Autonomous Vehicle Simulation in GTA-5

Project report submitted in partial fulfillment
of the requirements for the degree of

Bachelor of Technology
in
Computer Science and Engineering

by

Tushar Jain - 20UCS211

Under Guidance of
Mr. Vikas Bajpai

LNMIIT
The LNM Institute of
Information Technology

Department of Computer Science and Engineering
The LNM Institute of Information Technology, Jaipur

April 2023

The LNM Institute of Information Technology
Jaipur, India

CERTIFICATE

This is to certify that the project entitled Autonomous Vehicle Simulation in GTA-5, submitted by Tushar Jain (20UCS211) in partial fulfillment of the requirement of degree in Bachelor of Technology (B. Tech), is a bonafide record of work carried out by them at the Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, (Rajasthan) India, during the academic session 2022-2023 under my supervision and guidance and the same has not been submitted elsewhere for award of any other degree. In my/our opinion, this report is of standard required for the award of the degree of Bachelor of Technology (B. Tech).

Date

Adviser: Mr. Vikas Bajpai

Acknowledgments

We would like to express our gratitude to our instructor Mr. Vikas Bajpai, who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system. Without his constant support, motivation and guidance this project would not have been successful.

Abstract

300-500 word abstract to describe your project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Area of Work

Autonomous Vehicle are cars and trucks that can understand and navigate the environment with little to no help from a human driver. While consumers cannot yet buy a car that fully drives itself (Level 4 or 5 automation)[1], the technology is under development and improving quickly. The self-driving cars being worked on today have many different sensors and cameras to capture information about the world. So, a big challenge in the field of Computer Science and Engineering is to make a Artificial Intelligent network which takes the captured information and gives the best possible action a car should in that state of environment. My project aims to try to build a network capable of taking some of the actions and to able to drive a car efficiently.

1.2 Problem Addressed

An autonomous vehicle is made with the goal to make travel of an individual safer than non autonoumus ones. The environment being partially observable and undeterministic, an autonomous vehicle should be prepared to take action based on any outcome. So an artificial intelligent model which can take decision in any type of environment is desired. Thus we need a model that can learn and adapt on its own in any type of scenarios. This is why a neural network based model is ideal for the use case and forms the heart of any autonomous vehicle. Since driving a vehicle in not efficient, we need a simulator which can create a close-to real world environment. Grand Theft Auto - V fits this criteria perfectly. Thus the main objective is to design, train, and simulate a autonomous vehicle in GTA-V and try to make the model as efficient as possible.

1.3 Existing System

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

1.3.1 System 1

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 2

Literature Review

2.1 Sensors in Autonomous Vehicles

With the help of sensors, autonomous vehicles ensure no human interaction is needed while driving. A wide range of sensors are used in autonomous vehicles to build reliable vision. The sensors help the self-driving vehicle to detect hurdles or blockages in the driving environment and to move without causing fatalities.
Below are some primary sensors used in autonomous vehicles:

2.1.1 Camera

Cameras used in autonomous cars are specialized image sensors that detect the visible light spectrum reflected from objects. Cameras are the best sensor solution to give an accurate visual representation of an autonomous vehicle's surroundings. In autonomous vehicles, cameras are fixed on all four sides—front, rear, right, and left—to give a 360° view. These cameras use wide and narrow fields of view to perceive both short-range wide view and long-range arrow view. Super-wide lenses are used in autonomous vehicles for capturing a panoramic view that assists with parking. However, accurate camera visuals fail to give information regarding the distance of objects from autonomous vehicles.

2.1.2 LiDAR

LiDAR uses laser beams (light waves) to determine the distance between two objects. In autonomous vehicles, LiDAR is mounted on top of vehicles and is rotated at high speed while emitting laser beams. The laser beams reflect from the obstacles and travel back to the device. The time taken for this to happen is used to determine the distance, shape, and depth of the obstacles surrounding the autonomous vehicle.
Even though LiDAR can catch the position, shape, size, and depth of an obstacle, they can

get glitched by fake echoes showing far objects as near objects and vice versa. LiDAR fails to distinguish between multiple copies of laser signals and shows non-existent obstacles to autonomous vehicles. LiDAR does not function well in rain, snow, or fog.

2.1.3 RADAR

The principle of operation for LiDAR and RADAR are the same, but instead of the light waves used in LIDAR, RADAR relies on radio waves. The time taken by the radio waves to return from the obstacles to the device is used for calculating the distance, angle, and velocity of the obstacle in the surroundings of the autonomous vehicle.
RADAR in autonomous vehicles operates at the frequencies of 24, 74, 77, and 79 GHz, corresponding to short-range radars (SRR), medium-range radars (MRR), and long-range radars (LRR), respectively. They each have slightly different functions:

- SRR technology enables blind-spot monitoring, lane-keeping assistance, and parking assistance in autonomous vehicles.
- MRR sensors are used when obstacle detection is in the range of 100-150 meters with a beam angle varying between 30° to 160°.
- The automatic distance control and brake assistance are supported by LRR radar sensors.

RADAR technology in autonomous vehicles operates with millimeter waves and offers millimeter precision. The utilization of millimeter waves in autonomous vehicular RADAR ensures high resolution in obstacle detection and centimeter accuracy in position and movement determination. Compared to other sensor technologies in autonomous vehicles, RADAR works reliably under low visibility conditions such as cloudy weather, snow, rain, and fog.

2.1.4 Microphone

A microphone is a transducer that converts sound into an electrical signal. Microphones are used to give hearing abilities to autonomous vehicles. In autonomous vehicles, microphones can be used to detect horns, emergency sirens, etc.
Microphones in autonomous vehicles are also used to identify the condition of road, the vehicle is driving on. It is shown that different types of roads (like wet, sandy, snowy) have different spectrum hence plays an important role in the identification.

2.2 Navigation Systems in Autonomous Vehicles

Autonomous vehicles can use navigation systems to geolocate with numerical coordinates (e.g. latitude, longitude) representing their physical locations in space. They can also navigate by

combining real-time GPS coordinates with other digital map data (e.g. via Google Maps). Geolocation data often varies around a five-meter radius. To compensate for imprecise GPS data, self-driving cars can use unique data-processing techniques like particle filtering to improve location accuracy. Furthermore, these systems can also be used to make efficient judgement for vehicle to take turn.
Some geolocation services used nowadays includes Indian Regional Navigation Satellite System (*IRNSS* or *NavIC*), United States' Global Positioning System (*GPS*), Russia's Global Navigation Satellite System (*GLONASS*), China's *BeiDou* Navigation Satellite System and the European Union's *Galileo*.

2.3 Vehicle Control with Deep Learning

2.4 Workflow

I designed the following workflow for the project:

2.4.1 Forming a problem statement

2.4.2 Creating dataset

2.4.3 Data Preprocessing

2.4.4 Designing Neural Network

2.4.5 Training Neural Network

2.4.6 Simulation in GTA-V

Chapter 3

Proposed Work

3.1 Dataset

Since there is shortage of dataset, I have created a program to generate a dataset while a user plays the game. The data recorded consists of an image and 4 required labels. // A total of 10016 images and labels are recorded in the code.

3.1.1 Image

Image of resolution 1920 X 1080 is recorded while manually controlling vehicle in the simulator. Figure 3.1 shows a image recorded image.

FIGURE 3.1: Recorded Image

3.1.2 Labels

The dataset comprises of four labels:

- ***Throttle Value*** is the value of the acceleration the car is in. Throttle value 1.0 means forward acceleration, 0.0 means backward acceleration and 0.5 means no acceleration.
- ***Throttle Flag*** is a boolean value which states whether there is an acceleration.
- ***Steering Value*** is the value of the steering of the car. Steering value 1.0 means left steering, 0.0 means right steering and 0.5 means no steering.
- ***Steering Flag*** is a boolean value which states whether there is an steering.

3.2 Data Preprocessing

3.2.1 Generating Navigation Map

The Navigation System inbuilt in the game is extracted from the recorded image for coordinates (28,873) to (300,1045). Then since only the purple region is required to give the path, it is colored white ann the rest part is colored black. Then the image is resized to 128 X 128.

3.2.2 Generating Speed Map

Using a modification in GTA-V, it is possible to have a digital speedometer which is just above the Navigation Map. From there, the speed is cropped from (125,843) to (173,869). Then the image is made binary. Then the extracted image is resized to 128 X 128.

3.2.3 Resizing Image

Since the map and speed is extracted from the original image, the region in image containing map and speed is masked to not contail it in the image. The region masked is from (20,839) to (300,1060). Then the image is resized to 256 X 256 to make the image computationally feasible for Neaural Network training.
Figure 3.2 shows the processed images, map, speed.

FIGURE 3.2: Processed Images

3.2.4 Collecting generate images and labels

Then the processed images and map are pushed into a array and all the hence genrated array are pushed into another array to form training images named Xtrain and all the labels are pushed into another array to form training labels named Ytrain. Then the generated Xtrain and Ytrain are saved in .npz files.

3.3 Model Designing

With the reference from Vision Transformer[2] and Native Transformer[3] and transformations as per the dataset, I have devised a unique transformer, named DRAN.
Below is the description of the DRAN.

3.3.1 Patch Embedding

As mentioned in the vision transformer, the image before being passed to the transformer encoder needs to be converted into patches and positional embedding needs to be added to the generated patches.

For this, with the help of some online blogs, I have devised an algorithm for Patch Embedding as shown in figure 3.3.

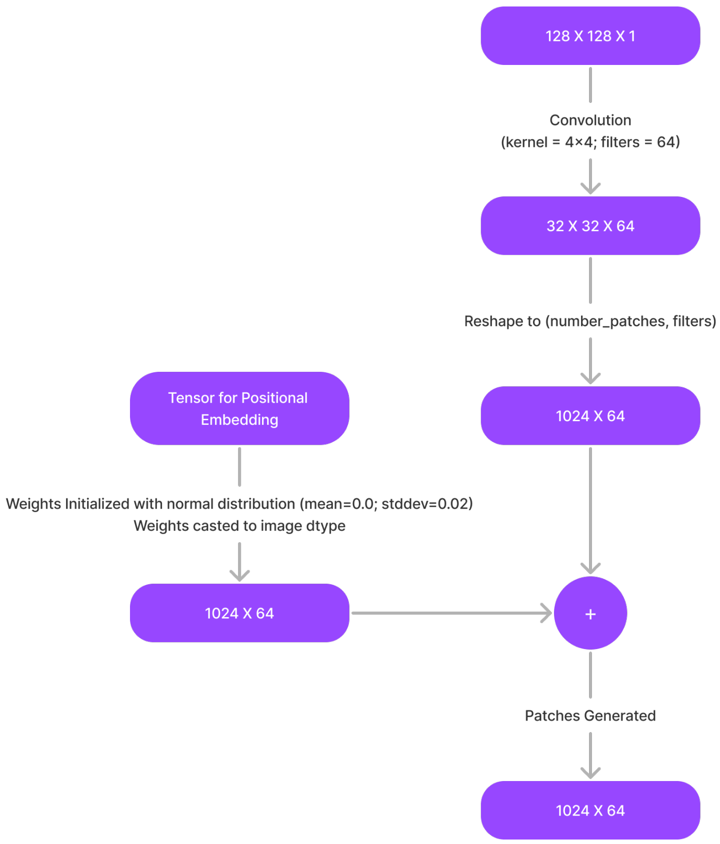

FIGURE 3.3: Patch Embedding Layer

3.3.2 Transformer Encoder

Encoder used in this transformer in a stack of 6 small identical units as explained in the original native transformer.

3.3.3 Model Decoder

There will be two transformer encoders in the final model so after some operations, data from two encoders will be combined and then further operations will be performed for generating output.
Figure 3.4, 3.5, 3.6 represents the structure of the decoder.

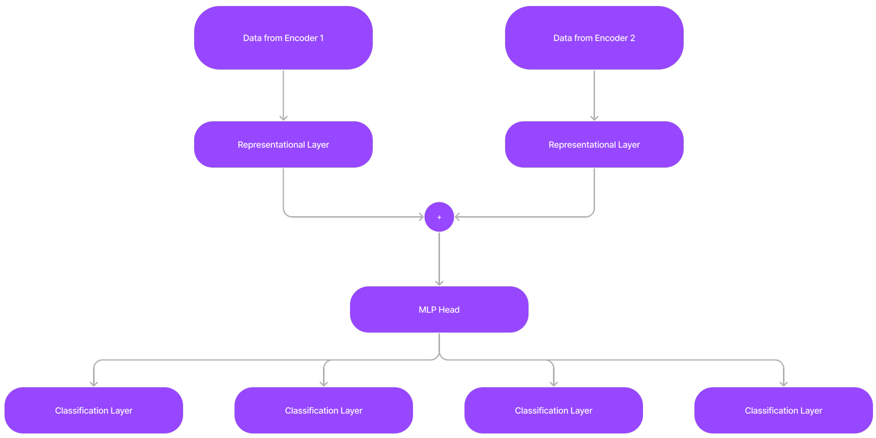

FIGURE 3.4: DRAN Decoder

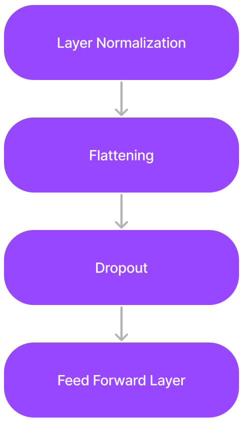

FIGURE 3.5: Representational Layer

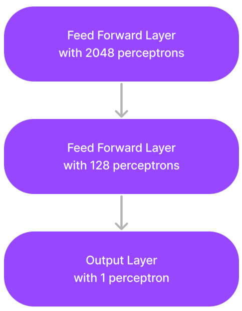

FIGURE 3.6: Classification Layer

3.3.4 Final Model

The main DRAN neural network is made with two input images one of the game camera recording, and other of the map processed in section 3.2.1.
There are 4 output labels of the neural network namely, Throttle Value, Throttle Flag, Steering Value, Steering Flag.
The final model designed is shown in figure 3.7 below.

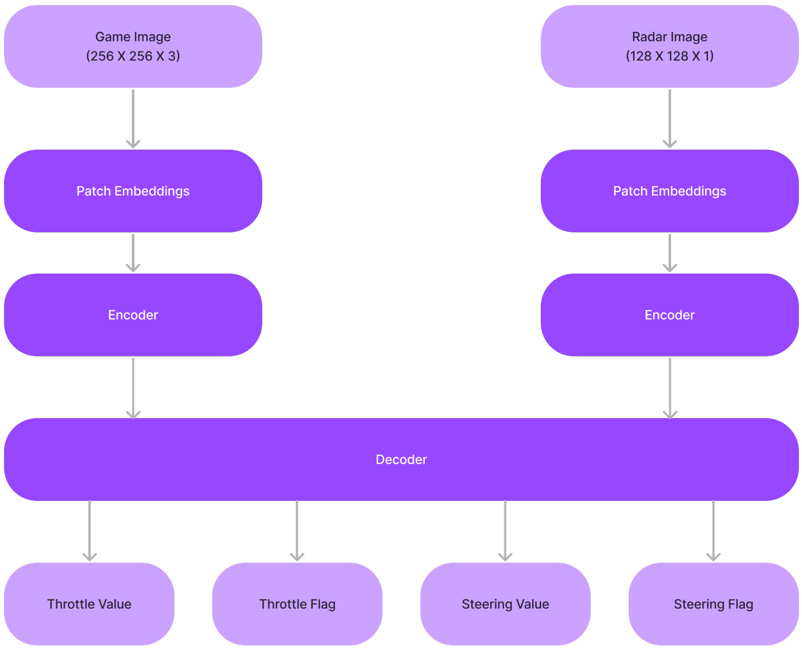

FIGURE 3.7: DRAN Neural Network

Chapter 4

Simulation and Results

Chapter 5

Conclusions and Future Work

Bibliography

[1] O.-R. A. D. O. Committee, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, apr 2021.

[2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” *arXiv e-prints*, p. arXiv:1706.03762, June 2017.