A Survey of Self-Driving Car Deep Learning Algorithms

Alexander Mehta

Fremont High School (Sunnyvale, CA), Grade 11

Gifted and Talented Institute, University of California, Irvine

mehtaa4@uci.edu

Abstract

This review paper covers commonly used deep learning techniques for self-driving vehicles. Due to increased research in deep learning algorithms and commercial advances in autonomous vehicle hardware such as LIDAR and cameras, deep learning algorithms for autonomous cars have become an area of rapid development and research due to increased capabilities of deep learning algorithms.

Deep learning research has focused on the use of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Deep Reinforcement Learning (DRL), Convolutional-Long Short-Term Memory Models (C-LSTMs), Graph Neural Networks (GNNs), and Hidden Markov Models and Chains (HMMs). This literature review covers four key applications of deep learning in autonomous driving: scene classification, lane management, path planning, and obstacle detection. These four applications are the main elements required for a fully self-driving vehicle to work safely and effectively over a long period of time.

Results indicate that for scene classification, recurrent network models outperform convolutional models due to the temporal benefits that convolutional neural networks do not possess. In lane management, continuous deep reinforcement learning had the best performance to keep the car in the lane over discrete learning. This allowed a vehicle to stay in a lane without swerving off a path. Path planning is best performed when traffic is represented by an undirected graph structure and fed into a graph neural network for path creation. Markov models were the strongest model for path decision-making. Obstacle detection was best performed by Convolutional-LSTMs over Convolutional Neural Networks and Recurrent Neural Networks due to temporal and image processing benefits.

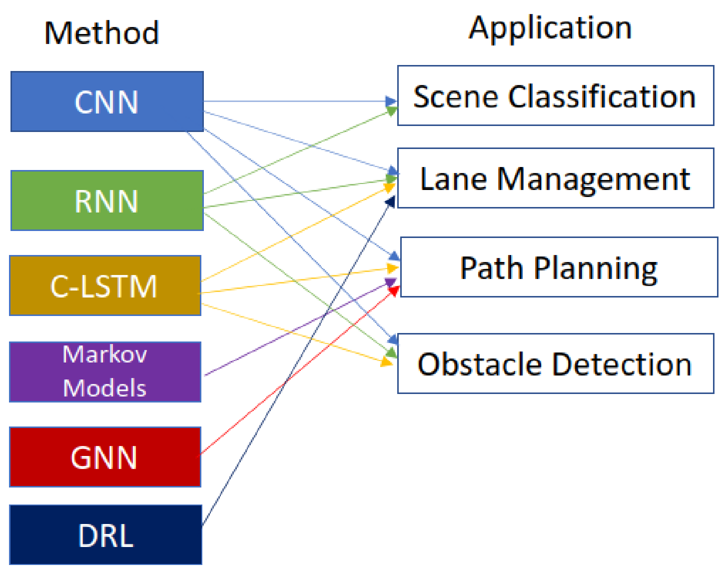
Introduction

Self-driving cars are one of the most researched topics in artificial intelligence due to the safety and efficiency benefits the automated world brings to our society. Automation brings the opportunity to prevent deaths caused by human error. Machines are designed for efficiency, making self-driving cars a key to a more fuel-efficient world.

Previous surveys of autonomous vehicle technology have focused on convolutional neural networks (CNNs), deep reinforcement learning (DRL), and recurrent neural networks (RNNs). This paper surveys new approaches in the C-LSTM architecture, Deep Reinforcement Learning improvements, graph neural networks, and Markov models and chains. This paper aims to provide insight into the current state of deep learning technologies for self-driving vehicles in the main fields of self-driving vehicles: obstacle detection, scene classification, lane recognition, motion control, traffic light and sign recognition, and path planning.

Self-driving car performance is separated into six levels from 0 to 5 by the NHTSA (National Highway Traffic Safety Administration). Level 0 represents a human-driven car. Level 1 has basic adaptive cruise control and emergency braking. Levels 2-5 are what many consider to be autonomous cars. Level 2 has partial autonomy with driver supervision and intervention to perform tasks. In level 3, the system is autonomous, but a driver can intervene. Level 4 self-driving cars can intervene when there is a system malfunction unlike level 3. Level 5 self-driving cars require zero human interaction and are considered the overall goal of self-driving. Levels of autonomous driving help understand the importance of different applications of self-driving vehicles.

Notable companies working on autonomous vehicles include Google, Tesla, Apple, and Nvidia. These companies have accelerated the hardware capabilities of autonomous vehicles. To capitalize on the improvements in hardware, academic research has focused on the development of newer machine learning techniques. The key focus of newer machine learning techniques is to utilize the enhancement in visual data to output sensible driving commands (i.e., accelerate, steer right, steer left) to the vehicle and to make the vehicle autonomous. Key techniques being developed include C-LSTMs, graph networks, Markov models and chains, and other new technologies covered in this literature review.



