An Experimental Analysis on Self Driving Car Using CNN

Dr Manju Bargavi S K[1], Sowjanya M[2]

[1]Professor & Department of MCA, Jain University School of Computer Science and IT, Bangalore, India,
[2]Student & Department of MCA, Jain University School of Computer Science and IT, Bangalore, India,

---

Abstract - *For the past decade, there has been a surge of interest in self-driving cars. This can be because of breakthroughs within the field of deep learning wherever deep neural networks square measure trained to perform tasks that generally need human intervention. CNN's apply models to spot patterns and options in pictures, creating them helpful within the field of pc Vision. Samples of these square measure object detection, image classification, image captioning, etc. during this project, we've trained a CNN victimization pictures captured by a simulated automotive to drive the automotive autonomously. The CNN learns distinctive options from the pictures and generates steering predictions permitting the automotive to drive while not somebody's. For testing functions and getting ready the dataset the Unity based mostly machine provided by Udacity was used.*

***Key Words*: Autonomous driving, Convolutional Neural Network (CNN), Deep learning, end-to-end learning, NVIDIA, Steering commands.**

1. INTRODUCTION

In recent years, autonomous driving algorithms persecution reasonably priced vehicle-mounted cameras have attracted increasing analysis endeavors from each, domain and trade. Varied levels of automation are outlined in autonomous driving. There's no automation in level zero. Somebody's driver controls the vehicle. Level one and a couple of square measure advanced driver help systems wherever somebody's driver still controls the system however a number of options like brake, stability management, etc. square measure machine-driven. Level three vehicles square measure autonomous, however, somebody's driver continues to be required to observe and intervene whenever necessary. Level four vehicles square measure totally autonomous however the automation is restricted to the operational style domain of the vehicle i.e. it doesn't cowl each driving state of affairs. Level five vehicles square measure expected to be totally autonomous, and their performance ought to be such as that of somebody's driver. We tend to square measure terribly off from achieving level Five autonomous vehicles within the close to future. However, level - 3/4 autonomous vehicles square measure probably changing into a reality within the close to future. Primary reasons for forceful technical achievements during these fields square measure technical breakthroughs and wonderful analysis being exhausted the sector of pc vision and machine learning and conjointly the cheap vehicle-mounted cameras which may either severally give unjust data or complement different sensors. Several vision-based drivers assist options are wide supported within the fashionable vehicles. A number of these options embrace pedestrian/bicycle detection, collision turning away by estimating the front automotive distance, lane departure warning, etc. However, during this project, we tend to target autonomous steering, that could be a comparatively unknown task within the field of pc vision and machine learning.

In this project, we tend to implement a convolutional neural network (CNN) to map raw pixels from the captured pictures to the steering commands for a self-driving automotive. With minimum coaching knowledge from the humans, the system learns to steer on the road, with or while not the lane markings.

2. RELATED WORK

The DAVE system was created by government agency [1] and used pictures from 2 cameras likewise as left and right steering commands to coach a model to drive. It demonstrates that the technique of end-to-end learning may be applied to autonomous driving. This implies that the intermediate options like the stop signs and lane markings don't need to be annotated or labeled for the system to be told. DAVE is associate degree early project within the field of autonomous driving. Within the context of current technology, a large portion relied on wireless knowledge exchange as a result of the vehicle couldn't carry the computers and power sources for the system, that contrasts the lighter instrumentation that exists nowadays. The design of this model was a CNN created of totally connected layers that stemmed from networks antecedents utilized in seeing. The ALVINN system [2] could be a 3-layer back-propagation network designed by a bunch at CMU to finish the task of lane-following. It trains on pictures from a camera and a distance live from an optical device vary finder to output the direction the vehicle ought to move. ALVINN's model uses one hidden layer back-propagation network.

We replicated a study by NVIDIA [3]. The system uses associate degree end-to-end approach wherever the information is 1st collected in multiple totally different environmental conditions. The information is then increased to form the system sturdy to driving off center and to totally different potential environments. Consecutive step is coaching the network on this knowledge. Recently, a paper by a handful of IEEE researchers introduced quite a totally different neural spec that additionally takes the temporal info under consideration [4]. They achieved this in applies by a mixture of normal vector-based Long memory (LSTM) and convolutional LSTM at totally different layers of the planned deep network. Consecutive frames typically have the same visual look, however delicate per element motions may be discovered once the optical flow is computed. Typical image convolutions, as those adopted by progressive image classification models, will shift on each spatial dimensions in a picture, which means that they're primarily 2-D. Since these convolutions care for static pictures or multichannel response maps, they're incapable of capturing temporal dynamics in videos. The authors adopted a spatio-temporal convolution (ST-Conv) that shifts in each spatial and temporal dimensions so applying the convolution in three dimensions as critical the normal 2-D method.

A similar paper additionally planned the concept to include temporal info within the model to be told the steering info [5]. During this paper the authors demonstrate quantitatively that a Convolutional Long memory repeated Neural Networks (C-LSTM) will considerably improve end-to-end learning performance in autonomous vehicle steering supported camera pictures. Impressed by the adequacy of CNN in visual feature extraction and therefore the potency of Long memory (LSTM) repeated Neural Networks in coping with long-range temporal dependencies our approach permits to model dynamic temporal dependencies within the context of steering angle estimation supported camera input.

3. DATA COLLECTION

We used Udacity's self-driving automobile machine for assembling the info. This machine is made in Unity and was utilized by Udacity for the Self-Driving Nanodegree program however was recently open-sourced [6]. It replicates what NVIDIA did within the simulation. We are able to collected all our information from the machine. Exploitation our keyboard to drive the automobile, we have a tendency to were able to instruct the simulated vehicle to show left, right, speed up and hamper. Another vital facet is that this machine may be used for coaching in addition as testing the model. Hence, it's 2 modes: (i) coaching mode, and (ii) Autonomous mode as shown in Fig. 1.

Fig. 1 Udacity Simulator

The coaching mode is employed to gather the info and also the autonomous mode is employed to check the model. In addition, there are two different varieties of tracks within the machine - the lake track and also the jungle track. The lake track is comparatively smaller and simple to handle the automobile in comparison with the jungle track as shown in Fig. 2 and Fig. 3. The machine captures information once the automobile is driven round the track exploitation left and right keys to manage the steering angles and up and down arrows to manage the speed.

Fig. 2 Udacity Simulator: The lake Track

From this, the machine generates a folder containing pictures and one CSV file. The image folder contains 3 pictures for each frame captured by the left, center and right camera and each row within the CSV file contains four metrics - steering angle, speed, throttle and brake, for each captured frame. Fig. 4, Fig. 5. and Fig. 6 show the left, center and right image, for one frame.

Fig . 4 Left Image

Fig. 5 Center Image

Fig. 6 Right Image

4. DATA PREPROCESSING

The data that we have a tendency to collect i.e. the captured pictures area unit preprocessed before coaching the model. Whereas preprocessing, the picture's area unit cropped to get rid of the sky and front portion of the automobile. The picture's area unit then reborn from RGB to YUV and resized to the input form utilized by the model. This is often done as a result of RGB isn't the simplest mapping for perception. YUV color-spaces could be a far more economical cryptography and reduces the information measure quite RGB capture will. After choosing the ultimate set of frames, the info is increased by adding artificial shifts and rotations to show the network a way to pass though a poor position or orientation. Whereas augmenting, we have a tendency to willy-nilly opt for right, left or center pictures, willy-nilly flip the pictures left/right and regulate the steering angle. The steering angle is adjusted by +0.2 for the left image and -0.2 for the proper image. Exploitation the left/right flipped pictures is helpful to coach the recovery driving state of affairs. We have a tendency to conjointly willy-nilly translate the image horizontally or vertically with steering angle adjustment. The horizontal translation may be helpful for handling eventualities with troublesome curves. The magnitude of those changes is decided willy-nilly from a standard distribution. The distribution contains a zero mean. We have a tendency to apply these augmentations employing a script from the Udacity repository.

Augmented pictures area unit adventitious into this set of pictures and their corresponding steering angles are adjusted with relevance augmentation performed. The first reason for this augmentation is to form the system a lot of strong, thus, learning the maximum amount as doable from the surroundings by exploitation various views with various settings.

5. DATA LEARNING MODEL

The deep learning model we have a tendency to ready relies on the analysis done by NVIDIA for his or her autonomous vehicle [3]. The model contains of the subsequent vital layers.

5.1 Convolutional Layer

Convolutional layer applies the convolution operate (filter) on the input image and produces a 3D output activation of neurons. This layer helps to seek out varied options of the image that is employed for classification. The amount of convolutional layers within the network depends on the kind of application. The initial convolutional layers within the networks facilitate observe the low level options of the image that area unit easy and also the convolutional layers more facilitate in police investigation the high-level options of the image.

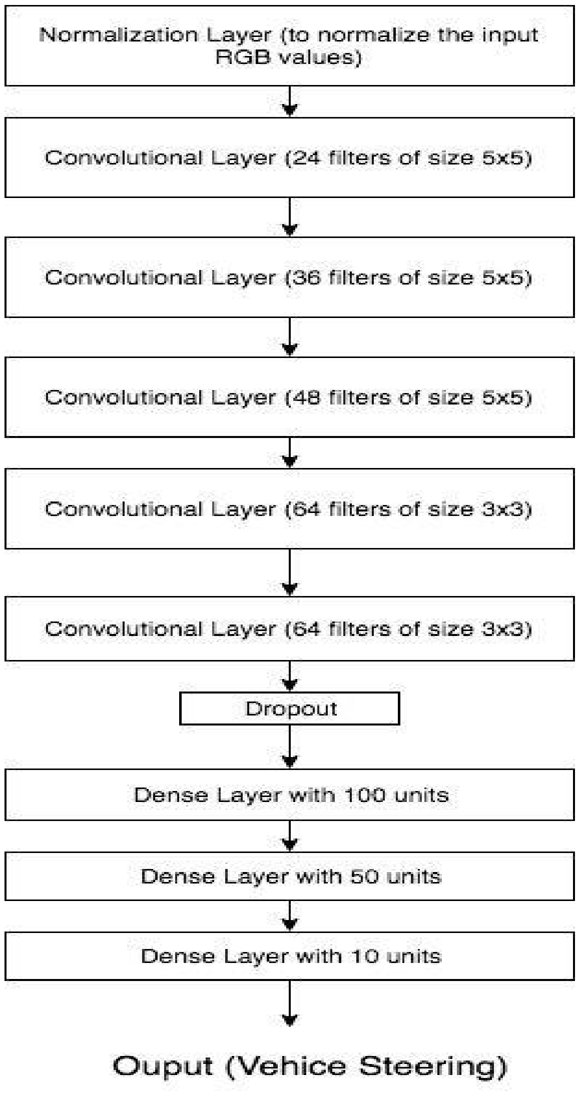

Fig. 7 CNN Architecture

5.2 Max Pooling Layer

Pooling/Down sampling layer helps in reducing the amount of parameters/weights within the network and helps in reducing the coaching time while not losing any specific feature info of the image. This layer produces a smaller image than the input image by down sampling the image exploitation pooling of neurons. There area unit differing kinds of pooling like gamma hydroxybutyrate pooling, average pooling, L2-norm pooling etc. however gamma hydroxybutyrate pooling is that the one that is most generally used.

5.3 Dense Layer

Dense layer or totally connected layer is that the same as traditional neural network layer during which all the neurons during this layer area unit connected to every neuron from previous layer. This layer is usually designed because the final layer within the convolutional network.

5.4 Neural network.

The entire design diagram is shown in Fig. seven There area unit five convolutional layers with varied range of filters and sizes and a dropout subsequently to handle overfitting. In the end, three dense layers were adventitious followed by the output layer. Adam optimizer was used for parameter optimization with a set learning rate. Batch size of one hundred was chosen and also the range of epochs of 50-60 was experimented with. On a machine while not GPU, sixteen GB ram Core i5 it took roughly 06 hours of coaching.

6. SYSTEM DESIGN

The high-level design of the system in Fig 8 shows when activity information augmentation on the input pictures, batches area unit created from them and fed to the CNN model for coaching. When the coaching is completed, the model is employed to perform prediction on the steering angle and send the predictions to the Udacity machine to drive the automobile in real time.

6.1 Performance analysis

For performance analysis throughout coaching, mean square error was used as a loss operate to stay a track of the performance of the model.

In reality eventualities, whereas driving on road, the subsequent metric has been planned in [3]. This metric has been named as autonomy. It's calculated by investigation the quantity of interventions, multiplying by vi seconds, dividing by the time period of the simulated check, and so subtracting the result from one.

Measure of autonomy, this could mean that the automotive was eighty-seven.5% autonomous. We tend to re-trained the model with additional information and also the vehicle ne'er drove off the trail, creating it 100 percent autonomous. The general driving behavior doesn't seem realistic. It seems to improve and forth between the sides of the lane.

During the primary run on the lake track, the vehicle doesn't leave the track, however it seems to hug the left facet of the lane which is that the direction that the curve within the track is popping towards. The vehicle drives to the left facet of the lane and once it runs getting ready to the left lane marking, the vehicle drives back within the direction of the middle of the lane till it's around halfway between the middle of the lane and also the left lane marking, then it defends to the left facet of the lane. This behavior repeats. Once re-training the model, the vehicle's driving seems far sleeker. It doesn't hug the left facet of the lane, however it seems to drive to 1 facet of the lane and move towards the opposite. This behavior conjointly repeats. In terms of autonomy, each of the models seem to be 100 percent autonomous.

After the second time we tend to train the model, we tend to noticed that the vehicle will drive autonomously, however the behavior wasn't natural. This can be one thing to stay in mind once purification the system.

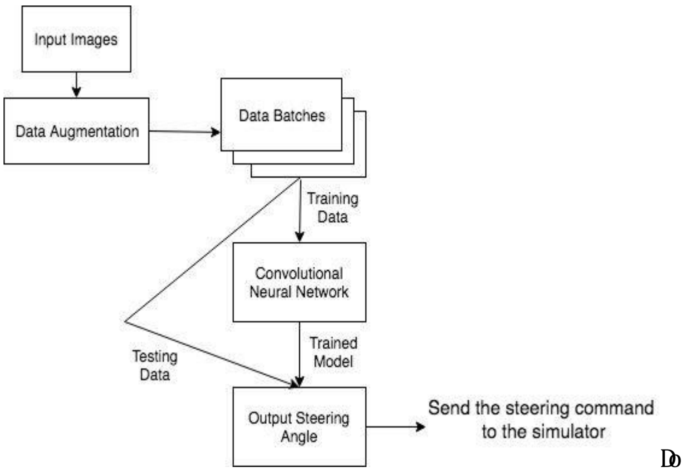

Fig. 8 High-level system architecture[7]

7. EXPERIMENTAL RESULTS

The simulation surroundings is using udacity self-driving car simulator and the system is based on NVIDIA model [10]. Figure 10 shows the learning surroundings formed in the simulation study and figure 6 shows trial images taken by left camera, centre camera and right camera in training method. The data composed is about 16,000 images and categorized with the type of road and driver movement such as on lane or turning. The simulation sample rate is 10 FPS and image resolution 320x160 pixels.

Before the learning progression, the image taken will be regularize by removing the sky and other avoidable parts of picture for simplifying the learning process. Some pre-processing of composed data images also comprise rotation, horizontal flip, vertical flip, and RGB to YUV transform to advance the quality of the classification end result.

The training procedure using 5x5 kernel for the first 3 convolution layer and 3x3 kernel for the last three convolution layer and utmost epochs 20 with 2.000 example for each epoch.

