Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

A Review of the Safety Survey of Intelligent Driving

Hao Wang *

University of Washington, Seattle, United States

* Corresponding Author Email: wh050713@uw.edu

Abstract. With the rapid advancement of artificial intelligence and sensing technologies, intelligent driving systems have transitioned from high-end model exclusivity to widespread adoption across diverse vehicle segments. This survey reviews three major object detection methods—traditional object detection, object detection based on Yolo, and object detection based on Transformer—and evaluates their respective strengths and limitations in autonomous driving contexts. We summarize representative improvements to R-CNN and Faster R-CNN that enhance small-object detection, occlusion handling, and anchor optimization, as well as YOLO variants that deliver real-time inference (up to 200 FPS) without sacrificing accuracy. Object detections which are based on Transformer, such as DETR, Deformable DETR, and Swin Transformer are shown to simplify pipelines, capture global context, and improve robustness against sparse or overlapping targets. Beyond algorithms, this paper identify critical challenges in perception latency, model generalization, data privacy, legal compliance, and hardware constraints. Finally, this paper outline future directions—lightweight end-to-end architectures via neural architecture search, self-supervised and federated learning for privacy-preserving adaptation, and heterogeneous edge-cloud hardware co-design—to accelerate the safe, efficient, and scalable deployment of intelligent driving technologies.

Keywords: Intelligent driving; object detection; YOLO; Transformer; R CNN.

1. Introduction

With the continuous development of technology, intelligent driving is no longer an exclusive feature of high-end models. More automakers are introducing intelligent driving systems to a wider range of models. Recently, several automakers have released their own intelligent driving systems. Byd announced that all its models are equipped with advanced intelligent driving systems as standard. Geely has launched the "Thousand Miles Vast" intelligent driving system based on AI "World Model" and other technologies. Changan's "Beidou Tianshu 2.0" plan proposes that it will launch 35 new digital and intelligent vehicles and so on in the next three years [1]. The advantages of intelligent driving mainly include reducing the occurrence of traffic accidents, using advanced artificial intelligence technology to assist or replace human driving, fundamentally eliminating people's bad driving behaviors, and lowering the risks of car travel. Secondly, many drivers suffer from lumbar intervertebral disc protrusion during long-term driving. Autonomous driving systems can help improve the driving environment of cars and provide people with superior driving space [2]. Having understood the development and advantages of intelligent driving technology, its definition also needs to be known to the public. Intelligent driving technology refers to the process where a car does not require human control during driving. Instead, it uses the on-board intelligent system to sense the surrounding environment and automatically plans the driving route to reach the destination based on the obtained information. From these, it can be known that intelligent driving technology is indeed a very promising technology. However, in its development process, many new problems and difficulties have also emerged. For example, how to precisely improve the safety of intelligent driving so that the probability of causing traffic accidents is much lower than that of human driving. This also makes the relevant object detection methods for assisting intelligent driving need to be continuously improved, such as the traditional object detection with low real-time performance and the object detection based on Yolo with high real-time performance that will be mentioned in this article, etc. Through the improvement of the target detection method, the actual performance can be increasingly

higher, and the safety of intelligent driving can also be enhanced. Furthermore, intelligent driving has also caused many new problems. The most typical example is how to determine the liability for traffic accidents caused by intelligent driving. Such problems keep emerging, which means that intelligent driving will also affect the subsequent formulation of relevant laws. Therefore, in this article, a review will be conducted from several aspects including various detection methods of intelligent driving, their existing limitations and prospects for the future. Overall, this article aims to summarize the current development and difficulties faced by intelligent driving, and put forward relatively reasonable suggestions for its future development.

2. Intelligent Driving Detection Methods

2.1. Traditional Object Detection

Here, R-CNN is taken as an example method. R-CNN, introduced in 2014, was a pioneering approach that combined region proposals with deep convolutional neural networks for object detection. The method involves generating around 2000 region proposals using Selective Search, then warping each region to a fixed size and passing it through a CNN (like AlexNet) to extract features. These features are subsequently classified using Support (SVMs), and bounding box regression is applied to refine the localization. While R-CNN significantly improved detection accuracy, it suffered from slow training and inference times due to the need to process each region proposal individually through the CNN [3]. Zhang et al. proposed the AD-Faster-RCNN model, aiming to enhance the detection ability for small targets and occluded targets. Based on the ResNet-50 backbone network, this model introduces deformable convolution and spatial attention mechanism to enhance the flexibility and focusing ability of feature extraction. Meanwhile, the Path Aggregation Balanced Feature Pyramid (PAB-FPN) structure was designed to achieve the effective fusion of multi-scale features. In the detection head part, the cascaded detector and boundary-aware positioning method were adopted, which improved the accuracy of bounding box regression. The experimental results show that on the COCO2017 dataset, Adfaster-RCNN has improved the average accuracy by 7.7% compared with the baseline model. On the BDD100k dataset, the average accuracy has increased by 4.1%, especially performing well in small targets and complex scenarios [4]. Thompson and Chen proposed an improved Faster-RCNN model for the problem of obstacle detection in driverless vehicles, using ResNet50 as the feature extraction network, aiming to improve the accuracy of obstacle detection in autonomous driving scenarios. Evaluated on the VOC2007 dataset, the average detection accuracy of this model has increased by 12.15% compared with the traditional methods. The results show that the model has outstanding performance in detecting various objects such as bicycles, buses and pedestrians, highlighting its potential for wide application in intelligent vehicle systems [5]. Carranza-Garcia et al. proposed an optimized Faster R-CNN model for the problems of unreasonable anchor box generation and class imbalance in autonomous driving scenarios. It divides the image into key regions through clustering and uses evolutionary algorithms to optimize the basic anchor points of each region. It also explores different reweighting strategies to solve the foreground - foreground class imbalance, indicating that the use of a simplified version of focus loss can significantly improve the detection of difficult and underrepresented objects in the two-stage detector. Its proposal was evaluated using the Waymo Open Dataset. The results showed that when the best single model was used, the average accuracy increased by 6.13% mAP, while when the integrated model was used, the average accuracy increased by 9.69% mAP. The proposed modifications to Faster R-CNN will not increase the computational cost and can be easily extended to optimize other anchor-based detection frameworks [6]. To sum up, these studies have significantly improved the target detection performance in autonomous driving scenarios through the structural optimization, feature enhancement and anchor box strategy improvement of the Faster R-CNN framework, especially in aspects such as small target detection, occlusion processing and category imbalance, providing strong support for achieving a safer and more efficient autonomous driving system. Of course, there are also many problems or shortcomings of it. For example, it has low computational efficiency, limited feature expression ability, poor robustness and generalization ability and so on.

2.2. Object Detection Based on Yolo

The YOLO series of algorithms achieve extremely high real-time performance by unifying the object detection problem into an end-to-end single-stage regression task - the original YOLO can run at around 45 FPS, while subsequent versions such as YOLOv7 can reach 155 FPS on standard hardware. Data Camp V7 document processing and data annotation fully meet the extremely highspeed requirements of application scenarios such as autonomous driving. The feature of YOLO lies in that it can divide the entire image into grids and simultaneously predict multiple bounding boxes and category probabilities with just one forward propagation. It uses global context information for detection, simplifying the training process and significantly reducing computational redundancy, achieving efficient detection that can be obtained with just one look [7][8]. YOLOv1 proposed by Redmon et al. aims to unify the object detection problem as an end-to-end regression task. It uses a network consisting of 24 layers of convolution and 2 layers of fully connected, divides the input image into 7×7 grids, and simultaneously predicts the bounding boxes and category probabilities within each grid. Thus, a forward propagation is achieved to complete the detection [9, 10]. Experiments show that YOLOv1 achieves a 63.4% mAP and an inference speed of approximately 45 FPS on the PASCAL VOC 2007 dataset, while its lightweight version, Fast YOLO, can reach 52.7% mAP and 155 FPS, significantly superior to the regionalization methods at that time. However, there are still deficiencies in positioning errors and the ability to detect small targets [11, 12]. The use of anchor frames to enhance the detection ability of targets of different sizes has significantly improved the recognition effect of small targets [13, 14]. On the COCO dataset, YOLOv3 (with 320×320 input) achieved a 28.2% mAP at a speed of 22 ms/ frame, which is approximately 3× faster than the 25 FPS of SSD at the same precision, offering both real-time performance and competitive detection accuracy [15]. Finally, Jocher et al. (2020) 's YOLOv5 introduces multiple data augmenting strategies such as the CSPDarknet backbone, PANet neck, Mosaic, and MixUp, and supports automatic learning of anchor boxes and semi-supervised hyperparameter search. Enable the network to enhance detection robustness while ensuring lightweight [16]. On the COCO validation set, YOLOv5s (640×640) can achieve approximately 50.5% mAP and realize a peak inference speed of ~200 FPS on NVIDIA A100, with significant improvements in both accuracy and speed compared to YOLOv3 [17, 18]. In conclusion, object detection based on Yolo is indeed superior to traditional object detection at present, which is specifically reflected in its ultimate real-time performance, lower computing and storage costs, good generalization ability, rich version iterations and community support and so on.

2.3. Object Detection Based on Transformer

In recent years, the Transformer-based object detection method has completely simplified the traditional detection pipeline through the end-to-end global attention mechanism and set prediction loss, significantly improving the detection accuracy and robustness. Carion et al. proposed DETR, an End-to-End Object Detection with Transformers models the detection task as a set prediction problem, adopted the Transformer encoder-decoder architecture, and was implemented through Hungarian matching — the corresponding global loss. Thereby eliminating manually designed components such as anchor and non-maximum suppression (NMS) [19]. The results show that On the COCO 2017 validation set, DETR achieves approximately 42 AP using the ResNET-50 backbone, which is comparable to the highly optimized Faster R-CNN baseline; The inference speed is approximately 50 ms/ frames (20 FPS), achieving a simple architecture and performance balance [20]. Zhu et al. proposed Deformable DETR: Deformable Transformers for End-to-End Object Detection, proposed a deformable attention module to address the problems of redundancy in attention calculation and slow convergence of DETR on high-resolution feature maps. This module samples only a few key positions around the reference points, thereby reducing the computational complexity and enhancing the spatial focusing ability [21]. The results show that Deformable DETR not only reduces the training period by 10 times on the COCO dataset, but also significantly improves in small object detection compared with DETR; Under the same training resources, the overall AP level surpasses the original DETR and significantly shortens the convergence time [22]. Liu et al. proposed Swin Transformer, Hierarchical Vision Transformer using Shifted Windows, designed a hierarchical

backbone network, dividing the image into fixed-size Windows and performing "window-moving" operations on the window positions in adjacent layers to consider both efficient self-attention within local Windows and information interaction across Windows, thus forming a scalable universal visual Transformer backbone [23, 24]. The results show that when used as a universal backbone in various detection frameworks (such as Mask R-CNN), Swin Transformer implements 58.7 box AP and 51.1 mask AP on COCO test-dev. It significantly outperforms the ResNet and traditional Transformer backbones of the same period, and achieves an excellent balance between speed and accuracy [25]. In conclusion, object detection based on Transformer also has significant advantages. For instance, it simplifies the detection process end-to-end, has global context awareness, can be extended to real-time and multi-task scenarios, and better handles sparse and occluded scenarios and so on.

3. Current Limitations and Future Prospects

Firstly, in the perception module of intelligent driving, deep models often face a trade-off between real-time performance and accuracy — high-precision networks such as large CNN or Transformer architectures require more computing resources, resulting in inference delay that is difficult to meet the closed-loop control requirements of 60 ms or less, while lightweight models often sacrifice some detection accuracy [26]. For example, although RT-BEV has achieved an improvement in BEV perception accuracy after designing the ROI-aware synchronizer, it still needs to repeatedly optimize between delay and throughput to ensure the real-time performance of vehicle safety decisions [27]. Secondly, transfer learning can reduce the annotation cost and training time in new scenarios through fine-tuning of pre-trained models, but its effect depends on the similarity of data distribution between the source domain and the target domain. If there are obvious domain differences (such as different climates and road characteristics), a large amount of on-site data is still required to restore performance [28]. Although federated learning provides a privacy protection solution for cross-fleet collaboration, in the context of communication bandwidth and heterogeneous device environments, the latency and accuracy degradation of model aggregation (due to noise introduced by differential privacy or compression strategies) remain urgent problems to be solved [29][30]. Next, let's talk about data and regulations. Firstly, the large amount of image, radar and path trajectory data collected by the autonomous driving system may contain sensitive information such as passenger positions and driving habits, which are vulnerable to the threats of model inversion and man-in-the-middle attacks. The existing encrypted storage and access control mechanisms are difficult to protect all risk points one by one in large-scale fleets [31]. Meanwhile, the laws and regulations concerning AV data have not yet fully kept pace with the industry's development: The EU's GDPR requires strict protection of personal location and biometric data, but lacks detailed guidance on cross-border data synchronization in the Internet of Vehicles. In the United States, it is jointly regulated by the Federal Trade Commission (FTC) and the state CCPA. The fragmentation of regulations has led to compliance challenges for global deployments [32][33]. Finally, regarding hardware and platforms, the current mainstream intelligent driving platforms are based on general-purpose GPU/CPU. Although they have strong computing universality, they have bottlenecks in terms of power consumption, thermal management and cost. Especially in the in-vehicle environment, the additional cooling system and high-power consumption can lead to a 10-15% increase in the overall vehicle energy consumption [34]. Specialized AI accelerators for the edge (such as NVIDIA DRIVE AGX Orin or Mobileye EyeQ) can provide hundreds of TOPS of inference capabilities, but the high cost and ecological lock-in make it difficult for small and medium-sized manufacturers to bear [35]. Though now intelligent driving is still in the face of many challenges and limitations, its future is still very bright. For example, by combining Neural Architecture Search (NAS), model pruning and dynamic reasoning technologies, high-precision real-time detection under gigabit-level computing power is achieved to build a lightweight end-to-end architecture, by using self-supervised learning to reduce the reliance on labeled data, and introducing adaptive quantization and asynchronous aggregation strategies in federated learning at the same time to balance privacy security and model performance, self-supervised transfer and efficient federated learning can be achieved, by promoting the formulation of AV data sharing and privacy protection norms by the International Organization

for Standardization, technical compatibility of regulations in multiple regions can be achieved, thereby establishing a unified cross-domain compliance framework, and by developing low-power dedicated accelerators and edge-cloud collaborative processing in coordination with 5G/6G V2X networks, the overall computing architecture is optimized, the energy consumption of the entire vehicle is reduced, and the system robustness is improved, thereby achieving heterogeneous hardware collaboration.

4. Conclusion

To sum up, intelligent driving is indeed one of the relatively hot technical topics at present, and it does have a very promising future. However, the challenges and difficulties it faces also follow one after another. From the explanation of its principles and methods in this article, it can be known that the improvement and enhancement of object detection methods can greatly improve the functions of intelligent driving. This means that in the future, if we want to quickly apply intelligent driving on a large scale to life and overcome the troubles it brings, the most crucial direction that needs to be studied is to improve the accuracy and real-time performance of object detection. This article also promotes the public's understanding of the advantages of intelligent driving by introducing its background, enabling people to comprehend the convenience it can bring and the most crucial technologies it relies on, and to understand the key factors driving intelligent driving. At the same time, it summarizes the difficulties and problems faced by intelligent driving, as well as its future development prospects, making people know that intelligent driving still needs improvement, but the prospects are bright.

References

[1] Guo T., Jiang Z., Ma H. Intelligent Driving Accelerates Forward. Economic Daily, 2020. https://cs.com.cn/cj2020/202503/t20250330_6482366.html.

[2] Fraedrich E., Heinrichs D., Bahamonde-Birke F. J., Cyganski R. Autonomous Driving, the Built Environment and Policy Implications. Transportation Research Part A: Policy and Practice, 2019, 122: 162–172.

[3] GeeksforGeeks. How Does R-CNN Work for Object Detection?, 2024 https://www.geeksforgeeks.org/ how-does-r-cnn-work-for-object-detection/.

[4] Zhou Y., Wen S., Wang D., Mu J., Richard I. Object Detection in Autonomous Driving Scenarios Based on an Improved Faster-RCNN. Applied Sciences, 2021, 11(24): 11630.

[5] Thompson E., Chen E. A Study on Obstacle Detection in Unmanned Driving Using an Improved Faster R-CNN Model. Journal of Computer Technology and Software, 2024, 3(5).

[6] Carranza-García M., Lara-Benítez P., García-Gutiérrez J., Riquelme J. C. Enhancing Object Detection for Autonomous Driving by Optimizing Anchor Generation and Addressing Class Imbalance. Neurocomputing, 2021, 449: 229–244.

[7] DataCamp. YOLO Object Detection Explained. https://www.datacamp.com/blog/yolo-object-detection-explained.

[8] Kundu R. YOLO: Algorithm for Object Detection Explained [+Examples]. V7 Labs. https://www.v7labs.com/blog/yolo-object-detection.

[9] Adarsh, P., Rathi, P., & Kumar, M. YOLO v3-Tiny: Object Detection and Recognition Using One Stage Improved Model. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 687-694). IEEE, 2020.

[10] Kandasamy S. YOLO for Computer Vision: A Deep Dive into Real-Time Object Detection. Juhomi, 2025. https://www.juhomi.com/post/yolo-for-computer-vision-a-deep-dive-into-real-time-object-detection.

[11] Sharma A. Understanding a Real-Time Object Detection Network: You Only Look Once (YOLOv1). PyImageSearch, 2022. https://pyimagesearch.com/2022/04/11/understanding-a-real-time-object-detection-network-you-only-look-once-yolov1/.

[12] Alif, M. A. R., & Hussain, M. YOLOv1 to YOLOv10: A Comprehensive Review of YOLO Variants and Their Application in the Agricultural Domain. arXiv preprint arXiv:2406.10139, 2024.

[13] Redmon, J., & Farhadi, A. Yolov3: An Incremental Improvement. arXiv preprint arXiv:1804.02767, 2018.

[14] Meel V. YOLOv3: Real-Time Object Detection Algorithm (Guide), 2022. https://viso.ai/deep-learning/yolov3-overview/.

[15] Padmanabula S. S., Puvvada R. C., Sistla V., Kolli V. K. K. Object Detection Using Stacked YOLOv3. Ingénierie des Systèmes d'Information, 2020, 25(5): 691–697.

[16] Solawetz J. What is YOLOv5? A Guide for Beginners, 2020. https://blog.roboflow.com/yolov5-improvements-and-evaluation/.

[17] NVIDIA Developer Forums. YOLOv3 is Very Slow. https://forums.developer.nvidia.com/t/yolov3-is-very-slow/74073.

[18] Ultralytics. Issue #2790: Inference Speed Improvements. https://github.com/ultralytics/yolov5/ issues/2790.

[19] Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-End Object Detection with Transformers. In: European Conference on Computer Vision. Cham: Springer, 2020: 213–229.

[20] Facebook Research. DETR: End-to-End Object Detection with Transformers. https://github.com/facebookresearch/detr.

[21] Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. arXiv preprint arXiv:2010.04159, 2020.

[22] Zhu X., Su W., Lu L., Li B., Wang X., Dai J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. In: International Conference on Learning Representations, 2021. https://iclr.cc/virtual/2021/oral/3448.

[23] Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Guo B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012–10022.

[24] Hugging Face. Swin Transformer—Model Doc. https://huggingface.co/docs/transformers/model_ doc/swin.

[25] Microsoft. Swin-Transformer. https://github.com/microsoft/Swin-Transformer.

[26] Spice B. New Perception Metric Balances Reaction Time, Accuracy: Both Elements Are Critical for Applications Such as Self-Driving Cars. Carnegie Mellon University, 2020. https://www.cs.cmu.edu/news/2020/new-perception-metric-balances-reaction-time-accuracy.

[27] Liu L., Lee J., Shin K. G. RT-BEV: Enhancing Real-Time BEV Perception for Autonomous Vehicles. In: 2024 IEEE Real-Time Systems Symposium (RTSS). IEEE, 2024: 267–279.

[28] Liu X., Li J., Ma J., Sun H., Xu Z., Zhang T., Yu H. Deep Transfer Learning for Intelligent Vehicle Perception: A Survey. Green Energy and Intelligent Transportastion, 2023, 2(12): 100125.

[29] Cui Y., Zhu J., Li J. FLAV: Federated Learning for Autonomous Vehicle Privacy Protection. Ad Hoc Networks, 2025, 166: 103685. https://doi.org/10.1016/j.adhoc.2024.103685.

[30] Xu H., Pan M., Huang X., Shen A. Federated Learning in Autonomous Vehicles Using Cross-Border Training. NVIDIA Developer Blog, 2024. https://developer.nvidia.com/blog/federated-learning-in-autonomous-vehicles-using-cross-border-training/.

[31] Fulbright, N. R. The Privacy Implications of Autonomous Vehicles. Data Protection Report, 2017.

[32] Mulder T., Vellinga N. E. Exploring Data Protection Challenges of Automated Driving. Computer Law & Security Review, 2021, 40: 105530. https://doi.org/10.1016/j.clsr.2021.105530.

[33] Brody A. J., Rolecki J. J., Stefan J. M. Navigating the Data Privacy Landscape for Autonomous and Connected Vehicles: Best Practices. The National Law Review, 2022. https://natlawreview.com/article/navigating-data-privacy-landscape-autonomous-and-connected-vehicles-best-practices.

[34] Xie J., Zhou X. Edge Computing for Real-Time Decision Making in Autonomous Driving: Review of Challenges, Solutions, and Future Trends. International Journal of Advanced Computer Science & Applications, 2024, 15(7).

[35] Rafie, M. Edge AI Computing Advancements Driving Autonomous Vehicle Potential. Global Semiconductor Alliance Tech Forum Article, 2021.