UDC 004.932: 004.896
DOI https://doi.org/10.32782/IT/2024-3-1

Stanislav AVRAMENKO
Postgraduate Student at the System Analysis and Control Department, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave, Dnipro, Ukraine, 49005, avramenko.st.y@nmu.one
ORCID: *0009-0001-7182-7852*

Timur ZHELDAK
Candidate of Technical Sciences, Associate Professor, Head of the System Analysis and Control Department, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave, Dnipro, Ukraine, 49005, zheldak.t.a@nmu.one
ORCID:*0000-0002-4728-5889*
Scopus Author ID: *55602208300*

To cite this article: Avramenko, S., Zheldak, T. (2024). Vyiavlennia obiektiv na osnovi hlybokoho navchannia dlia avtonomnoho keruvannia: zastosuvannia ta vidkryti problemy [Deep-learning based object detection for autonomous driving: aplications and open challenges]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 3, 3–13, doi: https://doi.org/10.32782/IT/2024-3-1

DEEP-LEARNING BASED OBJECT DETECTION FOR AUTONOMOUS DRIVING: APPLICATIONS AND OPEN CHALLENGES

Object detection is a critical component of autonomous driving systems, enabling accurate identification and localization of vehicles, pedestrians, cyclists, traffic signs, and other road objects. Deep learning techniques have revolutionized this field, propelling object detection capabilities to unprecedented levels. This paper presents a survey of state-of-the-art deep learning-based object detection methods tailored for autonomous driving applications using monocular camera input.

The purpose of this work is *to provide a unified perspective on modern deep learning approaches to object detection tailored for the unique requirements of autonomous driving. The monocular camera modality is chosen for its cost-effectiveness, widespread availability, and compatibility with existing automotive hardware. The focus is solely on deep learning techniques due to their ability to learn rich feature representations directly from data.*

The methodology *involves a systematic review of real-world applications and challenges, including pedestrian detection, traffic sign recognition, low-light conditions, and real-time performance requirements.*

The scientific novelty*. This survey consolidates the latest developments in camera-based object detection for autonomous driving, providing a comprehensive and up-to-date resource for researchers and practitioners. It offers insights into emerging techniques, such as attention mechanisms, multi-scale feature fusion, and model compression, which address critical challenges like occlusion handling, small object detection, and computational efficiency. Furthermore, the survey explores the potential of explainable AI and meta-learning techniques to enhance the transparency, interpretability, and generalization capabilities of object detectors in autonomous driving contexts.*

Conclusions*. Deep learning-based object detection has made significant strides in recent years, enabling robust and accurate perception for autonomous vehicles. However, challenges persist in real-world deployment, including handling diverse lighting conditions, adverse weather scenarios, and ensuring reliable performance under occlusions. This survey highlights promising research directions, such as incorporating attention mechanisms, temporal information, and multi-scale architectures, to address these challenges and pave the way for safer and more reliable autonomous driving systems.*

Key words: *object detection, autonomous driving, deep learning, transformers, attention mechanisms, occlusion handling, real-time performance.*

Станіслав АВРАМЕНКО
аспірант кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0009-0001-7182-7852*

Тімур ЖЕЛДАК
кандидат технічних наук, доцент, завідувач кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0002-4728-5889*
Scopus Author ID: *55602208300*

Бібліографічний опис статті: Авраменко, С., Желдак, Т. (2024). Виявлення об'єктів на основі глибокого навчання для автономного керування: застосування та відкриті проблеми. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 3–13, doi: https://doi.org/10.32782/IT/2024-3-1

ВИЯВЛЕННЯ ОБ'ЄКТІВ НА ОСНОВІ ГЛИБОКОГО НАВЧАННЯ ДЛЯ АВТОНОМНОГО КЕРУВАННЯ: ЗАСТОСУВАННЯ ТА ВІДКРИТІ ПРОБЛЕМИ

Виявлення об'єктів є критично важливим компонентом систем автономного водіння, що дозволяє точно ідентифікувати та локалізувати транспортні засоби, пішоходів, велосипедистів, дорожні знаки та інші дорожні об'єкти. Технології глибокого навчання зробили революцію в цій галузі, піднявши можливості виявлення об'єктів до безпрецедентного рівня. Ця стаття представляє огляд найсучасніших методів виявлення об'єктів на основі глибокого навчання, розроблених для додатків автономного водіння з використанням монокулярної камери.

Метою цієї роботи є *надання єдиного погляду на сучасні підходи глибокого навчання до виявлення об'єктів, спеціально розроблені для унікальних вимог автономного водіння. Модальність монокулярної камери обрано через її економічну ефективність, широку доступність і сумісність з існуючим автомобільним обладнанням. Основна увага зосереджена виключно на техніках глибокого навчання завдяки їхній здатності вивчати представлення багатьох властивостей безпосередньо з даних.*

Методологія *передбачає систематичний огляд реального застосування та проблем, включаючи виявлення пішоходів, розпізнавання дорожніх знаків, умови слабкого освітлення та вимоги до продуктивності в реальному часі.*

Наукова новизна*. Цей огляд об'єднує останні розробки в області виявлення об'єктів за допомогою камери для автономного водіння, надаючи всебічне і актуальне джерело для дослідників і практиків. Він пропонує розуміння нових методів, таких як механізми привернення уваги, багатомасштабне об'єднання властивостей і стиснення моделі, які вирішують критичні проблеми, такі як обробка загородження, виявлення малих об'єктів і ефективність обчислень. Крім того, огляд досліджує потенціал зрозумілого штучного інтелекту, і методів метанавчання для підвищення прозорості, інтерпретації та можливостей узагальнення детекторів об'єктів у контексті автономного водіння.*

Висновки*. За останні роки виявлення об'єктів на основі глибокого навчання досягло значних успіхів, забезпечивши надійне та точне сприйняття для автономних транспортних засобів. Однак під час розгортання в реальному світі залишаються проблеми, включаючи роботу з різними умовами освітлення, несприятливими погодними сценаріями та забезпечення надійної роботи в умовах загородження. Цей огляд підкреслює багатообіцяючі напрямки досліджень, такі як включення механізмів уваги, тимчасової інформації та багатомасштабних архітектур, щоб вирішити ці проблеми та прокласти шлях для безпечніших і надійніших систем автономного водіння.*

Ключові слова: *виявлення об'єктів, автономне водіння, глибоке навчання, трансформери, механізми уваги, загородження, продуктивність у реальному часі.*

Introduction. