Deep Learning for Autonomous Vehicle Vision Systems

By Dr. Yang Liu

Associate Professor of Computer Science, Shanghai Jiao Tong University, China

1. Introduction to Autonomous Vehicles

An autonomous vehicle, or self-driving car, refers to a car that can drive itself without human intervention. The development of autonomous vehicles has generated significant interest from tech companies that are developing autonomous electric cars. Companies that have their own autonomous car projects comprise Uber, and notable mentions from the automotive industry are Daimler AG and BMW with their joint project, as well as Ford Motor Company and the Volkswagen Group who also have separate autonomous car projects [1]. The enabling technologies of autonomous vehicles are computer vision, which is the field of teaching a computer how to interpret and understand the visual world, and machine learning, which is the act of programming a computer to learn or to adapt from the experiences [2].

Deep learning has had a significant impact in the automotive industry, particularly in computer vision for autonomous vehicles. Several use cases in the automotive industry stand to benefit significantly from deep learning techniques. These include image-based vehicle and lane detection; autonomous driving; data organization to identify networks, ECU classification and lane markings; social media analytics to classify road travel images; processing sensor data such as CANBUS, cloud point, and radar data to analyze car control, sensor data, and object-related issues of re-simulation; robotics-based vehicle assistant implementation; and connected vehicle services for customer insights, vehicle health, and related offerings [3].

1.1. History of Autonomous Vehicles

[4] Self-driving technologies have made a considerable progress in the past decade. Traditional methods consist of smart sensor fusion, while recent architectures focus on deep learning, thanks to the increase in deep learning algorithm performance. As a result, modern development in the self-driving car devolves to a minor extent on intuition. It is not atypical

for any state-of-the-art model to rely on an AI technique to extract salient features from the raw input and, subse-quently, to learn a mapping from sensory input to actuation output. The innovative front of AI-driven development is represented by increasingly complex end-to-end models that operate without the use of hard programmatic low-level components. Deep learning technology has a rapid development, this is partly due to the research of numerous inherent limitations and potentials. There are unsolved issues such as robustness to enhancement techniques, which directly influences the design of vision-based autonomous vehicle systems adequacy to tackle the issue that exist in the ambient real-world conditions.[5] The development of sensor-based systems has its roots in the early 90s,however modern systems are characterized by the pres-ence of artificial intelligence (AI) operations. The revolution of the last decade has seen the concomitant emergence of AI and big data management. In the field of driving, it paved the way to the transition from the driver assistance phase to the pure autonomy. Lane keeping assistance, automatic emergency braking, augmented reality are examples of the usefulness of AI applications in the autonomous vehicle framework. Learning from the success and failures in recent years, future development will witness a technological mam-borine; on one hand, new models, sensors and observation methodologies will be introduced, on the other hand new engineering strategies are to be developed to validate potential network struc-tures. It is possible to state that deep learning applications will conceivably be used for each new sensor or working scenario.

1.2. Importance of Vision Systems

In recent years, comprehensive autonomous driving technologies (ADTs) based on sensor-driven vision systems have also been propelled by direct perception (DP).^ Accordingly, DVP [direct visual perception] is employed to study direct mapping and end-to-end methods that can directly feed raw sensor inputs into route planning, obviating the need for distinguishing various perception and mapping technologies. [6] summarizes and further studies other ADT algorithms based on DP over the past few years. Previous generations of automobiles have been equipped with a variety of advanced driver-assistance systems capable of providing lane-keeping assistance, automatic parking assistance, automatic cruise control, emergency braking, active lane keeping, and other functions. Traditional rule-based and machine-learning methods have continued to serve as the main focus of early generations of vehicle vision applications. Nonetheless, deep learning (DL) methods are transforming the automotive industry. These advanced machine-learning techniques are beginning to be widely used in newly emerging automotive functions, including simulation software for scenario generation, traffic forecasting, and signal recognition using cameras.

The significance of vision systems in automated vehicles is underlined in many articles. [7] have recently discussed the importance of multi-modal sensory inputs, including light detection and ranging (LiDAR) and cameras. Developed through artificial intelligence (AI), these multi-modal vision-based perception (VBP) technologies have been integrated into vehicles to effectively process multi-modal information in the development of self-driving cars. Specifically, recent studies have demonstrated that VBP methods have become the main driving force behind motion tasks such as simulative analysis, self-driving cars, robotic vision, crowd motion prediction and re-identification, street scene flow processing, and multi-loop joint optimization of the pedestrian trajectory prediction and vehicle decision-making process. In addition, the rapid growth of visual processing power and widespread application of spatial-temporal graph-based convolutional neural networks have led to the expansion of various decision-making outputs beyond the original planning range, including acceleration, lane-keeping, and lane-following control.

2. Fundamentals of Deep Learning

In Dense Driver the real-valued distances are regressed for cars, whereas an occupancy grid (OG) of size 10 × 10 is discretely learned for collision detection. This OG is inspired by the target grid idea in 3D ADD-NN, and similar to the 2D Nav OG. OG's have the advantage of interpreting confident net classifications directly in terms of distances which are often the most important for safety critical decisions. We use the fusion GRU in encoder and decoder, whereas we used BiRNNs in that system. This contribution demonstrates that the fusion by GRU networks has significantly lower computational requirements and provides better generalization ability but with similar performance to BiRNNs. [1]

The ability of vision is crucial for automotive robotics. Cars equipped with vision sensors are required to make rob by decisions like Human drivers. This paper aims at providing a comprehensive survey of the recent advances in applying Deep Learning to Vision-based Autonomous Driving with a probabilistic vision system outlier rejector. It focuses on two important vision tasks in autonomous driving, i.e., object detection and scene understanding. Tackling the low-level vision challenges, a number of methods have been proposed for the robust detection of automotive objects from a vehicle in traffic. For scene understanding, a critical vision task for autonomous driving, this paper rigorous iterates over four crucial nodes in a Deep Learning enabled pipeline for pioneering reserach on image-based 3D object detection and lane detection. This is planned as an on-going survey, providing the most recent work in the domain of Deep Learning and Autonomous Driving. The associated GitHub repository contains further recent work, not in allowed project time-frames. [7]

2.1. Neural Networks

The purpose of navigation for aerial robots is to use navigation marks in the real world to locate navigation landmarks and choose the best path based on the navigation marks to arrive at the specified destination. Traditional navigation recognizes roadway signs and other recognition information through pre-imposed visual landmarks. This method has too many uncertainties and cannot guarantee its robustness. In the past few years, it has been used more widely. The related research field has been applied to plant protection and green pace in maize fields. The development of deep-learning-based dynamics technology has further increased the reliability of navigation technology, and the convolutional neural network-based body in body has been developed in this work. The recognition algorithm realizes autonomous maize column navigation by identifying the image taken from the front camera through the body wall. [6]

Most people have heard of "deep learning" and "neural networks" and are aware of the important problems they have been applied to in recent years, such as computer classification, speech recognition, and chemical molecule structures identification. Today, this computer vision-related technology is also closely related to vehicles—including Unmanned Aerial Vehicles (UAVs), Unmanned Ground Vehicles (UGVs), Adapted Ground Vehicles (AGVs), and hybrid aircraft. This article focuses primarily on autonomous driving and UAV, as shown in Figures [8]. In autonomous aerial robots, the navigation and path planning of the robot itself and the processing and interpretation of the visual information of the image formed by the camera are two key technologies. The mainstream navigation scheme of the robot is based on GPS and INS (Telerik inertial navigation system), but sensor fusion of radar and other sensors is used at the same time. The general real-time images are obtained by fixed-interval scanning. The pixel points in the picture represent the characteristics of the mapped scene in the form of gray-level pixels or RGB three-component color pixels. [9].

2.2. Convolutional Neural Networks

The success of CNNs for visual perception has motivated many research initiatives which attempt to exploit its discriminative power with respect to other higher level tasks in visual perception like vehicle to vehicle, vehicle to pedestrian, vehicle to bicycle, and vehicle-to-motorcycle collision prediction. Approaches based on deep learning like convolutional LSTM network, gated geometry CNN, ecto-CNN, MoVi-Net, and ConSeLoNet-highlight the increasing trend towards online deep learning-based approaches owing effective methods to capture long distance dependencies in driving scenarios with the ability to hierarchical positional dependencies over time [10].

Convolutional Neural Networks (CNNs) have revolutionized image processing in the deep learning era because of their ability to perform efficient and meaningful feature extraction [11]. In the case of autonomous driving, there has been substantial improvement in the detection, tracking and classification of various objects encountered in the driving scenario [12]. CNNs have the inbuilt capability to automatically learn the features which are necessary for classification, recognition, and detection of arbitrary objects. Because of their ability to learn the feature representations directly from the raw image data, they have become a clear choice for autonomous driving applications.

3. Deep Learning Applications in Autonomous Vehicles

In contrast to traditional object detection models like YOLO, Faster-RCNN, etc., recent deep learning architectures have demonstrated great improvements in accuracy and computational efficiency. For example, YOLO yields 68% average precision at 81 fps, outperforming all existing practical deep learning systems, and 57.9% mAP on the first frame of a video scene. DL has been employed to improve operations and target tracking in object detection, object type recognition, semantic segmentation, instance segmentation, etc [13]. OpenCV version 4.5.3 has included the pretrained YOLO models to detect around 100 types of objects, and this updated version reduced the latency over YOLOv2, YOLOv3, and YOLO-Tiny. LSA32/64/512 classification models have promised a real-time object classification performance with 55.1%, 60.7%, and 40.1% mAP in eight classes of the COCO dataset. Deep learning has improved feature extraction, classification (or regression), and fusion in object detection. YOLOv2 has enhanced the detection of small objects by making some modifications such as batch normalization, high-resolution classifier, and finer-grained features. CarderNet has been prepared to learn scale-invariant derivatives to handle highly overlapped multiple objects. Thing-based models outperform the traditional anchor-based/deformable-based detectors. PANet has enhanced the performance of MaskRCNN by making it independent of image resolution. Deep learning model EfficientDet has superimposed existing detectors with 13% mAP at real-time detection speed.

The application of AI in the automotive industry, particularly in autonomous vehicles (AVs), has garnered significant attention. An array of sensors, such as cameras, LiDAR, RADAR, etc., are used for the perception of AVs [[6]]. The sensor fusion algorithm is applied to combine the data from these diverse sensors to attain high-precision perception. Deep learning (DL) has the ability to handle the data from these sensors effectively and efficiently. It has enabled companies to achieve close-to-human level performance in a variety of tasks, such as object detection and recognition, end-to-end learning, semantic segmentation, Simultaneous Localization and Mapping (SLAM), etc.

3.1. Object Detection

One way to evaluate the method proposed for autonomous vehicles using convolutional neural networks and to improve the direction for future research and technological innovation is to involve cybersecurity issues in the context of autonomous vehicles. Traffic-sign detection and recognition – a key task of automotive vision systems – are less precise when dealing with adversarial examples [14]. Deep models for this specific object detection task lack robustness against extremely small attacks from an adversarial example. Therefore, building an understanding of the "behavior" of deep learned systems in this kind of "attack" could help us enhance the robustness of perception models for AV sensors.

Deep learning has revolutionized object detection in the automotive industry especially in the development of self-driving or autonomous cars. This is a key application of deep learning for Computer Vision in automotive industry. Traditionally, expensive sensors like LIDAR had been required to perceive the environment and specifically recognize traffic signs, traffic lights and other cars [3]. In addition to the typically lower capture resolution compared to high-quality camera sensors, LIDAR sensors also have other intrinsic limitations. To address these issues, there is an increasing trend to remove the need for such expensive sensors. As a result, deep learning has been used to replace LIDAR sensors with cameras by improving the object detection problem in the way of "detection" and "segmentation" challenges [2]. Different techniques (based on deep learning approaches) have had a remarkable impact on improving object detection in the automotive industry.

3.2. Lane Detection

Convolutional neural networks have demonstrated remarkable performance for real-time systems as they learn complex hierarchal features and end-to-end mapping directly from raw data without any feature engineering requirements [15]. In the last few years, based on powerful deep learning and especially CNNs, the research of the lane detection system received great attention. Keypoint detection means capturing an edge of lane marking to find distances between cameras and lane. By doing so, the distance between cameras and lanes is calculated from points on the left and right edges of the lane and matching its equation with camera pixel points gives the real distance in pixel corresponding to where pixels fall on that distance. Deep learning approaches in lane detection models, in general, outperformed traditional lane detection algorithms. Although using deep learning and end-to-end deep models would have good performance and robust systems, they have some drawbacks. Most trained deep models expect the similar or same input images at the inference time correspondingly to the training environment. Also, due to lack of transparency of training and learning processes, as well as larger model sizes, deep learning-based lane detection systems are not advisable for real-time systems.

Lane detection refers to the process of detecting lanes present on the road or a traffic scene. Lane detection is a crucial component in ADAS (advanced driver assistance systems) and autonomous vehicle systems [16]. A lane or lane marking is defined as a road's edge, or its lane center, or lane markings, white or yellow lane, pedestrian crossing, parking lane line, stop line, etc. Detection and extraction of lanes provide critical information about the roadway and can be used for multiple purposes like driver assistance, traffic analysis and understanding the context of the vehicle. Lane detection has been extensively researched by various researchers due to its importance in various applications. The lane detection approaches can be roughly categorized into model-based, appearance-based and deep learning-based. Model-based methods are those that are based on a priori information about the road in the form of a model. Model-based methods have shown good results and generalization capability for controlled driving context in good environmental conditions. Appearance-based methods do not require explicit models. These methods implicitly learn the important features from the image data. Appearance-based methods have shown promising results under specific environmental conditions such as controlled or well-visible driving context. However, these systems show the limitations in generalization capabilities [17]. Due to the consequences of incorrect lane line detection and positioning, there is a great need for more robust and precise lane finding models, especially under cluttered and limited visibility environmental conditions.

4. Challenges and Limitations

Other than these major challenges, image intake is prone to errors. Due to the problems mentioned under image quality issues, the perception part of the sensor fusion module needs to be particularly cautious of the image quality problems. Also, the exposure times in sensors generally influences how well the system can interpret colors or the image perception. The approach to a sensor fusion system with images as a perceptual capability can be much easier to integrate into a system as opposed to other sensor types. It can also ease the efforts in defining a system architecture where the safe fallbacks can be seamlessly deployed when necessary to ensure appropriate vehicle responsiveness [18].

Deep learning with cameras is the most commonly implemented method for visual perception with autonomous vehicles today [19]. The huge strides in quality of recognizing objects in images via classification models has pivoted vision research into using object detection-based models for autonomous vehicles. Following the recognition of objects, location of the detected objects with respect to the vehicle is needed, to use them for planning purposes. Using classical methods for perceiving the environment either affects the time and precision of the resulting models and the inferences, or require highly optimized models which are difficult to obtain [20].

4.1. Adverse Weather Conditions

How an algorithm deals with situations in the real world, such as when the scenario is dark, there is interference from the sun, low visibility due to fog or a strong storm with rain or snow, are examples of typical issues in the context of adverse weather conditions. Changing weather conditions are known to introduce significant distortions, including saturation, occlusions and motion blur, that can challenge robot vision systems, such as overall performance degradation, sensor-to-sensor cross-impacts and consequently decreased safety [21]. Having access to up-to-date detection performance can be a crucial necessity for maintaining safety levels, even human life. Therefore, it is essential that the adverse weather conditions are well described and that the relevant methods are investigated and solved for different vision applications and methods [22]. The vision system models should be capable of capturing complex situations like under low-light conditions, or when the visibility ranges from clear, for instance, to rainy. The simulation must integrate solutions sufficiently advanced to be able to gauge the system response to adverse weather conditions and assure safety with every specific scenario. Overall the technology must be able to detect potentially dangerous situations and the objectives are to drive reachability, look-ahead, and selection of driving actions under varying weather and light conditions, and to quantify the effects on driving risk and energy consumption [23].

4.2. Robustness and Safety

Safety and robustness requirements are a common denominator for both Fully Data-Driven systems and Hybrid systems, with the proviso that the remaining hybrid systems may be composed of more and less complex deep learning models [24]. In the case of entirely data-driven systems, it is difficult to guess what kind of failures the network may show, as the neural architectures trained by the system may not have clear structural rules that could be observed and verified. Due to this fact, these systems are more difficult to predict, especially in cases when they encounter a creating corner conditions in the real world. In addition, data-driven systems are generally more sensitive to the purity of the samples used for training and are also susceptible to the influence of incorrect examples for which the system is overfitted when the validation error starts to increase, i.e. the so-called datasets with adversarial examples.

During the development of vision-based autonomous vehicle systems, designers need to focus on the robustness and safety of the system [25]. These are the most critical issues for users, especially if the system is in charge of critical driving tasks. The following elements that can potentially endanger user safety will tend to receive additional focus, such as robustness to variations of the elements, including various environmental and lighting conditions, changes in camera parameters due to light or temperature, interferences caused by water, fog, rain, or dirt deposited on the camera, proper operation in case of occlusion of the road or moving objects (including humans and animals). It also seems important to reduce the sensitivity of the systems to adversarial input, which in the case of vision-based systems are especially easy to generate.

5. Future Directions

[13] Connectomics is a field in neuroscience which deals with the study of the brain's structural connectivity. It is argued that deep learning could be a valuable tool for the complex task of solving this problem. Most of the available brain connection identification methods can only deal with a limited number of samples and are plagued with high false detection rates. In this work, we present an alternative approach for this problem from the perspective of computer vision; albeit we are faced by a similar limited number of training data and considerable noise. This paper focuses on applications of deep learning for connectomics, the study of brain's structural connectivity. Just as computer vision tackles the problem of extracting meaningful information from image data, connectomics is concerned with extracting structural connections between cells or neuronal populations from raw image data obtained from brain samples.[1] ConnectCW is a program that we developed as part of this project, which has only one architecture but has been designed to deal with both, positive and negative queries, and includes a false connections filtration technique based on convolutional neural networks. As the basis for supervised learning, the valuable dataset from the MICC of seven configurations of queries (positive and negative) was collected, each with a real sample from two volume-upscaled cloverleaf graphs and with background noise. The filters for calculations of ROC-AUC and PR-AUC were developed in accordance with a number of true plaques set to the optimal value determined in advance. As evaluated, the performance of filters differ considerably depending on the respective distribution of positive and negative queries, i.e., which signal strength is used during the filtration. As an empirical analysis, regions with potential interest in the neural tissue were manually extracted from the reconstructed output and compared to the original true neocortical areas of identified neuromuscular junction (NMJ) datasets. For this reason, essentially the entire dataset was split randomly into two training and test datasets respectively, to allow us to check the reproducibility and generalization of our training results. The experiment with the ANOVA test illustrated that the performance of our program depends on significant inter-group differences in the distribution of training queries in the testing process.

5.1. Multi-Sensor Fusion

Implementation approaches that aim to design complete fusion components for the 3D Euclidian and the robust 2D cost-based space are realized comprehensively with multi-modal learning bridge CNNs. Camera and LiDAR characteristically output data in different formats, i.e., 2D and 3D, and span the RGB modality and high-precision LiDAR point cloud modality. Cameras are able to deliver extensive spatial contextual information and are decisive in the identification of small and distant objects, which LiDAR cannot perceive well. However, the lack of depth information is demonstrated in the image domain and makes the camera modality weak in handling lane detection. In contrast, point cloud data from LiDAR measures include distance, intensity and majorly provide 3D shape information. Positioning relatively distant objects is challenging for LiDAR because of their limited resolution. LiDAR typically captures the intensity in the cardinal direction of LiDAR, and that can be used to identify near-spherical objects, such as cars. Therefore, when integrating multiple sensors, cameras mainly handle the detection of small and distant objects, while LiDAR also captures depth information and underlines larger and nearby objects such as walls and cars. Non-standard dataset used randomly sampled paired data.

Camera-radar combination sensory fusion aims to effectively capture far- and near-range information, but radar objects can represent cars which then have near-spherical shapes and can be easily identified using cameras. Camera-LiDAR-radar data fusion enables the capture of extensive distance and velocity information and is therefore able to provide rich semantic information [26]. For instance, only a few camera pixels can indicate whether pedestrians are facing the vehicle's direction, and so inform us about their potential intention to cross the street. Camera-LiDAR-radar fusion may thus significantly enhance entity detection, while the increased distance and velocity resolution of objects represents a vital source of data for predicting future paths. Autonomous vehicles currently on the market employ multiple LiDARs, cameras, and radars to improve the reliability of environment perception. LiDAR data and camera-detected objects from different classes are eligible correspondence for their inclusion in data fusion algorithms, as both sensors provide similar object location in 3D space [27].

The field of autonomous vehicle vision is heavily defined by the need for systems that can deliver both accurate and timely information to the central perception and decision-making system [18]. The two main clans of autonomous sensing devices are LiDAR and vision, where the first is characterized by meager resolution and output rich in distance and intensity measures, while ROS supports large resolutions at a lower frame rate. Camera data can provide certain semantic information through 2D object classification and location, while LiDAR is capable of producing a realistic 3D localization of objects. Thus, to enrich the environment perception in autonomous vehicles, camera and LiDAR data are fused extensively.

5.2. Explainable AI

Explainable AI enables transparency in AI-based decision-making systems. It facilitates the development of trust and improvements in model performance. Explanations for AV decision-making assist four main classes of system stakeholders: the users, who need to build trust in the system; regulators, who are tasked with legal issues; system developers; and auditors. Hence, it is important to remove black boxes from the goal-based systems like AVs so they can provide explanations for their actions. Under the European General Data Protection Regulation (GDPR), a person has the right to receive an explanation for many kinds of automated decisions [28].

Complex systems like autonomous vehicles (AVs) could become hard to manage as they often rely on stumps of the human knowledge and are often totally opaque to the users. For this reason, the vehicles should be able to provide explanations for their actions and decisions. AV companies should protect the interests of multiple stakeholders while being able to ensure the any decision should be justified and based on accurate knowledge. The explainable AI can further increase the systems' safety and value [29].

6. Conclusion

Cluster-based data association techniques are also more and more used for simultaneous detection and tracking using Convolutional Neural Networks as it allows to perform end-to-end learning of data association, exploiting the natural clustering properties of natural scenes. With Deep Learning components, a simultaneous traffic light and particle-based data association tracking on 2D RGB-D data is maintained through a series of Bayesian updates and color-based exploration. In particular, the performance improvements in the technology have made them a competitive alternative to more traditional computer vision algorithms. The combination of mature convolutional neural networks for detection with generative adversarial networks for tracking is also becoming increasingly popular in the field of visual object tracking of ground-level sensor data [3].

[14] [30]In recent years, Deep Learning techniques have seen impressive performance improvements in numerous real-world problems related to computer vision for autonomous vehicles and advanced driver assistance systems. This domain has traditionally used ad hoc handcrafted features and specialized computer vision algorithms to detect and classify relevant objects or parameters for autonomous driving. The recent advent of Deep Learning has started to move such tailored designs towards more general end-to-end trainable systems that are capable to jointly believe object detection, semantic segmentation and tracking in the context of the scene. The learned features are expected to be efficiently shared and trained for the optimal perception tasks from the data with less human intervention. When using annotated data, powerful Convolutional Neural Networks have been trained end-to-end to simultaneously solve detection, tracking, segmentation, depth estimation, flow estimation, obstacle localization, topological scene understanding and lane detection from a single input sensor (Li and Hoiem 2018).

Reference:

1. Tatineni, Sumanth. "Exploring the Challenges and Prospects in Data Science and Information Professions." *International Journal of Management (IJM)* 12.2 (2021): 1009-1014.
2. Vemori, Vamsi. "Human-in-the-Loop Moral Decision-Making Frameworks for Situationally Aware Multi-Modal Autonomous Vehicle Networks: An Accessibility-Focused Approach." *Journal of Computational Intelligence and Robotics* 2.1 (2022): 54-87.
3. Shaik, Mahammad, Srinivasan Venkataramanan, and Ashok Kumar Reddy Sadhu. "Fortifying the Expanding Internet of Things Landscape: A Zero Trust Network Architecture Approach for Enhanced Security and Mitigating Resource Constraints." *Journal of Science & Technology* 1.1 (2020): 170-192.

4. Tatineni, Sumanth. "Climate Change Modeling and Analysis: Leveraging Big Data for Environmental Sustainability." *International Journal of Computer Engineering and Technology* 11.1 (2020).

5. Vemoori, V. “Towards Secure and Trustworthy Autonomous Vehicles: Leveraging Distributed Ledger Technology for Secure Communication and Exploring Explainable Artificial Intelligence for Robust Decision-Making and Comprehensive Testing”. *Journal of Science & Technology*, vol. 1, no. 1, Nov. 2020, pp. 130-7, https://thesciencebrigade.com/jst/article/view/224.