Obstacle detection

Obstacle detection is a key part of autonomous driving because it enables a vehicle to identify on-road hazards such as vehicles, traffic signs, animals, and other dangerous items on the road. By discovering the location of these objects, the autonomous vehicle performs a risk assessment of the roadway and avoids hazards to find a safe path. Level 3 to 5 self-driving cars are only achievable when the vehicle is capable of accurately detecting and identifying objects on the road promptly.

Obstacle detection works in two parts. The first is the acquisition of images of the vehicle's surroundings, typically acquired using cameras, RADAR, and LIDAR (Light Detection and Ranging) (Badue et al., 2019). An image recognition algorithm, typically based on deep learning, then parses the images to help identify objects on the road. The deep learning algorithm achieves this feat by being trained with prior image datasets containing labeled data on the common objects found on the road. Key deep learning techniques used for obstacle detection include convolutional neural networks (CNN), recurrent neural networks (RNN), and convolutional Long Short-Term Memory model (C-LSTM), of which C-LSTMs have garnered significant interest in the past two years.

CNN

In the early 2000s, DARPA (Defense Advanced Research Projects Agency) introduced obstacle challenges for self-driving cars. CNNs were the most successful technology and were the gold standard for obstacle detection due to their performance. CNNs primarily work by learning patterns to create filters to detect edges and other structures in images the network finds of interest. Despite being introduced over 20 years ago, CNNs are still one of the primary research interests and most effective methods in deep learning and obstacle detection. Parmar, Natarajan, and Sobha's recently developed CNN for obstacle detection boasts a 96.92% accuracy (Parmar et al. 2019). This was achieved by focusing on the depth of objects in a scene to better understand their position. Despite progress in CNNs, a key challenge is CNNs inability to effectively understand the concept of data in a time series (i.e., videos) since all data is inputted simultaneously.

RNN

Since the late 1990s, the recurrent neural network has been the standard for temporal data for its ability to provide the network a sense of time. RNNs understand the previously inputted data as previous information from a different time, whereas a convolutional network would view the data as inputted at once. In a comparison test performed by Parmar et al., ResNet, a type of recurrent neural network, outperforms an advanced CNN known as VGG by ~1-3% (Parmar et al., 2019). Despite the ability to understand time-oriented data, RNNs fall short when it comes to training time and lack the performance benefits of other temporal algorithms.

C-LSTM

The first LSTM was created in 1997, 20 years later, the abilities of CNNs and LSTMs were combined for use in video data. The LSTM has been used by researchers to solve the training and performance issues flaws in recurrent networks. The LSTM solves these issues by introducing a new operation to recurrent networks, the ability to forget previous data, making it easier for networks to determine what parts of a video are useful and what can be safely ignored. Researchers have combined the LSTM with the benefits of convolutional networks to combine the two into a single model known as the C-LSTM. The C-LSTM model incorporates the benefits of both CNNs and LSTMs. The C-LSTM showed a ~2% performance boost over a CNN in comparison tests (Cornacchia et al., 2018). Researchers have yet to focus on applying the new attention architecture. The use of attention has shown performance and time benefits in other applications of machine learning.

	CNNs	C-LSTM
Accuracy (6 classes)	88.59%	90.56%

Path Planning

Path planning is the ability to decide a trajectory a car must take to travel from one location to another safely. Proper path planning is essential to self-driving vehicles as it allows a car to change its path based on its environment. Research in other fields such as obstacle detection and lane recognition are beneficial to better path planning as the information learned from other models can influence safety precautions when path planning.

Path planning works by taking in positional input (from cameras, RADAR, or GPS), and information about the surrounding environment to create the optimal path, identifying any lane switches or turns along the way. Due to the complex nature of both the given information and the possible output, path planning requires deep learning techniques to detect underlying structures in decision-making instead of surface-level pattern detection. Leading deep learning algorithms for path planning are Hidden Markov Models, Markov Chains, C-LSTMs, and Graph Neural Networks, of which Graph Neural Networks (GNNs) have been of focus in the last year.

Markov Models

Markov chains were introduced for use in self-driving vehicles in 2011 by Althof & Mergel. Althof & Mergel utilized Markov chains to predict the risks of driving through a path in a given environment (Althof & Mergel, 2011). The Markov models are based on a set of mathematical processes that use probability to predict future events. For self-driving vehicles, Markov models predict the risk of crashing if a vehicle drives through a certain path. A key challenge of Althoff & Mergell's model was the overprediction of risk at higher speeds. Since then, researchers have developed a Hidden Markov model (HMM) to predict car behavior, achieving 83.34% mean accuracy (Wu et al., 2012). The use of the HMM to make decisions has been beneficial to the growth of stochastic methods of deep learning. The hidden Markov model provides additional variability, allowing for the coverage of a more complex set of possibilities, improving accuracy and preventing overprediction. Although the changes Hidden Markov models bring are significant, understanding and processing video data is not part of the existing model and must be done separately, which is time-consuming and error prone.

C-LSTM

C-LSTMs gained the attention of researchers in path planning after use in obstacle detection. Since then, the model has been modified to work with videos as well.

C-LSTMs use CNN created filters paired with an LSTM to understand time while effectively processing the image. Path finding is a challenge due to the fact that there are infinitely many possible correct paths from the model, so performance cannot be measured by a true/false metric but rather by how close the model was to the perfect solution. The existence of infinite correct solutions makes pathfinding a major challenge for researchers. Bai et al. have used C-LSTMs new method on a dataset open sourced by self-driving car company Comma.ai. Compared to the company's model, the C-LSTM had better accuracy and stability due to its use of time sequence and convolutions simultaneously (Bai et al., 2019).

GNN

In 2017, the graph neural network was first introduced with the original intent of protein analysis since proteins follow the shape of a connected graph. Since then, general research demonstrates graphs are effective structures to represent complex relationships, such as traffic and cars. Specifically, in 2021, a model proposed by Sheng et al. inputs a graph into a network and predicts the trajectories of cars at an intersection (Sheng et al., 2021). The model outperformed C-LSTMs and CNNs while having a smaller model size (leading to faster processing times). This is due to the fact that a graph represents a scene in a few data points, where a single frame in a video contains millions of data points. Shown in the table below are performance and efficiency benefits.

	Graph Network	Baseline Model	C-LSTM	CV
RMSE (lower better)	1.49	1.61	1.86	3.42
Processing Time (MS)	0.044	0.322 (7.3x)	N/A	N/A

Scene Classification

Scene classification is the ability to identify the environment of the car through video and LIDAR data. This is critical as a car must behave differently based on conditions such as weather. A driver must be more careful during snow or rain, and a self-driving vehicle must change its behavior based on the environmental conditions to have full level 5 autonomy. Factors such as visibility and weather influence how the vehicle operator drives.

Scene classification works by taking an image or video input and categorizing the environment, such as rain or fog. Since conditions are often not easily identifiable by common features, machine learning conditions make detection more effective. The main machine learning techniques are CNNs and RNNs. Both have seen major improvements and derivatives with higher stability and accuracy.

CNN

Since the 1980s, CNNs have been used to classify images. CNNs were utilized for scene classification before the emergence of its need in self-driving vehicles. Over the last 40 years, there have been major improvements in image tasks such as scene classification. CNNs help classify scenes by inputting labeled images of scenes into a convolutional network. A convolutional network learns through network training what the best filters are to understand a scene. CNNs improved significantly with a new technique known as global covariance pooling (GCP) (Wang et al., 2020). CNN pooling works by downscaling an image by averaging pixels, leading to data loss over multiple iterations. Using GCP, the downscaling happens while taking statistical significance into account, greatly improving accuracy and stability while still downsizing the image. The use of binocular inputs has greatly improved the field of view and accuracy in CNNs. Instead of a single camera as the input, binocular inputs access two or more cameras for a larger field of view and more information for the network. Zhong et al. construct a model with two separate convolutional networks, for each camera (Zhong et al., 2018). These networks are combined into a single network to get a more accurate result than a single image. Mancini et al.’s two-camera method outperformed the monocular (single camera) model proposed in the KITTI dataset (Mancini et al., 2016).

	Zhong et al.	Mancini et al.
RMSE (Lower is better)	4.451	6.863

RNN

Recurrent Neural Networks were introduced in the 1980s in a similar timeframe as CNNs. RNNs were not used for video data until researchers found that videos were best processed by understanding temporal structure. A recurrent model proposed by researchers in

2017 outperforms CNNs in scene classification (Phan et al., 2017). The model used both audio and visual data as inputs to classify the scene. The key issue behind RNNs is the longest train times and a training issue known as *exploding gradients*. Due to the complexity of an RNN, it's prone to large "exploding" errors when training, leading the network to have a hard time converging on an optimal solution. Further research must be performed on possible improvements to recurrent units and training methods such as gradient clipping to improve training accuracy.

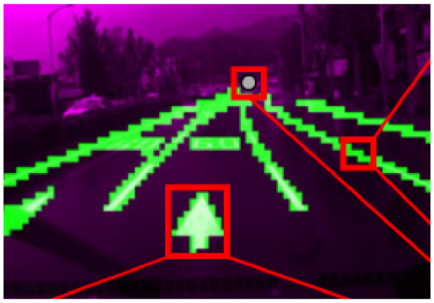
Lane Management

Lane management and recognition is the ability to detect the road's lanes and stay on a given path. Lane recognition also is the ability to switch lanes safely. Lane management and recognition have been used for all levels of self-driving. Current cruise control systems offer lane-keeping technology known as adaptive cruise control.

Lane management and recognition take input data from cameras and LIDAR sensors. These inputs are used to determine the location of lanes, and the car’s position relative to these lanes to keep the car in the lane. Common computer vision algorithms are inaccurate when in conditions such as rainstorms. Deep learning algorithms are implemented to improve accuracy in these conditions. Common deep learning techniques are CNNs and Deep Reinforcement Learning (DRL).

CNN

Since Caltech’s Lane Detection dataset was introduced in 2008, interest in CNNs as a primary architecture has grown, a technology shift from the more commonly used computer vision techniques. These models commonly output a mask over the lanes instead of classification (a more common output of a CNN). Researchers have focused on consistency during harsh environments such as rain, where images might not be as clear. Lee et al. propose a multi-task CNN that improves the performance of lane detection by finding not only lanes but the vanishing point – the point where the two lanes appear to become 1 point (where the two lanes vanish) (Lee et al., 2017). Using the vanishing point, it is easier to verify the location of hard-to-discern lanes in harsh conditions. The model (88.4%) performed ~16% better than Caltech's baseline CV model (72.3%). The new model outperformed the state-of-the-art CNN model (84.8%) by ~2%.



DRL

Deep reinforcement learning (DRL) for autonomous vehicles gained popularity in 2012 when Xu et al. proposed the first reinforcement learning driving system at ICRA (IEEE International Conference on Robotics and Automation) (Xu et al., 2017). DRL works by trial and error, attempting to reach a specified goal instead of a correct answer. In lane management, the goal is to either switch to the lane with minimal risk to others or stay in the lane without crashing into others. In 2016, a model introduced by Sallab et al. kept lanes indefinitely through deep reinforcement learning (Sallab et al., 2016). The model rewarded staying on the simulated track and gave less reward depending on the risk of how the car behaved. For example, running off the track gave the lowest reward. Sallab et al. compared two different algorithms for deep reinforcement learning, discrete and continuous. Discrete learning is similar to a discrete function, where the algorithm is working at time intervals and is not constantly in check of its lane. Discrete learning was not effective at keeping the car in lane due to the algorithm not updating its position constantly, causing the vehicle to drive off the lane until corrected at the last minute. Continuous deep learning, like a continuous function, performed adequately and constantly kept the car in the lane over a long period of time. Deep reinforcement learning's success in lane keeping has opened future research into lane switching, with a focus on improving efficiency while lowering accidents. Wang et al. found current models commonly focus on the error of the target vehicle, but the best way to train DRL is to focus on all cars to create the best-trained vehicle (Wang et al. 2019). Cooperation led to a better overall driving environment compared to a focus on the outcome of a single target vehicle.

Conclusion

One of the most researched topics in autonomy and deep learning are self-driving cars. Due to this, many new algorithms have been applied to self-driving cars. This review covered 6 main technologies: CNNs, RNNs, C-LSTMs, GNNs, DRL, and Markov Models and Chains. These technologies were covered in 4 different fields of application: obstacle detection, path planning, scene classification, and lane management. Although current models perform with high accuracy, improvements are still a major research area in the artificial intelligence community. Specifically, developments in C-LSTMs, DRL, and GNNs are still needed.

In conclusion, C-LSTMs are more effective than RNNs and CNNs at obstacle detection and path planning but are outperformed by GNNs in path planning. GNNs are the newest technology in artificial intelligence and autonomous vehicles and require more research to verify their effectiveness. Markov Models and Chains are useful tools for probability but are outperformed by other deep learning techniques that employ neural networks. In DRL, a shift of focus from single car algorithms to cooperation leads to greater accuracy. CNNs and RNNs still showed impressive results that allow them to be used as baseline models for improvements such as C-LSTMs. Future research into vehicle autonomy can greatly improve the safety and efficiency of transportation. Improvements in transportation decrease fatalities and allow for faster and longer travel.

References

Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixão, T., Mutz, F., Veronese, L., Oliveira-Santos, T., & De Souza, A. F. (2019). *Self-Driving Cars: A Survey* (arXiv:1901.04407; Issue arXiv:1901.04407). arXiv. https://doi.org/10.48550/arXiv.1901.04407

Parmar, Y., Natarajan, S., & Sobha, G. (2019). DeepRange: Deep-learning-based object detection and ranging in autonomous driving. IET Intelligent Transport Systems, 13(8), 1256–1264. https://doi.org/10.1049/iet-its.2018.5144

Cornacchia, M., Kakillioglu, B., Zheng, Y., & Velipasalar, S. (2018). Deep Learning-Based Obstacle Detection and Classification With Portable Uncalibrated Patterned Light. IEEE Sensors Journal, 18(20), 8416–8425. https://doi.org/10.1109/JSEN.2018.2865306

Althoff, M., & Mergel, A. (2011). Comparison of Markov Chain Abstraction and Monte Carlo Simulation for the Safety Assessment of Autonomous Cars. *IEEE Transactions on Intelligent Transportation Systems*, *12*(4), 1237–1247. https://doi.org/10.1109/TITS.2011.2157342

Wu, J., Cui, Z., Zhao, P., & Chen, J. (2012). Traffic Vehicle Behavior Prediction Using Hidden Markov Models. In J. Lei, F. L. Wang, H. Deng, & D. Miao (Eds.), *Artificial Intelligence and Computational Intelligence* (pp. 383–390). Springer. https://doi.org/10.1007/978-3-642-33478-8_48

Bai, Z., Cai, B., Shangguan, W., & Chai, L. (2019). *Deep Learning Based Motion Planning For Autonomous Vehicle Using Spatiotemporal LSTM Network* (arXiv:1903.01712; Issue arXiv:1903.01712). arXiv. http://arxiv.org/abs/1903.01712

Sheng, Z., Xu, Y., Xue, S., & Li, D. (2021). *Graph-Based Spatial-Temporal Convolutional Network for Vehicle Trajectory Prediction in Autonomous Driving* (arXiv:2109.12764; Issue arXiv:2109.12764). arXiv. http://arxiv.org/abs/2109.12764

Wang, Q., Xie, J., Zuo, W., Zhang, L., & Li, P. (2020). Deep CNNs Meet Global Covariance Pooling: Better Representation and Generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/TPAMI.2020.2974833

Zhong, Y., Li, H., & Dai, Y. (2018). *Open-World Stereo Video Matching with Deep RNN* (arXiv:1808.03959). arXiv. https://doi.org/10.48550/arXiv.1808.03959

Mancini, M., Costante, G., Valigi, P., & Ciarfuglia, T. A. (2016). *Fast Robust Monocular Depth Estimation for Obstacle Detection with Fully Convolutional Networks* (arXiv:1607.06349). arXiv. https://doi.org/10.48550/arXiv.1607.06349

Phan, H., Koch, P., Katzberg, F., Maass, M., Mazur, R., & Mertins, A. (2017). *Audio Scene Classification with Deep Recurrent Neural Networks* (arXiv:1703.04770). arXiv. http://arxiv.org/abs/1703.04770

Lee, S., Kim, J., Yoon, J. S., Shin, S., Bailo, O., Kim, N., Lee, T.-H., Hong, H. S., Han, S.-H., & Kweon, I. S. (2017). *VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition* (arXiv:1710.06288). arXiv. http://arxiv.org/abs/1710.06288

Xu, W., Wei, J., Dolan, J. M., Zhao, H., & Zha, H. (2012). A real-time motion planner with trajectory optimization for autonomous vehicles. 2012 IEEE International Conference on Robotics and Automation, 2061–2067. https://doi.org/10.1109/ICRA.2012.6225063

Sallab, A. E., Abdou, M., Perot, E., & Yogamani, S. (2016). *End-to-End Deep Reinforcement Learning for Lane Keeping Assist* (arXiv:1612.04340). arXiv. https://doi.org/10.48550/arXiv.1612.04340

Wang, G., Hu, J., Li, Z., & Li, L. (2019). *Cooperative Lane Changing via Deep Reinforcement Learning* (arXiv:1906.08662). arXiv. https://doi.org/10.48550/arXiv.1906.08662

Elallid, B. B., Benamar, N., Hafid, A. S., Rachidi, T., & Mrani, N. (2022). A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving. Journal of King Saud University - Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2022.03.013

Ni, J., Chen, Y., Chen, Y., Zhu, J., Ali, D., & Cao, W. (2020). A Survey on Theories and Applications for Self-Driving Cars Based on Deep Learning Methods. Applied Sciences, 10. https://doi.org/10.3390/app10082749