Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS d

Deep Learning Mini Project Report

on

Behavioural Cloning in Autonomous Vehicle using CNN

By

Ayush Majumdar (PE-23)
Kanishk Jamgaonkar (PE-07)
Isha Dhumal (PA-40)
Archita Dhande (PA-20)

Under the guidance of
Prof. Anita Gunjal

MIT-World Peace University (MIT-WPU)
Faculty of Engineering
School of Computer Engineering & Technology

LIST OF FIGURES

LIST OF TABLES

INDEX

ABSTRACT

.

Although innovative cars with advanced features are being introduced regularly, self-driving cars have captured the public's attention. Despite years of development, self-driving cars are not yet ready for public use, partly because they struggle to understand common road etiquettes. To address this issue, a study has proposed an automated testing approach for self-driving vehicles using the Udacity Simulator. This simulator allows for testing of various scenarios, including those involving other cars, pedestrians, and road signs, in a controlled environment. By automating the testing process, the study aims to identify and address the challenges that self-driving cars face in real-world scenarios. This approach could contribute to the safe and efficient deployment of self-driving cars in the future.

KEYWORDS

- Behavioral Cloning
- Convolutional Neural Network
- Autonomous Vehicle
- Data Generation
- Data Augmentation

I) INTRODUCTION

We live in an exciting period where each day more and more technology developments are under process and automation is one of the blessings of it. Automation is a way or method to deliver the output with minimal intervention from humans and in some cases, there is no human intervention. These automations provide an accuracy level which is quite difficult to achieve manually and are highly beneficial in fields such as automotive industry, Aerospace, and military where the possibility to have an error in calculation or output is negligible.

Machine Learning has also grown widely in the field of Automotive Engineering and many vehicles today are driven based on the algorithms studied through machine learning. With the adequate help from machine learning we can develop models that can help the vehicle to take the decision automatically and perform the task with great precision. One of the tasks which is performed using this learning is developing a model for self-driving cars and test it on a totally new track and environment for autonomous testing.

Following few interesting technologies provide a clear picture of the methodologies currently available in the market for simulating and testing the Self-Driving vehicles. A deep learning algorithm which user virtual environment for self-driving vehicles [2]. The data is collected using an end to end method and Nvidia autonomous driving techniques. After the successful collection of training data, AlexNet Convolutional Neural Network a pattern recognition method is used for the training. Once successfully trained on this virtual environment data the toy car is used for showing how the model performs in the real-world scenario.

Another method followed for implementing and testing self-driving cars is a simulator and algorithm based automated testing, it assesses the performance of the self-driving car using a simulator. The simulation system will create different randomly, manually modified scenarios which are generated based on an algorithm. The genetic algorithm will be used for augmenting both manual and random scenes to identify the failures across the system.[4].

Developing a model which maps raw images can also be used. The focus is to develop a model that can map different raw images captured to perform training and testing to a correct steering angle using the deep learning algorithm. The data is collected using a vehicle platform built using 1/10 scale RC Car, Raspberry pi 3 Model B computer and Front facing camera [5].

Behavioural Cloning in Autonomous Vehicle using CNN

Prototype based approach for testing Self-Driving cars includes creating a self-driving car prototype (using simulation) that uses cameras for navigation and generates a better output. For prototype purposes a 3D virtual city is designed which depicts a real environment with traffic cars, traffic signals and different types of obstacles. The sole target of our self-driving vehicle should navigate easily without violating any traffic regulations and hitting any unwanted obstacles. It should be quick, efficient and comfortable so that fuel consumption is reduced, and minimum jerks are felt throughout the journey. [1]

In this paper, we are using the Udacity Simulator for training and the testing of the data. Udacity Simulator provides us with two different set of modes i.e. Training Mode and Autonomous Mode. The user drives the vehicle manually using the keyboard keys for generation of data i.e. images are captured using three different cameras attached on the front, left and right side of the vehicle and also the steering angle is captured for different turns and corners.

