season. The total number of training samples for each category is as follows: 1,671 for "dry", 1,628 for "wet", and 1,003 for "water". While the validation set comprises 200 "dry", 400 "wet", and 150 "water" samples. Our preliminary training results indicated that this slightly imbalanced distribution does not significantly degrade the model's classification performance.

To enhance the robustness of the model against varying real-world illumination conditions, brightness perturbation augmentation is applied to the original training set. Specifically, one brightness-enhanced and one brightness-reduced version are generated for each original sample. This procedure triples the effective training sample size, yielding a total of 12,906 road patches for training. Additionally, horizontal flipping does not alter the physical category or semantic meaning of the road surface friction condition. Therefore, a random horizontal flip with a probability of 0.5 is applied to each training sample during the training phase. This enhancement further improves the model's generalization capabilities.

6.1.2 Comparative evaluation design

To comprehensively evaluate the performance of our proposed MobileNet-V3 + CATS framework, we selected three representative models as baselines: standard MobileNet-V3, ShuffleNet-V2 equipped with CATS, and Dilated Residual Networks (DRN) equipped with CATS. Comparing our method with the standard MobileNet-V3 serves as an essential ablation study to validate the effectiveness of the proposed CATS module in mitigating misclassifications between similar "wet" and "water" classes. ShuffleNet-V2 is chosen due to its highly efficient, lightweight convolutional architectures designed specifically for edge deployment. Moreover, DRN is selected as a heavier baseline because it is adept at maintaining high spatial resolution and retaining rich structural information without losing the receptive field. We incorporate CATS into these two diverse network backbones to obtain fair evaluation results.

To quantitatively measure the classification performance, Overall Accuracy (OA) and Macro $F1$ score are adopted as our primary evaluation metrics. OA provides an intuitive assessment of the model's global prediction capability, which is defined as:

$$OA = \frac{\sum_{i=1}^{C} TP_i}{N} \times 100\% \tag{42}$$

where C is the total number of classes, TP_i represents the True Positives for class i. And N is the total number of evaluated samples. While OA reflects general performance, it can be biased toward majority classes in imbalanced datasets. To address this, we additionally utilize the Macro $F1$ score, which calculates the arithmetic mean of the $F1$ scores across all classes independently. The Macro $F1$ score ensures that minority classes are given equal weight, thus providing a more comprehensive evaluation of the model's robustness. It is formulated as:

$$\text{Macro } F1 = \frac{1}{C} \sum_{i=1}^{C} \left(\frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \right) \times 100\% \tag{43}$$

where $Precision_i$ and $Recall_i$ respectively denote the precision and recall for the i class.

Table 4 Comparisons of classification performance.

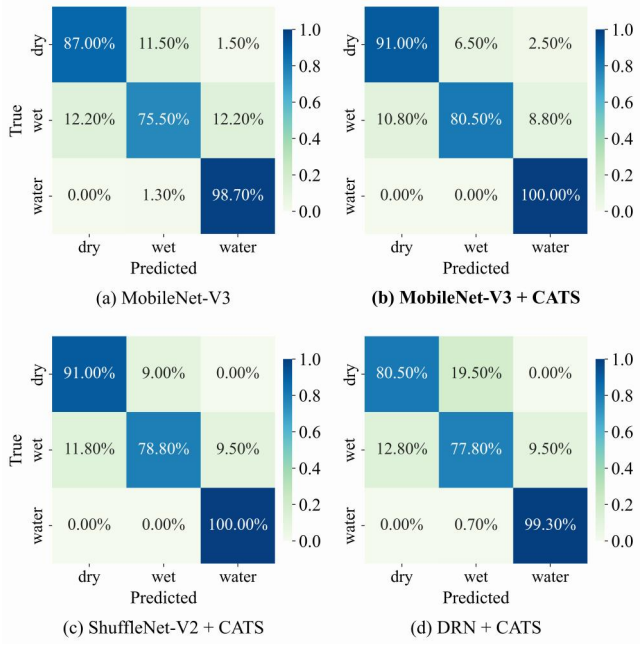
Method	OA (%)	Macro $F1$ (%)
MobileNet-V3	83.20	83.31
MobileNet-V3 + CATS	87.20	87.17
ShuffleNet-V2 + CATS	86.27	86.52
DRN + CATS	82.80	83.14

6.1.3 Results and analysis

Table 4 compares the classification performance of models trained under different frameworks, showing that MobileNet-V3 + CATS achieves the best result. ShuffleNet-V2 + CATS reaches an OA of 86.27% and a Macro $F1$ of 86.52%, which is close to our classifier, but its model complexity is higher than MobileNet-V3. In contrast, DRN + CATS shows the weakest performance. This is because DRN is more suitable for tasks that need dense pixel-level prediction, such as semantic segmentation. For the classification task on 224×224 road ROI patches in this study, the high spatial resolution and deeper structure of DRN do not bring performance gains; instead, they may introduce redundant information and make the model harder to fit with limited training samples. Overall, MobileNet-V3 with CATS achieves the best classification accuracy among the three while keeping a lightweight structure, offering a good balance between efficiency and accuracy.

The confusion matrices of all methods are shown in Fig.5, which are used to check whether the CATS module can reduce confusion between "wet" and "water". Taking MobileNet-V3

Just Accepted



(a) MobileNet-V3 **(b) MobileNet-V3 + CATS**

(c) ShuffleNet-V2 + CATS (d) DRN + CATS

Fig. 5 Comparisons of the confusion matrix.

as an example, without CATS, 1.30% of "water" samples are misclassified as "wet". After adding CATS, this rate drops to 0%, and the confusion is nearly eliminated. Meanwhile, the rate of "wet" samples misclassified as "water" also drops from 12.20% to 8.80%, but this direction of confusion remains. This suggests that CATS is more effective in helping the model learn discriminative features for the "water" class, which greatly reduces the chance of it being mistaken for "wet". A similar pattern is observed in ShuffleNet-V2 + CATS and DRN + CATS, where the "water"-to-"wet" misclassification rate also stays close to 0%. This suggests that the effect of CATS in reducing "water"-to-"wet" confusion does not depend on a specific backbone structure, showing good generality.

6.2 Experiments on TRFC fusion estimation

To verify the accuracy, response speed to sudden changes, convergence speed, robustness, and stability of the Fusion method, we design various scenarios to conduct comparative experiments for TRFC estimation.

6.2.1 Experimental setup

The experiments are conducted on a Hongqi E-QM5 vehicle, whose key characteristics are listed in Table 5. During the road tests, the proposed estimation framework is executed online on an onboard computing platform running the Ubuntu operating system. The forward-facing camera acquires 1920×1080 images of the road ahead at 30 Hz, while vehicle-state signals are acquired from the IMU and ESP system at around 100 Hz.

Table 5 Vehicle configuration.

Parameters	Notation	Nominal values	Unit
Vehicle mass	m	1750	kg
Distance from CG to front axle	l_f	1.196	m
Distance from CG to rear axle	l_r	1.794	m
Height of CG	h_g	0.55	m
Front track width	t_f	1.654	m
Rear track width	t_r	1.655	m
Tire size	-	225/55 R17	-
Powertrain layout	-	Front-engine and front-wheel drive	-



(a) Test vehicle (b) Test field

Fig. 6 Test vehicle and field.

The visual perception module generates conditional PFF estimates at approximately 10 Hz. In parallel, the dynamics-based A-CKF is executed at approximately 100 Hz using synchronized vehicle-state measurements. The confidence-aware Fusion method generates the final wheel-level TRFC estimates at 10 Hz. All sensor signals and estimation results are time-stamped and recorded during the tests. All experiments are carried out on a three-lane, straight asphalt road segment with a length of 1.1 km. The test vehicle, along with the test field, is shown in Fig. 6.

To effectively evaluate the respective performance of the visual perception module and the dynamic estimator, as well as the overall performance of their fusion model under specific scenarios, the experimental scenarios are designed considering two aspects: the intensity of lateral excitation acting on the vehicle and the difficulty of distinguishing road types. From the perspective of lateral excitation, high-excitation maneuvers, such as slalom and double lane change, are set up, while low-excitation maneuvers that make it difficult for the dynamic estimator to estimate the peak TRFC are also included. For the small-excitation scenarios, which primarily test the capability of the visual perception module, split-μ conditions composed of different road surface types are configured.

The baseline estimation models employed for comparison with the proposed Fusion method are the classic UKF, the

Just Accepted

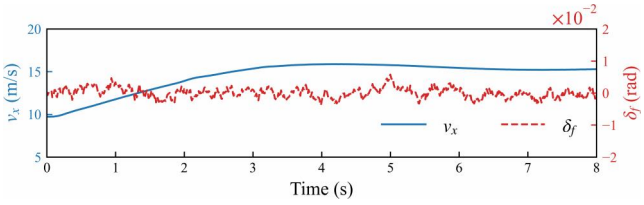


Fig. 7 Vehicle speed and steering wheel angle in S1.

A-CKF, and the PFF. UKF and A-CKF are purely dynamics-based, whereas PFF is purely vision-based. The TRFC reference true values are calculated by the ABS brake tests on the corresponding road surfaces.

As for the ROI partitioning when constructing the PFF, we consider the ego lane with the left and right adjacent lanes. Specifically, $N_{lat} = 3$, $n_d = 2$, $W_{lane} = 3.75$ m, giving $d_j = \pm 4.6875, \pm 2.8125, \pm 0.9375$ m. Longitudinally, $[s_{near}, s_{far}] = [5, 80]$ m and $N_{long} = 4$, with bin boundaries at 5, 20, 35, 55, and 80 m, yielding $N_{lat} n_d \times N_{long} = 24$ ROI patches per field. And the resolution of the grid cell is $\Delta s = 2$ m, $\Delta d = 1$ m.

6.2.2 *Scenario 1: straight driving on dry-wet split-μ road surface*

A split-μ straight-line driving test is conducted at a closed proving ground. The left wheels travel on dry asphalt, while the right wheels travel on wet asphalt. Before the main experiment, ABS braking tests are conducted on both surfaces to obtain the reference ground truth. The reference TRFC are approximately 0.78 for the left dry surface and 0.65 for the right wet surface. This experiment is referred to as Scenario 1 (S1) in the following content. Fig. 7 shows the longitudinal vehicle speed and steering-wheel angle in S1. From 0 to 3 s, the vehicle accelerates from 10 m/s to approximately 15 m/s. It then maintains an approximately constant speed. The steering wheel remains stable during the test, and no intentional steering input is applied. Fig. 8 shows the vehicle states during the test. The lateral velocity, lateral acceleration, and yaw rate remain at low levels. Therefore, S1 represents a typical condition with limited lateral dynamic excitation.

At the initial instant of S1, the forward-facing camera captures the image shown in Fig. 9. Fig. 9 (b) and (c) present the ROI extraction and validity verification procedures, respectively. The valid ROI sub-blocks are subsequently classified, as illustrated in Fig. 9 (d). Three categories of conditional PFF are then obtained, as shown in Fig. 9 (e). The results show that the proposed PFF algorithm can effectively distinguish dry and wet asphalt ROIs, and generate a grid-level friction distribution for the road ahead. Fig. 10 presents the classification confidence and results provided by the visual perception module for the road surfaces beneath the four

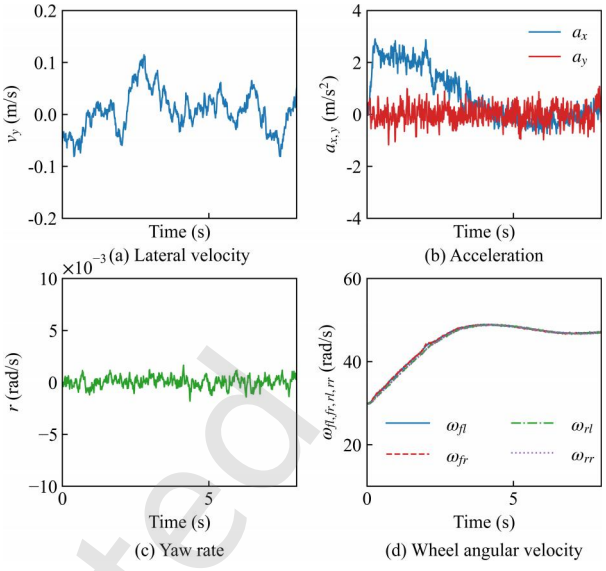


Fig. 8 Vehicle dynamic responses in S1.

wheels. The wet surface on the right side is identified with high confidence throughout the test. In contrast, some ROIs on the dry surface on the left side are misclassified as wet. However, these misclassifications have low confidence values. Therefore, the confidence-based fusion strategy reduces their influence on the fusion results.

The TRFC estimation results for S1 are shown in Fig. 11. The purely dynamics-based UKF and A-CKF methods have difficulty obtaining accurate estimates under the limited lateral excitation in S1. In particular, the UKF does not provide effective updates for either rear wheel, and its estimates remain at the initial value of 0.8 throughout the test. The A-CKF responds to the longitudinal excitation during the initial acceleration stage. Its estimates begin to update from 0.8, but their accuracy remains limited. These results indicate that the dynamics-based methods provide insufficient TRFC estimation performance under the S1 condition.

In comparison, the vision-based PFF provides accurate and stable friction updates after 3 s. During the first 3 s, the TRFC estimation errors for the left front and left rear tires mainly result from two factors. First, during the early stage of this interval, the ROIs located from 5 m to 20 m ahead of the initial vehicle position do not satisfy the geometric validity requirement, as shown in Fig. 9(d). Therefore, friction information cannot be obtained from these regions. Second, the dry surface on the left side is initially misclassified as wet, as shown in Fig. 10. The corresponding confidence values are low. As a result, the vision-based estimates fluctuate and remain relatively low during the early stage. Overall, the vision-based PFF compensates for the limited effectiveness of the dynamics-based estimates. Consequently, the proposed

Just Accepted

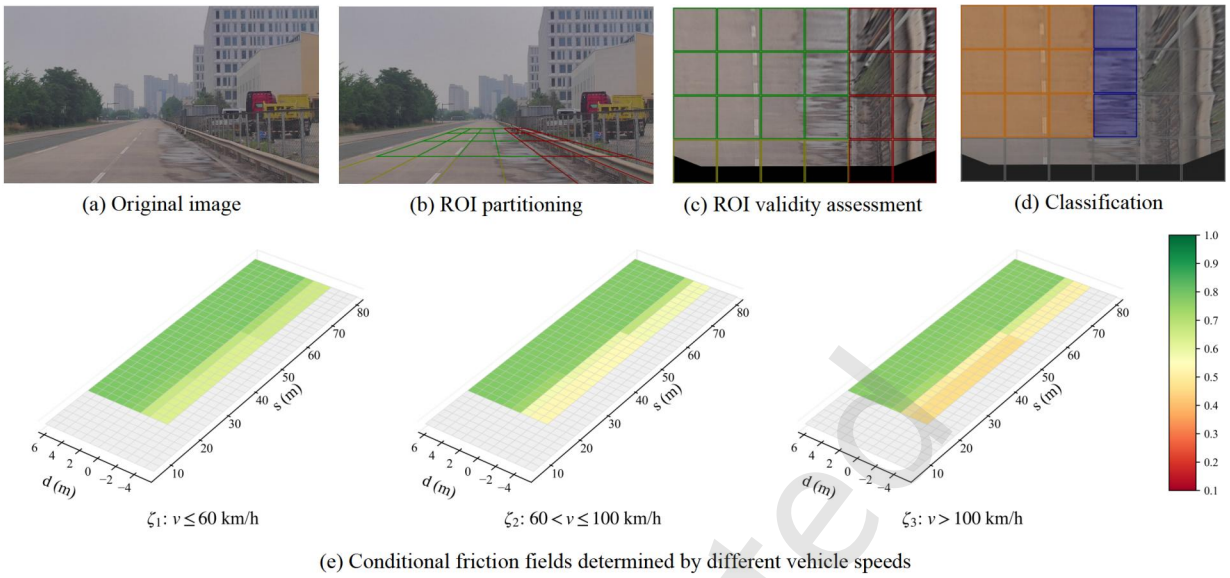


Fig. 9 Visual perception results at the initial time of S1. The bounding box colors in (b) and (c) denote the ROI patch assessment results: **green** indicates validity in both road area ratio and geometric usability; **yellow** signifies geometric invalidity during the FPV-to-BEV transformation; and **red** represents non-road areas with a road area ratio below the prescribed threshold. In (d), the **orange**, **blue**, and **purple** masks correspond to ROIs classified as "dry", "wet", and "water" surfaces, respectively. In (e), the color reflects the magnitude of μ within each grid cell. This visualization scheme is consistently applied across all scenarios.

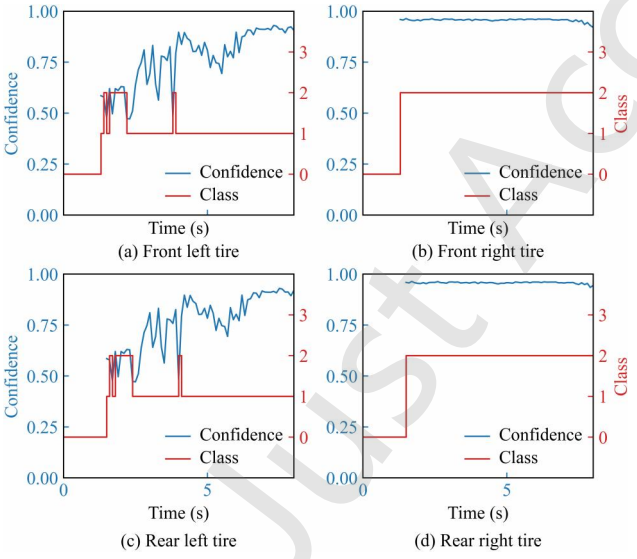


Fig. 10 Classification confidence and results in S1. For the classification results, types 0, 1, 2, and 3 indicate "unknown", "dry", "wet", and "water", respectively. This scheme is consistently applied across all scenarios.

fusion method provides reliable TRFC estimates.

6.2.3 *Scenario 2: straight driving on wet-water split-μ road surface*

A second split-μ test is conducted on a different road section at the same proving ground. This test is referred to as Scenario 2 (S2). The left wheels travel on water asphalt, while the right wheels travel on wet asphalt. ABS braking tests are conducted to obtain the reference TRFC values, which are approximately 0.55 for the left water surface and 0.38 for the right wet surface. Fig. 12 shows the longitudinal vehicle speed and steering-wheel angle in S2. From 0 to 3.5 s, the vehicle accelerates from 15.4 m/s to 21.5 m/s. The vehicle speed then remains between 21 m/s and 25 m/s. Fig. 13 shows the vehicle states during the test, showing that the vehicle's lateral velocity, lateral acceleration, and yaw rate remain at low levels. A slight tendency of sideslip occurs during acceleration, but its magnitude is limited. Therefore, S2 also represents a condition with limited lateral dynamic excitation.

At the initial instant of S2, the forward-facing camera captures the image shown in Fig. 14 (a). Fig .14 (b)–(d) illustrate the ROI extraction, validity verification, and classification procedures, respectively. Three categories of conditional PFF are obtained, as shown in Fig .14 (e). The results show that the proposed PFF algorithm effectively distinguishes water-covered and wet asphalt ROIs. A comparison of Figs.15 (a) and (b) indicates that the water surface is identified effectively, with high confidence throughout the test. In contrast, wet ROIs are more frequently misclassified as water than dry ROIs are misclassified as wet in S1. This result is consistent with the classifier performance of the visual perception module shown in Fig. 5 (b). However, both classifications correspond to lower TRFC values. Moreover, the misclassifications do not persist for long periods. Therefore, their influence on the final estimates of the Fusion method remains limited.

The TRFC estimation results for S2 are shown in Fig. 16.

Just Accepted

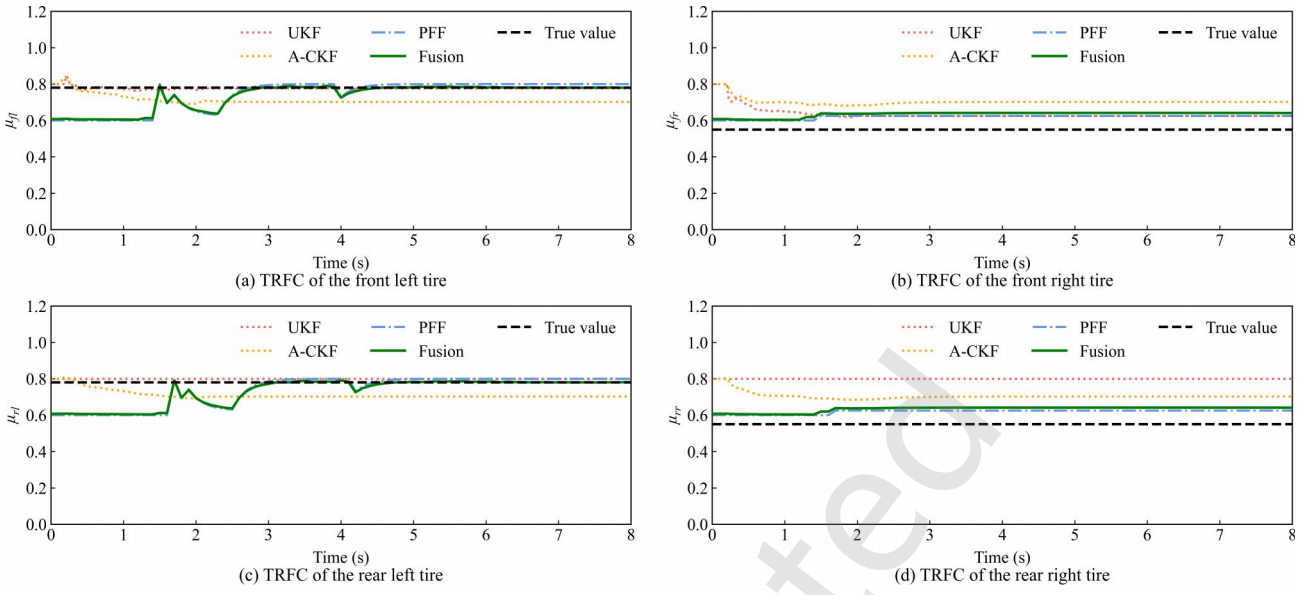


Fig. 11 TRFC estimation results in S1.

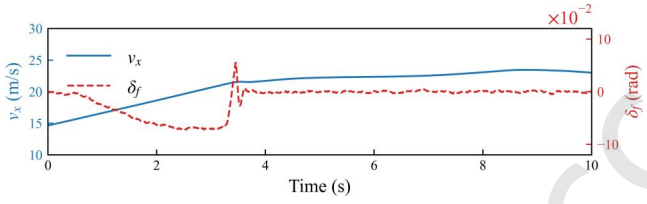


Fig. 12 Vehicle speed and steering wheel angle in S2.

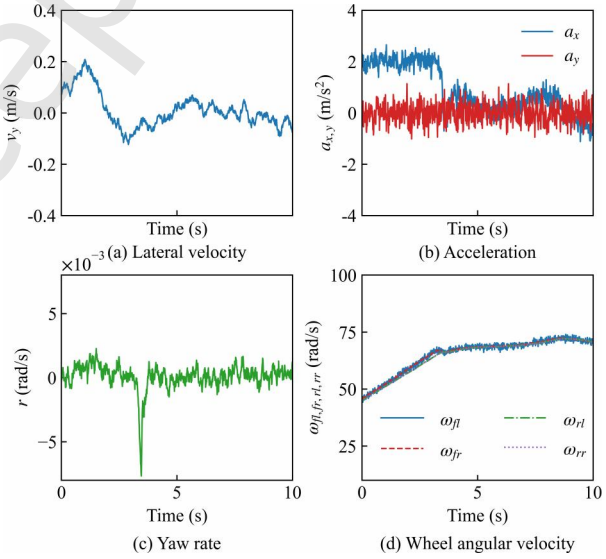


Fig. 13 Vehicle dynamic responses in S2.

Similar to S1, the UKF cannot accurately track the TRFC values of the left and right rear tires. However, the influence of this issue is more serious in S2. In S1, the reference TRFC values are 0.78 and 0.65, whereas those in S2 are 0.55 and 0.38. Therefore, the UKF estimate remaining at its default value of 0.8 may lead to substantial errors and could adversely affect driving safety. In comparison, the proposed A-CKF performs better. It responds clearly at approximately 0.2 s during vehicle acceleration. Under the lower-μ side with a TRFC of 0.38, its estimates are also closer to the reference values. The vision-based PFF provides accurate and stable estimates after 1 s and effectively reduces the estimation error. Overall, the Fusion method provides the best estimates among the evaluated methods in S2, mainly due to the contribution of the visual perception results.

6.2.4 *Scenario 3: slalom on uniform low-friction road surface*

A slalom maneuvering test is conducted on a low-friction road surface at a closed proving ground. The road surface is considered to have approximately uniform friction, and ABS braking tests yield a reference TRFC of 0.58. This experiment is referred to as Scenario 3 (S3) in the following content. Fig. 17 shows the longitudinal vehicle speed and steering-wheel angle in S3. During the whole 0–14 s test period, the vehicle speed varies to some extent, but remains between 10 m/s and 14 m/s during the slalom maneuver. To negotiate the cones, large and repeated steering inputs are applied. Fig. 18 shows the vehicle states during the test. It can be seen that the lateral velocity, lateral acceleration, and yaw rate remain at high levels in this scenario. Therefore, S3 represents a typical condition with strong and constant lateral dynamic excitation.

Fig. 19 (a) shows the image captured by the forward-facing camera at the initial instant of S3. As shown in Figs. 19 (c) and (d), certain ROI ahead of the vehicle fails the validity verification because road obstacles reduce the proportion of

Just Accepted

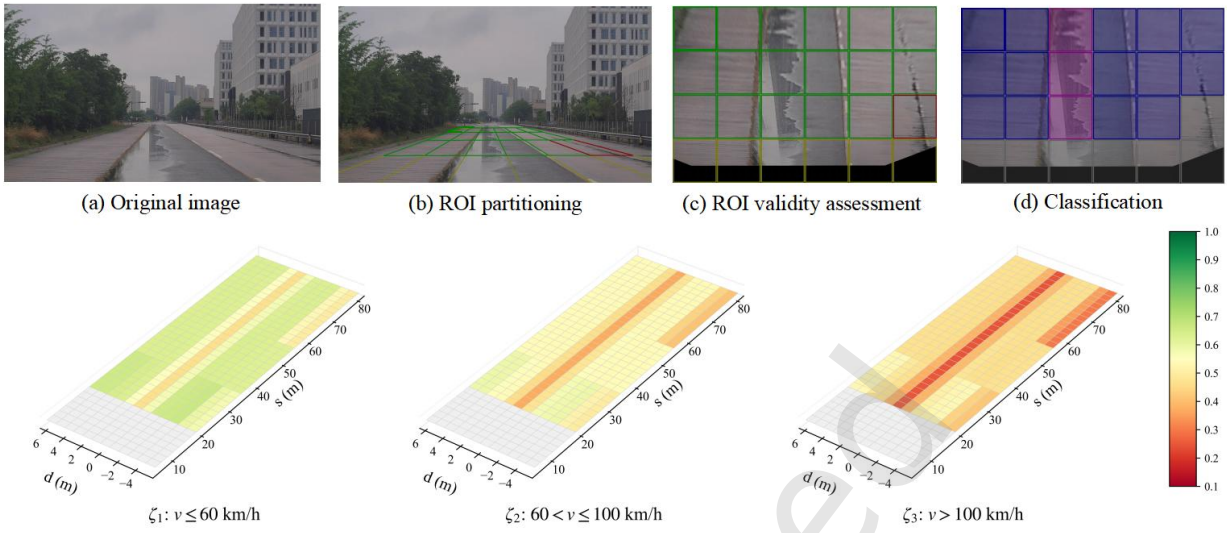


Fig. 14 Visual perception results at the initial time of S2.

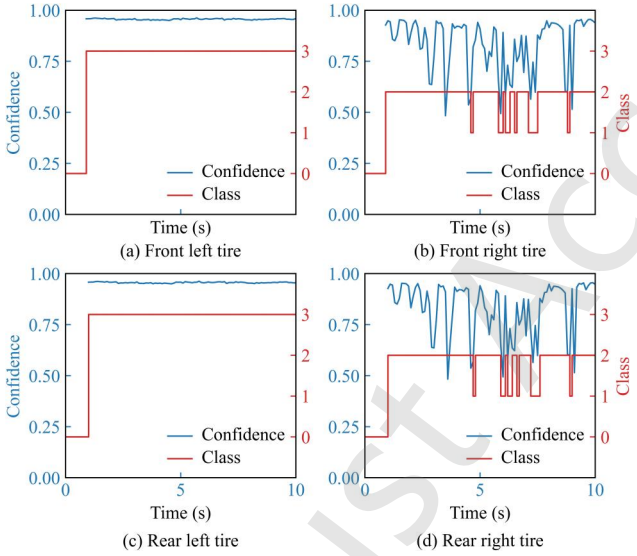


Fig. 15 Classification confidence and results in S2.

road-surface pixels within this ROI. This effect is also reflected in the three categories of conditional PFF shown in Fig. 19 (e). Fig.20 presents the valid anchor, the reference-line lateral jump, and the resulting field-lock gate during the construction of PFF. Owing to the constant and intensive lateral maneuver, the field-lock gate remains at zero throughout most of the test process. Therefore, the vision-based PFF is assigned limited weight in the Fusion framework. Specifically, the valid anchor remains at zero for most of the test and exceeds the ratio threshold 0.5 only occasionally when the vehicle heading is aligned with the road direction. Although the reference-line lateral jump remains at zero during several intervals, these intervals do not indicate successful reference-line updates. Instead, the most recently valid reference line is retained. Once the valid anchor exceeds the update ratio threshold, the reference line is updated, and the reference-line lateral jump immediately exceeds the jump threshold.

The TRFC estimation results for S3 are shown in Fig. 21. Under sustained strong lateral excitation, the purely dynamics-based UKF and the A-CKF can rapidly converge toward the reference TRFC. The UKF still responds more slowly for the rear tires than for the front tires. Moreover, as shown in Fig. 21 (d), a substantial step error reaching more than 1.2 occurs and persists for approximately 2 s. In comparison, the proposed A-CKF performs better. Although its estimates fluctuate during the test, they remain close to the reference value of 0.58. In addition, the A-CKF's estimates for all four tires converge rapidly, within approximately 0.5 s. Therefore, when the field-lock gate remains at zero, the dynamics-based estimation contributes more to the proposed Fusion method.

As for the vision-based PFF, the estimate updates from the default value of 0.6 to 0.64 after 3.2 s and remains at this value until 7.8 s. This update mainly results from the field-lock gate briefly remaining one during the initial 0–0.2 s interval, as shown in Fig. 20. During this interval, the friction information of the road segment located 20–80 m ahead of the initial vehicle position is generated and stored, as is illustrated in Fig. 19 (e). When the vehicle subsequently reaches this road segment at around 3.2 s, the PFF deploys the stored history information to estimate the real-time TRFC of four tires. However, these grids' friction is not updated for a long period. Thus the history information has very low confidence. Consequently, the Fusion method still assigns greater weight

Just Accepted

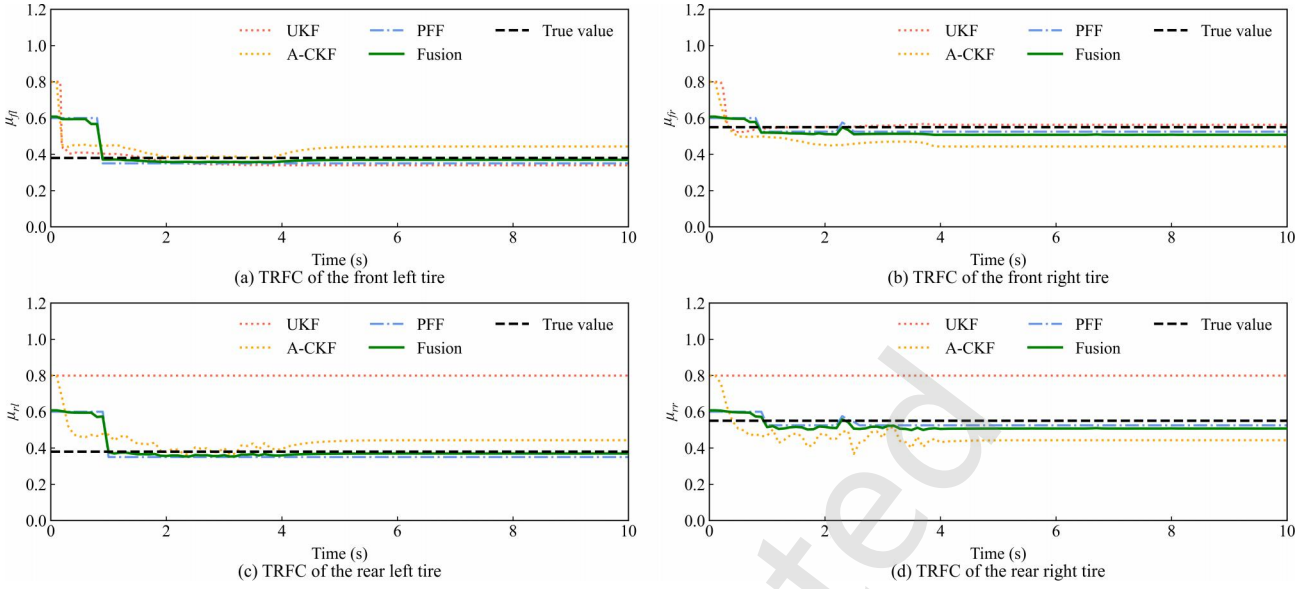


Fig. 16 TRFC estimation results in S2.

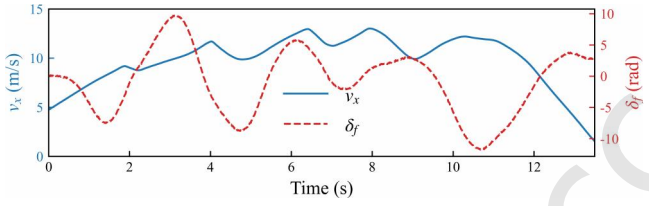


Fig. 17 Vehicle speed and steering wheel angle in S3.

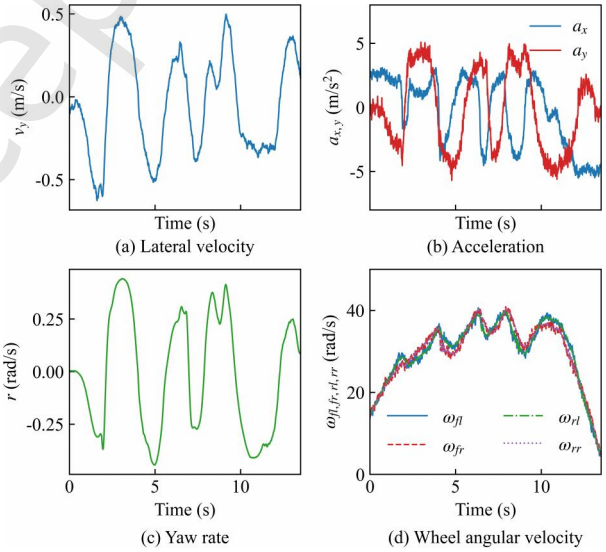


Fig. 18 Vehicle dynamic responses in S3.

to the A-CKF estimates.

The results of S1 and S2 show that, under limited lateral excitation, the Fusion method relies heavily on the visual perception module. Consequently, it provides stable and accurate estimates even when the A-CKF estimates have relatively large errors. In contrast, results in S3 demonstrate that, under sustained strong lateral excitation, where the field-lock gate remains close to zero and the vision-based estimation information is not trusted, the Fusion method relies more heavily on the dynamics-based A-CKF and still provides stable and reliable TRFC estimates.

6.2.5 *Scenario 4: double lane change on joint road surface*

Scenario 4, referred to as S4 in the following content, involves a double lane-change maneuver on a joint road surface. From the initial vehicle position to 115 m, the road surface is wet asphalt. Beyond 115 m, the surface changes to water asphalt. ABS braking tests provide reference TRFC values of approximately 0.60 and 0.42 for the two road sections, respectively. Fig. 22 shows the longitudinal vehicle speed and steering-wheel angle throughout the test. The vehicle speed remains approximately constant at 13.6 m/s. The first lane-change maneuver occurs between 4 s and 6 s, during which the vehicle moves from the center lane to the left lane. The second maneuver occurs between 10.5 s and 14 s, during which the vehicle returns from the left lane to the original center lane. Fig. 23 illustrates the vehicle states during the test, showing that the lateral velocity, lateral acceleration, and yaw rate remain at relatively high levels. Therefore, S4 represents a typical condition with strong lateral dynamic excitation.

At the initial instant of S4, the forward-facing camera captures the image shown in Fig. 24 (a). At this instant, the selected ROIs contain only the wet road section, whereas the water-covered section is outside the PFF construction range, as demonstrated in Figs.24 (c) and (d). In addition, no lane

Just Accepted

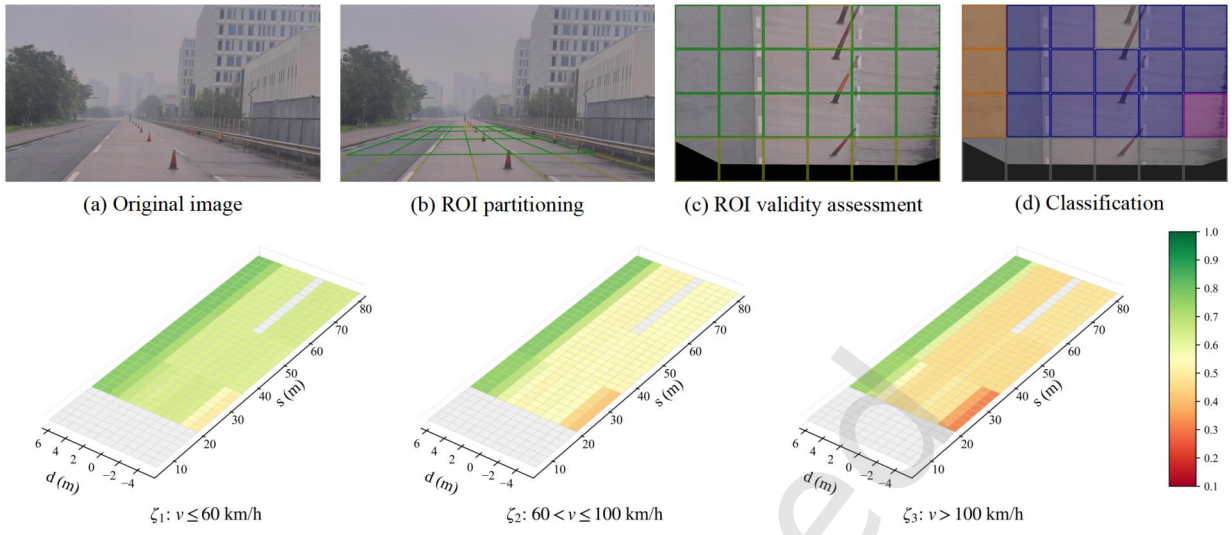


Fig. 19 Visual perception results at the initial time of S3.

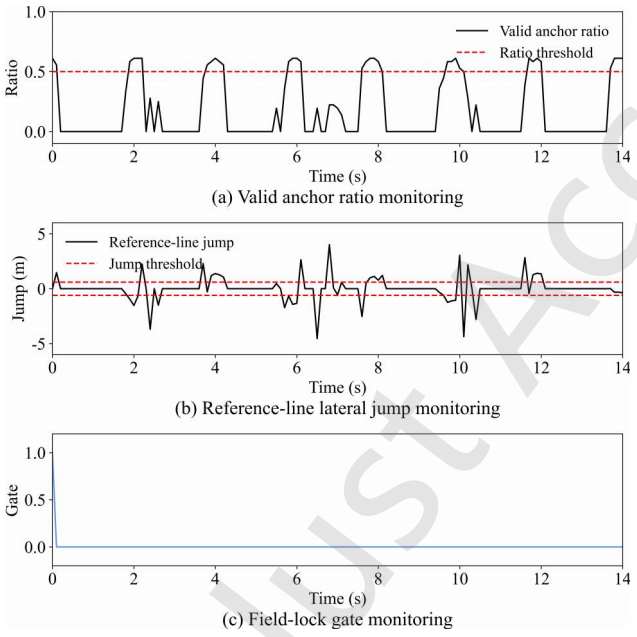


Fig. 20 Indicators related to the field-lock gate in S3.

exists on the right side of the vehicle at the initial position, and this region corresponds to a non-road area. Therefore, the initial conditional PFF does not contain friction information for the right-side region. After the first lane-change maneuver, the original center lane then becomes the left-side lane relative to the vehicle, and the PFF will subsequently generate friction information for the grids on the right side.

Fig. 25 presents the monitoring of the valid anchor, the reference-line lateral jump, and the resulting field-lock gate. During both lane-change maneuvers, the valid anchor ratio falls substantially below the ratio threshold and may even decrease to zero. Meanwhile, the reference-line lateral jump can reach 4.7 m when the steering-wheel angle is large. As a result, the field-lock gate decreases to zero during the lane changes. At other times, the gate remains equal to one, and the PFF keeps updating normally.

The TRFC estimation results for S4 are shown in Fig. 26. At approximately 8.4 s, the vehicle enters the water road surface, and the reference TRFC decreases abruptly from 0.60 to 0.42. For the purely dynamics-based UKF and A-CKF, however, the ability to track this TRFC change depends on the availability of sufficient lateral excitation. Without adequate lateral excitation, the dynamics-based estimators cannot converge toward the new TRFC value, even after the vehicle enters the lower-μ road section. In can be seen that the A-CKF provides higher estimation accuracy than the UKF. However, the update times of both methods are primarily associated with the two lane-change maneuvers. For example, as shown in Fig. 26 (a), the A-CKF correctly identifies the TRFC of the first wet road section at approximately 4 s, during the first lane change. It identifies the TRFC of the second water-covered road section at approximately 11 s, during the second lane change. Therefore, its response to the abrupt TRFC change is delayed by approximately 2.6 s.

The vision-based PFF effectively compensates for the delayed updates of the dynamics-based estimators. The PFF begins to update at approximately 1.7 s and provides friction information for the road ahead. During the first lane-change maneuver, the field-lock gate becomes zero. Thus, the PFF estimate is based on the most recently retained valid estimate rather than a newly updated PFF. At this time, the lateral

Just Accepted

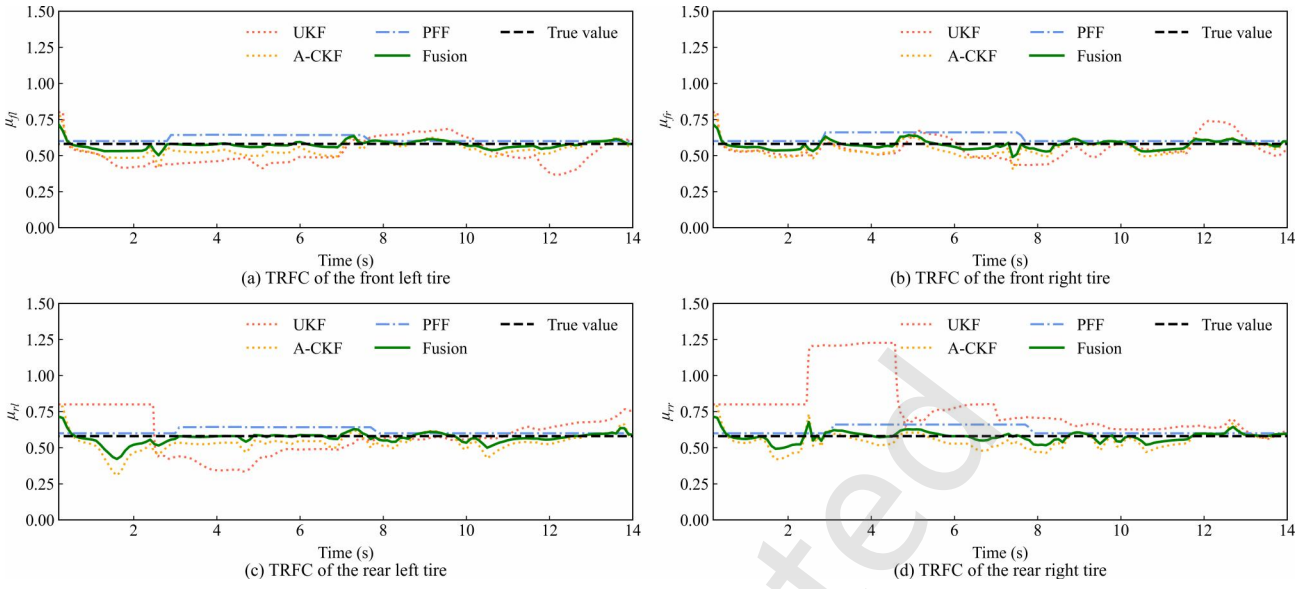


Fig. 21 TRFC estimation results in S3.

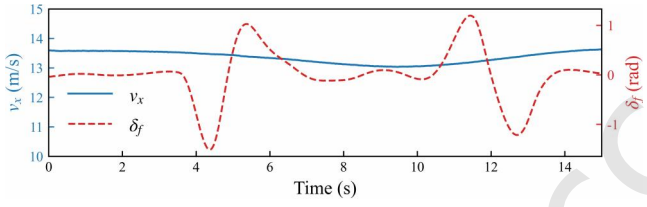


Fig. 22 Vehicle speed and steering wheel angle in S4.

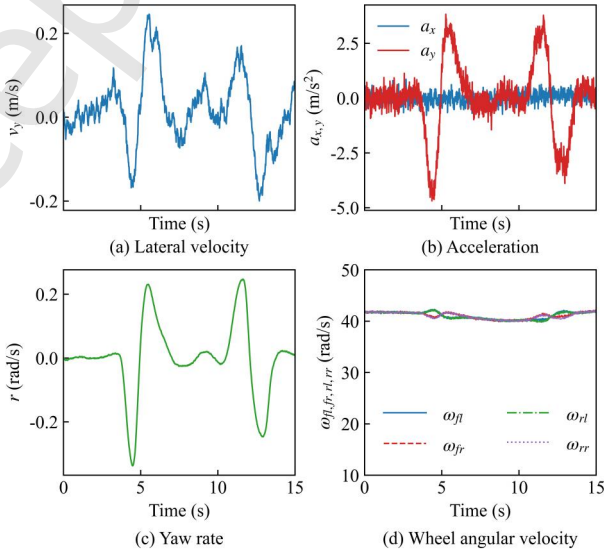


Fig. 23 Vehicle dynamic responses in S4.

excitation is sufficiently strong, and the A-CKF provides accurate estimates. Therefore, the Fusion method assigns greater weight to the dynamics-based estimation. After the first lane-change maneuver, the vehicle travels stably in the new lane. The PFF then rapidly resumes updating and obtains the friction information associated with the upcoming TRFC change. Consequently, at 8.4 s, although the dynamics-based estimators cannot immediately track the abrupt friction reduction because of insufficient dynamic excitation, the Fusion method assigns greater weight to the PFF estimate and correctly responds to the TRFC change. Meanwhile, during the second lane-change maneuver, the PFF update again becomes unreliable, and the Fusion result also relies more heavily on the A-CKF estimate.

Overall, S4 represents a composite condition involving an abrupt TRFC transition and complex lateral driving maneuvers. The proposed Fusion method effectively exploits the complementary strengths of the dynamics-based and vision-based estimations. By adaptively weighting the two sources according to their respective reliability, it produces TRFC estimates that are closer to the reference values, outperforming the baselines. These results demonstrate the robustness and stability of the proposed method.

6.2.6 *Overall peformance evaluation*

The overall performance across all scenarios is reported in Table 6. Notably, the purely visual estimation results are excluded, which inherently present step-like state transitions rather than continuous estimates. Applying conventional continuous time-series error metrics would not yield a physically meaningful comparison alongside the dynamic estimators and the proposed Fusion method. Typical evaluation metrics are deployed, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) [54, 55]. The three metrics are defined as:

$$\mathrm{MAE}_{\mu}^{j} = \frac{1}{N}\sum_{t=0}^{N-1}|e_{j,t}|, \tag{44a}$$

Just Accepted

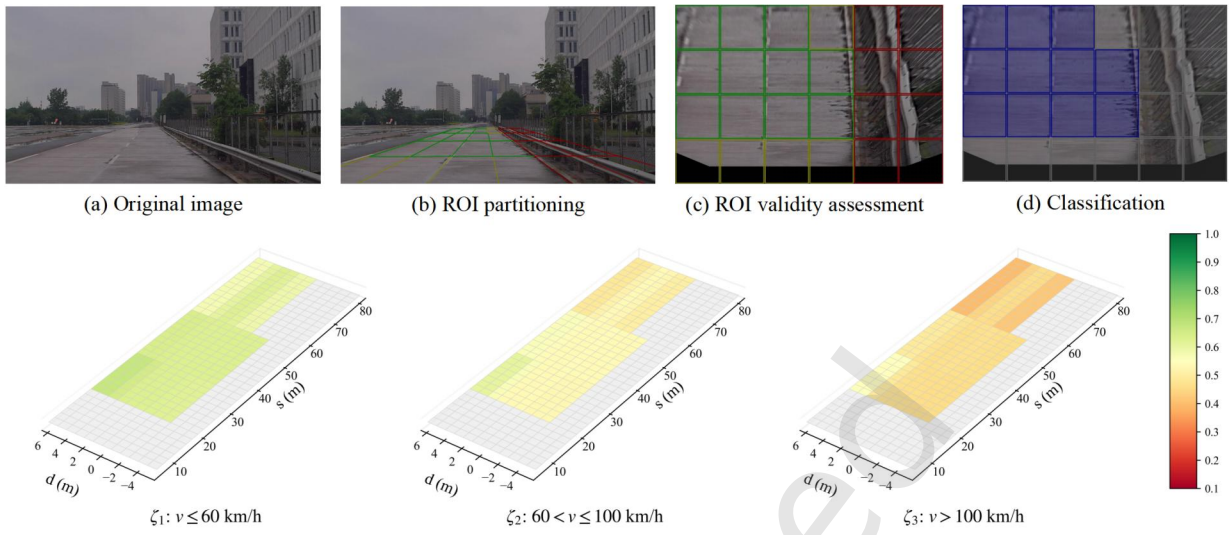


Fig. 24 Visual perception results at the initial time of S4.

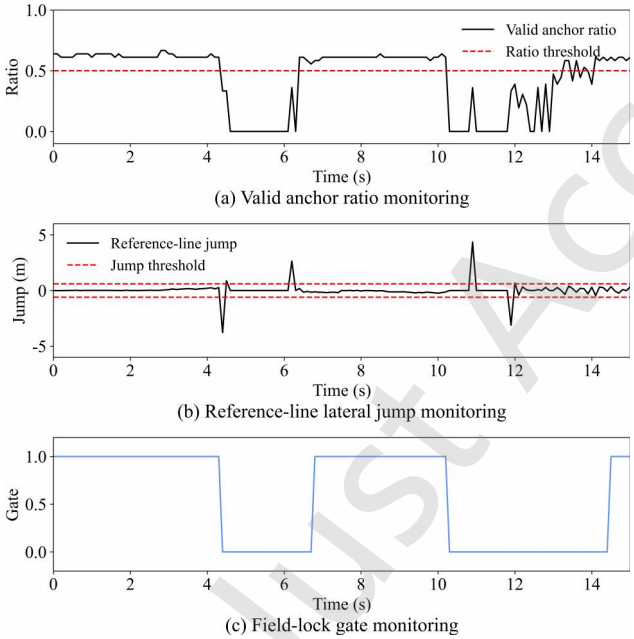


Fig. 25 Indicators related to the field-lock gate in S4.

$$\mathrm{RMSE}_{\mu}^{j} = \sqrt{\frac{1}{N}\sum_{t=0}^{N-1} e_{j,t}^{2}}, \tag{44b}$$

$$\mathrm{MAPE}_{\mu}^{j} = \frac{100\%}{N}\sum_{t=0}^{N-1}\left|\frac{e_{j,t}}{\mu_{j,t}^{true}}\right| \tag{44c}$$

where $e_{j,t} = \hat{\mu}_{j,t} - \mu_{j,t}^{true}$ indicates the estimation error of tire j at time step t, and N is the total number of estimation samples. The overall performance for each metric is obtained by averaging over the four tires, i.e., $\mathrm{X}_{\mu}^{\mathrm{avg}} = \frac{1}{4}\sum_{j}\mathrm{X}_{\mu}^{j}$ for $\mathrm{X} \in \{\mathrm{MAE}, \mathrm{RMSE}, \mathrm{MAPE}\}$.

As summarized in the evaluation results, the proposed Fusion method consistently outperforms the baseline approaches across all scenarios in terms of MAE, RMSE, and MAPE. Specifically, the Fusion framework achieves the lowest estimation error, as shown by its MAE (0.03) and MAPE (4.48%) in S1. In contrast, two purely dynamic baselines exhibit significant performance degradation, such as the UKF estimator yielding an MAE of 0.18 and an MAPE of 42.60% in S2. Furthermore, the proposed approach shows general stability and robustness, as its MAE and RMSE maintain a tight margin in all scenarios. Given that RMSE heavily penalizes large transient errors, this proximity indicates that the proposed Fusion method is able to effectively suppress extreme estimation deviations.

Furthermore, the evaluation results highlight the adaptability of the Fusion method to different dynamic excitations and environmental disturbances. Among the four scenarios, S2 is the most challenging even for the proposed Fusion method. In this scenario, all evaluation metrics are relatively high; for instance, the MAPE of the Fusion method reaches 7.98% in S2. This can be attributed partly to the fact that the reference TRFC in S2 is the lowest among the four scenarios, with values of 0.38 on the left side and 0.55 on the right side. Thus, an estimation error of approximately 0.02–0.03 constitutes a larger proportion of the true value. On the other hand, S2 is indeed challenging for purely dynamics-based estimators, as the UKF performs extremely poorly, yielding a MAPE of 42.60%. As shown in Fig. 16 (c) and (d), the UKF fails to update the rear tires' TRFC and retains the default initial value of 0.8 throughout the entire scenario.

Although UKF and A-CKF appear to perform better in

Just Accepted

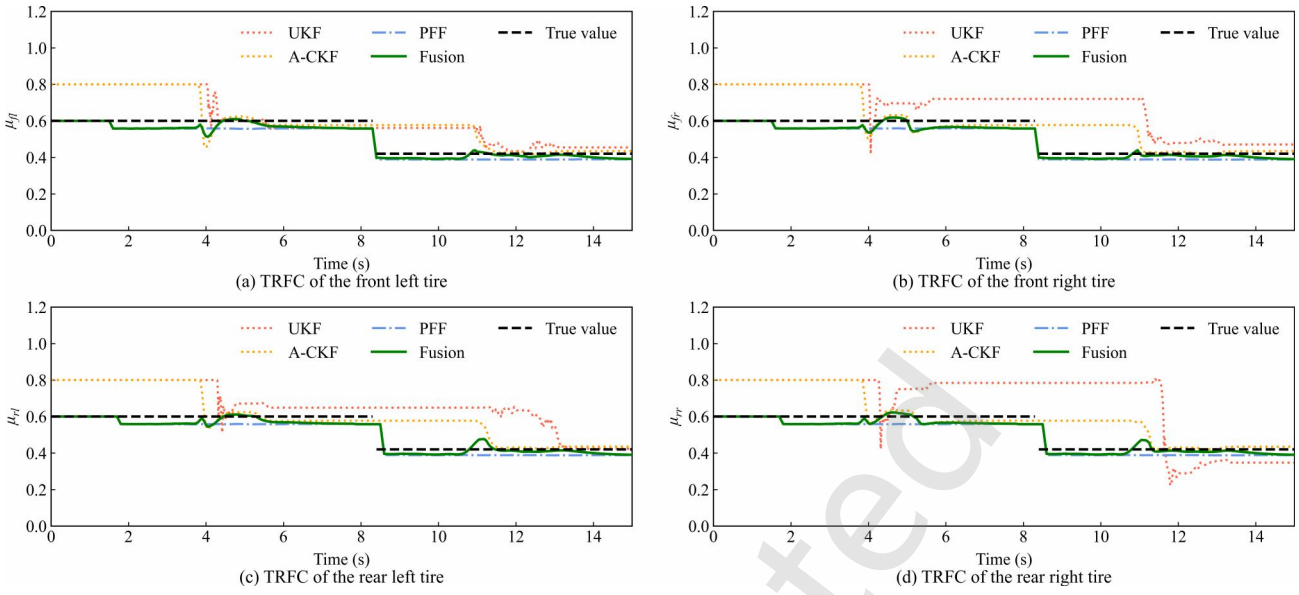


Fig. 26 TRFC estimation results in S4.

Table 6 Comparisons of TRFC estimation performance across different scenarios.

Method	S1			S2			S3			S4		
	MAE_μ^{avg}	$RMSE_\mu^{avg}$	$MAPE_\mu^{avg}$	MAE_μ^{avg}	$RMSE_\mu^{avg}$	$MAPE_\mu^{avg}$	MAE_μ^{avg}	$RMSE_\mu^{avg}$	$MAPE_\mu^{avg}$	MAE_μ^{avg}	$RMSE_\mu^{avg}$	$MAPE_\mu^{avg}$
UKF	0.05 ± 0.01	0.05	7.68	0.18 ± 0.02	0.19	42.60	0.12 ± 0.09	0.15	20.89	0.15 ± 0.09	0.17	29.50
A-CKF	0.06 ± 0.02	0.07	8.71	0.08 ± 0.05	0.09	16.99	0.06 ± 0.04	0.07	9.51	0.09 ± 0.08	0.12	17.55
Fusion	0.03 ± 0.04	0.05	4.48	0.04 ± 0.03	0.05	7.98	0.03 ± 0.02	0.04	4.60	0.02 ± 0.01	0.03	4.75

Note: MAE is expressed as Mean $\pm$ Std, and the unit of MAPE is %.

the similarly straight driving S1, this result is incidental since the left side's true value is close to their initial value of 0.8. Therefore, a comparison of the two dynamics-based estimators and the Fusion method across S1 and S2 shows that incorporating the vision-based PFF mitigates the limitations of dynamics-based estimation under low-excitation conditions, thereby enabling more robust TRFC estimation.

As for S3 and S4, which involve substantial lateral excitation, the proposed Fusion method maintains reliable performance. While A-CKF achieves a relatively low MAPE of 9.51% in S3, its MAPE increases substantially to 17.55% in S4. In contrast, the Fusion method yields comparable MAPE values of 4.60% and 4.75% in S3 and S4, respectively. This discrepancy arises from the different temporal characteristics of lateral excitation in the two scenarios.

In S3, the slalom maneuver continuously generates strong lateral excitation, enabling the A-CKF to track the TRFC accurately throughout the scenario. In S4, however, sufficient lateral excitation is available primarily during the two lane-change maneuvers, occurring between 4–5 s and 11–12 s. Therefore, the A-CKF can update its TRFC estimates mainly during these intervals and cannot respond effectively to friction changes occurring outside them. The visual perception module compensates for this limitation when dynamic excitation is insufficient, providing a preview of upcoming friction changes. Consequently, the Fusion method maintains robust performance in safety-critical scenarios such as S3 and S4, which are characterized by strong-excitation lateral maneuvers and varying friction conditions.

6.3 Discussion and limitations

Although the proposed fusion framework is validated on representative highway scenarios, its generalization capability and underlying limitations warrant further discussion. This section discusses the proposed system's transferability to different vehicles, its robustness across various environments, the existing limitations, and potential future work.

In the proposed framework, the core vehicle and tire models are derived from fundamental vehicle mechanics and are decoupled from specific vehicle hardware. Deploying the framework to a different vehicle platform primarily requires updating the calibration matrix for the IPM in the visual mod-

Just Accepted

ule according to the onboard camera's extrinsic parameters. Vehicle configuration presented in Table 5 should also be updated in the dynamics-based module. Notably, conventional CKF estimators often suffer from tedious, empirical re-tuning of covariance matrices across different vehicle platforms and operating maneuvers. In our framework, instead, the adaptive noise adjustment mechanism dynamically schedules the noise covariances, thereby largely mitigating the burden of manual hyperparameter tuning when porting to new operations.

The dynamics-based A-CKF module is inherently immune to ambient illumination and visibility, ensuring continuous basic functionality of the proposed framework. For the vision module, environmental variations primarily affect image segmentation and surface classification. On the one hand, the road segmentation model is trained on the large-scale BDD100K dataset containing diverse illumination conditions, while the self-bulit classification dataset incorporates images across multiple seasons under sunny, overcast, and rainy days. Through these strategies, efforts are made to reduce the disturbances from road environments and enhance the vision module's robustness.

On the other hand, in scenarios involving severe environmental disturbances (e.g., poor night-time lighting or heavy fog) where visual classification reliability significantly decreases, the proposed dual-confidence fusion mechanism provides an effective safety guarantee. As the visual observation uncertainty $P_{w,t}^{\text{vis}}$ rises, the visual reliability $\bar{\rho}_{w,t}^{\text{vis}}$ automatically decays, smoothly shifting the fusion weight $\gamma_{w,t}$ toward dynamics-dominant estimation.

Meanwhile, a primary limitation of this study lies in the spatiotemporal fusion mechanism, especially under aggressive vehicle maneuvers. The spatial alignment between the vision and dynamics modules relies on the predicted wheel positions. In highway driving where the lateral slip and heading deviation are relatively small, this kinematic projection is sufficiently accurate. However, during severe lateral maneuvers like large steering inputs, neglecting the vehicle sideslip angle and heading variation may introduce errors when matching the wheel positions with the PFF grids. Thus, future studies will consider integrating vehicle sideslip angle compensation into wheel position projection.

7 Conclusion

This paper presents a predictive TRFC estimation framework for autonomous driving, which couples an A-CKF estimator with a vision-based friction perception module through a dual-confidence fusion mechanism. The proposed Fusion method addresses the difficulty of accurately estimating peak TRFC with the capability to obtain road friction conditions ahead of the vehicle. Because the two estimation branches are reliable under complementary driving scenarios, the dual-confidence strategy arbitrates between them, allowing the fused estimation to remain smooth and physically consistent as the TRFC changes and the vehicle transitions between different driving maneuvers.

Real-vehicle tests across four typical scenarios, covering split-μ and joint road surfaces, confirm the benefit of this dynamics-vision hybrid design. For instance, the Fusion method reduces the MAPE from 29.5% obtained with a conventional UKF to 4.75%, and it consistently outperforms the dynamics-only A-CKF and the vision-only PFF evaluated separately, particularly in scenarios where lateral excitation is transient or road conditions vary spatially. The experimental results indicate that fusing vehicle dynamics with visual road-condition preview is an effective and practical route to dependable TRFC estimation for trajectory planning in autonomous driving.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Reference

[1] X. Tang, M. Yang, T. Wen, P. Jia, L. Cui, M. Luo, K. Sheng, B. Zhang, K. Jiang, and D. Yang, Priorfusion: Unified integration of priors for robust road perception in autonomous driving, *Communications in Transportation Research*, vol. 5, p. 100229, 2025.

[2] X. Bai, P. Dong, J. Wang, Y. Huang, H. Yu, and Y. Ren, Advglow: Covert adversarial attacks against autonomous driving perception, *Journal of Intelligent and Connected Vehicles*, vol. 8, no. 4, pp. 9 210 067–1–9 210 067–13, 2025.

[3] S. Tao, Z. Ju, L. Li, H. Zhang, and W. Pedrycz, Tire road friction coefficient estimation for individual wheel based on two robust pmi observers and a multilayer perceptron, *IEEE Transactions on Vehicular Technology*, pp. 1–12, 2024.

[4] M. Zhang, J. Xu, Y. Gao, D. Shen, and Z. Xu, Can combined virtual-real testing speed up autonomous vehicle testing? findings from aeb field experiments, *Communications in Transportation Research*, vol. 5, p. 100216, 2025.

[5] T. K. Kurse, G. Gebresenbet, G. F. Daba, and N. T. Tefera, Experimental determination of factors causing crashes involving automated vehicles, *Multimodal Transportation*, vol. 4, no. 1, p. 100186, 2025.

[6] S. Najafi, G. W. Flintsch, and A. Medina, Linking roadway crashes and tire–pavement friction: A case study, *International*

Just Accepted

Journal of Pavement Engineering, vol. 18, no. 2, pp. 119–127, 2017.

[7] Z. Han and H. O. Sharif, Investigation of the relationship between rainfall and fatal crashes in texas, 1994–2018, *Sustainability*, vol. 12, no. 19, p. 7976, 2020.

[8] S. S. Channamallu, D. Almaskati, S. Kermanshachi, and A. Pamidimukkala, Autonomous vehicle safety: An advanced bagging classifier technique for crash injury prediction, *Multimodal Transportation*, vol. 4, no. 1, p. 100189, 2025.

[9] K. Huang, C. Jiang, M.-m. Qiu, D. Wu, and B.-z. Zhang, Research on intelligent vehicle lane changing and obstacle avoidance control based on road adhesion coefficient, *Journal of Vibration and Control*, vol. 28, no. 21-22, pp. 3269–3285, 2022.

[10] J. Zhang, B.-L. Ye, X. Wang, L. Li, and B. Song, A trajectory planning and tracking method based on deep hierarchical reinforcement learning, *Journal of Intelligent and Connected Vehicles*, vol. 8, no. 2, pp. 9 210 056–1–9 210 056–9, 2025.

[11] Z. Song, W. Li, S. Zhang, H. Ding, N. Xu, and Y. Huang, Enhanced motion planning strategy for automated lane change on slippery roads, *IEEE/ASME Transactions on Mechatronics*, pp. 1–12, 2026.

[12] Y. Guan, H. Liao, C. Wang, B. Wang, J. Zhang, J. Hu, and Z. Li, Domain-enhanced dual-branch model for efficient and interpretable accident anticipation, *Communications in Transportation Research*, vol. 5, p. 100214, 2025.

[13] Y. Yan, C. Zhang, J. Wu, C. Sun, Q. Meng, C. Chen, J. Wang, and G. Wang, Autonomous driving on mountain roads via an adaptive deep reinforcement learning approach, *Journal of Intelligent and Connected Vehicles*, vol. 8, no. 4, p. 9210069, 2025.

[14] J. Liu, Y. Liu, L. Wang, and X. Qu, A survey and comprehensive taxonomy of tire-road adhesion coefficient estimation for intelligent vehicles, *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 9, pp. 12 819–12 833, 2025.

[15] N. Xu, Y. Huang, H. Askari, and Z. Tang, Tire slip angle estimation based on the intelligent tire technology, *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2239–2249, 2021.

[16] S. Yang, Y. Chen, R. Shi, R. Wang, Y. Cao, and J. Lu, A survey of intelligent tires for tire-road interaction recognition toward autonomous vehicles, *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 520–532, 2022.

[17] J. Alonso, J. López, I. Pavón, M. Recuero, C. Asensio, G. Arcas, and A. Bravo, On-board wet road surface identification using tyre/road noise and support vector machines, *Applied Acoustics*, vol. 76, pp. 407–415, 2014.

[18] M.-H. Kim, J. Park, and S. Choi, Road type identification ahead of the tire using d-cnn and reflected ultrasonic signals, *International Journal of Automotive technology*, vol. 22, no. 1, pp. 47–54, 2021.

[19] L. Wang, X. Fan, X. Yu, Z. Huang, and K. Zhao, A review of road surface recognition and tyre-road friction coefficient estimation methods, *International Journal of Heavy Vehicle Systems*, vol. 32, no. 3, pp. 327–367, 2025.

[20] Y. Mo, P. Zhang, Z. Chen, and B. Ran, A method of vehicle-infrastructure cooperative perception based vehicle state information fusion using improved kalman filter, *Multimedia tools and applications*, pp. 1–18, 2022.

[21] Z. Quan, X. Sun, T. Gu, Y. Cai, and B. Li, Adaptive ckf-based vehicle sideslip angle estimation using time-synchronized lateral forces from optimized intelligent tires, *IEEE/ASME Transactions on Mechatronics*, pp. 1–12, 2026.

[22] Q. Chen, B. Yu, C. Zhong, Z. Jiang, D. You, and Y. Cai, Study on vehicle state and tire–road friction coefficient estimation based on maximum correntropy generalized high-degree cubature kalman filter, *Transactions of the Institute of Measurement and Control*, vol. 47, no. 14, pp. 2918–2929, 2025.

[23] J. Ye, Z. Zhang, J. Jin, R. Su, and B. Huang, Estimation of tire-road friction coefficient with adaptive tire stiffness based on rc-sckf, *Nonlinear Dynamics*, vol. 112, no. 2, pp. 945–960, 2024.

[24] Y. Han, Y. Lu, H. Wang, and Y. Wang, A new preprocessment method for road peak adhesion coefficient fusion estimation, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 238, no. 1, pp. 224–240, 2024.

[25] B. Li, L. Zhang, C. Zhao, J. Lu, Q. Meng, Z. Zhang, and H. Chen, Capturing rapid vehicle dynamics: A refined adaptive kalman filter for enhanced tire–road friction estimation, *IEEE Transactions on Industrial Electronics*, 2024.

[26] R. Zhang, Y. Feng, P. Shi, L. Zhao, Y. Du, and Y. Liu, Tire-road friction coefficient estimation for distributed drive electric vehicles using pmsm sensorless control, *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 8672–8685, 2023.

[27] R. Marotta, V. Ivanov, S. Strano, M. Terzo, and C. Tordela, Estimation of the tire-road interaction forces by using pacejka's formulas with combined slips and camber angles, SAE Technical Paper, Tech. Rep., 2023.

[28] Z. Zhang, L. Zheng, H. Wu, Z. Zhang, Y. Li, and Y. Liang, An estimation scheme of road friction coefficient based on novel tyre model and improved sckf, *Vehicle System Dynamics*, vol. 60, no. 8, pp. 2775–2804, 2022.

[29] G. Li, Q. Liu, S. Zhao, W. Qiao, and X. Ren, Automatic crack recognition for concrete bridges using a fully convolutional neural network and naive bayes data fusion based on a visual detection system, *Measurement Science and Technology*, vol. 31, no. 7, p. 075403, 2020.

[30] O. A. Egaji, G. Evans, M. G. Griffiths, and G. Islas, Real-time machine learning-based approach for pothole detection, *Expert Systems with Applications*, vol. 184, p. 115562, 2021.

[31] F. Kortmann, K. Talits, P. Fassmeyer, A. Warnecke, N. Meier, J. Heger, P. Drews, and B. Funk, Detecting various road damage types in global countries utilizing faster r-cnn, in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 5563–5571.

Just Accepted

[32] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama, and Y. Sekimoto, Deep learning-based road damage detection and classification for multiple countries, *Automation in Construction*, vol. 132, p. 103935, 2021.

[33] B. Leng, D. Jin, L. Xiong, X. Yang, and Z. Yu, Estimation of tire-road peak adhesion coefficient for intelligent electric vehicles based on camera and tire dynamics information fusion, *Mechanical Systems and Signal Processing*, vol. 150, p. 107275, 2021.

[34] B. Leng, D. Jin, X. Hou, C. Tian, L. Xiong, and Z. Yu, Tire-road peak adhesion coefficient estimation method based on fusion of vehicle dynamics and machine vision, *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21 740–21 752, 2022.

[35] C. Tian, B. Leng, X. Hou, Y. Huang, W. Zhao, D. Jin, L. Xiong, and J. Zhao, Robust identification of road surface condition based on ego-vehicle trajectory reckoning, *Automotive Innovation*, vol. 5, no. 4, pp. 376–387, 2022.

[36] C. Tian, B. Leng, X. Hou, L. Xiong, and C. Huang, Multi-sensor fusion based estimation of tire-road peak adhesion coefficient considering model uncertainty, *Remote Sensing*, vol. 14, no. 21, p. 5583, 2022.

[37] B. Leng, C. Tian, X. Hou, L. Xiong, W. Zhao, and Z. Yu, Tire-road peak adhesion coefficient estimation based on multisource information assessment, *IEEE Transactions on Intelligent Vehicles*, 2023.

[38] J. Liu, L. Wang, Y. Liu, and X. Qu, Hybrid dynamics-data estimation of tire-road adhesion coefficient under water film effect, in *2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2025, pp. 1373–1378.

[39] Y. Huang, Y. Zhang, X. Wang, H. Wei, K. Deng, L. Li, and J. Song, On the tyre–road friction coefficient fusion identification framework utilising computer vision-based and vehicle dynamics-based road surface information, *Mechanical Systems and Signal Processing*, vol. 235, p. 112821, 2025.

[40] S. Zhao, J. Zhang, Y. Jiang, C. He, and J. Han, Tire-road friction coefficients adaptive estimation through image and vehicle dynamics integration, *Mechanical Systems and Signal Processing*, vol. 224, p. 112039, 2025.

[41] W. Wang, S. Bai, H. Ding, D. Pi, and G. Yin, A dual-event-driven fusion estimation method for road friction coefficient based on image and vehicle dynamics, *IEEE Transactions on Vehicular Technology*, 2026.

[42] H. Guo, X. Zhao, J. Liu, Q. Dai, H. Liu, and H. Chen, A fusion estimation of the peak tire-road friction coefficient based on road images and dynamic information, *Mechanical Systems and Signal Processing*, vol. 189, p. 110029, 2023.

[43] Z. Du, A. Skar, M. Pettinari, and X. Zhu, Pavement friction evaluation based on vehicle dynamics and vision data using a multi-feature fusion network, *Transportation Research Record*, vol. 2677, no. 11, pp. 219–236, 2023.

[44] Y. Wang, G. Yin, P. Hang, J. Zhao, Y. Lin, and C. Huang, Fundamental estimation for tire road friction coefficient: A model-based learning framework, *IEEE Transactions on Vehicular Technology*, 2024.

[45] G. Zhou, W. Yin, Y. Liu, Q. Gong, Z. Qi, J. Chen, and L. Wang, An attention-based hybrid model combining graph convolutional networks and gated recurrent units for ground-level ozone spatiotemporal prediction, *Tsinghua Science and Technology*, 2026.

[46] S. Zhao, J. Zhang, Y. Jiang, C. He, and J. Han, Tire-road friction coefficients adaptive estimation through image and vehicle dynamics integration, *Mechanical Systems and Signal Processing*, vol. 224, p. 112039, 2025.

[47] D. Puscul, A. Schönlieb, A. Ramsauer, and C. Lex, On-road water detection and water film height estimation using a time-of-flight camera, in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024, pp. 1383–1388.

[48] L. Zhao, H. Zhao, and J. Cai, Tire-pavement friction modeling considering pavement texture and water film, *International Journal of Transportation Science and Technology*, vol. 14, pp. 99–109, 2024.

[49] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation, *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3051–3068, 2021.

[50] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang *et al.*, Searching for mobilenetv3, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.

[51] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, Bdd100k: A diverse driving dataset for heterogeneous multitask learning, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 2633–2642.

[52] Z. Qin, H. Wang, and X. Li, Ultra fast structure-aware deep lane detection, in *European conference on computer vision*. Springer, 2020, pp. 276–291.

[53] J. Ling, F. Yang, J. Zhang, P. Li, M. I. Uddin, and T. Cao, Water-film depth assessment for pavements of roads and airport runways: A review, *Construction and Building Materials*, vol. 392, p. 132054, 2023.

[54] C. Ma, X. Huang, Y. Zhao, T. Wang, and B. Du, Gru-lstm model based on the ssa for short-term traffic flow prediction, *Journal of Intelligent and Connected Vehicles*, vol. 8, no. 1, pp. 9 210 051–1–9 210 051–10, 2025.

[55] B. Feng and Z. Ding, Application-oriented cloud workload prediction: A survey and new perspectives, *Tsinghua Science and Technology*, vol. 30, no. 1, pp. 34–54, 2024.

Just Accepted