DOI: 10.54254/2755-2721/101/20241011

Research on target detection and recognition system of autonomous driving based on deep learning

Haotian Ye

Beijing-Dublin International college, University of Technology, Beijing, China

Gui-fang.wang@dsm-firmenich.com

Abstract. Autonomous driving technology relies heavily on advanced target detection and recognition systems to ensure safety and efficiency. This essay explores how deep learning has revolutionized these systems in four key areas: lane detection, detection of obscured vehicle parts, blind spot detection, and multi-target detection. Traditional methods often struggle under challenging conditions, but deep learning approaches, including Convolutional Neural Networks (CNNs) and Fully Convolutional Networks (FCNs), offer superior performance by analyzing complex features from raw data. Techniques like the Detection of Incomplete Vehicles using Deep Learning and Image Inpainting (DIDA) enhance safety by reconstructing obscured vehicle parts. In blind spot detection, models such as Sep-Res-SE blocks provide an effective, cost-efficient alternative to radar systems. The Adaptive Perceive SSD (AP-SSD) framework improves multi-target detection accuracy and real-time tracking by incorporating advanced feature extraction and temporal analysis. The essay concludes with future research directions aimed at refining real-time capabilities, expanding datasets, and exploring collaborative learning to further enhance autonomous driving technology.

Keywords: Autonomous driving, target detection, target recognition, deep learning.

1. Introduction

Autonomous driving represents a transformative leap in transportation technology, promising safer roads, improved traffic flow, and enhanced mobility for all. Central to the realization of autonomous vehicles (AVs) is the development of robust target detection and recognition systems. These systems, powered by advancements in deep learning, play a critical role in enabling vehicles to perceive and understand their environment, thereby making informed decisions in real-time. The evolution of autonomous driving from theoretical concepts to practical applications underscores the rapid pace of technological advancement. Early autonomous systems relied heavily on rule-based algorithms and sensor fusion techniques. However, these approaches often struggled to adapt to the complexities of real-world driving scenarios, where environmental conditions, traffic dynamics, and unforeseen obstacles demand sophisticated sensing capabilities. In recent years, the advent of deep learning has revolutionized the field of computer vision, offering unprecedented capabilities in object detection, classification, and scene understanding. Convolutional Neural Networks (CNNs), in particular, have emerged as a cornerstone technology for autonomous driving systems, enabling accurate and robust perception under diverse conditions, including varying lighting, weather conditions, and complex traffic environments [1]. This paper provides a comprehensive overview of the current state of target detection and recognition systems for autonomous driving, with a specific focus on the transformative impact of deep learning techniques. Through a systematic exploration of key technologies such as lane detection, detection of obscured vehicle parts, blind spot detection, and multi-object detection, this paper elucidates how deep learning enhances both the safety and efficiency of autonomous driving systems.

2. Lane detection

Lane detection is one of the fundamental functions of autonomous navigation which provides critical information of positioning with the road infrastructure and lane boundaries to the vehicle. Traditional lane detection methods often rely more on geometric modelling and feature extraction techniques that might be concerned as inefficient and unreliable under sophisticated circumstances such as adverse weather conditions or poor luminance scenarios. On the contrary, deep learning might be able to confront these issues as it has advanced accuracy and reliability when integrated in lane detection systems. It leverages CNN architectures to learn and process complex features directly prom raw pixel data. This approach enables robust performance across diverse environmental conditions, including daytime, nighttime, and inclement weather scenarios [2]. Recent innovations in deep learning for lane detection include the development of Fully Convolutional Networks (FCNs), which preserve spatial information through dense upsampling layers. FCNs are particularly effective in achieving pixel-level accuracy in lane marking detection, crucial for precise vehicle trajectory planning and lane departure warning systems [2]. Hybrid CNN-RNN architectures further enhance lane detection capabilities by incorporating recurrent neural networks to model temporal dependencies. This enables the system to predict lane trajectories and adapt to dynamic driving scenarios, such as lane changes and complex road geometries [2]. An American military scientist Boyd stated the OODA loop theory in the 1970s. It concluded modern aerial combat process into a continuous cycle with 4 stages: observe, orient, decide and act [3]. In his theory, completing this loop faster leads to greater efficiency. In lane detection, it might be similar to this theory as processing and recognizing captured images faster significantly aids the performance of the system overall, which reveals massive potential integrating deep learning approaches [4, 5]. For example, set a scenario where road is in bad condition, lane marks could be fragmentary. The vehicle enters this road with proper navigation at first. While lane marks suddenly became vague or disappear, indeterminacy appears. If the lane detection system utilizes traditional methods, it might lose the correct orientation as telling the true way from damaged lane marks could be hard for those methods. On the other hand, deep learning integrated system with FCNs will have a great chance surviving it due to its formidable ability to analyze patterns from pixels. Together with inertial navigation system (INS), which is commonly seen on jets or guided missiles that predicts the trajectory augment autonomous navigation capabilities in mountain roads and areas that has poor road infrastructures [6, 7]. Real-time performance remains a critical consideration for autonomous driving applications, prompting ongoing research into lightweight CNN variants optimized for deployment in resource-constrained environments [2]. Future advancements in lane detection will focus on improving instantaneous response times and robustness across diverse driving conditions, while also exploring adaptive learning techniques tailored to specific regional and environmental factors.

3. Detection of hidden parts of vehicles

To enhance the safety and reliability of autonomous driving systems, it is vital to ensure accurate detection of obscured vehicle parts. In 2016, the first ever known fatal accident of auto pilot system malfunction happened on a Tesla model S [8]. In this accident, a side of a white car was not detected by all of the sensors and camaras on the Tesla model S due to the similarity of it to the bright sky. After that, there are more crashes caused by sensor failures on autonomous vehicles. Thus, traditional optical detection methods often struggle to identify vehicles when their color blends with the background or when parts are obscured by environmental factors such as foliage or structures could be informed [4]. In one article about autonomous driving, the authors did an experiment, shown in Figure 1 [9].

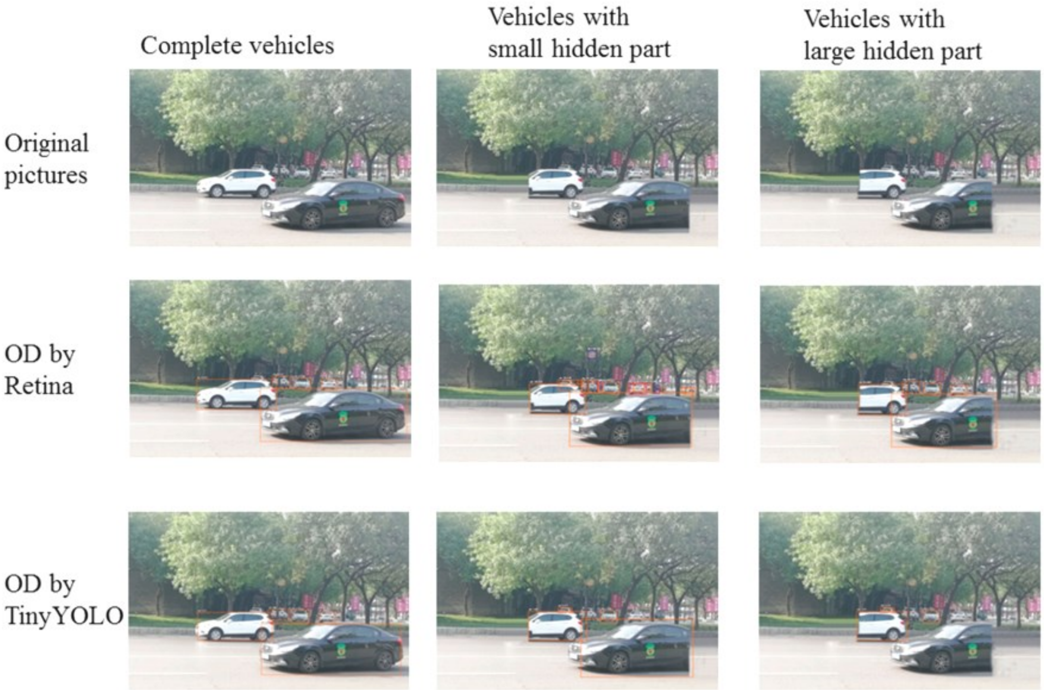

Figure 1. The experiment of recognizing part of vehicles by multiple different algorithms [9].

In this figure the authors took a picture in real life with 2 cars in it. Then they used some blending techniques to erase part of the 2 cars while the back ground remaining the same to simulate the scene of the crash where part of the car blends with the environment. Though it might look ridiculous as perhaps no car looks like the shape in this picture, it could be seen as a creditable example because in real life scenarios the situation might only be more complex. The second and third row were identified by Retina and Tint YOLO which are 2 advanced object detection algorithms, the red frame shows the algorithm recognized the object as a car. In the results, even these latest recognition methods could not include the hidden part of the vehicles, one even failed to correctly recognize the whole car. Therefore, the importance of recognizing hidden parts of vehicles should be stressed. In the paper the authors also utilized a method called "Detection of Incomplete Vehicles using Deep Learning and Image Inpainting" (DIDA), which represents a significant advancement in addressing these challenges. By integrating deep learning techniques with image inpainting algorithms, DIDA reconstructs obscured vehicle parts based on contextual information from surrounding pixels [9-11]. In recent years, significant improvements were done in image generation and image classification, as well as object detection with the research on deep neural network [12-14]. The authors demonstrated how the missing part of an image could be filled utilizing DIDA, which is shown in Figure 2 [9][15-17] .

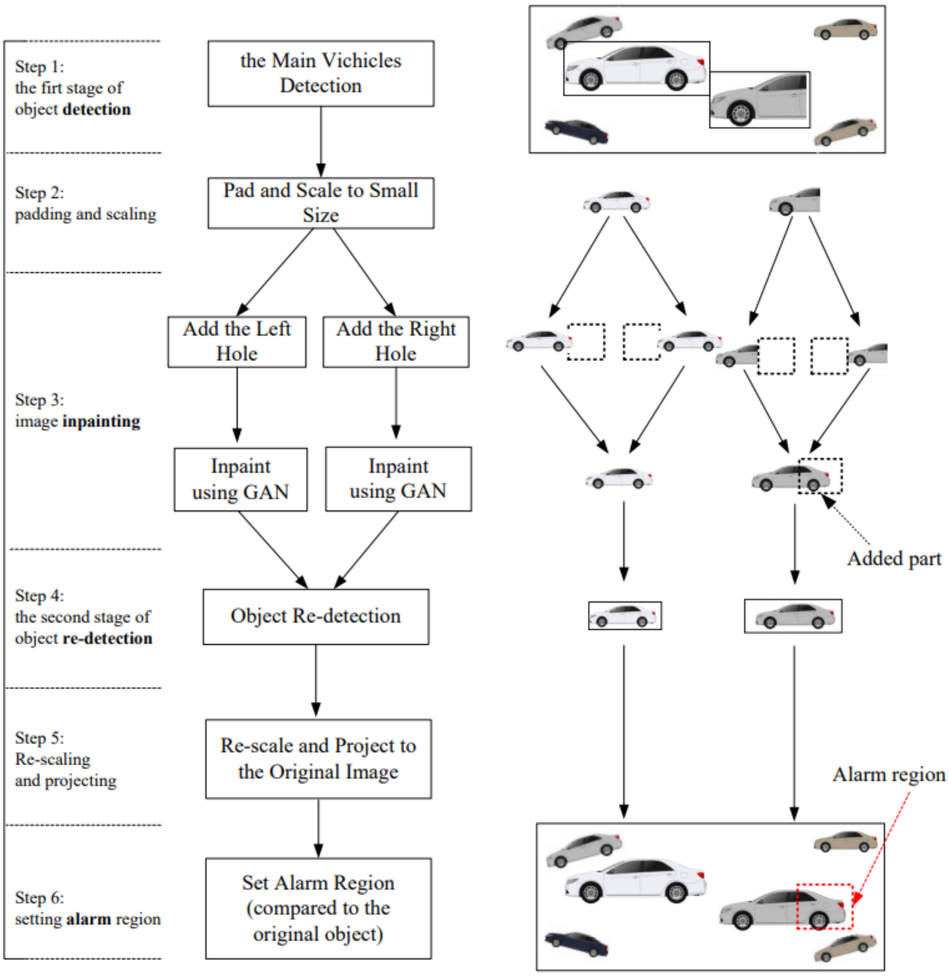

Figure 2. The process of complementing missing structures of vehicles with DIDA [9].

In this process, 6 steps were included, utilizing the capability of optical sensors to capture approximately 20-30 frames per second, excluding far vehicles to ensure there is enough performance for nearby vehicles that might potentially cause threats. Then the algorithm sets a hypothesis which all cars captured could in theory have hidden parts, where deep neural network compares and calculates the true missing part of certain vehicles. Finally, the hidden parts will be filled (if there is one) and an alarm frame will be set to tell the autonomous driving system to avoid entering that specific area. Training datasets for DIDA encompass a wide range of challenging scenarios, including vehicles partially hidden by natural or man-made obstructions. This extensive training enables the system to accurately predict and visualize obscured vehicle parts, thereby enhancing overall detection accuracy and reliability. Initial object detection in the DIDA framework typically involves the use of modified neural network architectures, such as the RetinaNet, which are optimized for detecting partially visible objects. Subsequent image inpainting processes then reconstruct missing vehicle parts based on learned contextual cues, ensuring comprehensive scene understanding and hazard anticipation [9]. The experiment carried out by the authors involves in more than 50,000 pictures taken in Xi'an, China, all under complex traffic conditions [9]. The result show that this DIDA algorithm has an accuracy of 91%, which is compelling. Moreover, the DIDA method has no potential risks due to its principles of only adding parts to vehicles rather than deleting them. In the worst scenario, it might only slow the traffic a bit. Another possible approach that adds above DIDA could be integrating a database of vehicles to it.

By comparing detected vehicles with the database utilizing the performance of deep learning neural network, the accuracy might increase while keeping the process real-time. Options such as frequent appearing vehicles in certain region could also be selected to optimize the performance of the autonomous driving system. The application of DIDA extends beyond simple object detection to encompass real-time decision-making capabilities within autonomous driving systems. By mitigating the risks associated with incomplete object recognition, DIDA contributes to safer and more efficient navigation in complex driving environments.

4. Blind spot detection

Identifying objects in a vehicle's blind spots is essential for avoiding accidents and ensuring safe lane changes in autonomous driving situations. Conventional blind spot detection methods mainly use X-band microwave radar sensors, some are mechanically scanned while others are using electronic scanned array ones, which can be costly and less efficient at spotting small or hidden objects, especially in city settings or poor weather, and limited in range [18, 19]. Deep learning methods provide a budget-friendly and efficient option compared to radar-based systems by utilizing camera-based frameworks to improve detection precision and dependability. Architectures such as the Sep-Res-SE block combine depthwise separable convolutions, residual learning techniques, and squeeze-and-excitation (SE) modules to boost object detection performance while reducing computational demands [20]. In recent years, the development of deep convolutional neural network (R-CNN) has been successful, which shows some superiority against traditional models such as you only look once (YOLO) and single shot mutibox detector (SSD) as it could process images with less reliance on parameters and layers, thus reducing requirements on hardware and saves performance for extra recognition speed and real-time capabilities [16, 17][20-23]. As an experiment, a high performance Sekonix camara was mounted on a Linkoln MKZ sedan. Blind spot areas were then illustrated by 4 squares that are 4 by 2 meters, which is demonstrated in Figure 3 [20]. CIFAR-10 data set was chosen to assess the performance of this machine learning method, which has 60000 sets of images in size of 32 * 32 [20]. In Figure 4, the example of how the blind spot detection of this setting works.

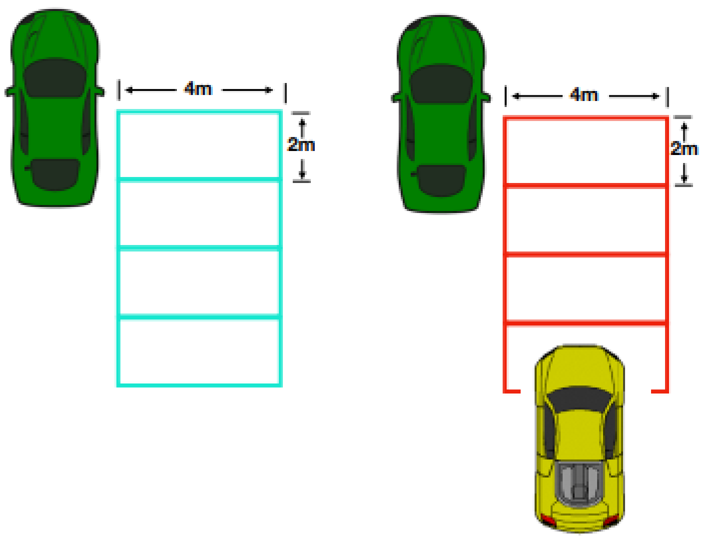

Figure 3. The demonstration planform of blind spot grid of the test vehicle [20].

DOI: 10.54254/2755-2721/101/20241011

Figure 4. The example of 3D blind area grid [20].

In the figure, the frame stands for blind spot areas where it remains green if no cars entered it. When there are cars entering parts of the area, it turns into red [20]. As a result, 5000 images without vehicles in the blind spot together with 3874 images with vehicles in the blind spot were captured. The Sep-Res-SE model obtained the same level and even better in some segments of performance compared to other methods, which only featured few layers and parameters, making a significant contribution on real-time embedded systems such as autonomous driving [20]. These innovations enable deep learning models to reliably detect and track objects within a vehicle's blind spots across a variety of environmental conditions, including low visibility and congested traffic scenarios. By leveraging large-scale datasets specifically annotated for blind spot detection, these models can adaptively adjust to dynamic driving situations, ensuring robust performance and timely hazard identification. Experimental validations elucidate the superiority of deep learning-based blind spot detection systems over conventional radar-based methodologies, particularly under conditions of visual clutter or compromised visibility. The incorporation of adaptive techniques, such as dynamic region magnification and threshold modulation, significantly augments detection accuracy and system responsiveness, thereby enhancing overall safety and operational efficacy [20, 22]. Future research in blind spot detection will focus on refining algorithmic architectures to achieve real-time responsiveness as well as scalability across diverse driving environments. Continued advancements in deep learning techniques, coupled with the expansion of annotated datasets, will further accelerate the adoption of reliable blind spot detection systems in autonomous driving applications.

5. Multi-target detection

Navigating complex traffic situations demands that autonomous vehicles simultaneously detect and track numerous objects. Conventional single-object detection approaches frequently fall short in real-time multi-object detection and tracking, especially in environments with heavy vehicle traffic, varying speeds, and erratic behaviors. The Adaptive Perceive SSD (AP-SSD) framework represents a significant advancement in enhancing the capabilities of SSD architectures for multi-target detection tasks. By integrating innovations such as a Gabor convolution kernel library, AP-SSD enhances feature extraction capabilities across multiple scales and color variants, improving detection accuracy in diverse traffic scenes. The Dynamic Region Zoom-in Network (DRZN) enhances computational efficiency by selectively focusing on different areas of images, which minimizes processing demands while maintaining detection accuracy. Additionally, adaptive threshold techniques modify confidence levels according to scene complexity and object features, ensuring reliable performance in practical driving conditions]. Long Short-Term Memory (LSTM) networks are essential for linking temporal information and tracking objects through consecutive frames. This function is critical for real-time decision-making in autonomous driving, allowing vehicles to predict and react to changing traffic conditions effectively. Figure 5 shows an example of a time aware based framework on video object detection, which clearly demonstrated that by using LSTM and DRZN, object detection with videos could be accurate as this method could precisely predict the travelling orientation and lane of cars and pedestrians [21].

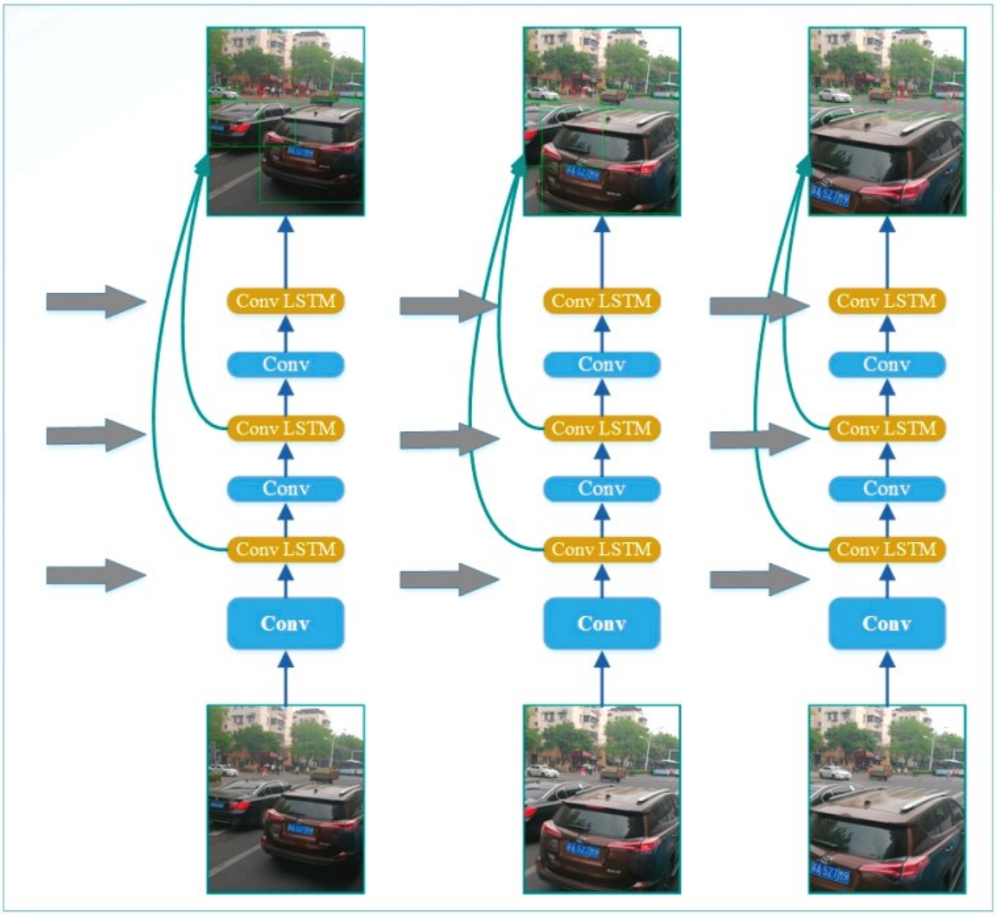

Figure 5. LSTM and DRZN powered video object detection framework [21].

It not only achieved a high accuracy in this experiment but also saves cost for the embed system for the autonomous driving function that has multiple functions which needs more performance as dire [1, 21]. This method not only provides a promising approach for detection of slow, small multi targets, if integrate with thermal vision camaras to enhance detection capabilities in dark or highly complex traffic zones, the accuracy and universality of this method could be even higher as thermal vision could provide contrast for human and vehicles to the background by the heat they emit [18]. Experimental evaluations demonstrate the superior performance of the AP-SSD framework in terms of Average Precision and Mean Average Precision, validating its efficacy in multi-object detection tasks under challenging real-world conditions. By improving detection precision and scalability, AP-SSD enhances the overall safety and operational efficiency of autonomous driving systems.

6. Conclusion

In summary, this paper has thoroughly examined how deep learning techniques are applied in target detection and recognition for autonomous driving systems. The techniques discussed—such as lane detection, recognizing obscured vehicle parts, blind spot detection, and multi-object detection—demonstrate how deep learning significantly enhances both safety and operational efficiency in these systems. Looking forward, future research will aim to further refine algorithmic architectures to improve real-time responsiveness and robustness under various driving conditions. Enhancing dataset diversity and scale will be vital for developing models that perform well across different global driving environments, including varying road infrastructures, weather conditions, and traffic patterns. Moreover, investigating collaborative learning and federated learning methods could improve model scalability and adaptability while maintaining data privacy and security. Advancements in these areas are expected to drive progress in autonomous driving technology, leading to wider adoption and safer integration into everyday transportation systems. By harnessing deep learning capabilities, autonomous driving systems have the potential to achieve unmatched safety, efficiency, and reliability, ultimately revolutionizing mobility and urban transportation. As these technologies advance, they are poised to transform transportation, making roads safer and more accessible for everyone.

References

[1] Kang K, Park N, Park J and Abdel-Aty M 2024 Deep Q-network learning-based active speed management under autonomous driving environments Comput.-Aided Civ. Infrastruct. Eng. 23
[2] Singal G, Singhal H, Kushwaha R, Veeramsetty V, Badal T and Lamba S 2023 RoadWay: lane detection for autonomous driving vehicles via deep learning Multimedia Tools Appl. 82 4965-4978
[3] Wang L, Wu Y, Zhang S, Huang Y and Zhang X 2023 Research on Combat Effectiveness Based on Internet of Things IEEE 167
[4] Laurell C and Sandstrom C 2022 Social Media Analytics as an Enabler for External Search and Open Foresight-The Case of Tesla's Autopilot and Regulatory Scrutiny of Autonomous Driving IEEE Trans. Eng. Manage. 69 564-571
[5] Lee S D, Tzeng C Y, Kehr Y Z, Huang C C and Kang C K 2010 Autopilot System Based on Color Recognition Algorithm and Internal Model Control Scheme for Controlling Approaching Maneuvers of a Small Boat IEEE J. Ocean. Eng. 35 376-387
[6] Tseng Y W, Hung T W, Pan C L and Wu R C 2019 Motion control system of unmanned railcars based on image recognition Appl. Syst. Innov. 2 1-16
[7] Hoque S, Xu S, Maiti A, Wei Y and Arafat M Y 2023 Deep learning for 6D pose estimation of objects — A case study for autonomous driving Expert Syst. Appl. 223 119838
[8] The Tesla Team A Tragic Loss Available online: https://www.tesla.com/blog/tragic-loss
[9] Xu Y, Wang H, Liu X, He H R, Gu Q and Sun W 2019 Learning to see the hidden part of the vehicle in the autopilot scene Electronics (Basel) 8 331
[10] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X 2016 Improved Techniques for Training GANs In: Advances in Neural Information Processing Systems 2234-2242
[11] Smirnov E A, Timoshenko D M and Andrianov S N 2014 Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks AASRI Procedia 6 89-94
[12] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Going deeper with convolutions IEEE 1

DOI: 10.54254/2755-2721/101/20241011

[13] Simonyan K and Zisserman A 2015 Very Deep Convolutional Networks for Large-Scale Image Recognition Cornell University Library arXiv.org Ithaca
[14] He K, Zhang X, Ren S and Sun J 2016 Deep Residual Learning for Image Recognition IEEE 770
[15] Lin T, Goyal P, Girshick R, He K and Dollar P 2017 Focal Loss for Dense Object Detection IEEE 2999
[16] Redmon J and Farhadi A 2017 YOLO9000: Better, Faster, Stronger IEEE 6517
[17] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You Only Look Once: Unified, Real-Time Object Detection IEEE 779
[18] Ramteke A Y, Ramteke P, Dhongade A, Modak U and Thakre L P 2024 Blind Spot Detection for Autonomous Driving Using RADAR Technique J. Phys. Conf. Ser. 2763 12015
[19] Jichkar R, Paraskar S, Parteki R, Ghosh M, Deotale T, Pathan A S, Bawankar S and Thakare L P 2023 5g: An Emerging Technology And Its Advancement IEEE 1
[20] Zhao Y, Bai L, Lyu Y and Huang X 2019 Camera-based blind spot detection with a general purpose lightweight neural network Electronics (Basel) 8 233
[21] Wang X, Hua X, Xiao F, Li Y, Hu X and Sun P 2018 Multi-object detection in traffic scenes based on improved SSD Electronics (Basel) 7 302
[22] Chang C, Liou Y, Huang Y, Shen S, Yu P, Chuang T and Chiou S 2022 YOLO based deep learning on needle-type dashboard recognition for autopilot maneuvering system Meas. Control (Lond.) 55 567-582
[23] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C 2016 SSD: Single shot multibox detector 21