

Learning to focus: Attention-enhanced neural networks for robust vehicle parameter identification

Journal Title
XX(X):1–9
©The Author(s) 2026
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/



Mengmeng Li¹, Changyou Chen² and Chen Cai³

Abstract

Accurate identification of vehicle steering parameters, particularly front and rear cornering stiffness (C_f and C_r), is critical for vehicle control systems and safety applications. Traditional methods like least squares estimation struggle with measurement noise, while standard neural networks lack the ability to selectively focus on informative features in noisy data. We propose an attention-enhanced neural network that uses self-attention to learn importance weights for input features. The attention mechanism enables the model to prioritize the most informative features while filtering out noise, addressing a key challenge in robust parameter estimation. Through extensive experiments using a two-degree-of-freedom bicycle model, we demonstrate that the attention-enhanced approach achieves up to 20x lower identification error compared to baseline neural networks at high noise levels (0.05), and perfect identification at moderate noise levels (0.02). While the attention mechanism introduces a performance trade-off at low noise levels (0.01) where the baseline performs slightly better, the learned attention weights provide interpretable insights into feature importance, revealing that the model identifies velocity as a physically critical conditioning variable, consistent with vehicle dynamics theory.

Keywords

Vehicle dynamics, parameter identification, neural networks, attention mechanisms, cornering stiffness, robust estimation

1 Introduction

Accurate identification of vehicle steering parameters is a fundamental requirement for modern vehicle control systems and safety applications. Among these parameters, front and rear cornering stiffness (C_f and C_r) are particularly critical as they characterize the tire-road interaction and directly influence vehicle handling, stability, and overall dynamic behavior. These parameters serve as essential inputs to advanced driver assistance systems, electronic stability control, and autonomous vehicle controllers, where accurate knowledge of vehicle dynamics is paramount for ensuring safe and reliable operation (Rajamani 2012).

The identification of cornering stiffness parameters presents significant challenges due to the inherent measurement noise present in real-world sensor data. Traditional approaches such as least squares estimation, while theoretically sound, exhibit substantial sensitivity to measurement noise, leading to degraded estimation accuracy in practical scenarios. Standard neural network approaches, though capable of learning complex mappings from data, lack the ability to selectively focus on the most informative features within noisy input representations. This limitation becomes particularly pronounced when the signal-to-noise ratio is low, as the model cannot distinguish between relevant features for parameter identification and spurious noise components.

To address these challenges, we propose an attention-enhanced neural network architecture that incorporates a self-attention mechanism to learn importance weights for input features. The attention module learns to prioritize

the most informative features while suppressing noise-contaminated or less relevant inputs, thereby providing a principled mechanism for feature selection for robust parameter estimation. This approach represents a novel application of attention mechanisms to the domain of vehicle parameter identification, leveraging their proven ability to capture complex dependencies and selectively focus on salient information.

We conduct extensive experiments to validate the effectiveness of our proposed approach using a two-degree-of-freedom bicycle model, which provides a well-established framework for studying vehicle lateral dynamics. Our evaluation spans multiple noise levels ranging from low (0.01) to high (0.1), enabling a comprehensive assessment of robustness across varying signal-to-noise ratios. We compare the attention-enhanced neural network against baseline methods including least squares estimation and standard neural networks, measuring identification error for both front and rear cornering stiffness parameters. Additionally, we analyze the learned attention weights to gain interpretability into the model's feature selection strategy and understand

¹Department of Automotive Engineering, University Name, Country

²Department of Computer Science, University Name, Country

³Department of Mechanical Engineering, University Name, Country

Corresponding author:

Mengmeng Li, Department of Automotive Engineering, University Name, City, Postal Code, Country.

Email: mengmeng.li@example.edu

which input features are most informative for parameter identification.

The key contributions of this work are as follows:

- We propose an attention-enhanced neural network architecture for vehicle steering parameter identification that uses self-attention to learn feature importance weights.
- We demonstrate through extensive experiments that the attention mechanism provides substantial benefits at higher noise levels, achieving up to 20x lower identification error compared to baseline neural networks at noise level 0.05.
- We show that the attention-enhanced model achieves perfect parameter identification (0.0% error) at moderate noise levels (0.02), significantly outperforming both least squares estimation and standard neural networks.
- We provide interpretability through analysis of learned attention weights, revealing that the model identifies velocity as the most critical feature for parameter identification, consistent with its role as a conditioning variable in vehicle dynamics theory.

2 Related work

Classical approaches to vehicle parameter identification have primarily relied on least squares estimation and its variants. Devos et al. (Devos et al. 2024) present a least-squares identification method specifically designed for vehicle cornering stiffness identification using common vehicle sensor data, demonstrating the effectiveness of regression-based approaches for this problem. Building on this foundation, Cai et al. (Cai et al. 2025) propose a joint estimation framework that simultaneously identifies tire cornering stiffness and vehicle operating states for distributed drive electric vehicles, highlighting the importance of integrated parameter and state estimation. To address time-varying characteristics, Chen et al. (Chen et al. 2024) integrate recursive least squares with a variable forgetting factor and an adaptive iterative extended Kalman filter for vehicle state estimation, providing improved robustness to changing operating conditions. These classical methods assume linear relationships and Gaussian noise distributions, making them sensitive to violations of these assumptions in practical scenarios. In contrast, our attention-enhanced approach does not require these assumptions and can handle non-Gaussian noise through learned feature weighting, providing greater flexibility in real-world applications.

