DETECTION OF OBJECTS BY SIGHT FOR AUTONOMOUS VEHICLES USING DEEP LEARNING

[1]Mr. D. HIMAGIRI, [2]K. DIKSHA, [3]H. SANDEEP, [4]K. SATHVIK, [5]K. PRAJWALA

[1](Assistant Professor) ,CSE. J.B. Institute of Engineering & Technology, Hyderabad

[2345]B,tech scholar ,CSE. J.B. Institute of Engineering & Technology, Hyderabad

ABSTRACT

Autonomous vehicles (AV) are expected to improve, reshape, and revolutionize the future of ground transportation. It is anticipated that ordinary vehicles will one day be replaced with smart vehicles that are able to make decisions and perform driving tasks on their own. In order to achieve this objective, self-driving vehicles are equipped with sensors that are used to sense and perceive both their surroundings and the faraway environment, using further advances in communication technologies, such as 5G. In the meantime, local perception, as with human beings, will continue to be an effective means for controlling the vehicle at short range. In the other hand, extended perception allows for anticipation of distant events and produces smarter behavior to guide the vehicle to its destination while respecting a set of criteria (safety, energy management, traffic optimization, comfort). In spite of the remarkable advancements of sensor technologies in terms of their effectiveness and applicability for AV systems in recent years, sensors can still fail because of noise, ambient conditions, or manufacturing defects, among other factors; hence, it is not advisable to rely on a single sensor for any of the autonomous driving tasks. The practical solution is to incorporate multiple competitive and complementary sensors that work synergistically to overcome their individual shortcomings. This article provides a comprehensive review of the state-of-the-art methods utilized to improve the performance of AV systems in short-range or local vehicle environments. Specifically, it focuses on recent studies that use deep learning sensor fusion algorithms for perception, localization, and mapping.

I. INTRODUCTION

Autonomous Vehicles (AV) is a fast-growing technology and has sought the attention of many global vehicle companies. This technology enables the vehicle to be on autopilot and navigate itself with little or zero human input.

Autonomous driving is one of the most anticipated technologies of the 21st century and one of the most active research topics at the moment. Autonomous driving attempts navigating roadways without human intervention by sensing and reacting to the vehicles immediate environment. It includes major challenges for Computer Vision and Machine Learning. Object detection is one of the most important requirements for autonomous navigation and consists of localization and classification of objects. Therefore, accurate object detection algorithms are needed. Solutions to these problems often compromise speed, accuracy, or simplicity. Recent state-of-the-art deep learning models that address the problem of object detection include Region-Based Convolutional Neural Networks (R-CNN) and their improved versions Fast R-CNN and Faster R-CNN, designed for model performance and first introduced in 2013. A second model for object detection introduced in 2015 is YOLO, designed for speed and real-time use.

One of the critical aspects of Autonomous Navigation is Object Detection. Effective detection is essential as they need to detect road elements and pedestrians before they can understand and respond to their surroundings. Some of the challenges of object detection in AVs are detecting objects in low-light conditions and slick surfaces. Other challenges would be poor resolution of data obtained by specific devices.

However, the problem with radar is that in some cases, they cannot distinguish pedestrians, especially children. These days auto manufacturers have implemented Light Detection and Ranging (LiDAR) sensor for AVs due to the advantage of detecting objects in low light conditions as camera-based systems offer dense light projection but lack distance information. However, even with the best performing sensors, most systems associated with AVs lack accuracy for complete self-drive because of the limitation of algorithms. The limitation for AVs to detect objects precisely will probably endanger the safety of both vehicle occupants and surrounding pedestrians.

In recent years, the development and integration of autonomous vehicles into our transportation systems have emerged as a transformative technological frontier. These vehicles, equipped with advanced sensing technologies and computational capabilities, hold the promise of revolutionizing the way we commute, enhance road safety, and optimize traffic flow. At the core of this revolution is the ability of autonomous vehicles to perceive and interpret their surroundings accurately, a task crucial for safe navigation and decision-making. One of the fundamental challenges in achieving this perception is the real-time detection of objects by sight, a complex task that has witnessed significant advancements with the advent of deep learning techniques.

Traditionally, object detection in computer vision relied on handcrafted features and rule-based algorithms. However, the limitations of these approaches became evident in the face of the intricate and dynamic nature of real-world environments. The emergence of deep learning, particularly convolutional neural networks (CNNs), has marked a paradigm shift in the field of computer vision, offering unprecedented capabilities in image recognition, classification, and object detection.

