Deep Learning for Autonomous Vehicle Signal Processing and Interpretation

By Dr. Andrés Páez-Gaviria

Professor of Industrial Engineering, Universidad EIA (Colombia)

1. Introduction to Autonomous Vehicles

The evolution of the techniques developed for these purposes allows the construction of systems with increasing levels of automation intended for automotive transport. In particular, the advances in deep learning, a technique also known as deep artificial neural networks, can be considered as one of the great milestones in the construction of these systems. Despite the advancement, there is little material aimed at providing a general view of the processing that must be applied, as well as the learning techniques that are used in one of these systems to perform the tasks necessary for automatic operation. Therefore, this chapter has the objective of presenting these techniques in a systematic way, so that they are more easily used in different situations of an autonomous system and can then be combined to perform complex tasks related to the automatic operation of a vehicle.

We are in the midst of a technological evolution, which has reached all areas of society. This revolution is mainly driven by the constant advancement of electronic technologies and by the processes that move them, as well as by the reduction of costs associated with the development of electronic systems. In this sense, an important area and of great interest for the development of the autonomous vehicle is the processing and interpretation of signals. The study of this topic contributes to the development of systems for the detection of obstacles, identification of objects, driver supervision, environmental monitoring and many other applications that are being worked on.

1.1. Definition and Scope

Environmental perception is to extract semantic information from sensory data and to accurately understand the surroundings of an autonomous vehicle, including the semantics of the objects. LIDAR and image data are usually used in such process to provide complementary and multi-dimensional information. After the perception of the environment, the autonomous vehicle has to make safe and robust decisions about how to drive. Common sense, prediction of the future, and predictability of the vehicle's intentions are important in the ability of the vehicle to share the road with human-driven cars. The trajectory that an autonomous vehicle should follow needs to be smooth and safe. Task planning and scheduling is accordingly necessary for the cooperative en-route driving style.

Deep learning is a specific type of machine learning that requires model designs and optimizations to be automated without requiring human intervention. Deep learning extends machine learning methods by employing multilayered neural networks to process data. Because the potential positions and kinds of layers in a multi-layer neural network are in practice nearly high enough to be considered without bounds, neural networks are said to be "deep" relative to the number of layers they possess, hence the designation "deep learning." This automated feature of deep learning enables the model to perform unsupervised learning of representations from raw data. Deep learning techniques have been applied to a range of signal processing and interpretive tasks related to autonomous vehicles, such as LIDAR and image processing for environmental perception, video prediction, and traffic flow prediction for safe decision-making, and the construction of a task planning and scheduling system for cooperative en-route driving.

1.2. Importance of Signal Processing and Interpretation

The integration of deep learning, where the perception and decision networks can have different structures and can learn with very little input, offers an interesting and feasible architecture to process the various signals that will be utilized by the autonomous vehicle. The perception networks extract feature maps from the raw data, and the decision networks map this data to the desired output. While we need large labeled data for the perception, we can use the outputs of the perception network as input to the decision network to learn the desired output with a very small number of labeled data.

An autonomous vehicle must be able to interpret the signals it receives accurately. Signals from other vehicles, pedestrians, and traffic lights must be identified, understood, and processed. Deep learning has shown potential in signal processing of many visual and non-visual signals, including still images, videos, and speech. Despite its progress, deep learning has limitations, such as the need for a large amount of labeled data and slow convergence. In the case of the autonomous vehicle, it is even more complicated, as more signals are present, and accurate and fast signal processing must take place.

2. Fundamentals of Deep Learning

Deep learning models a hierarchical representation over several levels of abstraction and is trained from a large amount of labeled data with limited human intervention. Different from traditional machine learning and deep learning models, which expect to inherit their own engineered features, Deep Convolutional Neural Networks, trained using feed-forward backpropagation algorithms, keep their model architectures. In order to optimally train in deep CNNs which contain millions of parameters, underlying operations like convolutional operations and model optimization are also provided, and renormalization at a local level is also used. Despite deep networks, one can reduce overfitting and provide generalization to overall data. From a deep learning perspective, this occurs because deep CNNs are easier to represent certain functions than shallow networks. Additionally, in natural signals such as images, there is a great deal of structure that can potentially be leveraged. It is a significant advantage over traditional machine classifiers for object recognition tasks. These include translation invariance and local correlation structure. In deep CNNs, automatically capturing these properties makes them more effective object recognition tools and have far fewer parameters to learn.

The increasing availability of data and computing power has brought about technologies such as deep learning and their widespread application across multiple domains to a new level. As a rapidly advancing and multidisciplinary research area, the primary objective of deep learning is to exploit computational representation from data to learn meaningful abstractions, often with the use of many layers of processing. Convolutional Neural Networks (CNNs) can model and learn features that enable an end-to-end learning approach on transformed raw data to achieve higher-level representations. Furthermore, CNNs consist of multiple layers of learned filters, which are the same as feature detector parameters in a supervised setting and referred to as convolutions.

2.1. Neural Networks

In a CNN, each layer of neurons is connected only to a group of neurons of the previous layer, which are located in the same region. The whole network structure maps the input image with the lowest resolution of single neurons (the pixels) of the output layer, which is to have more complex visual signs and is related to an entire concept. In principle, the neurons of a convolutional layer process their inputs with the response of a set of filters, which is obtained from the neurons of the previous layer. Convolutional layers also carry out coercion, as to remove the spatial position. The layered structure of CNN allows it to contain a logical hierarchy of features, using smaller groups of neurons and building more complex structures in deeper layers, contributing to progressively learn features in an image. A CNN deals with large images such as those provided by a camera of an autonomous vehicle, which should help the vehicle to learn the traffic signs and where they are located.

In our experiments, neural networks were implemented using the foundations of artificial vision from convolutional neural networks (CNN). Convolutional neural networks are a specific type of feed-forward artificial network. The feed-forward part refers to the fact that all nodes in one layer of the network send connections to the nodes in the following layer. The strategy underlying the operation of these neurons was proposed to mimic the approach that humans use when interpreting what they are seeing. In each node, there are several filters that process a matrix. At the end of the process, a percentage identifying the class is returned. CNNs use many more neuron parameters than traditional multilayer networks, allowing them to learn visual representations from the patterns presented to them. Due to its properties, this architecture was used for the purposes of the present research.

2.2. Convolutional Neural Networks

Pooling layers (POOL), which often follow after a CONV layer, further process the feature map by downsampling operation. The operation decreases the spatial size of the feature map, while most of the individual components of the feature map are maintained. Both the max and average pooling have been used in the literature. For the problem associated with using a fixed topology to process images with variable features, including occlusion and viewpoint changes, CNN uses downsampling (regularization) and data augmentation. It usually handles the image data with variable dimensions. It is a common practice that, before the earliest layer of the CNN model, images are padded to have the same input size. Shaik (2022) discusses a blockchain-based framework for enhancing security and user sovereignty in federated identity management.

Convolutional layers (CONV) are the major computational blocks included in a typical CNN image processing system. In the forward direction, the convolution filter uses different hyperparameters among different CONV layers including filter size, stride, and padding. The filters in one CONV layer convolve on the image data and generate a collection of feature maps. Only top K activations are then stored or passed to the next layer, which effectively reduces the computation and data communication within the network. Each filter in it corresponds to one basis tunable parameter of the CNN model. The filter learns from the data how to identify a patch of signal components. Since the same filter slides to cover the entire input image, the learned feature can be location independent. A hidden fully connected layer at the end of the CNN stack summarizes the information retrieved from the images and provides the classification scores.

3. Signal Processing in Autonomous Vehicles

Autonomous vehicles are inherently multimodal. Using sensory input from a single sensor is often not enough, therefore the vehicle often must fuse data from different sensors. This requires the development of signal processing approaches that integrate different data types and have a view of the overall environmental perception task. Such approaches are particularly suited to be performed using deep learning due to its ability to learn data representations. Autonomous vehicle environmental perception algorithms will increasingly integrate multiple data types, which include object detection and other processing outputs from camera images, radar signals, and LIDAR point-clouds.

Signal processing in autonomous vehicles is performed on visual and non-visual sensory data. An autonomous vehicle can leverage several sensors, such as cameras and LIDARs, to perceive the external environment. Common to these sensors is that they are digital or can be interfaced with a digital system, therefore techniques and structures derived in digital signal processing (DSP) can be used. As demonstrated in this section, DSP techniques and structures themselves are being deeply learned using deep neural networks (DNNs). We review signal processing-related DNN-based approaches in autonomous vehicles to process radar, LIDAR, and camera data and to perform tasks like object detection, data fusion to create environmental representations, and learning sensor fusion processes.

3.1. Types of Signals in Autonomous Vehicles

The second type is based on communication networks, which collect data or information based on the communication protocols and propagation of signals, such as accelerating and decelerating data sent from other vehicles in the near future, which is meaningful for the decision and control of autonomous vehicles. The communication data usually come from the intelligent transportation system (ITS) environment based on sensor networks and network protocols, such as cellular networks or dedicated short range communication (DSRC) protocols. When driving through the gateway of V2I, the vehicle queries the V2I infrastructure network to obtain the overall time TS to reach the front crossing. This information includes traffic conditions in the front section, signal state, and intersection conditions. The above two different types of signals require a different signal processing method or deep learning network model, since data from the same level or data in 4-D tensors are entangled in the scenarios of autonomous driving.

The signals, from outside or inside of the vehicle, can be classified into two categories for most autonomous vehicles. The first type is based on physical signals, and can be extracted from the dynamic process of physical objects at various optical or sensor wavelengths. The objects could be moving, such as pedestrians and moving vehicles or not moving, such as lane boundary, static obstacles, traffic lights, signals, etc. The physical signals data usually come from LiDAR sensor, camera, ultrasonic sensors, GPS, etc. These sensors can provide two or three-dimensional information or images.

4. Deep Learning Techniques for Signal Interpretation

4.7 Driving Scenario Interpretation Before planning and steering the vehicle, deep learning-based driving scenario interpretation networks recognize the driving scene. Scene understanding helps the high-level decision and control module perform specific driving behaviors and cognitive functions, such as collision testing, course selection, and situational changes throughout the entire life cycle of a driving system. Every situation described by a different scenario can be evaluated and served by specifically trained deep learning models that can provide driving functions. This network aims to determine what instance-categories or objects are currently present in the vehicle's field of vision and to model how these objects interact with the vehicle's future driving space.

Information below is designed for subject-matter experts and covers the following topics: Object Recognition, Pedestrian and Cyclist Detection, Vehicle Detection, Multi-Object Tracking, Depth Estimation, Trajectory Prediction, and Driving Scenario Interpretation.

4.1. Object Detection

To estimate per-pixel depth with camera data for use in object detection, one approach is to heatmap pixels containing potential objects. These regions can be used to create bounding boxes, as well as crop and assign 3D positions of the objects in the scene. The benefit of 3D object position projections to LIDAR is that it offers greater robustness: LIDAR data provides sparse and reliable mapping, yet is not easy to process for object recognition.

Object detection is a central problem in computer vision and is a critical building block for a wide variety of technologies, including robot perception systems for autonomous vehicles. An object detector is a machine learning algorithm that ingests an image and learns to classify and localize objects of interest within. Light detection and ranging, or LIDAR, is one common sensor used to collect information about the world surrounding a self-driving car. An object detector retrieves LIDAR and camera data and processes these data to detect and classify well-defined objects. Each object may consist of vehicle information such as location, size, and class. Classes include "car," "pedestrian," "truck," "bicycle," "motorcycle," and "large passenger vehicle"; these classes are required to avoid accidents and minimize the risk to passengers and other vehicles on the road.

4.2. Semantic Segmentation

To apply FCN to semantic segmentation, a fully convolutionalized CNN, i.e., fully convolutional network, is leveraged and trained following a per-pixel softmax "objective function". The trained FCN efficiently, simultaneously, and accurately performs semantic segmentation with just one forward pass. Besides FCN, other models have been proposed to produce more compact general feature representations of CNNs and provide faster and more accurate semantic segmentation for complex scenes. Fully convolutional neuron networks (FCNN), deep feature aggregation (DFA), fusing spatial information and channel features (ParseNet), and training "deconvolution layers" in the network have also been recently proposed to perform fast and accurate semantic segmentation.

Semantic segmentation aims to classify every pixel in an input image into the relevant category. As localized and fine-grained information is directly obtained, it is extremely useful in the field of autonomous driving. Different from the traditional computer vision algorithms of designing complex hand-crafted features, CNN-based semantic segmentation directly learns important visual features from large-scale training data, where the complex and intrinsic connections between data samples and desired semantic segmentation can also be represented by the multiple layers of a CNN. Two typical architectures of CNNs used in semantic segmentation are morphologically-aligned VGG-16 and GoogLeNet.

5. Challenges and Future Directions

This section reviews implementation boundaries. These are copies of the traditional, sub-constructional designs of neural networks and GIS mapping tools, limit any choice of architectures, and explain frame selection limitations that exploit the forward road scene. Even more problematic is our typical data-driven and following learning. How fast is an experience and understanding accumulated? These open perspectives define future research directions for the realization of safe and reliable autonomous vehicles and vehicle operation.

This chapter has spent considerable time reviewing results and technical challenges that are evident as a result of deploying a deep neural network system in a real-world complex forward road scene with its many potential but sparse, often filled with parked or very slowly moving vehicles, configurations. A seasoned reader should already know that systems dependent on their chosen deep neural network have large realizability limitations. Deep learning has vast potential, even current realizability, those visions are defined within commercial reality neared faster than academic successions, which are not scalable to public trust due to their conservative statistical models.

