

Enhancing Accuracy and Robustness of Steering Angle Prediction with Attention Mechanism

Swetha Nadella^{1,*}, Pramiti Barua^{1,*}, Jeremy C. Hagler¹, David J. Lamb¹, and Qing Tian^{1,2}

¹Dept. of Computer Science, Bowling Green State University, Bowling Green, OH 43403, USA

²Dept. of Computer Science, University of Alabama at Birmingham, Birmingham, AL 35294, USA

{nswetha, pbarua, jhagler, djlamb}@bgsu.edu, qtian@uab.edu

Abstract—In this paper, our focus is on enhancing steering angle prediction for autonomous driving tasks. We initiate our exploration by investigating two veins of widely adopted deep convolutional neural architectures, namely ResNets and InceptionNets. Within both families, we systematically evaluate various model sizes to understand their impact on performance. Notably, our key contribution lies in the incorporation of an attention mechanism to augment steering angle prediction accuracy and robustness. By introducing attention, our models gain the ability to selectively focus on crucial regions within the input data, leading to improved predictive outcomes. Our findings showcase that our attention-enhanced models not only achieve state-of-the-art results in terms of steering angle Mean Squared Error (MSE) but also exhibit enhanced adversarial robustness, addressing critical concerns in real-world deployment. For example, in our experiments on the Kaggle SAP and our created publicly available datasets, attention can lead to over 6% error reduction in steering angle prediction and boost model robustness by up to 56.09%.

Index Terms—Steering angle prediction, attention, adversarial robustness

I. INTRODUCTION

As we approach the era of autonomous vehicles, precise and reliable prediction of steering angles is essential in ensuring safe and efficient self-driving operations of a vehicle. The accuracy of this steering angle prediction directly impacts the vehicle's ability to navigate complex and dynamic environments. Deep learning models have emerged as powerful tools to tackle this challenge [1]. Although deep learning models can increase the accuracy of steering angle prediction, it is important to acknowledge their shortcomings, including inefficiency and susceptibility to adversarial attacks. Such issues prevent deep neural networks' wide deployment in real-world steering angle prediction. This paper first embarks on a comprehensive exploration of the two most popular veins of deep convolutional neural networks (i.e., InceptionNet and ResNet variations) for the task of steering angle prediction. Specifically, we explore the two types of networks positioned towards the compact end of the complexity spectrum. More importantly, in this paper, we introduce an attention mechanism to the task of steering angle prediction, which can enhance both the prediction accuracy and robustness. For example, for ResNet32, the introduction of attention leads

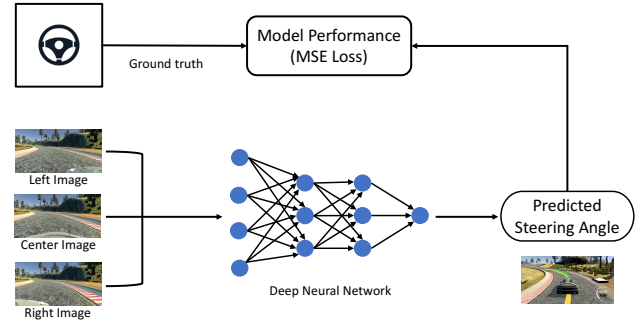


Fig. 1: Illustration of deep-learning-based steering angle prediction utilizing visual data captured by left, central, and right cameras.

to a 6.83% error reduction in steering angle prediction and a robustness increase of up to 53.96%. Qualitatively, we demonstrate that through our incorporation of the attention mechanism, our models gain the capability to selectively concentrate on vital regions within the input images.

II. RELATED WORK

Autonomous driving is a rapidly evolving field that can potentially revolutionize the future of mobility. The current state of the work done in steering angle prediction is early compared to other areas like visual recognition and detection. Steering angle prediction continues to rely more heavily on visual processing and understanding. For example, Tesla has attempted to use pure vision in their cars [2]. Steering angle prediction needs to be accurate, efficient, and robust. Gidado *et al.* [3] survey different deep learning (DL) methods for steering angle prediction. Most works in this direction focus on improving the accuracy with large models, but few explore the accuracy-efficiency trade-off and robustness issues. While there are continuous changes in the field, the ResNet architecture remains a key player in steering angle prediction [4; 5; 6], utilizing input from cameras, lidars, and/or radars. Oussama and Mohamed [4] showed that a ResNet50-based architecture gives a satisfactory prediction of the due wheel angle. However, for on-board steering angle prediction, the cost to train and deploy ResNet50 is high, due to its large number of parameters [7]. Navarro *et*

*These authors contributed equally.

al. [1] presents complex models and shows that these new end-to-end models have better optimization from the use of additional data information, like speed. Al-Qizwini *et al.* [8] found InceptionNet as a best model for various autonomous driving tasks. McNeely-White *et al.* [9] discovered that InceptionNet can extract qualitatively equivalent information as its ResNet counterpart. Based on one single image, He *et al.* [10] utilized an aggregated set of large VGG16 models to extract features for steering angle prediction. The LSTMs utilized further contributed to the approach's inefficiency. On an unpublished dataset, Zhu *et al.* [11] proposed a multi-task MapNet to simultaneously search the passable area, detect the intersection, and predict the steering angle with many expensive 5×5 conv filters.

Attention has been used in a variety of computer vision tasks, and there exists a varied interpretation of its definition. In contrast to the costly attention mechanism in modern transformers [12], this paper explores attention mechanisms for convolutional neural networks (CNNs), aiming to improve steering angle prediction on the edge. Specifically, we focus on task-dependent soft attention [13; 14] for CNNs. Wang *et al.* [15] introduced a residual attention network that emphasizes the most informative regions of an image to enhance classification accuracy and efficiency. The attention mechanism works by adaptively weighing different features of the image, based on the context and spatial location. Within each attention module, a bottom-up feed-forward structure generates feature maps with strong semantic information at lower resolutions, while a top-down network generates dense features at the pixel level to serve as the attention mask. Inserting skip connections between the bottom and top feature maps is found to be beneficial in applications like image segmentation [16]. Due to its effectiveness, such attention mechanisms have been adopted in many deep classification models. However, no works have investigated such attention mechanisms in the regression task of steering angle prediction. It is worth mentioning that He *et al.* [10] adopted a different sparse attention mechanism, which is impractical for edge applications due to its cumbersome VGG16 backbone and its reliance on LSTMs. Also, it is based on a single input image.

The robustness of steering angle predictors is also of paramount importance. Prior studies have highlighted the impact of adversarial attacks in the realm of image classification [17]. This inherent susceptibility raises concerns about the safety and reliability of neural networks, particularly in the safety-critical field of self-driving cars [18]. While adversarial attacks on neural networks have been extensively explored in different fields, mainly in image classification, the investigation of these attacks on autonomous steering angle prediction is limited, if any. In this paper, particularly in the context of steering angle prediction, we evaluate the vulnerability of neural networks to adversarial attacks such as Fast Gradient Sign Method (FGSM) [17] and Projected Gradient Descent (PGD) [19]. We can successfully improve

the robustness of steering angle prediction by incorporating an attention mechanism.

III. METHODOLOGY

Figure 1 offers a high-level overview of our utilization of deep learning models and visual data for steering angle prediction. Our deep models are created to process visual information from the left, central, and right cameras, enabling the inference of steering angles. Given the limited resources for on-board steering angle prediction, we focus our attention on compact deep models in this paper. We will begin by providing a comparison of the two classic families of deep CNN models for steering angle prediction. Subsequently, we will introduce an attention mechanism to enhance the precision and robustness of steering angle prediction models.

A. Classic Deep CNN Models for Steering Angle Prediction

Our first goal is to find a compact model with an accuracy score that is comparable to or better than its larger counterparts. This paper explores the two most popular veins of CNNs (ResNet [20] and InceptionNet [21] variations) with compact capacities. We present the outcomes of various sizes for each of the two families. For example, we built and trained eight ResNet architectures; ResNet20, ResNet22, ResNet24, ResNet26, ResNet28, ResNet30, ResNet32, and ResNet34. We built seven InceptionNet models to match similar numbers of parameters so that we could compare the architectures at different complexity levels. The number of parameters and detailed configurations for each of the ResNet and InceptionNet models are illustrated in Table I and II. We will provide the comparison results in Sec. IV.