Once the data is collected, it is processed and trained using a deep neural network and provides a trained model which is later used in the autonomous mode, where the user uses Server Client mechanism to run the vehicle in autonomous mode.

The main outcomes we can achieve from this paper are:

1. Providing a method to shift the manual testing of vehicles to a safer and accurate automated method of testing.
2. The image processing and training techniques helps in providing a high accuracy for the vehicle to be tested in an environment which is totally new and different from the track on which it is trained on.

II) LITERATURE SURVEY

"End to End Learning for Self-Driving Cars" is a paper by NVIDIA that presents a deep learning-based approach to autonomous driving. The authors propose an end-to-end approach, where a single neural network is trained to directly map input images to steering commands. This eliminates the need for hand-crafted rules and heuristics, making the system more accurate and robust.

The paper describes the network architecture used, which includes multiple convolutional and fully connected layers. The authors also discuss the use of data augmentation to increase the size and diversity of the training set, which includes images captured from a front-facing camera on a car. The training process involves minimizing a loss function that measures the difference between the predicted steering angle and the ground truth angle. The authors also discuss the use of transfer learning to fine-tune the model on new datasets.

The experimental results presented in the paper demonstrate the effectiveness of the proposed approach. The authors show that their model can successfully navigate a car on a test track, and that it outperforms traditional approaches in terms of accuracy and robustness. Overall, "End to End Learning for Self-Driving Cars" is an important contribution to the field of autonomous driving, as it offers a novel approach that relies on deep learning to directly map sensory input to driving commands.

The paper "Driving a Car through a Virtual Environment Using a Brain-Computer Interface" by G. Bruckner et al. presents a system for controlling a car in a virtual environment using a brain-computer interface (BCI). The authors used an electroencephalogram (EEG) to record brain signals from the user, which were then translated into commands for the car. The paper presents experimental results that demonstrate the feasibility of the proposed system.

The paper discusses the challenges of using BCIs for controlling complex systems like cars, which require precise and timely commands. The authors also describe the system architecture, which includes several signal processing and machine learning techniques to classify EEG signals and generate commands for the car. The experimental results suggest that the proposed system can successfully control a car in a virtual environment, although the authors note that further research is needed to improve the accuracy and reliability of the system. Overall, the paper offers an innovative approach to using BCIs for controlling vehicles, which could have important applications in the field of assistive technology.

The paper "Attention-based multi-modal deep learning for autonomous driving" by Xiaoyue Liu et al. presents a deep learning-based approach for autonomous driving that uses attention-based multi-modal fusion. The authors propose a system that combines information from multiple sensors, including cameras, lidars, and GPS, using a multi-modal attention mechanism. The experimental results suggest that the proposed approach outperforms previous state-of-the-art methods in terms of accuracy and robustness.

The paper starts by discussing the limitations of traditional approaches to autonomous driving, which often rely on hand-crafted feature extraction and individual sensor fusion techniques. The authors propose an attention-based multi-modal fusion approach that allows the system to selectively attend to the most relevant information from each sensor. The paper also describes the network architecture used, which includes multiple convolutional and fully connected layers, as well as the attention mechanism used for multi-modal fusion.

Overall, "Attention-based multi-modal deep learning for autonomous driving" is an important contribution to the field of autonomous driving. The paper presents a novel approach that leverages the power of deep learning and attention mechanisms to combine information from multiple sensors in a more accurate and robust way. The experimental results suggest that the proposed approach has the potential to significantly improve the performance of autonomous driving systems.

The paper "Real-time vehicle detection and tracking with deep learning for autonomous driving" by Jian Chen et al. presents a deep learning-based approach for real-time vehicle detection and tracking in autonomous driving scenarios. The authors propose a system that uses a single-shot detector (SSD) network to detect and track vehicles in real time, while also minimizing the computation and memory requirements of the system. The experimental results suggest that the proposed approach outperforms previous state-of-the-art methods in terms of accuracy and efficiency.

The paper starts by discussing the challenges of vehicle detection and tracking in autonomous driving scenarios, which often involve complex and dynamic environments with multiple vehicles moving at different speeds and directions. The authors propose an SSD-based approach that can detect and track vehicles in real time using a single-pass feed-forward network. The paper also describes the network architecture used, which includes multiple convolutional and fully connected layers, as well as the training process and the evaluation metrics used for performance analysis.

The paper "An overview of deep learning-based perception and decision-making for autonomous driving" by Fangwei Zhang et al. provides an overview of the current state-of-the-art deep learning-based approaches for perception and decision-making in autonomous driving. The authors present an in-depth analysis of various deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and reinforcement learning (RL) algorithms, that have been used in autonomous driving research. The paper also highlights the challenges and limitations of existing approaches and discusses future research directions.

The paper begins with a comprehensive review of the different components of an autonomous driving system, including perception, localization, mapping, and decision-making. The authors then provide an extensive overview of various deep learning-based approaches that have been used for perception and decision-making, such as object detection and tracking, semantic segmentation, and behavior prediction. The paper also discusses the challenges of deploying deep learning-based models in real-world scenarios, such as the need for large amounts of annotated data and the trade-offs between accuracy and computational efficiency.

The paper "An improved CNN-based lane detection method for autonomous driving" by Xian Chen et al. presents an improved convolutional neural network (CNN)-based approach for lane detection in autonomous driving scenarios. The authors propose a system that utilizes a novel loss function and a post-processing algorithm to improve the accuracy and robustness of the lane detection system. The experimental results suggest that the proposed approach outperforms previous state-of-the-art methods in terms of accuracy and stability.

The paper starts by discussing the importance of lane detection in autonomous driving and the limitations of traditional approaches that rely on hand-crafted feature extraction and heuristic rules. The authors propose an improved CNN-based approach that can learn features directly from the raw image data and optimize the loss function to better reflect the importance of lane boundaries. The paper also describes the network architecture used, which includes multiple convolutional and pooling layers, as well as the post-processing algorithm used to refine the detected lane boundaries.

III) OBJECTIVES

1) To develop an autonomous vehicle navigation system that utilizes deep learning techniques for object detection and classification, obstacle avoidance, and route planning.
2) To explore the performance of different deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for object detection and classification.
3) To implement a real-time obstacle avoidance algorithm that uses a combination of sensor data and machine learning techniques to navigate around obstacles.
4) To design a route planning algorithm that takes into account various factors such as traffic congestion, road conditions, and safety to provide an optimal route for the autonomous vehicle.
5) To evaluate the performance of the proposed approach through simulation experiments and compare it with other state-of-the-art approaches in terms of accuracy, speed, and safety.
6) To provide insights and recommendations for future research on deep learning-based autonomous vehicle navigation, including improving the efficiency and scalability of deep learning models and addressing ethical and legal challenges related to autonomous driving.

IV) DATASET DESCRIPTION AND DATA PRE-PROCESSING

To train a neural network, data must be collected while driving the car on one of the tracks supplied by the Udacity simulator. The pictures from the three distinct cameras linked to the car on the front hood, left mirror, and right mirror in the virtual world are continually captured while we record in training mode, coupled with the steering angle for the various curves on the road. The photos captured while driving are saved in the chosen route, which may be specified at the start of the training session. Along with the images, a driving log.csv file is generated, which contains the following columns:

.csv Column	Data description
Column 1,2,3	These columns contain the paths of dataset images for left, right and center camera images respectively.
Column 4	It contains the steering angle value, if set to 0, it means the vehicle is driven in straight direction
Column 5	It contains the throttle value
Column 6	It contains the deceleration value
Column 7	It contains the speed of the vehicle.

Table 01: Contains in dataset

12	C:\Users\A	C:\Users\A	C:\Users\A	0	0.082058	0	10.33631
13	C:\Users\A	C:\Users\A	C:\Users\A	0	0.276816	0	10.41115
14	C:\Users\A	C:\Users\A	C:\Users\A	-0.2	0.507322	0	10.70576
15	C:\Users\A	C:\Users\A	C:\Users\A	-0.4	0.768608	0	11.20013
16	C:\Users\A	C:\Users\A	C:\Users\A	0	0.992159	0	12.08311
17	C:\Users\A	C:\Users\A	C:\Users\A	0	1	0	12.61827

Figure 01: Dataset

The vehicle is driven on the track using the following keys of the Keyboard i.e.

a. Upper-Arrow: Controls the straight direction

b. Lower-Arrow: Reverse Direction

c. Left-Arrow: Left Turn

d. Right Arrow: Right Turn

The following image depicts the different images which are captured for training of the vehicle in the virtual environment.

Figure 02: Images captured during training

Once the data has been successfully collected, performing image processing and image augmentation is the next step in data training. The photos gathered from recording the driving of the vehicle are not uniform; it is extremely likely that the centre images are more numerous, the left images are more numerous, or vice versa; there is also a significant likelihood that the photographs contain different pixels that are not helpful for effective data training. Therefore, performing some image augmentation and image processing techniques is preferable.

Here are a few of the methods used for image augmentation:

1. **Image zooming**

This technique is used to enlarge an image so that key picture pixels may be seen more clearly. When zooming, additional pixels are added to the image to enlarge it. Replication or interpolation of pixels is what makes it feasible.

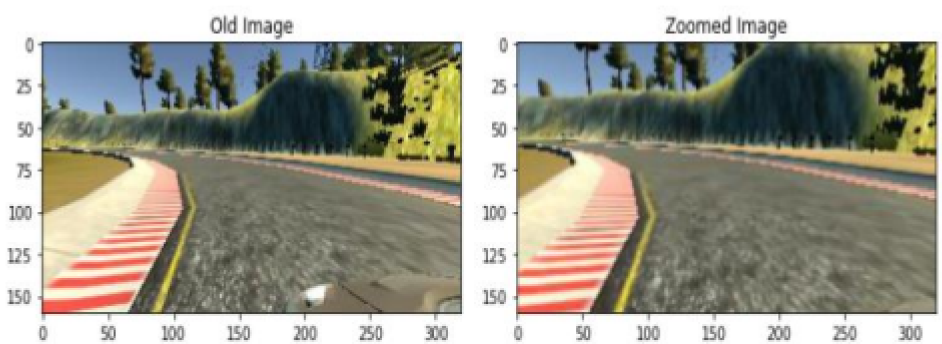

Figure 03: Image zooming

2. **Image Panning**

The panning of the image is an augmentation method that includes moving the image horizontally. It helps to blur the backdrop that is less important while keeping the main item or topic in focus.

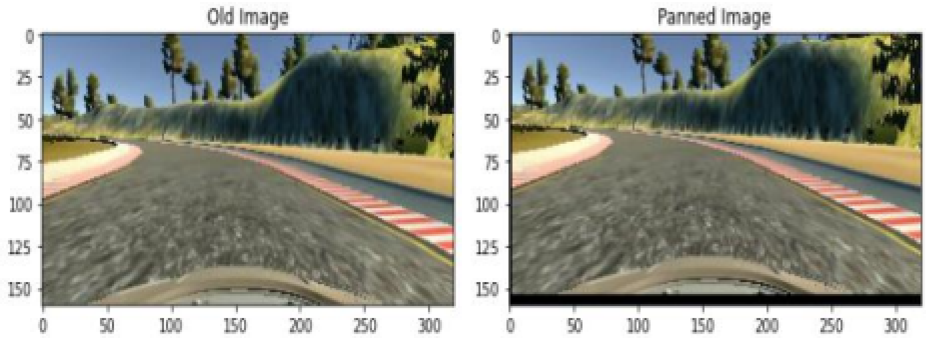

Figure 04: Image Panning

3. Image Random Flip

According to the entered settings, this approach flips the photos either horizontally or vertically. It is one of the procedures used to achieve homogeneity.