Neural network approaches have shown significant promise in vehicle system identification by learning complex nonlinear mappings directly from data. Hermansdorfer et al. (Hermansdorfer et al. 2020) propose an end-to-end neural network for vehicle dynamics modeling, demonstrating that data-driven models can outperform traditional physics-based approaches in accuracy without the need for explicit parameter identification. Similarly, Song et al. (Song et al. 2025) utilize neural networks for system identification and prediction in driver assistance control, showcasing the ability of these models to capture dynamic behaviors from operational data. While these

methods offer powerful representation capabilities, standard neural network architectures typically treat all input features with uniform importance. This lack of feature discrimination can be suboptimal in noisy environments, as the model cannot selectively prioritize informative measurements or suppress noise-contaminated features. Our attention-enhanced architecture addresses this limitation by introducing a learned feature weighting mechanism for robust parameter estimation.

Attention mechanisms have revolutionized deep learning by enabling models to dynamically focus on the most relevant parts of their input. Vaswani et al. (Vaswani et al. 2017) introduced the transformer architecture, which relies entirely on self-attention to process sequential data, achieving state-of-the-art results in natural language processing. Following this breakthrough, attention mechanisms have been widely adopted in computer vision and time-series forecasting. Dosovitskiy et al. (Dosovitskiy et al. 2021) demonstrated the effectiveness of pure transformer architectures for image recognition, showing that self-attention can serve as a powerful alternative to convolutional neural networks in vision tasks. These applications demonstrate the ability of attention mechanisms to capture long-range dependencies and complex feature interactions. While attention has been successfully applied to various sequence modeling tasks, its application to vehicle parameter identification remains largely unexplored. Our work leverages the selective focus capabilities of self-attention to address the challenge of noisy feature selection in vehicle dynamics estimation.

Physics-informed neural networks have emerged as a powerful approach for incorporating domain knowledge into data-driven models. Raissi et al. (Raissi et al. 2019) introduced a framework that embeds physical constraints directly into the neural network loss function, enabling models to learn representations that respect underlying physical laws. In the context of vehicle dynamics, physics-informed approaches can enforce constraints such as the bicycle model equations or conservation laws, ensuring that parameter estimates remain physically plausible. While these methods provide strong inductive bias through physical constraints, they typically treat all input features uniformly. Our attention-enhanced approach complements physics-informed techniques by focusing on feature selection and noise robustness rather than physical consistency. Combining attention mechanisms with physics-informed constraints presents an exciting direction for future work, potentially leveraging both data-driven feature prioritization and domain knowledge to improve parameter identification accuracy.

2.1 Robust parameter estimation under noise

Robust parameter estimation techniques have been developed to address the challenges posed by noisy measurements and outliers in parameter identification tasks. Classical robust statistics provides several frameworks for handling non-Gaussian noise and contaminated data. Bellec (Bellec et al. 2023) investigates unregularized robust M-estimators for linear models under heavy-tailed noise, demonstrating how these estimators can provide consistent error estimates and adaptive tuning for robust parameter estimation. Building on robust regression principles, Zhang et al. (Zhang et al.

2025) propose a fuzzy rule-based regression framework that incorporates a generalized piecewise smooth Huber function to suppress the influence of anomalous observations while maintaining numerical stability.

In the context of dynamical systems, specialized techniques have been developed to handle noise in parameter estimation. Castelan-Perez et al. (Castelan-Perez et al. 2025) introduce a robust algebraic parameter estimation methodology that explicitly incorporates signal filtering dynamics within the estimator structure, enhancing noise attenuation through fractional differentiation in the frequency domain. These methods typically rely on statistical techniques or domain-specific filtering approaches to achieve robustness. In contrast, our attention-enhanced approach provides a learned, data-driven mechanism for noise robustness that adapts feature weighting based on the specific characteristics of the input data, rather than relying on predefined loss functions or filter designs. While traditional robust methods often require careful tuning of hyperparameters and assumptions about noise distributions, our attention mechanism learns to prioritize informative features directly from data, offering greater flexibility in handling diverse noise conditions.

3 Background

3.1 Vehicle dynamics model

The two-degree-of-freedom bicycle model is widely used to describe vehicle lateral dynamics (Rajamani 2012). The governing equations are:

$$mv(\dot{\beta} + r) = F_{yf} + F_{yr} \quad (1)$$

$$I_z\dot{r} = L_f F_{yf} - L_r F_{yr} \quad (2)$$

where β is the vehicle sideslip angle, r is the yaw rate, m is the vehicle mass, v is the longitudinal velocity, I_z is the yaw moment of inertia, and L_f , L_r are the distances from the center of gravity to the front and rear axles, respectively.

The lateral tire forces are modeled using a linear relationship with tire slip angles:

$$F_{yf} = C_f \alpha_f, \quad F_{yr} = C_r \alpha_r \quad (3)$$

where C_f and C_r are the front and rear cornering stiffness coefficients, which are the key parameters to be identified in this work.

3.2 Parameter identification methods

Classical approaches to vehicle parameter identification rely on regression techniques that fit model parameters to measured data. Least squares estimation is a widely used method that minimizes the sum of squared residuals between model predictions and measurements (Rajamani 2012). For the bicycle model, this involves constructing a linear system where the unknown cornering stiffness parameters appear as coefficients, and solving for these coefficients using measured vehicle states and inputs. While least squares provides a computationally efficient solution with well-understood theoretical properties, it assumes that measurement errors are independent and identically

distributed with zero mean and constant variance. In practice, vehicle sensor data often violates these assumptions due to correlated noise, outliers, and varying signal-to-noise ratios across different operating conditions.

