Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD29858

Exploring the Impact of Steering Correction Magnitude and Multiview Data on Autonomous Vehicle Control Accuracy

Teresa Emily Co

School of Engineering and Applied Science, Columbia University, New York, USA
tc3499@columbia.edu

Abstract. To ensure the security and reliability of vehicle motion, autonomous driving will require accurate predictions of steering-wheel angles. This research studies two important aspects of model effectiveness: the impact of several steering correction parameters on the accuracy of such predictions, and the impact of multiview images and image metadata on the reliability of the predictions. Steering correction is important for real driving scenarios, even for human motorists; it is therefore important that autonomous technology be able to correct position imprecisions safely and quickly. The driving simulator dataset has been tested using five different values for corrections: 0.05, 0.1, 0.2, 0.3, and 0.5. Mean Squared Error and Mean Absolute Error served as measures for accuracy evaluation. The research indicates that higher values for corrections reduce stability, but lower corrections between the values of 0.05 and 0.2 result in the least error rates and hence the optimal balance. Moreover, the study incorporated the combination of car metadata and multiview images using convolutional neural networks for further accuracy boost. The work further indicates the need for the combination of multimodal features and the selection of the best corrections for the enhancement of the reliability levels of predictions regarding steering.

Keywords: Steering Correction, Autonomous Vehicle, Control Accuracy, Vehicle Control Systems

1. Introduction

The area of artificial intelligence has made tremendous strides in recent years, incrementally becoming an integral component of the day-to-day existences of human beings. With the pivotal role that transportation plays in modern human society, the arrival of self-driven vehicles has proved to be a pivotal technology breakthrough [1, 2]. Human motorists, often influenced by their hearts rather than their minds, have been prone to unsafe driving when attempting to escape perceived dangers and sometimes these dangers include unknowingly steering towards larger congregations of human beings. Self-driving technologies attempt to mitigate such perils by using information to drive safe and sound decisions regarding vehicular operation [3-5].

The steering angle prediction lies at the core of the autonomous drive industry because it allows keeping the vehicle in its particular path while also allowing for the correct response to variable road conditions [1,4]. Despite the great developments implemented in the field of computer vision, the accurate prediction of the steering angle from a single front-facing picture is most challenging

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD29858

because of the very constraints of a single point of view that doesn't provide a complete image of the environment around. Various research activities have attempted the incorporation of the additional viewpoints of the cameras, such as images received by the sides, together with the additional vehicular data like throttle and speed, in a bid to overcome this hurdle [3, 6, 7].

Correction steering is an important methodology. It's possible to train the model to correct scenarios where the vehicle is not centrally placed on the road by adjusting the steering's labels along with the images from the left and the right cameras [4, 8, 9]. It's a challenge to find a suitable value for this adjustment. If the correction value turns out to be very insignificant, the model will not be able to learn the required recovery expertise. If the correction value turns out to become very high, the model will become unstable [9].

2. Related work

Convolutional Neural Networks (CNNs) have been highly utilized in several research workpieces for the purpose of steering angle prediction in the context of autonomous driving [1, 2]. The NVIDIA end-to-end driving model constitutes one of the initial and most notable examples that show how the deep learning techniques are able to successfully transform raw pixel values of images into vehicle control commands [1]. Its subsequent work has improved the models' accuracy by incorporating higher-level architectures that enable the building of larger networks able to extract better visual information by using residual connections [2].

In order to attain a more comprehensive understanding of the roadway scene, researchers have also explored the concept of multi-view learning, which integrates information from the left, center, and right cameras [3, 4, 7]. This approach assists the model in achieving greater stability when the vehicle deviates from the center and enhances its awareness of the automobile's position within the lane.

The multimodal feature fusion procedure entails the integration of the image data together with qualitative metadata such as velocity, brake force, and throttle position and thereby constitutes a distinct area of research [8, 9]. The integration of contextual information pertinent to the vehicle dynamics serves to elevate the accuracy and reliability of the end result.