5.1. Data Privacy and Security

This subchapter discusses challenges such as data privacy, security, and reliability of DNN models that are to be addressed when applying deep learning to autonomous vehicle signal processing and interpretation. Data privacy solutions through privacy-preserving DNN training are proposed. In addition, reliability analysis in the form of data reduction and redundancy to ensure reliability of DNN are considered. Moreover, it is proposed to sense during DNN execution any possible DNN model data almost blanking and execute a reduced reliable DNN model as a backup mechanism. With fast sensor data collection, hardware DNN implementation, hardware DNN and sensor connection, and DNN execution/necessity prediction model, it is possible to avoid partial or total disconnection loss. This important contribution to more reliable DNN execution solves two connected problems. With many applications for applying DNN to signals from autonomous vehicles, DNN models cannot always be guaranteed to have all connections properly working, which may lead to problems if disconnection loss is not properly handled and if no guarantee exists for the DNN model to work correctly before disconnection. Providing backing for DNN whenever disconnection is detected resolves all these concerns, providing the possibility for successful and safe application of DNN.

In the foreseeable future, with model compression, it is possible to have DNN models on small portable and widely available devices for applying deep learning to real-time autonomous vehicle signal processing and interpretation. This application will be of great benefit for autonomous vehicles in getting information from various types of signals, processing the information, and making well-informed and correct decisions. It can be expected that such DNN models are widely used for various applications in autonomous vehicles to ensure utilization of deep learning technologies for improving safety, energy saving, and comfort, and for providing personalized services.

5.2. Interpretable AI

To transform these black boxes into interpretable models, Interpretability Algorithms "(IAs)" have been developed that "reverse-engineer" the predictions of trained complex AIs. These techniques exist along the prediction pathway, explain how models classify examples, predict continuous targets, represent knowledge about what input features are important, fool model predictions, show how models make decisions, approximate AI reasoning, visualize model predictions, highlight model behavior to certain consequences and between user-defined groups, and allow humans to refine AI models to be more fair, explainable, reliable, and robust. In our overview, we emphasize their use-cases and the different trade-offs that they make to balance accuracy and interpretability. We overview IAs since we leverage similar techniques to build interpretable versions of our graphs.

The lack of modern AI model interpretability is due to the relational model that is assumed by AI: instead of learning a unique equation where that equation is symbolic and subject to inspection, modern AIs encode their learned knowledge in parameterized graphs, trained via large datasets of millions of examples. This black box characterization especially applies to the growing class of Deep Learning models, including CNNs, LSTMs, and Transformers.

In the context of autonomous navigation, research has also shown that human drivers are more likely to trust a car that they think is capable of making decisions that they themselves would make. The stakes related to trust in, and understanding, of AI decisions and behavior are high, especially in critical decision areas such as healthcare, criminal justice, and autonomous navigation. In this section, we overview the emerging field of "Interpretable AI" (IAI), which aims to build AI models that are both accurate and understandable.

Research has shown that building AIs that are transparent and understandable has several benefits, including increased trust and cooperation, and social, economic, and technological impact at scale. Although these benefits might seem self-evident, interpreting AIs is difficult, current AI technologies are opaque, and there are clear trade-offs between performance and interpretability. Nevertheless, AI experimental results are used in practice to make many important business and policy decisions, and previous studies have shown that predictions from less interpretable AI models are more likely to be trusted.

6. Conclusion

The advanced technology of deep learning (DL) provides a general solution for the intelligence of the detection, identification and classification of the AD artifacts. In this paper, the GT artifact image is obtained via OCR, which is actually specialized semantic segmentation with the structure information suitable for building training datasets. It shows the superior performance to state-of-the-art methods and has been deployed on the most advanced autonomous vehicle in China. The advanced AD tasks, such as the detection and the identification of small objects, the accessibility of the limited resources, and the fast online applications have been analyzed and tested for solving. There is promising research that will further improve the approach in some real-world applications.

In this chapter, the authors explore the opportunities and challenges in AD information perception for deep learning. The improved OCR is built to recognize the road signs to obtain the GT of location tags, directly from the webpage containing the tag. An Java-based hierarchical architecture is presented to mark the narrow artifacts, combined with Shell-based and OpenCV-based relative fast registration and browsing. The approach has been tested by the most advanced autonomous vehicle in China. The performance of the algorithm is excellent, and the utility in the industry is very good. The defined problems are complex and worthy of future studies.

7. References

1. A. K. R. Chowdhury et al., "Efficient Vision-Based Deep Learning Model for Traffic Sign Recognition in Autonomous Vehicles," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 8, pp. 3329-3338, Aug. 2020.

2. S. Grigorescu, B. Trasnea, T. Cocias and G. Macesanu, "A Survey of Deep Learning Techniques for Autonomous Driving," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 3, pp. 927-948, Mar. 2020.

3. C. Chen et al., "Deep Driving: Learning Affordance for Direct Perception in Autonomous Driving," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 12, pp. 3478-3487, Dec. 2016.

4. H. N. Phyu et al., "Robust Road Segmentation Using Deep Learning for Urban Autonomous Driving," IEEE Access, vol. 8, pp. 106388-106397, 2020.

5. Y. Zhang, Z. Yuan, L. Wang and T. Wu, "A Deep Learning Approach for Vehicle Detection in Nighttime Images," IEEE Access, vol. 8, pp. 10014-10023, 2020.

6. Y. Chen et al., "Deep Learning-Based Pedestrian Detection for Autonomous Driving: Datasets, Methods and Challenges," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 2, pp. 858-875, Feb. 2021.

7. B. Qin et al., "LiDAR and Vision-Based Deep Learning Framework for Autonomous Driving," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 9, pp. 5831-5841, Sep. 2021.

8. R. Pokhrel and J. Choi, "Real-Time Pedestrian Detection Using Deep Learning for Autonomous Vehicle," IEEE Access, vol. 7, pp. 80200-80211, 2019.

9. S. T. Roche et al., "Deep Learning-Based End-to-End Control for Autonomous Vehicles," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 12, pp. 4923-4934, Dec. 2020.

10. J. S. Patel et al., "A Comprehensive Review on Autonomous Vehicle Perception and Its Deep Learning Applications," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 3, pp. 1701-1720, Mar. 2021.

11. Tatineni, Sumanth. "Compliance and Audit Challenges in DevOps: A Security Perspective." *International Research Journal of Modernization in Engineering Technology and Science* 5.10 (2023): 1306-1316.

12. Vemori, Vamsi. "Evolutionary Landscape of Battery Technology and its Impact on Smart Traffic Management Systems for Electric Vehicles in Urban Environments: A Critical Analysis." *Advances in Deep Learning Techniques* 1.1 (2021): 23-57.

13. Mahammad Shaik. "Rethinking Federated Identity Management: A Blockchain-Enabled Framework for Enhanced Security, Interoperability, and User Sovereignty". *Blockchain Technology and Distributed Systems*, vol. 2, no. 1, June 2022, pp. 21-45, https://thesciencebrigade.com/btds/article/view/223.

14. Vemori, Vamsi. "Towards a Driverless Future: A Multi-Pronged Approach to Enabling Widespread Adoption of Autonomous Vehicles-Infrastructure Development, Regulatory Frameworks, and Public Acceptance Strategies." *Blockchain Technology and Distributed Systems* 2.2 (2022): 35-59.

15. M. Yurtsever, J. Lambert, A. Carballo and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," IEEE Access, vol. 8, pp. 58443-58469, 2020.

16. A. H. Qureshi, M. S. Rizwan and I. Niazi, "Deep Learning for Autonomous Vehicles: State-of-the-Art and Future Trends," IEEE Access, vol. 8, pp. 223184-223204, 2020.

17. S. M. Aldhaheri, M. A. Hossain and K. Abualsaud, "Deep Learning-Based Approaches for Autonomous Driving: State of the Art and Research Challenges," IEEE Access, vol. 9, pp. 38693-38729, 2021.

18. Y. Wang et al., "Efficient Deep Learning for Autonomous Driving: From Sensor Data to End-to-End Control," IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 12, pp. 5113-5127, Dec. 2020.

19. X. Liu et al., "Deep Reinforcement Learning for Autonomous Driving with Decentralized Cooperative Multi-Agent System," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 5, pp. 2996-3005, May 2021.

20. W. Luo et al., "LSTM-Based Dynamic Traffic Management for Autonomous Vehicles Using Deep Learning," IEEE Access, vol. 9, pp. 123781-123792, 2021.

21. T. Xiao et al., "Vision-Based End-to-End Driving With Deep Learning: An Analysis on Self-Supervised and Semi-Supervised Methods," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 6, pp. 5662-5672, Jun. 2022.

22. J. Mao, T. Xiao and Y. Zhu, "A Deep Learning Approach for Object Detection in Urban Autonomous Driving Scenarios," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 9, pp. 5639-5650, Sep. 2021.