TABLE I: ResNet Model Parameters & Block Layers

ResNet Model	Parameters	Block Layers
ResNet20	12.8 million	(2,2,3,2)
ResNet22	13.1 million	(2,3,3,2)
ResNet24	14.3 million	(2,3,4,2)
ResNet26	17.9 million	(3,3,3,3)
ResNet28	19.1 million	(3,3,4,3)
ResNet30	19.4 million	(3,4,4,3)
ResNet32	20.6 million	(3,4,5,3)
ResNet34	21.7 million	(3,4,6,3)

TABLE II: InceptionNet Model Parameters & Block Layers

InceptionNet Model	Parameters	Block Layers
InceptionNet	13.0 million	(2,5,2)
InceptionNet a	14.5 million	(5,5,2)
InceptionNet b	15.5 million	(2,8,2)
InceptionNet c	17.0 million	(5,8,2)
InceptionNet d	17.6 million	(2,5,5)
InceptionNet e	20.2 million	(2,8,5)
InceptionNet f	21.7 million	(5,8,5)

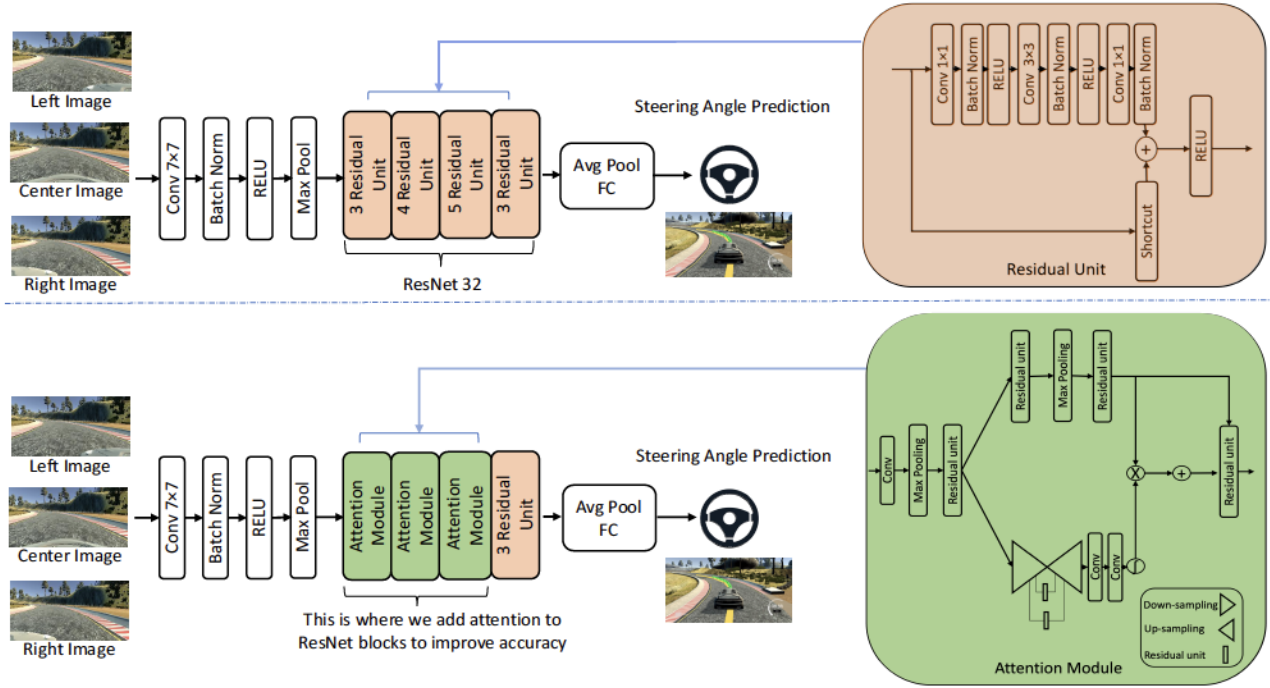


Fig. 2: Comparison of ResNet32 and ResNet32 with attention (two models are of similar sizes). The attention module consists of two basic branches: the trunk branch and the attention mask branch. More details are in Sec. III-B.

B. Attention-aware Deep Steering Angle Prediction

In this section, we draw inspiration from the success of attention mechanisms (e.g., [15]) and propose to integrate attention into the regression task of steering angle prediction. As an example, we take the best-performing ResNet32 as the baseline model for our experiments. Figure 2 compares the basic ResNet32 network and our ResNet32 network with attention modules added, where both networks have the same number of residual units and a similar number of parameters for fair comparison. As shown in Figure 2, the attention module consists of two branches, i.e., the trunk branch and the attention branch. The trunk branch works as a standard feature extractor. The attention branch learns an attention mask of the same dimension as the trunk output. To capture global context and high-level semantic meaning, the input features undergo several layers of down-sampling. Subsequently, these resultant global features are symmetrically up-sampled to their original sizes to generate the attention mask. By computing the dot product between the mask and the output of the trunk branch, attention is focused on regions relevant to the task at hand. Our attention-equipped network comprises three attention blocks designed to progressively enhance attention, each consisting of three stages: region selection, feature amplification, and feature fusion. The region selection operation identifies and chooses valuable regions while disregarding unimportant ones using bottom-up top-down attention, while the feature amplification operation re-weights the feature maps from the trunk by the attention mask generated. The final feature fusion stage

sums the original trunk features with the attention-weighted features to improve their representation.

C. Model Robustness Assessment

The attention module, as explained previously, incorporates trunk and attention mask branches, facilitating an effective attention mechanism that selectively emphasizes essential features in the input data. In addition to enhancing accuracy through attention mechanisms, our research also delves into the models' ability to resist adversarial attacks, such as Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD). We evaluate how adversarial attacks impact our best-performing models, both with and without the attention mechanism. Through these adversarial tests, we aim to assess whether the attention mechanism enhances the model's security and its capacity to cope with adversarial manipulations [17; 19]. Again, to ensure a fair comparison, the networks being compared are configured with the same number of residual units and a similar number of parameters.

IV. EXPERIMENTS AND RESULTS

A. Environmental Setup

Experiments were run on the Ohio Supercomputer Center's Pitzer cluster with NVIDIA Tesla V100 GPUs. The deep learning framework used was PyTorch. Mean squared error (MSE) was used as the training loss:

$$MSE = \frac{1}{n} \sum_{n=1}^n (r_i - \hat{r}_i)^2, \quad (1)$$

where r_i and $\hat{r}_i$ are respectively ground-truth and predicted steering angles (in radians) for sample i , and n is the total number of samples. Models were trained for 50 epochs and the batch size tested for all models varied from 16 to 128 in search of the best outcome. Generally, the batch sizes of 32 were found to have the best results.

B. Quantitative Results

Our first analysis was with a publicly available Kaggle Steering Angle Prediction dataset of approximately 97,330 images [22], which we denote as the Kaggle SAP dataset. A performance comparison among different ResNet and Inception architectures across varying levels of low-end complexities is shown in Fig. 3. As we can see, ResNet32 achieved the lowest error of 0.05293. This is a competitive score, comparable with or surpassing that reported in recent studies [23; 24]. The Inception nets performed worse after 18M parameters and had the lowest error of 0.05846. In our experiments, ResNet models also converged faster than their Inception counterparts.

To demonstrate the effectiveness of our attention mechanism, we first conducted a comparison between the baseline ResNet32 and our attention-enhanced ResNet32 variant, employing the Kaggle SAP dataset. The outcomes depicted in Figure 4a exhibit a significant decrease in validation MSE following our integration of attention modules into the baseline model, occurring around the 70th epoch of training. Furthermore, it is noteworthy that the validation loss of our attention-equipped model consistently decreases throughout the entire training process, whereas we observe some abrupt fluctuations around 60-80 epochs for ResNet32 without attention. To ensure that the results were not just an anomaly of ResNet32 on the particular Kaggle SAP dataset, we evaluated the performance of another architecture (ResNet26) with and without attention modules on a different dataset. Due to the scarcity of steering angle prediction datasets, the new dataset was created by the authors using the Udacity self-driving-car simulator [25], which consists of approximately 10,765 images. We refer to this dataset as the custom dataset and publish it at [26]. Figure 4b displays the results, plotting the validation MSE against the number of training epochs. According to Figure 4b, there is a notable decrease in MSE observed in the attention-enhanced ResNet26 model as the number of epochs progresses. Apart from epoch 70, where both models perform comparably, the inclusion of attention modules is seen to enhance the performance of the ResNet26 base. After 70 epochs, the loss of our model begins to decay smoothly, while we can still see some fluctuations in the MSE score for the ResNet26 without attention modules. With the introduced attention, at epoch 100, we can observe a 6.83% reduction in MSE for the ResNet32 on the Kaggle SAP dataset and a 6.09% reduction for the ResNet26 on the custom dataset. This is intuitively depicted in Figure 4c, indicating that the incorporation of attention modules can clearly improve steering angle prediction.

In addition to improved accuracy, we are also interested in the introduced attention mechanism's influence on adversarial robustness. To this end, we employed two well-established adversarial attack techniques: the Fast Gradient Sign Method (FGSM) [17] and Projected Gradient Descent (PGD) [19]. While numerous studies address adversarial challenges in autonomous driving [27], our research marks the first attempt to evaluate the adversarial robustness of attention-equipped deep models specifically for steering angle prediction. The results of our adversarial attack experiments, including the MSE scores of ResNet32 and ResNet26 with and without integrated attention modules under FGSM and PGD attacks, are presented in Table III and Table IV. Epsilon (eps) values control the magnitude of adversarial perturbations. While the MSE loss does increase under attacks, our findings suggest that the inclusion of attention modules notably mitigates the extent of this increase. As we can see, attention boosts the model robustness by up to 53.95% on ResNet32 (Table III) and as high as 56.09% on ResNet26 (Table IV).

Model	FGSM		PGD	
	eps=0.01	eps=3	eps=0.01	eps=3
w/o attention	0.214	0.336	4.763	5.853
w/ attention	0.200	0.291	2.581	2.695
change	6.54%	13.39%	45.81%	53.95%

TABLE III: Adversarial attack results on ResNet32 with and without attention. The value reported for each case is the MSE error of steering angle prediction.

Model	FGSM		PGD	
	eps=0.5	eps=0.7	eps=0.5	eps=0.7
w/o attention	0.253	0.574	9.464	9.636
w/ attention	0.183	0.252	5.558	5.616
change	27.66%	56.09%	41.27%	41.71%

TABLE IV: Adversarial attack results on ResNet26 with and without attention. The value reported for each case is the MSE error of steering angle prediction.

C. Qualitative Results

To better understand the influence of our introduced attention modules, we investigated the changes in saliency maps. The results (Fig. 5) provide insight into where the model is focused on in an image. According to the results, we can see that the attention modules assist ResNet in directing its focus toward the road ahead, prioritizing it over irrelevant areas. Sometimes (e.g., Row 2 and Row 4 in Fig. 5), without our attention mechanism, the original ResNet model focuses only on one side of the road, and with the attention mechanism added, the model can pay attention to the center and the other side of the road to a greater extent. Fig. 5 qualitatively explains our approach's efficacy. In the future, we will explore transformer models and their attention for steering angle prediction. Nevertheless, network compression techniques (e.g., [28]) are needed for edge deployment.

MSE of series of ResNet & Inception architectures on Kaggle SAP

ResNet Model	MSE	InceptionNet Model	MSE
ResNet20	0.06557	InceptionNet	0.06044
ResNet22	0.06371	InceptionNet a	0.05945
ResNet24	0.06721	InceptionNet b	0.05849
ResNet26	0.05832	InceptionNet c	0.05846
ResNet28	0.05552	InceptionNet d	0.05961
ResNet30	0.05641	InceptionNet e	0.05802
ResNet32	0.05293	InceptionNet f	0.05907
ResNet34	0.05425		

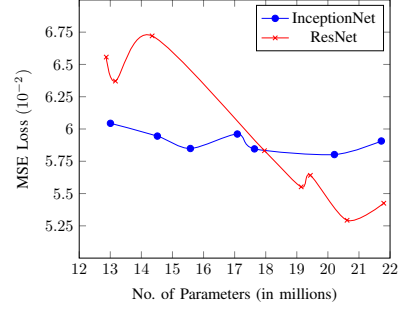
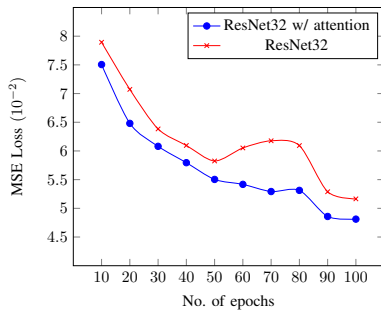
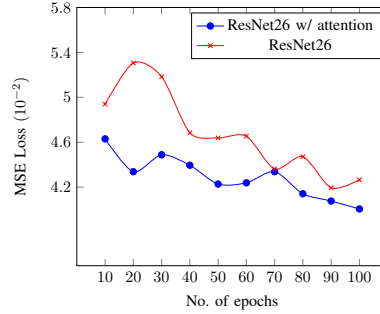


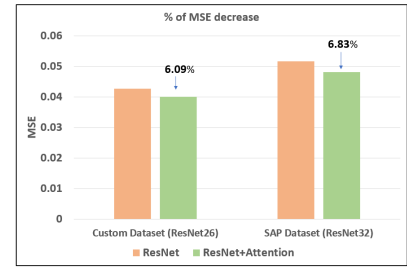
Fig. 3: ResNets vs. InceptionNets in terms of MSE loss on the Kaggle SAP dataset. The figure on the right plots the MSE losses of the tested models against their parameter complexity measures.



(a) ResNet32 on Kaggle SAP



(b) ResNet26 on Custom Data



(c) MSE improvement w/ attention

Fig. 4: Validation MSE of base ResNets vs our ResNets with attention. We utilize two base models (ResNet32 and ResNet26) on two datasets (Kaggle SAP and custom data). (c) illustrates the percentage reduction in MSE at the 100th epoch.

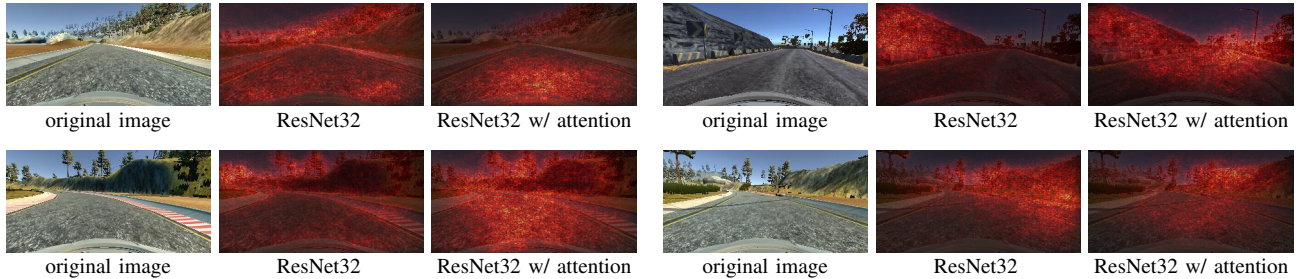


Fig. 5: Saliency maps representing the focus areas of the model with and without our introduced attention modules. Black signifies no focus, red indicates some focus, and yellow suggests the highest level of focus. Here, a saliency map is generated by taking the gradient of the model output with respect to the original input image. For ResNet32 and ResNet32 with attention, the saliency maps are blended with the original image with a blending ratio of 0.75.

V. CONCLUSION

Targeting the task of steering angle prediction, this paper first examines the two main veins of CNN architectures and their intra-variations of various complexity (i.e., InceptionNets and ResNets). In addition, we have introduced attention modules to deep steering angle predictors, which has increased the models' performance by decreasing the MSE up to 6.83% on the Kaggle SAP dataset and up to 6.09% on the custom dataset. The incorporation of our attention modules also demonstrates a promising potential to improve the adversarial robustness of steering angle predictors. For example, it can enhance the adversarial robustness

of ResNet32 by up to 53.96% against the PGD attack.

ACKNOWLEDGMENT

This research was enabled in part by support provided by the National Science Foundation (under Award No. 2153404 and 2412285) and the Ohio Supercomputer Center (<https://www.osc.edu/>). It draws upon some undergraduate and master's projects conducted at Bowling Green State University (BGSU) under the supervision of the PI (last author) when he was with BGSU.

REFERENCES

- [1] P. J. Navarro, L. Miller, F. Rosique, C. Fernández-Isla, and A. Gila-Navarro, "End-to-end deep neural network architectures for speed and steering wheel angle prediction in autonomous driving," *Electronics*, vol. 10, no. 11, p. 1266, 2021.
- [2] J. Klender, "Tesla tells new fsd beta member their cars will use pure vision, axing radar altogether," <https://www.teslarati.com/tesla-fsd-beta-members-pure-vision-no-radar-beta-pool-expansion/>, October 2021, last accessed 27-July-2022.
- [3] U. M. Gidado, H. Chiroma, N. Aljojo, S. Abubakar, S. I. Popoola, and M. A. Al-Garadi, "A survey on deep learning for steering angle prediction in autonomous vehicles," *IEEE Access*, vol. 8, pp. 163 797–163 817, 2020.
- [4] A. Oussama and T. Mohamed, "A literature review of steering angle prediction algorithms for self-driving cars," in *Int'l Conference on Advanced Intelligent Systems for Sustainable Development*, 2019, pp. 30–38.
- [5] M. K. Islam, M. N. Yeasmin, C. Kaushal, M. Al Amin, M. R. Islam, and M. I. H. Showrov, "Comparative analysis of steering angle prediction for automated object using deep neural network," in *Int'l Conf. on Reliability, Infocom Technologies and Optimization*, 2021, pp. 1–7.
- [6] F. Munir, S. Azam, B.-G. Lee, and M. Jeon, "Multi-modal fusion for sensorimotor coordination in steering angle prediction," *arXiv preprint arXiv:2202.05500*, 2022.
- [7] S. Du, H. Guo, and A. Simpson, "Self-driving car steering angle prediction based on image recognition," *arXiv preprint arXiv:1912.05440*, 2019.
- [8] M. Al-Qizwini, I. Barjasteh, H. Al-Qassab, and H. Radha, "Deep learning algorithm for autonomous driving using googlenet," in *IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 89–96.
- [9] D. McNeely-White, J. R. Beveridge, and B. A. Draper, "Inception and resnet features are (almost) equivalent," *Cognitive Systems Research*, vol. 59, pp. 312–318, 2020.
- [10] S. He, D. Kangin, Y. Mi, and N. Pugeault, "Aggregated sparse attention for steering angle prediction," in *24th IEEE International Conference on Pattern Recognition*, 2018, pp. 2398–2403.
- [11] K. Zhu, W. Chen, W. Zhang, R. Song, and Y. Li, "Autonomous robot navigation based on multi-camera perception," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 5879–5885.
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [13] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3640–3649.
- [14] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [15] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6450–6458.
- [16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [17] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.
- [18] Y. Deng, X. Zheng, T. Zhang, C. Chen, G. Lou, and M. Kim, "An analysis of adversarial attacks and defenses on autonomous driving models," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2020, pp. 1–10.
- [19] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [22] Kaggle, "Self-driving car simulator," <https://www.kaggle.com/datasets/zaynena/selfdriving-car-simulator>, 2019, last accessed 20220724.
- [23] N. Ijaz and Y. Wang, "Automatic steering angle and direction prediction for autonomous driving using deep learning," in *IEEE Int'l Symposium on Computer Science and Intelligent Controls*, 2021, pp. 280–283.
- [24] C. Oinar and E. Kim, "Self-driving car steering angle prediction: Let transformer be a car again," *arXiv preprint arXiv:2204.12748*, 2022.
- [25] <https://github.com/udacity/self-driving-car-sim>, 2016.
- [26] P. Barua, "Attention-aware automatic steering angle prediction," <https://github.com/PramitiBarua/Attention-based-Steering-Angle-Prediction>, 2023.
- [27] J. Im Choi and Q. Tian, "Adversarial attack and defense of yolo detectors in autonomous driving scenarios," in *IEEE Intelligent Vehicles Symposium (IV)*, 2022, pp. 1011–1017.
- [28] Q. Tian, T. Arbel, and J. J. Clark, "Grow-push-prune: Aligning deep discriminants for effective structural network compression," *Computer Vision and Image Understanding*, vol. 231, p. 103682, 2023.