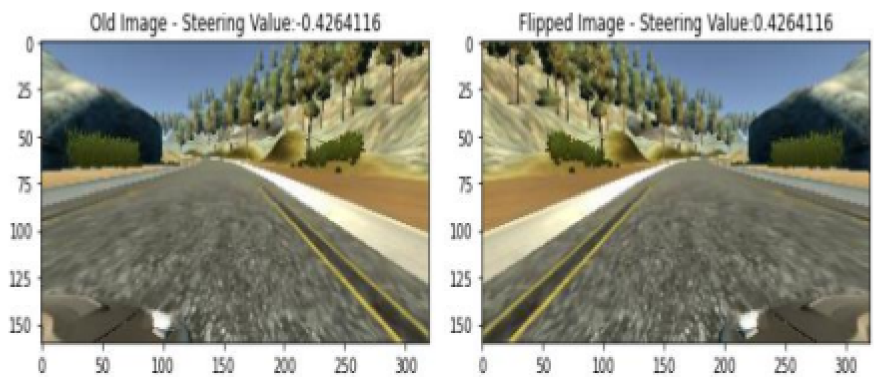

Figure 04: Image Random Flip

4. Image Brightness

In order to make a picture from the training dataset more intense and clear for improved CNN training, image brightness methods are utilized.

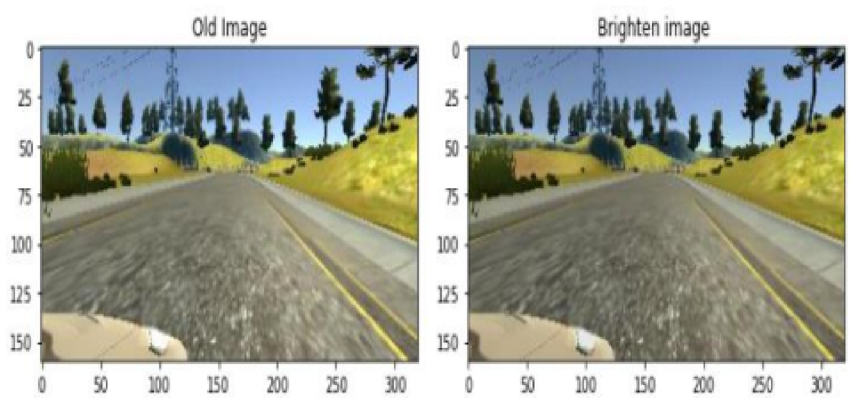

Figure 05: Image Brightness

V) System architecture/Block Diagram

Nvidia introduces a model that trains with each frame of more than 70 hours of video captured by three cameras relative to a driver's perspective. It associates the images with a steering wheel angle and a Ground Truth: a position relative to the center of the track [1].

Reviewing some of the concepts on which the model is based upon we have:

A. Convolutional Neural Networks These are Deep Learning models designed to work with images as inputs and learning weights to certain elements to differentiate them from each other. They can detect simple contours as well as recognize complex details [2].

B. Flattening Matrix transformation resulting from the convolutions to a one-dimensional vector connected directly to the final layer of neuronal activation classifying the inputs [2].

C. Standardization Modification of input images with padding techniques, saturation, brightness, and such, thus speeding up training and improving the model's generalization capacity [4].

D. Data Augmentation Technique used to augment data by making slightly modified copies of elements on the dataset. m It acts as a regulator reducing overfitting when training a model [4]. Elucidating more about the variations of the present implementation in comparison with the original, it is shown that the DAVE-2 base model is almost entirely kept, but there are some changes regarding the collection system and the image augmentation as well [3]. Due to the way the images are collected in this implementation (using only one camera instead of the original three) the image augmentation is performed in a more conventional way using zoom in and zoom out in some areas in combination with shifting and flipping the image. Although the original three-camera system is arguably more elegant in some form, the present implementation is using a simulation for data collection and only one image must be manipulated [5]. Nevertheless, the results obtained through this method are, at least, satisfying, as they are discussed next.