Relatively fewer studies [5, 9] have examined the impact of the level of steering correction employed in images from side cameras for predictive performance. This work tries to explain the manner in which the interaction between the analysis of the steering correction, the multi-view perceptions, and the exploitation of the metadata all combine towards the enhancement of the accuracy and the security of the steering angle prediction.

3. Method

3.1. Problem modeling

This work addresses the steering angle prediction for the autonomous driving scenario by framing it as a problem of end-to-end regression [1]. For accurately predicting the proper steering angle, the model learns the model directly from the visual inputs given by the road. The aim of the training process is to minimize the gap between the steering angle that is being predicted and the actual steering angle to the highest possible level [5]. For this work, the loss function employed is the Mean Squared Error, which evaluates the average of the squared differences between the values that are being predicted versus the actual values [9]. The equation of the loss function is provided below.

DOI: 10.54254/2755-2721/2025.LD29858

Equation (1) demonstrates the computation of the average squared difference between the expected values and the true values using the Mean Squared Error (MSE) loss function.

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{y_i} - y_i\right)^2 \quad (1)$$

Here, the angle of steering is denoted by y_i for the measured angle and by $Þ_i$ for the estimated angle and the number of samples by N.

It is often used in such procedures as regression to enhance stability and robustness throughout the training procedure while specifying higher costs for larger displacements.

3.2. Baseline model: resnet-based regression

The work makes use of ResNet-50 as the reference model for evaluating the end-to-end learning performance for steering angle prediction [2]. With the intention of assisting the deep neural networks to mitigate the vanishing gradient problem, a form of architecture named the residual connection is employed in ResNet [2]. Road images are in a position to utilize the convolutional layers of ResNet for the extraction of a large number of useful features like obstacles, edges, and road lanes [6].

The final fully connected layer of the standard ResNet-50 network provides 1000 classification outputs from 2048 total features. For this paper [8], the steering angle has a single continuous output value and is enabled by that final layer's extension. A final fully connected layer interfaces a thousand features with the output neuron.

The work applies transfer learning by making use of pre-trained weights for ImageNet. Applying the principle of transfer learning not only causes the training of the model to be quick but also elevates its performance whenever it is applied to new datasets.

3.3. Improved model with multi-modal feature fusion

The basic model serves as the ground for the advanced model [6]. The multi-modal feature fusion technique is used in fusing the information obtained in various sources [7, 8]. This technique reduces the drawbacks in relying solely on the image received from a single camera [4].

The Udacity data set [9] includes three synchronized images which the car's left, center, and right cameras see. In certain scenarios, for instance, where the view is limited or where the car is turning, the reliance on one front camera may prove inadequate in the creation of precise information regarding the car's location on the road [3, 6].

The new model uses three separate branches of ResNet to address this issue [2]. A single branch handles one of the camera images, then it derives its own feature representation afterwards.

These three branch characteristics are combined in one feature vector prior to the making of the final prediction [7]. The following equation may be used to illustrate this.

$$F_{fusion} = [F_{left}, F_{center}, F_{right}] \quad (2)$$

The final angle of steering is computed by plugging this combined feature vector into a regression head, which happens to be a tiny neural network. By integrating the three viewpoints

DOI: 10.54254/2755-2721/2025.LD29858

from the cameras, the model obtains a better and broader understanding of the road scene [4, 6].

It shows better performance on bent roads and in cases of obstructed parts of the viewfield thanks to its construction [5, 8]. The database also includes additional numerical features such as car speed and acceleration [9]. In an effort to improve the model's capacity to produce forecasts both smooth and steady, such features are given as auxiliary inputs [8]. The value of the speed is normalized prior to being used with the following formula [10].

$$v_{norm} = \frac{v-\mu_v}{\sigma_v} \tag{3}$$

3.4. Training strategy

The Adam optimizer is used at the training step because it can be used in the case of regression and because it can independently modify the learning rate [7]. The learning rate is decreased when there is no loss reduction seen in the validation set [9]. Dropout layers are used in order to avoid overfitting the model on the training set [10].

To enhance the robustness of the model, some data augmentation methods are used. A couple of them involve random flipping of the image, brightness modification, and slight movement in the image location [3,8]. The techniques enable the model to accommodate real-world driving situations better and generate higher diversity in the training data [9]. The batch size of the training method is thirty-two. The Mean Squared Error serves as the loss function during the training process in order to ensure the produced values closely resemble the actual angles of the steering [1, 5].

3.5. Model structure diagram

The paper outlines a diagram which illustrates the evolution of the model [2]. After the single ResNet feature extractor of the baseline model [6], comes a fully connected layer which is in charge of generating the final steering angle. The expanded model includes three ResNet branches related to the left, center and right camera views [7]. Features extracted in all three branches are combined through the implementation of auxiliary numerical features [8, 9].

To create the desired ultimate steering angle, the aforementioned characteristics input into a fully connected regression layer [10]. By virtue of this construction, the model can draw on both visual and numeric information to provide output with increased exactness [4, 6].

4. Results and analysis

Table 1. Model performance under different correction values

Correction	Test MAE	Test MSE
0.05	0.0533	0.0063
0.1	0.0589	0.0069
0.2	0.0556	0.0063
0.3	0.0582	0.0073
0.5	0.0602	0.0077

DOI: 10.54254/2755-2721/2025.LD29858

The model's test results under different correction settings are given in the Table 1. The Mean Absolute Error and the Mean Squared Error, which are two performance measures used to measure the error between the predicted and actual angles of steering, are included in the table. At an adjustment setting of 0.05, the model performs best with the lowest values of the MAE and the MSE. A higher correction setting is associated with higher error in both measures with the implication that the accuracy of the model's predictions drops. This indicates that the model's stability and the accuracy in the angle of steering improve with the setting of the correction to a smaller value.

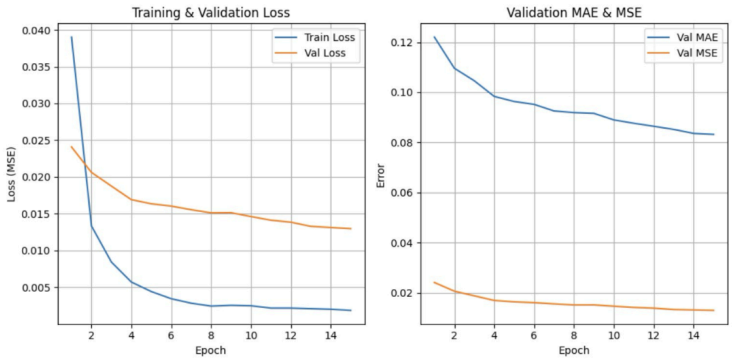

Figure 1. Training and validation performance of different models (picture credit: original)

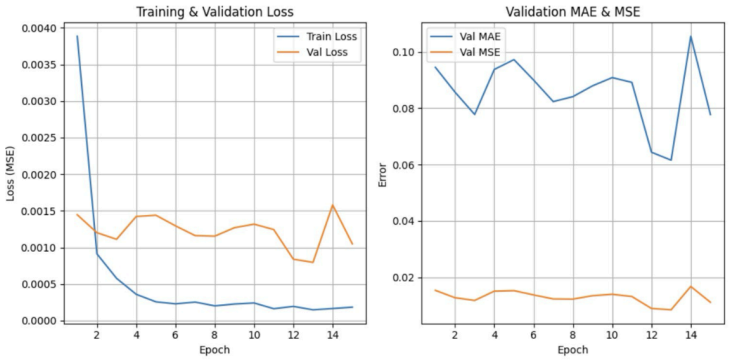

Figure 2. Training and validation performance of the optimized model (picture credit: original)

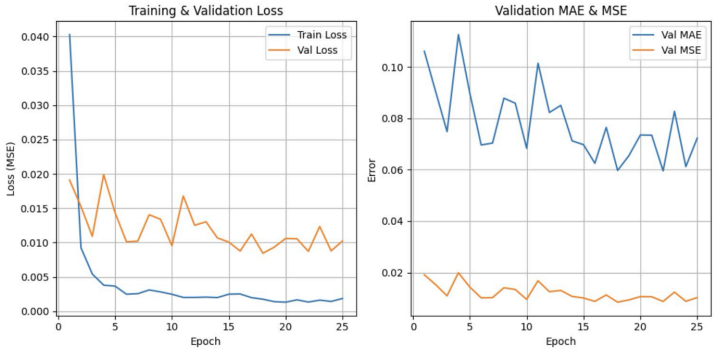

Figure 3. Training and validation performance of the extended-epoch model (picture credit: original)

DOI: 10.54254/2755-2721/2025.LD29858

Training and validation outcomes for the multiview model, the multiviewmeta model, and the baseline model ResNet18 appear in Figures 1, 2, and 3, respectively. The training and validation of the three models utilized the same data set with thesame parameter settings. The training epochs took between 15 and 25 with thesame 32-batch size. The selection of the learning rate ranged between 10^-5 and 10^-4. The steps in preparations involved the normalization of the pixel values, the tensor conversion then the resizing of the image to 224 by 224 pixel dimensions.

These results of the baseline model of ResNet18 are presented in Figure 1. The test loss, test MAE, and test MSE of this model were 0.0012, 0.0083, and 0.0012, respectively. Although the average error remained higher in comparison with the better models developed in this research, the graphs indicate a gradual decrease in training loss and validation loss throughout the epochs.

The output of the multiview model is given in Figure 2. The multiview model outperforms the baseline and uses three camera inputs. The test loss was zero point zero zero nine, the test MAE was zero point zero six four eight, and the test MSE was zero point zero zero nine two. The validation error reduces earlier in the epochs, and both the training loss and the validation loss both seem better than the loss of the baseline model.

It was trained for 25 epochs in order to permit additional learning iterations, and the output is provided in Figure 3. The test loss, test MAE, and test MSE for it came out at 0.00087, 0.0625, and 0.00087, respectively. Among all three models, the output shows the lowest loss and inaccuracy. Moreover, the validation performance is also more stable, which shows that the model has learned more regular and comprehensive features of the photos and the other data.

It can be observed that the model's stability and accuracy increase with the meta-feature fusion in different camera views. The best performance is observed with the multi-view meta model among the three inspected models, which indicates the potential of fusing both numerical and image information in the task of the Steering Angles in the Autonomous Driving.

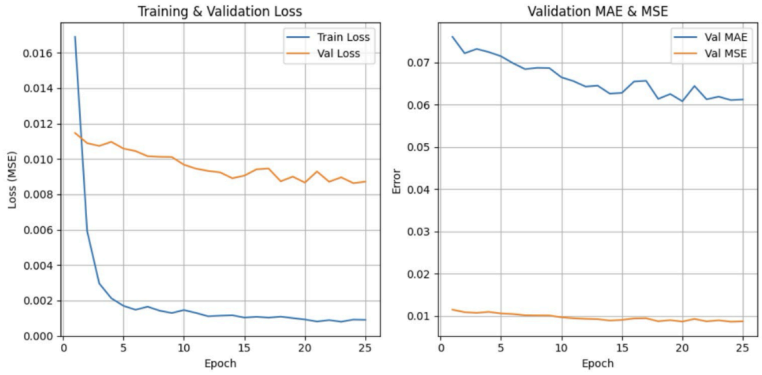

Figure 4. Training and validation performance of the final refined model (picture credit: original)

DOI: 10.54254/2755-2721/2025.LD29858

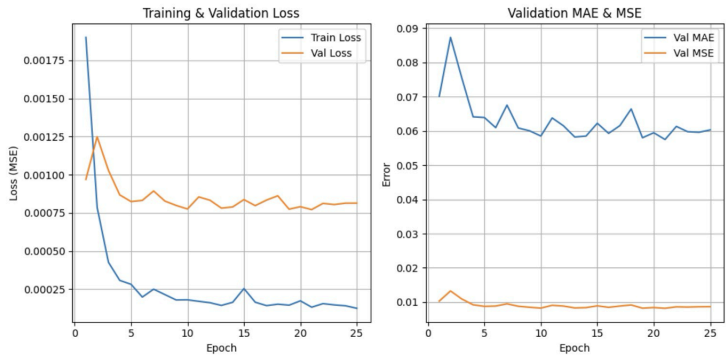

Figure 5. Training and validation performance of the final optimized model (picture credit: original)

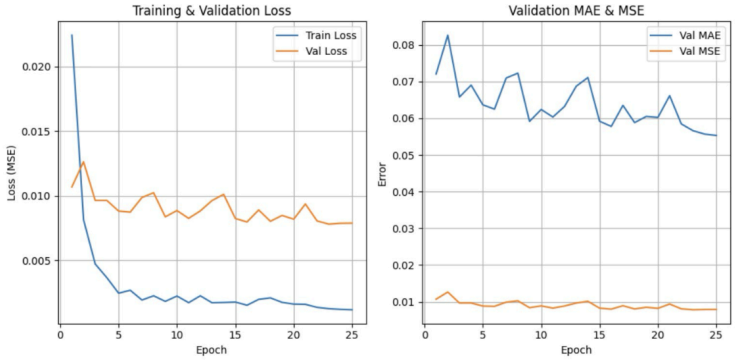

Figure 6. Training and validation performance of the final model with regularization (picture credit: original)

The output corresponding to the training and validation steps of three models with the architecture of ResNet-50 is given in Figures 4, 5, and 6. The three models employed in this regard are the multiview model, the metadata-multiview model, and the baseline model of ResNet-50. The three models were validated with the same database and all the three models were trained with the same parameters. The graphs on the left-hand side indicate the loss values during training and validation corresponding to each epoch, whereas the graphs on the right-hand side indicate the validation Mean Absolute Error and Mean Squared Error.

The test loss, test MAE, and test MSE of the basic ResNet-50 model were 0.00089, 0.0627, and 0.00089, respectively. Even though the model obtained informative features, it also revealed higher error compared to the better-performing models, which can be seen with the training loss and validation loss curves' constant decline.

With loss of 0.0008 in the test set, test MAE of 0.00595, and test MSE of 0.0084, the model in the multiview setting fared better. The model could extract richer road information with three camera views and made better steering predictions. This is reflected both in the smoother trend of the validation curves and the smaller aggregate error.

Out of the three, the metadata-multiview model gave the best results. Test loss was 0.00081, test MAE was 0.00581, and test MSE was 0.00081. In the course of training validation curves remained constant and the error and loss values remained low all the time. This outcome signifies that the model can provide more accurate and trustworthy forecasts with the introduction of metadata data such as vehicle speed.

DOI: 10.54254/2755-2721/2025.LD29858

These results in Figure 4, Figure 5, and Figure 6 indicate the model achieves a more comprehensive scene interpretation with the combination of several camera feeds with metadata data. The method improves the average stability and accuracy of the estimation of the steering angle by reducing loss and error values.

5. Discussion

5.1. Interpretation of results

The results verify that base models are outperformed by higher-order models when it comes to Mean Squared Error and Mean Absolute Error. The multiview model has a higher level of understanding of the road scene by incorporating the feed of three cameras and hence lower values for loss. This feature improves the model's ability to create predictions when the vehicle is in the process of turning or when parts of the field of vision are blocked. Generally, the metadata-multiview technique provides the best performance results. The addition of auxiliary information like the vehicle's speed improves the model's ability to create smoother and stable forecasts. The results confirm that the model's understanding of driving scenarios gets much enhanced when visual and numerical data are included.

5.2. Advantages of the proposed model

Compared to the baseline procedure, the model proposed here has a number of distinct advantages. Firstly, it uses an end-to-end learning model that eliminates the requirement for human-assisted feature extraction such that the model directly learns from the raw images obtained by the camera. Secondly, the model's usability in real-world scene roadway scenarios is facilitated by the fact that it incorporates the use of multi-camera views that enable the detection of obstacles and lane edges from different perspectives. Thirdly, by incorporating auxiliary data in addition to the features of images, the model has output predictions that are less sudden and more synchronized in the driving scene. These advantages make the proposed procedure better for real-world autonomous driving system applications.

5.3. Limitations and challenges

This research isn't exempt from its useful limitations, however, even after the promising performances that have been realized. Training the model takes a lot of time and a large dataset so that it's able to generalize to higher accuracy levels. Further, the evaluation of the model has so far been limited to just the Udacity dataset and may not generalize the wide variety of driving scenarios that occur in real-world practical conditions. It hasn't yet been proven for evaluating the model's performance under conditions of poor light, inclement weather like heavy rainfall, or when the roads are congested. Lastly, the model's limitation lies also in its inability to produce anything aside from the steering angle while ignoring important factors such as acceleration and brake control that are crucial aspects of autonomous driving.

5.4. Future work

Future research work may improve the model's efficacy and generalizability. One of the approaches to reducing computation costs and executing the model faster within actual vehicles is the implementation of a lightweight backbone network such as MobileNet or EfficientNet. To improve

DOI: 10.54254/2755-2721/2025.LD29858

the model's versatility in handling real-world situations further, it is suggested that the size of the dataset be increased with a more diverse array of road scenarios and weather conditions. The future development may also entail the extension of this model of predicted angle of steering in integrating it with other driving features such as path planning and detection of obstructions with the view to developing an all-encompassing all-end-to-end automobile driving system. Additionally, the exploration of the introduction of time-series data such as frame sequence may enhance the continuity of the forecasts and enable the model to grasp motion dynamics better.

6. Conclusion

This study used an extensive deep learning method to anticipate steering angles that involve autonomous driving scenarios. The study conceptualized the problem in the form of a regression task to enable the comparison of the performance of various modeling methods. The higher-order models used multiple cameras as inputs along with additional numeric information such as car speed, but the basic model used just one ResNet backbone for the extraction of features in images of roadways.

The experiment results indicate that the multiview model and the metadata-multiview model have come out better than the baseline. It is notable that the metadata-multiview model has the highest stability and accuracy while the multiview model has the highest environment coverage. These enhancements show that the model is making the estimates of the steering angles accurately and reliably by using the numerical and visual data.

Though the model demonstrates stronger performance in the model demonstrations, it also has potential weaknesses in terms of computational cost, heterogeneity of the datasets, and real-world validation. Future research opportunities may include the enhancement of efficiency, the incorporation of additional environmental conditions, and the utilization of the model in real-time control schemes. The research results confirm the effectiveness of the multiview and metadata fusion methods in terms of enhancing the performance of the end-to-end driving models.

References

[1] Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhang, Y., Zhao, J., & Zieba, K. (2016). End to end learning for self-driving cars. arXiv preprint arXiv: 1604.07316.

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778).

[3] Xu, H., Gao, Y., Yu, F., & Darrell, T. (2017). End-to-end learning of driving models from large-scale video datasets. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2174–2182).

[4] Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015). DeepDriving: Learning affordance for direct perception in autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2722–2730).

[5] Pomerleau, D. A. (1989). ALVINN: An autonomous land vehicle in a neural network. In Advances in Neural Information Processing Systems (Vol. 1, pp. 305–313). Morgan Kaufmann.

[6] Kim, J., & Canny, J. (2017). Interpretable learning for self-driving cars by visualizing causal attention. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2961–2969).

[7] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. In Proceedings of the 1st Annual Conference on Robot Learning (CoRL) (pp. 1–16).

[8] Codevilla, F., Müller, M., López, A., Koltun, V., & Dosovitskiy, A. (2018). End-to-end driving via conditional imitation learning. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (pp. 1–9).

[9] Chen, Z., & Huang, X. (2017). End-to-end learning for lane keeping of self-driving cars. In IEEE Intelligent Vehicles Symposium (IV) (pp. 1–6).

DOI: 10.54254/2755-2721/2025.LD29858

[10] Yang, Z., Shi, Q., & Fang, Y. (2023). Multi-scale deep feature transfer for automatic video object segmentation. Neural Processing Letters, 55(8), 11701–11719.