The motivation behind this major project stems from the critical role that object detection plays in the success and safety of autonomous vehicles. Accurate and efficient detection of objects in the vehicle's surroundings, such as pedestrians, vehicles, cyclists, and obstacles, is paramount for making informed decisions in real-time. The consequences of false positives or negatives in object detection could range from inefficient traffic flow to life-threatening accidents, underscoring the significance of developing robust and dependable detection systems.

Moreover, the increasing integration of autonomous vehicles into urban landscapes and the ever-expanding array of environmental conditions necessitate adaptive and intelligent perception systems. Deep learning, with its ability to automatically learn hierarchical representations from data, provides a promising avenue to address these challenges. By leveraging deep neural networks, this project aims to enhance the perceptual capabilities of autonomous vehicles, enabling them to navigate complex scenarios with a level of precision and reliability that was once deemed unattainable.

1.1 EXISTING SYSTEMS

The existing system for object detection in autonomous vehicles typically relies on traditional computer vision techniques or a combination of classical methods with machine learning approaches. These may include techniques like Haar cascades, and other feature engineering methods. While these approaches have been successful to some extent, they often struggle with complex and dynamic environment.

It has a low resolution and high false positive rate. This results in inaccurate detection and classification of objects on the road. Current object detection systems for autonomous vehicles rely on a combination of cameras, lidar, and radar sensors to detect objects in the vehicle's surroundings. These systems use computer vision algorithms to process the data collected by the sensors and identify objects such as other vehicles, pedestrians, and obstacles.

1.1.1 DISADVANTAGES

- **Limited Generalization**: Traditional methods may struggle to generalize well in complex and dynamic environments with varying lighting conditions, viewpoints, and object orientations.
- **Manual Feature Engineering**: These methods often require manual feature engineering, making them less adaptive to new and evolving scenarios.
- **Limited Accuracy**: In complex scenarios, the accuracy of object detection may not be sufficient for the high safety requirements of autonomous vehicles.

1.2 PROPOSED SYSTEMS

The proposed object detection system for autonomous vehicles has a high resolution and low false positive rate. This results in more accurate detection and classification of objects on the road. Our proposed system utilizes Deep learning algorithms to accurately detect and classify objects in real-time. These algorithms have been trained on vast amounts of data and are capable of identifying even the most complex objects with high precision. Our system processes data in real-time, allowing for immediate detection and response to potential hazards on the road. This ensures the safety of passengers and pedestrians alike.

1.2.1 ADVANTAGES

- **End-to-End Learning**: Deep learning models can learn hierarchical features directly from raw data, eliminating the need for manual feature engineering.
- **Better Generalization**: Deep learning models tend to generalize well across diverse scenarios, making them suitable for complex and dynamic environments.

Adaptability: Neural networks can adapt to new data and scenarios, making them more robust in changing conditions...

II. LITERATURE SURVEY

Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceed-ings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [Google Scholar][1]

Huang, G.; Liu, Z.; Der Maaten, L.-V.; Weinberger, K.-Q. Densely Connected Convolu-tional Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [Google Scholar][2]

Zoph, B.; Vasudevan, V.-K.; Shlens, J.; Le, Q.-V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710. [Google Scholar][3]

Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. IEEE T. Pattern Anal. 2017, 32, 99–113. [Google Scholar][4]

Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Ob-ject Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 20–23 June 2014; pp. 580–587. [Google Scholar][5]

He, K.; Zhang, X.; Ren, S.; Sun., J. Intelligence, M. Spatial Pyramid Pooling in Deep Con-volutional Network for Visual Recognition. IEEE T. Pattern. Anal. 2015, 37, 1904–1916. [Google Scholar] [CrossRef] [Green Version][6]

Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [Google Scholar][7]

Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Conference and Workshop on Neural In-formation Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1137–1149. [Google Scholar][8]

Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162. [Google Scholar][9]

Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125. [Google Scholar][10]

Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788. [Google Scholar][11]

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. SSD: Single shot multibox detec-tor. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37. [Google Scholar][12]

Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.M.; Dollar, P. Focal loss for dense object detec-tion. IEEE Trans. Pattern Ana. Mach. Intell. 2020, 42, 318–327. [Google Scholar] [CrossRef] [Green Version][13]

Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [Google Scholar][14]

Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. Available online: https://arxiv.org/abs/1804.02767 (accessed on 8 April 2018).[15]

Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detec-tion. In Proceedings of the 2019 IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9627–9636. [Google Scholar][16]

Yang, Y.; Deng, H. GC-YOLOv3: You Only Look Once with Global Context Block. Elec-tronics 2020, 9, 1235. [Google Scholar] [CrossRef][17]

Chen, Y.; Han, C.; Wang, N.; Zhang, Z. Revisiting Feature Alignment for One-stage Ob-ject Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pat-tern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [Google Scholar][18]

Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for ob-ject detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019. [Google Scholar][19]

Uijlings, J.; Sande, K.-E.; Gevers, T. Selective Search for Object Recognition. INT. J. Comput. Vision 2013, 104, 154–171. [Google Scholar] [CrossRef] [Green Version][20]

III. SYSTEM DESIGN

3.1 Proposed system architecture:

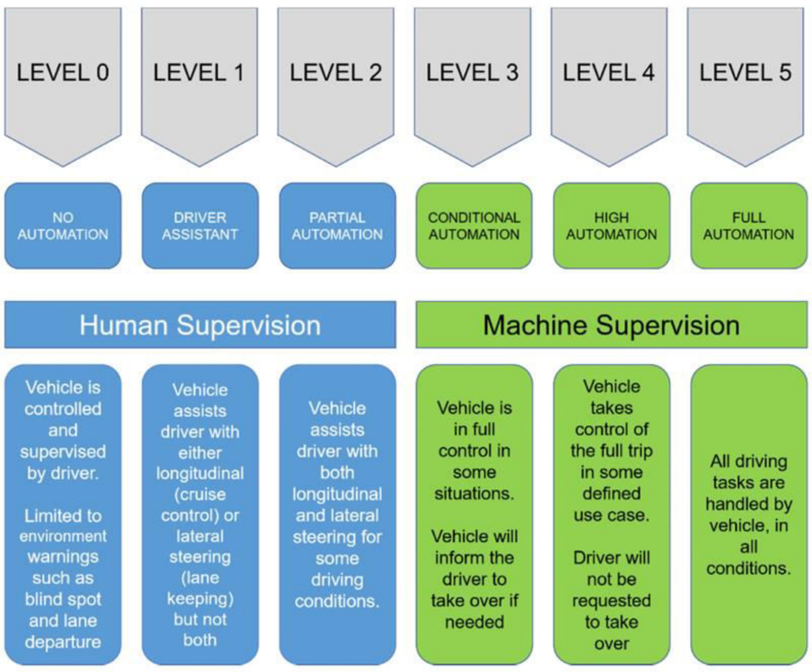

Figure 3.1 : The six levels of autonomous vehicles

There are six different levels of automated vehicles, starting from level 0 where the driver is in full control of the vehicle, and ending with level 5 where the vehicle is in full control of all driving aspects. Currently, it can be confidently stated that levels 2 and 3 are being adopted in some of the commercial cars, such as GM's Cruise, Tesla's Autopilot, and BMW. Several autonomous features are already being performed in these cars, such as adaptive cruise control, automatic braking, and lane-keeping aid systems.Although different AV systems may differ slightly from one to another, they all need to present a solution for the autonomous navigation problem, which is generally divided into four main elements: perception, localization and mapping, path planning, and control. In perception, the vehicle utilizes a group of onboard sensors to detect, understand, and interpret the surrounding environment, including static and dynamic obstacles, such as other moving vehicles, pedestrians, road signs, traffic signals, and road curbs. Localization and mapping tasks attempt to locate the vehicle globally with respect to world coordinates. Additionally, they are responsible for building a map of the vehicle's surroundings and continuously tracking the vehicle's location with respect to that map. Subsequently, path planning exploits the output of the previous two tasks in order to adopt the optimal and safest feasible route for the AV to reach its destination, while considering all other possible obstacles on the road. Lastly, based on the selected path, the control element outputs the necessary values of acceleration, torque, and steering angle for the vehicle to follow that selected path. Additionally, multiple studies consider adding connected vehicle technologies , such as vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) technologies, where essential information is shared to create an enhanced cooperative driving environment . This extended and improved cooperative perception allows vehicles to predict the behavior of the key environmental components (obstacles, roads, ego-vehicles, environment, driver behavior) efficiently and to anticipate any possible hazardous events.

3.2 Architecture Diagram:

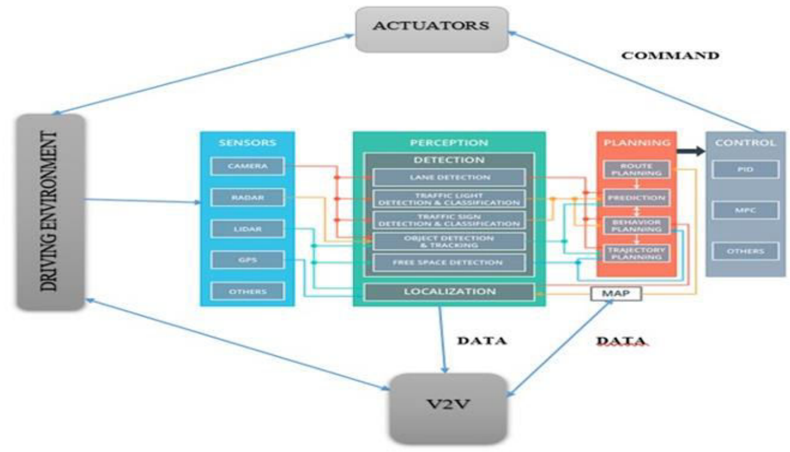

Figure 3.2: Architecture Diagram

IV. OUTPUT SCREENS

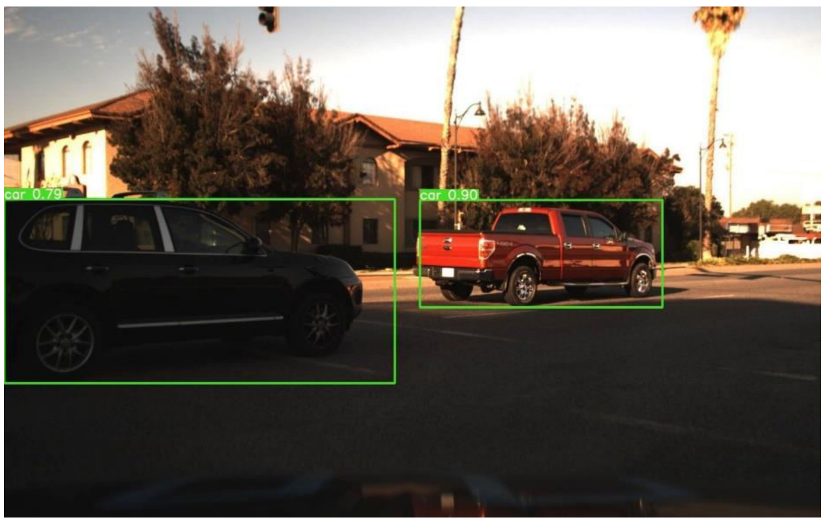

Figure 4.1: Detection Of Cars

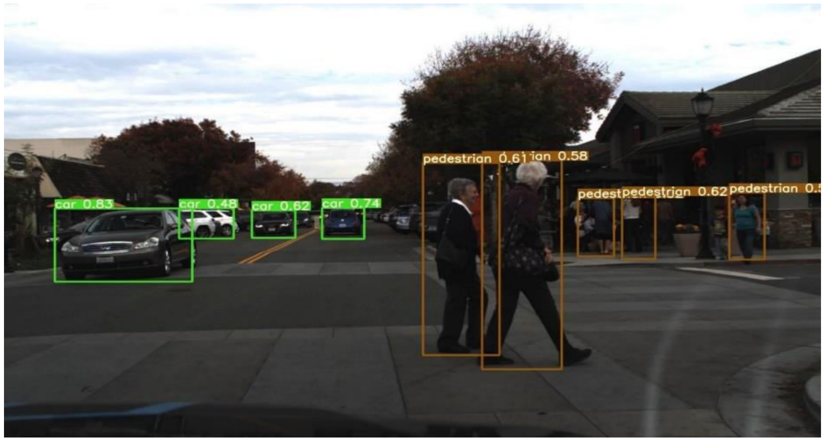

Figure 4.2: Detection Of Cars And Pedestrians

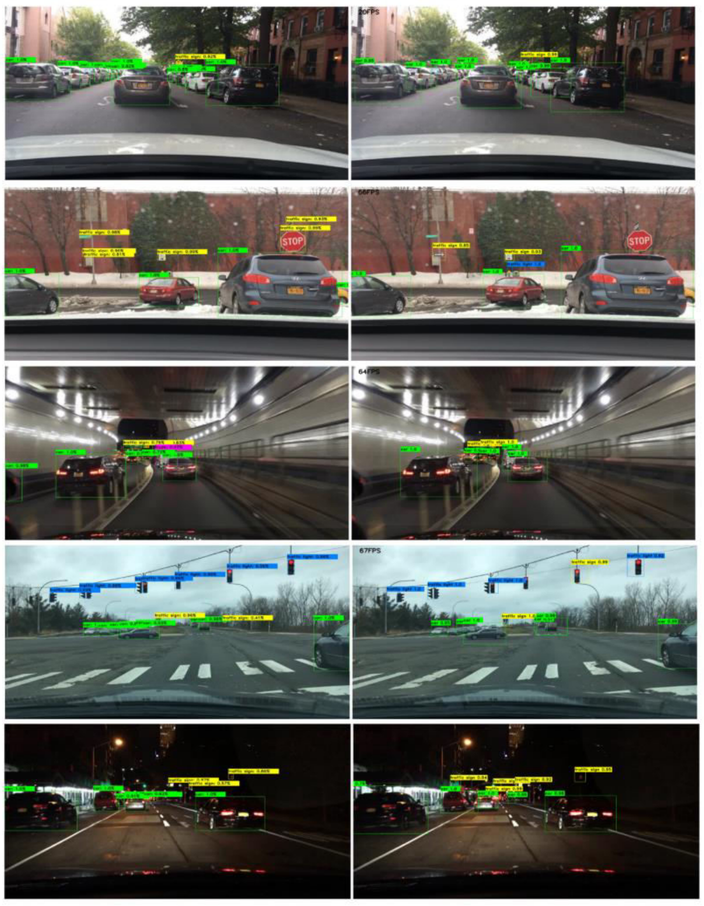

Figure 4.3: Detection Of Traffic Lights And Cars

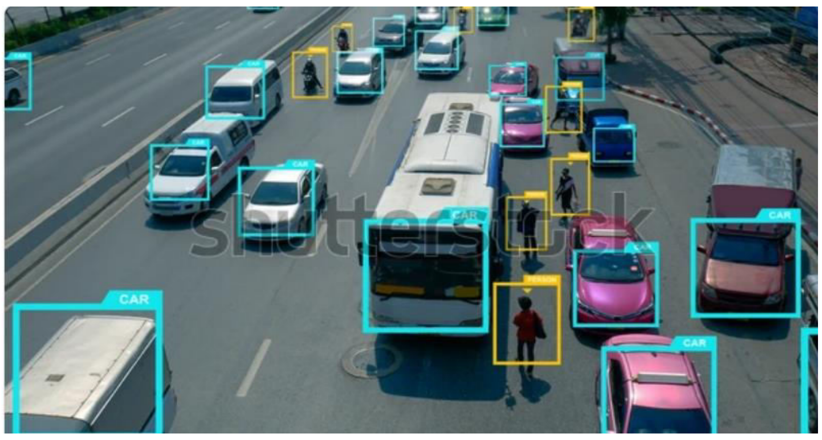

Figure 4.4: Detection Of Vehicles

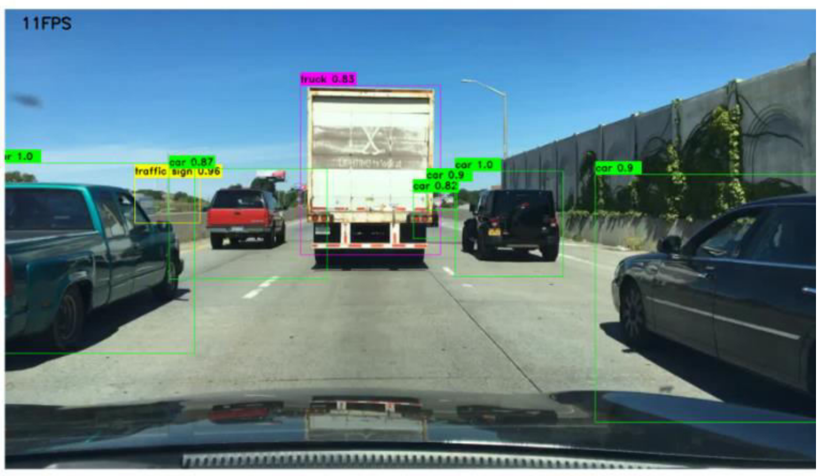

Figure 4.5: Detection Of Trucks And Other Vehicles

Figure 4.6: Detection Of Objects Around The Vehicle

Figure 4.7: Precise Detection Of Minute Objects

V. CONCLUSION

The objective has been to study the perception problem in the contexts of real-time object detection for autonomous vehicles. Self-driving systems are commonly categorized into three subsystems; perception, planning, and control, where the perception system is responsible for translating raw sensor data into a model of the surrounding environment. To study this problem, a cutting-edge real-time object detection deep neural network called SSD was trained and evaluated on both real and virtual driving-scene data.

VI. FUTURE ENHANCEMENT

Improved Accuracy: Deep learning models for object detection are continuously evolving to achieve higher accuracy and better performance. Future research will likely focus on developing more advanced architectures, better optimization techniques, and larger, more diverse datasets to further improve accuracy.

Real-Time Detection: There is a growing demand for real-time object detection systems in various applications such as autonomous vehicles, surveillance, and augmented reality. Future advancements will aim to make object detection models faster and more efficient to enable real-time deployment on resource-constrained devices.

Robustness to Variability: Object detection models need to be robust to variations in object appearance, scale, pose, lighting conditions, and occlusions. Future research will focus on developing algorithms that can handle these challenges more effectively, possibly through better data augmentation techniques, domain adaptation, or more robust architectures.

VII. REFERENCES

[1] Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [Google Scholar]

[2] Huang, G.; Liu, Z.; Der Maaten, L.-V.; Weinberger, K.-Q. Densely Connected Convolutional Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [Google Scholar]

[3] Zoph, B.; Vasudevan, V.-K.; Shlens, J.; Le, Q.-V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710. [Google Scholar]

[4] Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. IEEE T. Pattern Anal. 2017, 32, 99–113. [Google Scholar]

[5] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 20–23 June 2014; pp. 580–587. [Google Scholar]

[6] He, K.; Zhang, X.; Ren, S.; Sun., J. Intelligence, M. Spatial Pyramid Pooling in Deep Convolutional Network for Visual Recognition. IEEE T. Pattern. Anal. 2015, 37, 1904–1916. [Google Scholar] [CrossRef] [Green Version]

[7] Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [Google Scholar]

[8] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Conference and Workshop on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1137–1149. [Google Scholar]

[9] Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162. [Google Scholar]

[10] Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125. [Google Scholar]

[11] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788. [Google Scholar]

[12] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. SSD: Single shot multibox detector. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37. [Google Scholar]

[13] Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.M.; Dollar, P. Focal loss for dense object detection. IEEE Trans. Pattern Ana. Mach. Intell. 2020, 42, 318–327. [Google Scholar] [CrossRef] [Green Version]

[14] Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [Google Scholar]

[15] Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. Available online: https://arxiv.org/abs/1804.02767 (accessed on 8 April 2018).

[16] Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the 2019 IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9627–9636. [Google Scholar]

[17] Yang, Y.; Deng, H. GC-YOLOv3: You Only Look Once with Global Context Block. Electronics 2020, 9, 1235. [Google Scholar] [CrossRef]

[18] Chen, Y.; Han, C.; Wang, N.; Zhang, Z. Revisiting Feature Alignment for One-stage Object Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [Google Scholar]

[19] Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019. [Google Scholar]

[20] Uijlings, J.; Sande, K.-E.; Gevers, T. Selective Search for Object Recognition. INT. J. Comput. Vision 2013, 104, 154–171. [Google Scholar] [CrossRef] [Green Version]

[21] Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015. [Google Scholar]

[22] Badrinarayanan, V.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2481–2495. [Google Scholar] [CrossRef]

[23] Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer: Berlin/Heidelberg, Germany, 2015; pp. 3019–3037. [Google Scholar]

[24] Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. arXiv 2016, arXiv:1606.02147. [Google Scholar]

[25] Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [Google Scholar]

[26] Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 518–534. [Google Scholar]

[27] Wu, H.; Zhang, J.; Huang, K.; Liang, K. Computer Vision and Pattern Recognition. FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation. Available online: https://arxiv.org/abs/1903.11816 (accessed on 28 March 2019).

[28] Khan, A.; Ilyas, T.; Umraiz, M.; Mannan, Z.-I.; Kim, H. CED-Net: Crops and Weeds Segmentation for Smart Farming Using a Small Cascaded Encoder-Decoder Architecture. Electronics 2020, 9, 1602. [Google Scholar] [CrossRef]

[29] Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [Google Scholar]

[30] Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [Google Scholar]

.