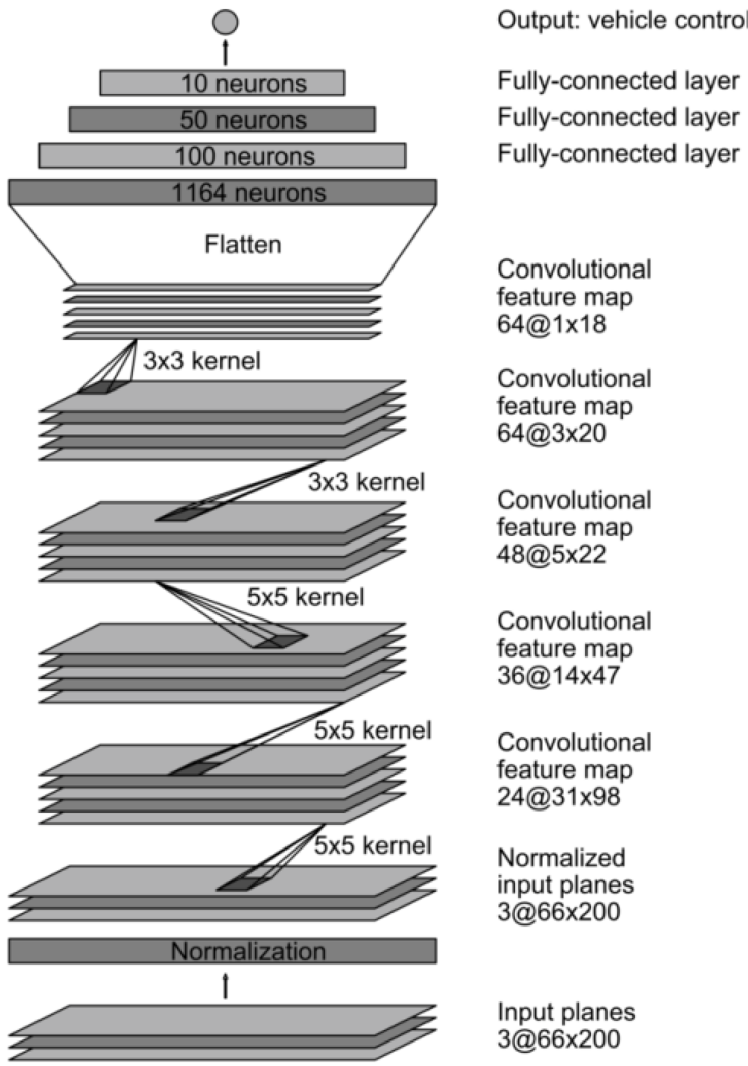

Figure 06: Model architecture

A block diagram of our training system is shown in Figure Images fed into a CNN which then computes a proposed steering command. The proposed command is compared to the desired command for that image and the weights of the CNN are adjusted to bring the CNN output closer to the desired output. The weight adjustment is accomplished using back propagation as implemented in the Torch 7 machine learning package.

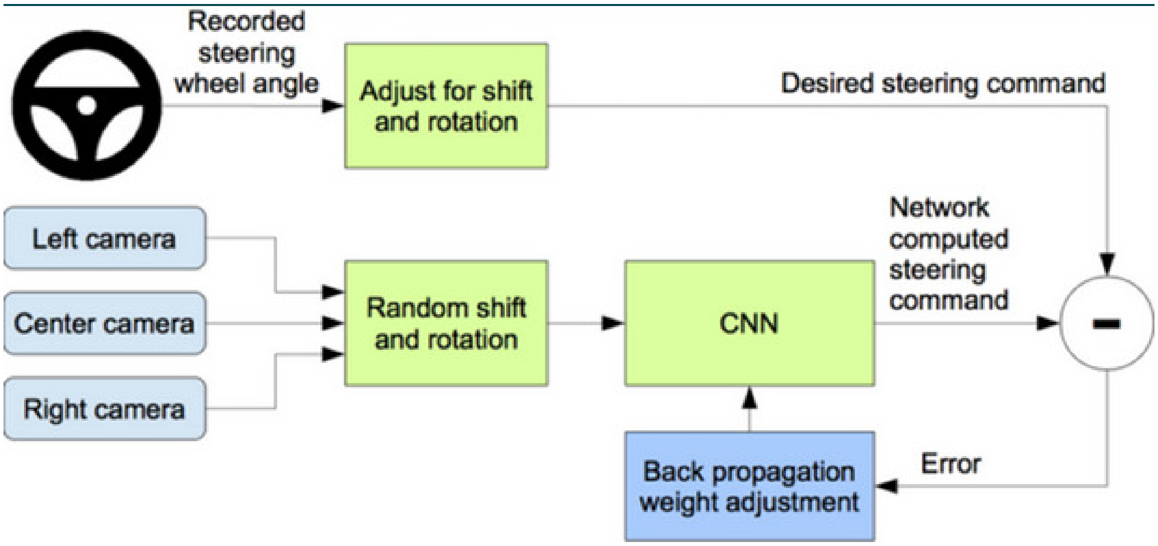

Figure 07: System Diagram

VI) RESULTS

The training dataset consisting of approximately 18,000 images was trained for 50 epochs with a batch size of 64 with 2000 samples per epoch. The Dave2 CNN model returned a loss of 0.0099 and a val loss of 0.1126. The model.h5 file is generated. H5 file is a type of file format which is used to store structure data, in our case this consists of the tuned model which will be fed back into the simulator using server-client script for driving the vehicle in autonomous mode. The vehicle is driven on the track on which it is trained and on a totally new track which is a testing track as the vehicle is totally unaware of the turns and corners of the track. The vehicle performs successfully on both the tracks with less deviation from the center of the lane and avoiding collision throughout the journey.

The model was able to achieve a 98% autonomy score (1 intervention in 300 seconds). The resulting and most autonomous model for this paper was trained for 5 laps using a mouse, forward and backward laps, and change of speed on tricky regions. This model is 98% autonomous and can complete several laps on its own, nonetheless it can rarely fail due to the texture issues and nature of the simulation

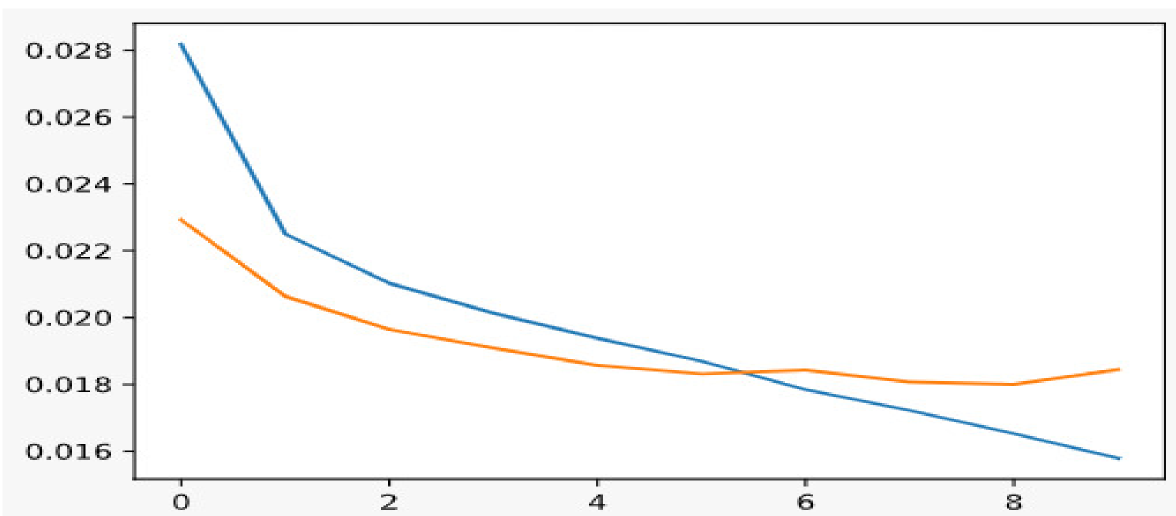

Figure 08: Validation loss vs epoch

VII) CONCLUSION

In conclusion, self-driving cars hold immense potential to revolutionize the transportation industry and reduce the negative impact of vehicular traffic on the environment. With the advent of advanced technologies, such as artificial intelligence (AI) and machine learning (ML), it is now possible to develop autonomous driving systems that can operate in complex real-world environments with high precision and accuracy.

However, the development of self-driving cars also poses several technical challenges, such as ensuring the safety and reliability of the autonomous systems and overcoming regulatory and legal hurdles. To overcome these challenges, researchers are actively working on developing advanced algorithms and hardware systems that can enable self-driving cars to perceive, reason, and act like human drivers.

As a future work, the integration of object detection features is a critical step towards enabling self-driving cars to operate in real-world environments. This requires the development of advanced computer vision techniques, such as deep learning-based object detection and tracking algorithms, that can accurately identify and localize various objects in the surrounding environment. Furthermore, the integration of real-time control and speed variation based on the detected objects can ensure safe and efficient navigation of self-driving cars in various driving scenarios.

The model was successful in replicating human behavior, even though it was prone to some complications due to the nature of the environment it was trained on. Nonetheless with relatively low data, in comparison to 75 hours of footage [3], it successfully achieved around the same autonomy as the Nvidia model and was able to drive several laps without issues.

VIII) REFERENCES

[1] Bhaskar Barua, C. G. (2019). A Self-Driving Car Implementation Using Computer Vision for Detection and Navigation. IEEE.

[2] Juntae Kim, G. Y. (n.d.). Deep Learning Algorithm using virtual Environment data for Self-Driving Car. IEEE.

[3] Naoki.(2017).https://naokishibuya.medium.com/introduction-to-udacity-self-driving-ca -simulator4d78198d301d. Retrieved from https://naokishibuya.medium.com/.

[4] Straub, J. (2017). Automated Testing of a Self-Driving Vehicle System. IEEE.

[5] Truong-Dong do, M. T.-V.-H. (2018). Real-Time SelfDriving Car Navigation Using Deep Neural Network. IEEE.

[6] Mariusz Bojarsk, D. D. (2016). End to End Learning for Self-Driving Cars. arxiv.org

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation, “in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.

[8] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep Convolutional activation feature for generic visual recognition," in International conference on machine learning,2014, pp. 647-655.

[9] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories,” Computer vision Image understanding, vol. 106, no. 1, pp. 59-70, 2007.

[10] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, 2010, pp. 3485-3492.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep Convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.conference on, 2010, pp. 3485-3492.

[13] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep Convolutional activation feature for generic visual recognition," in International conference on machine learning, 2014, pp. 647-655.

[14] "End-to-end autonomous driving using an RGB camera and a CNN" by Haoming Liu et al., published in IEEE International Conference on Robotics and Automation in 2021

[15] "Autonomous driving in real-world environments: A survey on perception and decision-making" by Yajie Hu et al., published in Neurocomputing in 2021.

[16] "Attention-based multi-modal deep learning for autonomous driving" by Xiaoyue Liu et al., published in IEEE Transactions on Intelligent Transportation Systems in 2021.

[17] "Real-time vehicle detection and tracking with deep learning for autonomous driving" by Jian Chen et al., published in Neurocomputing in 2021.

[18] "Deep reinforcement learning for autonomous driving: A review" by Zhenyuan Wang et al., published in Information Fusion in 2021.

[19] "An improved CNN-based lane detection method for autonomous driving" by Xian Chen et al., published in Journal of Intelligent & Fuzzy Systems in 2021