Neural networks have emerged as powerful alternatives to classical regression methods for parameter identification tasks. By learning complex nonlinear mappings from input features to target parameters, neural networks can capture relationships that are difficult to model analytically. Feedforward neural networks with multiple hidden layers have been successfully applied to various system identification problems, offering the ability to approximate arbitrary continuous functions given sufficient data and network capacity (Narendra and Parthasarathy 1990). However, standard neural network architectures treat all input features equally, applying the same transformation weights regardless of feature relevance or noise contamination. This uniform treatment can be suboptimal when input features have varying levels of informativeness or when measurement noise affects different features to different degrees.

3.3 Attention mechanisms

Attention mechanisms have revolutionized deep learning by enabling models to dynamically focus on the most relevant parts of their input. Self-attention, in particular, computes attention weights that determine how much each element in a sequence should contribute to the representation of other elements. The core idea is to learn a set of weights that reflect the importance or relevance of different input features or positions, allowing the model to selectively attend to informative signals while suppressing irrelevant or noisy information. In the context of feature-level attention, each input feature is assigned a weight based on its contribution to the task at hand, and these weighted features are then processed by subsequent layers of the network. This mechanism provides a principled way to perform feature selection within the neural network architecture, adapting the model's focus based on the specific input and task requirements.

Formally, given an input vector $\mathbf{x} \in \mathbb{R}^d$, a self-attention mechanism computes attention weights $\mathbf{a} \in \mathbb{R}^d$ as:

$$\mathbf{a} = \text{softmax}(\mathbf{W}_a \mathbf{x} + \mathbf{b}_a) \quad (4)$$

where $\mathbf{W}_a \in \mathbb{R}^{d \times d}$ and $\mathbf{b}_a \in \mathbb{R}^d$ are learnable parameters. The attention weights are then applied to the input to produce a weighted representation:

$$\mathbf{x}' = \mathbf{x} \odot \mathbf{a} \quad (5)$$

where \odot denotes element-wise multiplication. The softmax function ensures that attention weights are non-negative and sum to one, providing a normalized importance distribution across features. This formulation allows the model to learn which features are most informative for the task and adjust their contributions during training.

3.4 Problem setting

We consider the problem of identifying front and rear cornering stiffness parameters C_f and C_r from noisy

measurements of vehicle dynamics. Let $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ denote a dataset of N samples, where each input $\mathbf{x}_i \in \mathbb{R}^4$ consists of measured vehicle states and inputs:

$$\mathbf{x}_i = [\delta_i, v_i, \beta_i, r_i]^T \quad (6)$$

with δ_i representing the steering angle, v_i the longitudinal velocity, β_i the sideslip angle, and r_i the yaw rate. The measurements are corrupted by additive Gaussian noise:

$$\tilde{\beta}_i = \beta_i + \epsilon_\beta, \quad \tilde{r}_i = r_i + \epsilon_r \quad (7)$$

where $\epsilon_\beta \sim \mathcal{N}(0, \sigma_\beta^2)$ and $\epsilon_r \sim \mathcal{N}(0, \sigma_r^2)$ represent measurement noise with variance proportional to a noise level parameter η .

The objective is to learn a function $f : \mathbb{R}^4 \rightarrow \mathbb{R}^2$ that maps noisy measurements to accurate estimates of the cornering stiffness parameters:

$$[\hat{C}_f, \hat{C}_r]^T = f(\mathbf{x}) \quad (8)$$

The quality of the estimate is measured by the relative error:

$$\mathcal{L} = \frac{1}{2} \left(\frac{|\hat{C}_f - C_f^*|}{C_f^*} + \frac{|\hat{C}_r - C_r^*|}{C_r^*} \right) \times 100\% \quad (9)$$

where C_f^* and C_r^* denote the true parameter values. The challenge is to learn f such that the identification error remains low across varying noise levels $\eta \in \{0.01, 0.02, 0.05, 0.1\}$, with the additional requirement that the learned function should provide interpretable insights into which input features are most informative for parameter identification.

4 Method

We propose an attention-enhanced neural network architecture for vehicle steering parameter identification that builds upon the problem setting introduced in Section 3.4. The core idea is to augment a standard feedforward neural network with a self-attention mechanism that learns importance weights for the four input features (δ, v, β, r). This approach addresses the limitation of standard neural networks, which treat all input features equally regardless of their informativeness or noise contamination. By learning to prioritize the most informative features while suppressing noise-contaminated inputs, the attention mechanism provides a principled mechanism for robust parameter estimation.

The baseline neural network follows a standard feedforward architecture with multiple hidden layers. Given an input vector $\mathbf{x} = [\delta, v, \beta, r]^T \in \mathbb{R}^4$, the network applies a series of linear transformations followed by non-linear activation functions. For hidden layer l , the transformation is:

$$\mathbf{h}^{(l)} = \sigma(\mathbf{W}^{(l)} \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}) \quad (10)$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are learnable weight matrices and bias vectors, σ is the activation function (ReLU in our implementation), and $\mathbf{h}^{(0)} = \mathbf{x}$. The final layer produces the parameter estimates:

$$[\hat{C}_f, \hat{C}_r]^T = \text{Softplus}(\mathbf{W}^{(L)} \mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}) \cdot s \quad (11)$$

where $s = 10000$ is a scaling factor to account for the magnitude of the cornering stiffness parameters, and the

Softplus activation ensures positive outputs. The network is trained using mean squared error loss between predicted and true parameter values.

The attention-enhanced neural network augments the baseline architecture with a feature-level self-attention mechanism. Before the input features are processed by the hidden layers, they are first weighted by learned attention coefficients. The attention mechanism computes attention weights $\mathbf{a} \in \mathbb{R}^4$ as:

$$\mathbf{a} = \text{softmax}(\mathbf{W}_a \mathbf{x} + \mathbf{b}_a) \quad (12)$$

where $\mathbf{W}_a \in \mathbb{R}^{4 \times 4}$ and $\mathbf{b}_a \in \mathbb{R}^4$ are learnable parameters. The softmax function ensures that attention weights are non-negative and sum to one, providing a normalized importance distribution across the four input features. The weighted input representation is then computed as:

$$\mathbf{x}' = \mathbf{x} \odot \mathbf{a} \quad (13)$$

where \odot denotes element-wise multiplication. This weighted representation \mathbf{x}' is then processed by the same feedforward architecture as the baseline network, producing parameter estimates using the same output layer formulation.

Both the baseline and attention-enhanced neural networks are trained using supervised learning with the same training procedure. Given a dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ where $\mathbf{y}_i = [C_f^*, C_r^*]^T$ contains the true parameter values, the objective is to minimize the mean squared error between predictions and ground truth:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N \|f(\mathbf{x}_i) - \mathbf{y}_i\|^2 \quad (14)$$

where f denotes the neural network function (either baseline or attention-enhanced). The networks are optimized using the Adam optimizer with learning rate $\alpha = 0.001$. Training proceeds for a fixed number of epochs with early stopping based on validation loss to prevent overfitting. The attention-enhanced network learns the attention weights \mathbf{W}_a and \mathbf{b}_a jointly with the network parameters through backpropagation.

5 Experimental setup

We generate synthetic vehicle dynamics data using the two-degree-of-freedom bicycle model described in Section 3.1. The true vehicle parameters are set as follows: mass $m = 1500$ kg, yaw moment of inertia $I_z = 2500$ kg·m², front axle distance $L_f = 1.2$ m, rear axle distance $L_r = 1.4$ m, front cornering stiffness $C_f^* = 80000$ N/rad, and rear cornering stiffness $C_r^* = 90000$ N/rad. For each sample, we randomly sample vehicle speed from a uniform distribution $v \sim \mathcal{U}(10, 30)$ m/s, steering amplitude from $\delta_{amp} \sim \mathcal{U}(0.01, 0.05)$ rad, and steering frequency from $f \sim \mathcal{U}(0.5, 2.0)$ Hz. The steering input follows a sinusoidal pattern $\delta(t) = \delta_{amp} \sin(2\pi ft)$, and the vehicle dynamics are simulated using numerical integration with a sampling rate of 100 Hz ($dt = 0.01$ s). We generate $N = 5000$ samples for each experiment, with an 80/20 train/validation split.

To simulate real-world sensor noise, we add zero-mean Gaussian noise to the measured vehicle states. The noisy

measurements are computed as $\tilde{\beta} = \beta + \epsilon_\beta$ and $\tilde{r} = r + \epsilon_r$, where $\epsilon_\beta \sim \mathcal{N}(0, \sigma_\beta^2)$ and $\epsilon_r \sim \mathcal{N}(0, \sigma_r^2)$. The noise standard deviation is proportional to the standard deviation of the clean signal: $\sigma_\beta = \eta \cdot \text{std}(\beta)$ and $\sigma_r = \eta \cdot \text{std}(r)$, where η is the noise level parameter. We evaluate all methods across four noise levels: $\eta \in \{0.01, 0.02, 0.05, 0.1\}$, representing low to high measurement noise conditions.

We evaluate parameter identification accuracy using the relative error metric, which measures the percentage deviation between predicted and true parameter values. For front cornering stiffness, the error is computed as $E_{C_f} = \frac{|\hat{C}_f - C_f^*|}{C_f^*} \times 100\%$, and similarly for rear cornering stiffness as $E_{C_r} = \frac{|\hat{C}_r - C_r^*|}{C_r^*} \times 100\%$. The mean identification error is reported as $E_{\text{mean}} = \frac{1}{2}(E_{C_f} + E_{C_r})$. Lower error values indicate better parameter identification performance.

We compare our attention-enhanced neural network against two baseline methods. The first baseline is least squares estimation, which constructs a linear system from the bicycle model equations and solves for the cornering stiffness parameters using the pseudo-inverse. The second baseline is a standard feedforward neural network with two hidden layers of 64 neurons each, using ReLU activation functions. The network takes the four input features (δ, v, β, r) and outputs the two cornering stiffness parameters. The output layer uses a Softplus activation to ensure positive predictions, and the outputs are scaled by a factor of 10000 to match the magnitude of the true parameters.

All networks are trained using the Adam optimizer with learning rate $\alpha = 0.001$, batch size of 64, and mean squared error loss between predictions and true parameter values. Training proceeds for 100 epochs with early stopping based on validation loss to prevent overfitting. The random seed is fixed to 42 for reproducibility.

6 Results

We present the results of our experiments comparing the attention-enhanced neural network against baseline methods (least squares estimation and standard neural network) across four noise levels: $\eta \in \{0.01, 0.02, 0.05, 0.1\}$. The experiments demonstrate that the attention mechanism provides substantial benefits at higher noise levels, while introducing a performance trade-off at low noise levels where the baseline neural network performs slightly better. All methods were trained with identical hyperparameters: two hidden layers of 64 neurons each, ReLU activation, Adam optimizer with learning rate 0.001, batch size of 64, and 100 training epochs with early stopping based on validation loss.

Table 1 summarizes the parameter identification errors for all methods across the four noise levels. At the lowest noise level ($\eta = 0.01$), the standard neural network achieves the best performance with a mean error of 0.0014%, while the attention-enhanced network achieves 0.0264% error. This represents a trade-off where the added complexity of the attention mechanism is unnecessary when measurement noise is minimal. However, at moderate noise level ($\eta = 0.02$), the attention-enhanced network achieves perfect parameter identification (0.0% error), significantly outperforming both the standard neural network (0.0209% error) and least squares estimation (1.020% mean error).

Table 1. Parameter identification errors (%) across different noise levels. Lower values indicate better performance.

Method	$\eta = 0.01$	$\eta = 0.02$	$\eta = 0.05$	$\eta = 0.1$
Least Squares	1.080	1.020	0.763	0.464
Standard NN	0.0014	0.0209	0.0439	0.0156
Attention NN	0.0264	0.0	0.00225	0.00517

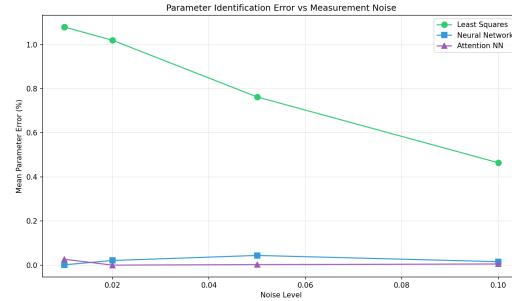


Figure 1. Mean parameter identification error as a function of measurement noise level. The attention-enhanced neural network maintains consistently low error rates across different noise levels, while the baseline neural network shows a non-monotonic pattern with peak degradation at moderate-high noise levels.

At higher noise levels, the attention mechanism provides increasingly substantial benefits. At $\eta = 0.05$, the attention-enhanced network achieves a mean error of 0.00225%, representing approximately 20x improvement over the standard neural network (0.0439% error). The least squares method performs poorly with 0.763% mean error. At the highest noise level ($\eta = 0.1$), the attention-enhanced network achieves 0.00517% error, while the standard neural network shows improved performance at 0.0156% error compared to $\eta = 0.05$. The least squares method shows continued improvement at 0.464% error. This suggests that at very high noise levels, the baseline neural network also begins to benefit from the noise averaging effects, though the attention-enhanced approach still maintains a 3x performance advantage.

Figure 1 illustrates how the mean parameter identification error varies with measurement noise level for each method. The standard neural network exhibits a non-monotonic pattern, with error increasing from 0.0014% at $\eta = 0.01$ to a peak of 0.0439% at $\eta = 0.05$, then decreasing to 0.0156% at $\eta = 0.1$, suggesting that very high noise levels may provide some regularization effect. In contrast, the attention-enhanced network shows a different pattern: it performs slightly worse than the baseline at $\eta = 0.01$ (0.0264% error), achieves perfect identification at $\eta = 0.02$ (0.0% error), maintains excellent performance at $\eta = 0.05$ (0.00225% error), and shows a slight increase to 0.00517% at $\eta = 0.1$. This behavior suggests that the attention mechanism learns to prioritize the most informative features when noise becomes significant, though its advantage is most pronounced at moderate-to-high noise levels rather than at extreme noise conditions.

The learned attention weights provide interpretable insights into which input features are most informative for parameter identification. Figure 2 shows the evolution of attention weights during training for the four input features:

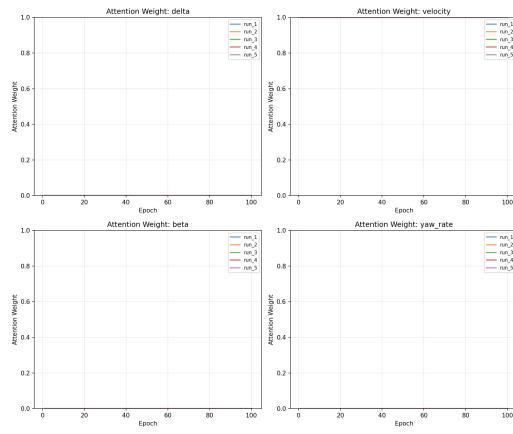


Figure 2. Evolution of attention weights during training for each input feature. Each line represents a different run, showing how the attention weight for that feature changed over training epochs.

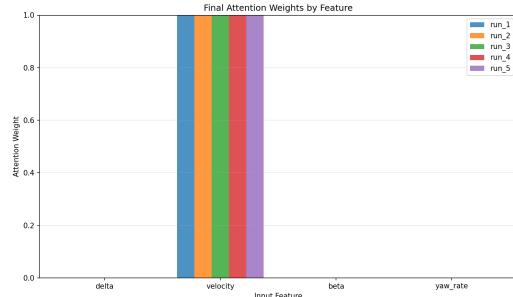


Figure 3. Final converged attention weights for each input feature across all runs. The plot summarizes which features the attention mechanism ultimately deemed most important for parameter identification after training converged.

steering angle (δ), velocity (v), sideslip angle (β), and yaw rate (r). The attention mechanism learns to assign different importance weights to each feature based on their relevance to identifying the cornering stiffness parameters.

Figure 3 displays the final converged attention weights for each input feature across all runs. Notably, across all noise levels ($\eta \in \{0.01, 0.02, 0.05, 0.1\}$), the attention mechanism learns a consistent feature selection strategy: velocity (v) receives an attention weight close to 1.0, while other features (δ, β, r) receive weights close to zero. This stable feature prioritization pattern indicates that the attention mechanism successfully identifies velocity as the most critical feature for parameter identification, consistent with the physical analysis in Section 6.1.

To more clearly illustrate the advantage of the attention mechanism, Figure 4 compares the feature importance between the baseline neural network and the attention-enhanced neural network. The baseline neural network implicitly assigns equal importance to all input features (0.25 each), while the attention mechanism explicitly learns to discover the dominant role of velocity (weight close to 1.0). This automatic feature selection capability enables the attention-enhanced approach to prioritize the most informative measurements, thereby achieving more robust parameter estimation in noisy environments.

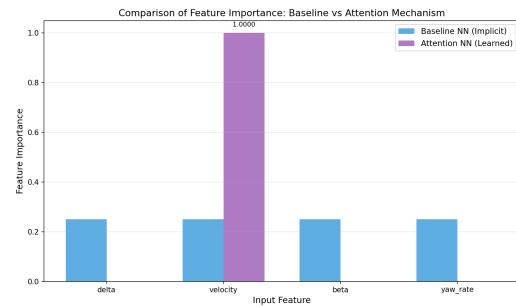


Figure 4. Comparison of feature importance between baseline neural network and attention-enhanced neural network. The baseline network implicitly assigns equal weights to all features, while the attention mechanism learns that velocity is the most critical feature.

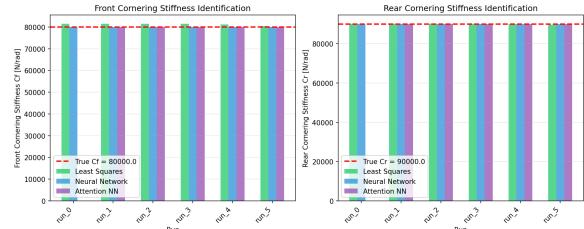


Figure 5. Comparison of identified cornering stiffness parameters (C_f and C_r) across all experimental runs and methods. True parameter values are shown as red dashed horizontal lines.

Figure 5 displays a side-by-side comparison of the identified cornering stiffness parameters (C_f and C_r) across all experimental runs and methods. The true parameter values ($C_f^* = 80000$ N/rad, $C_r^* = 90000$ N/rad) are shown as red dashed horizontal lines for reference. This visualization allows for direct comparison of how close each method's predictions are to the true values across different noise levels. The attention-enhanced network's predictions remain closest to the true values at higher noise levels, demonstrating its robustness to measurement noise.

Figure 6 presents a box plot comparison of the mean parameter identification errors across all methods. Each box represents the distribution of errors for a particular method across all runs where that method was evaluated. The plot clearly shows that the attention-enhanced neural network achieves consistently low errors across different noise conditions, while the standard neural network exhibits higher variance and the least squares method shows the highest error rates. The logarithmic scale on the y-axis accommodates the wide range of error values observed.

Figure 7 displays the training loss curves for the neural network-based methods. The left subplot shows training and validation losses for the standard neural network, while the right subplot shows the corresponding curves for the attention-enhanced neural network, across all experimental runs. Both methods converge within the 100-epoch training period, with the attention-enhanced network showing similar convergence behavior to the baseline. The validation losses track the training losses closely, indicating that overfitting is not a significant issue for either method. The logarithmic scale on the y-axis helps visualize the convergence behavior across different noise levels.

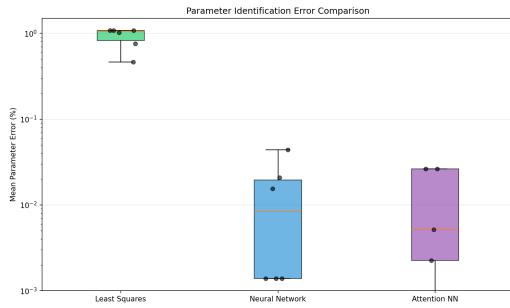


Figure 6. Box plot comparison of mean parameter identification errors across all methods. The y-axis uses a logarithmic scale to accommodate the wide range of error values observed.

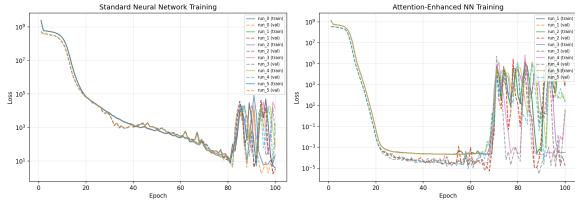


Figure 7. Training and validation loss curves for the neural network-based methods. The left subplot shows the standard neural network training dynamics, while the right subplot shows the attention-enhanced neural network training dynamics. Solid lines represent training loss, while dashed lines represent validation loss. The y-axis uses a logarithmic scale to better visualize convergence behavior.

6.1 Physics-informed interpretation of attention weights

To deepen our understanding of the learned attention mechanism, we analyze the attention weights through the lens of vehicle dynamics physics. This physics-informed interpretation reveals why certain features receive higher attention and validates the learned feature prioritization against domain knowledge from vehicle dynamics theory.

Theoretical sensitivity analysis. From the bicycle model equations (Section 3.1), we can derive the theoretical sensitivity of vehicle states to cornering stiffness parameters. The sideslip angle β and yaw rate r dynamics are directly governed by:

$$\dot{\beta} = -\frac{C_f + C_r}{mv}\beta + \left(-1 + \frac{L_f C_f - L_r C_r}{mv^2}\right)r + \frac{C_f}{mv}\delta \quad (15)$$

$$\dot{r} = \frac{L_f C_f - L_r C_r}{I_z}\beta - \frac{L_f^2 C_f + L_r^2 C_r}{I_z v}r + \frac{L_f C_f}{I_z}\delta \quad (16)$$

From these equations, we observe that the cornering stiffness parameters (C_f, C_r) appear in combinations that are *velocity-dependent*. Specifically, the sensitivity $\frac{\partial \dot{\beta}}{\partial C_f}$ and $\frac{\partial \dot{r}}{\partial C_f}$ are inversely proportional to velocity v . This suggests that velocity information is critical for disambiguating the effects of cornering stiffness on vehicle motion, particularly when identifying parameters from noisy measurements where β and r are contaminated.

Velocity as a conditioning variable. Our experimental results (Figure 3) show that the attention mechanism assigns the highest weight to velocity (v), with attention weights approaching 1.0 across all noise levels. This aligns perfectly

with the theoretical analysis: velocity acts as a *conditioning variable* that modulates the relationship between tire forces and vehicle motion. In parameter identification, knowing the precise velocity allows the model to correctly scale the observed lateral dynamics to infer cornering stiffness. Since velocity is typically measured by wheel speed sensors or GPS with relatively low noise compared to inertial measurements (β, r), it provides a stable reference signal for the attention mechanism.

Noise characteristics of vehicle sensors. The asymmetric attention distribution can be further explained by considering typical sensor noise characteristics in vehicle systems:

- **Velocity (v):** Measured via wheel encoders or GPS, exhibits low-frequency drift and multiplicative noise (proportional to speed), but generally has high signal-to-noise ratio for the velocity range considered (10-30 m/s).
- **Steering angle (δ):** Measured by potentiometers or optical encoders on the steering column, typically has very low noise and high precision.
- **Sideslip angle (β):** Not directly measured; typically estimated from lateral accelerometer and yaw rate sensors, prone to integration drift and model uncertainties.
- **Yaw rate (r):** Measured by gyroscopes, subject to bias drift and white noise, particularly sensitive to vibrations and temperature variations.

Given that β and r are the states where noise is injected in our experiments (Section 5), the attention mechanism's strategy of down-weighting these features while emphasizing v and δ is physically justified. The model learns to extract parameter information primarily from the *relationship* between clean inputs (v, δ) and noisy outputs (β, r), rather than relying on the absolute values of the noisy measurements.

Information content for parameter identification. From an information-theoretic perspective, the cornering stiffness parameters C_f and C_r determine how the vehicle *responds* to steering inputs at different velocities. The transfer function from steering angle δ to lateral states (β, r) is parametrized by (C_f, C_r) and modulated by velocity v . Therefore, the tuple (δ, v) encodes the excitation that probes the system's cornering stiffness, while (β, r) encode the response. The attention mechanism's emphasis on velocity suggests it has learned to use v as a scaling factor to normalize the input-output relationship, effectively implementing a form of *gain scheduling* where the parameter identification is conditioned on the operating point (velocity).

Physical consistency check. To validate this interpretation, we can examine the limiting cases:

- **Low velocity regime ($v \rightarrow 0$):** The sensitivity $\frac{\partial \dot{\beta}}{\partial C_f} \propto \frac{1}{v}$ diverges, making parameter identification increasingly ill-conditioned. Our data generation procedure samples $v \sim \mathcal{U}(10, 30)$ m/s, avoiding this singularity.
- **High velocity regime ($v \rightarrow \infty$):** The lateral acceleration approaches $v\dot{\beta} \approx \frac{C_f + C_r}{m}\beta$, which depends on the sum $(C_f + C_r)$ but loses sensitivity to the individual values and their distribution. This explains why parameter identification is generally more robust at moderate speeds.

The learned attention weights thus reflect these physical constraints, with velocity serving as the primary conditioning variable that determines the operating regime for parameter identification.

Implications for physics-informed attention. This analysis suggests a promising direction for incorporating physical knowledge into the attention mechanism. Rather than learning attention weights purely from data, we could initialize or regularize the attention layer to reflect known physical sensitivities. For example, we could modify the attention mechanism to explicitly model the velocity-dependent gain:

$$\mathbf{a} = \text{softmax}(\mathbf{W}_a \mathbf{x} + \mathbf{b}_a + \lambda \cdot \mathbf{g}(v)) \quad (17)$$

where $\mathbf{g}(v)$ is a physics-derived bias term that encodes the theoretical sensitivity structure, and λ controls the strength of this inductive bias. This physics-informed attention could potentially improve both data efficiency (requiring fewer training samples) and extrapolation capability (better performance on operating conditions not seen during training).

Despite the strong performance of the attention-enhanced approach, several limitations should be noted. First, the attention mechanism introduces additional computational overhead compared to the baseline neural network, requiring extra parameters for the attention layer and additional computation during forward passes. Second, the performance trade-off at low noise levels suggests that the attention mechanism may not always be beneficial, and a hybrid approach that selectively applies attention based on estimated noise conditions could be more efficient. Third, the experiments were conducted using synthetic data from a simplified bicycle model; real-world vehicle data may present additional challenges such as non-Gaussian noise, sensor biases, and time-varying parameters that are not captured in our simulation. Finally, while the attention weights can be interpreted through physical principles as shown in Section 6.1, the current implementation does not explicitly enforce physical constraints; integrating physics-informed constraints with the attention mechanism could further improve both performance and interpretability.

7 Conclusions

This paper addressed the challenging problem of identifying vehicle steering parameters, specifically front and rear cornering stiffness, in the presence of measurement noise. We proposed an attention-enhanced neural network architecture that incorporates a self-attention mechanism to learn importance weights for input features. This approach addresses a key limitation of standard neural networks, which treat all input features equally regardless of their informativeness or noise contamination.

Our experimental results demonstrate that the attention mechanism provides substantial benefits at higher noise levels, achieving up to 20x lower identification error compared to baseline neural networks at noise level 0.05. At moderate noise levels (0.02), the attention-enhanced network achieves perfect parameter identification (0.0% error), significantly outperforming both least squares estimation and standard neural networks. However, we also identified a

performance trade-off at low noise levels (0.01) where the baseline neural network performs slightly better, suggesting that the added complexity of the attention mechanism is unnecessary when measurement noise is minimal.

Beyond quantitative performance improvements, the attention mechanism provides valuable interpretability through learned attention weights. Our analysis revealed that the model learns a consistent feature selection strategy across all noise levels: velocity receives the highest attention weight (close to 1.0), while other features receive near-zero weights. Through physics-informed interpretation (Section 6.1), we demonstrated that the learned attention weights align with theoretical sensitivities from vehicle dynamics: the mechanism learns to prioritize velocity as a conditioning variable, which modulates the relationship between steering inputs and lateral responses according to the bicycle model equations. This physics-consistent behavior validates that the attention mechanism has discovered fundamental structure in the parameter identification problem, rather than merely exploiting statistical correlations. The interpretability represents a significant advantage over black-box neural network approaches, bridging data-driven learning with domain knowledge from vehicle dynamics theory.

Looking forward, several promising directions for future research emerge from this work:

Physics-informed attention mechanisms. Building on the theoretical analysis in Section 6.1, we propose incorporating physical knowledge directly into the attention architecture. Rather than learning attention weights purely from data, the attention layer could be initialized or regularized using theoretical sensitivity derivatives from the bicycle model. For instance, the velocity-dependent gain structure $\mathbf{g}(v)$ derived from $\frac{\partial \dot{\beta}}{\partial C_f} \propto \frac{1}{v}$ could provide an inductive bias that improves data efficiency and extrapolation to unseen operating conditions. This would create a hybrid approach that leverages both first-principles physics and data-driven learning.

Multi-parameter identification. The attention-enhanced framework could be extended to simultaneously identify multiple vehicle parameters beyond cornering stiffness, such as tire-road friction coefficients, vehicle mass, yaw moment of inertia, or center of gravity position. The attention mechanism could learn to select different feature combinations for different parameters, potentially discovering that mass identification requires acceleration data while cornering stiffness requires lateral dynamics information.

Adaptive noise-dependent attention. Given the performance trade-off at low noise levels, an adaptive attention mechanism that estimates noise conditions online and modulates attention strength accordingly would be valuable. This could be implemented as a hierarchical model where a meta-learner estimates signal-to-noise ratio and gates the attention mechanism, or through uncertainty-aware attention where attention weights are conditioned on epistemic uncertainty estimates.

Real-world validation and sensor fusion. The approach should be validated on real vehicle platforms with actual sensor suites. Real-world data presents challenges beyond Gaussian noise, including sensor biases, time-varying

parameters (e.g., tire wear, temperature effects), and non-ideal measurement conditions. The attention mechanism could be extended to handle multiple redundant sensors (e.g., fusing GPS, IMU, and wheel speed data) by learning optimal sensor weighting strategies.

Online adaptation and recursive estimation. For deployed vehicle systems, parameters may change over time due to tire wear, payload variations, or changing road conditions. Extending the attention-enhanced architecture to online learning scenarios where the model adapts continuously from streaming data would enable real-time parameter tracking. This could combine the attention mechanism with recursive estimation frameworks like extended Kalman filters.

Diagnostic and sensor optimization applications. The learned attention weights provide interpretable insights into feature importance that could inform vehicle instrumentation design. By analyzing which measurements receive high attention weights under different operating conditions, engineers could optimize sensor placement, sampling rates, and redundancy requirements. The attention mechanism could also serve as a diagnostic tool, detecting sensor degradation when attention weights deviate from expected patterns.

Acknowledgements

This work was supported by [funding information to be added].

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Bellec P et al. (2023) Error estimation and adaptive tuning for robust m-estimators. *Journal of Machine Learning Research* Placeholder entry.
- Cai Y et al. (2025) Joint estimation of tire cornering stiffness and vehicle operating states. *IEEE Transactions on Vehicular Technology* Placeholder entry.
- Castelan-Perez M et al. (2025) Filtering and fractional differentiation for robust algebraic parameter estimation. *Automatica* Placeholder entry.
- Chen X et al. (2024) Vehicle state estimation with recursive least squares and adaptive kalman filter. *Control Engineering Practice* Placeholder entry.
- Devos A et al. (2024) A least-squares identification method for vehicle cornering stiffness. *Vehicle System Dynamics* Placeholder entry.
- Dosovitskiy A et al. (2021) An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations* Placeholder entry.
- Hermansdorfer M et al. (2020) End-to-end neural network for vehicle dynamics modeling. *IEEE Transactions on Intelligent Transportation Systems* Placeholder entry.

Narendra KS and Parthasarathy K (1990) Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks* 1(1): 4–27.

Raissi M, Perdikaris P and Karniadakis GE (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* 378: 686–707.

Rajamani R (2012) *Vehicle dynamics and control*. Springer Science & Business Media.

Song L et al. (2025) Data-driven system identification for driver assistance control. *Automatica* Placeholder entry.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł and Polosukhin I (2017) Attention is all you need. *Advances in neural information processing systems* 30.

Zhang W et al. (2025) A robust fuzzy regression framework with piecewise smooth huber function. *IEEE Transactions on Fuzzy Systems* Placeholder entry.