After the training finish about 2 or 3 laps, the simulator then switch to the autonomous mode and then base the model created by the CNN, the car will run unconventionally and can calculate for every road situation.

The simulation is implemented on a system with i5 processor and 12G memory. The learning takes about 4-5 hours.

Fig 9 Driving mode

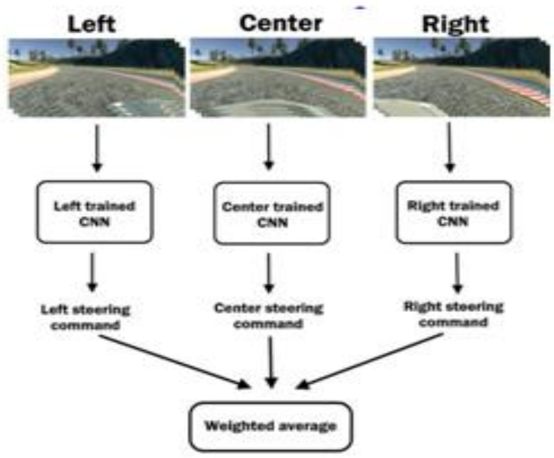

Fig.10. Model Prediction

8. CONCLUSION

The projected technique of autonomous car using CNN can run effortlessly without error and very stable without fluctuation in environmental simulation using udacity self-driving car simulator as shown in figure 9. CNN can learn road situation from three cameras (left, centre and right) and also make a model for driving in autonomous mode restricted by one centre camera. We also verified that CNN's are able to learn the complete task of lane and road following with no manual disintegration into road or lane marking detection, semantic abstraction, path planning, and control. An interesting limitation to this is that the system was able to successfully drive on the roads that it had been trained on. To get the enhanced performance, next trialing will include the complexity in the simulation surroundings and rain model with supplementary noise and distortion at data image set. We have many ideas to enhance the act of the self-driving automotive, out of that one was obligatory thanks to lack of your time. The primary plan would be to attribute the feature of speed to the CNN so once the machine in autonomous mode, it's persecution the predictable speed creating the movement seem to be supplementary realistic.

REFERENCES

[1] LeCun, Y., et al. *DAVE: Autonomous off-road vehicle control using end-to-end learning*. Technical Report DARPA-IPTO Final Report, Courant Institute/CBLL, 2004.

[2] Pomerleau, Dean A. "Alvinn: An autonomous land vehicle in a neural network." *Advances in neural information processing systems*. 1989.

[3] Bojarski, Mariusz, et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).

[4] Chi, Lu, and Yadong Mu. "Deep Steering: Learning End-to-End Driving Model from Spatial and Temporal Visual Cues." *arXiv preprint arXiv:1708.03798* (2017).

[5] Eraqi, Hesham M., Mohamed N. Moustafa, and Jens Honer. "End-to-End Deep Learning for Steering Autonomous Vehicles Considering Temporal Dependencies." *arXiv preprintarXiv:1710.03804* (2017).

[6] International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE) Vol 5, Issue 4, April 2018 Prototype of Autonomous Car Using Raspberry Pi.

[7] Spice B, "Carnegie Mellon, Argo AI Form Center for Autonomous Vehicle Research with $15M Multiyear Grant," Carnegie Mellon University, 2019.

[8] Matt Richardson, Shawn Wallace, Getting Started with Raspberry Pi, 2nd Edition, Published by Maker Media, Inc., USA, 2014. Book ISBN: 978-1-457- 18612-7

[9] Deep learning for Video classification and captioning byZuxuan Wu, Ting Yao, Yanwei Fu, Yu-Gang Jiang https://keras.io/backend

[10] Visualizing and Understanding Convolutional Networks by Matthew D. Zeiler, Rob Fergus.

[11] Design and Implementation of Autonomous Car using Raspberry Pi International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 9, March 2015

[12] Dropout: A Simple Way to Prevent Neural Networks fromOverfitting by Nitish Srivastava, Geoffrey Hinton , Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov.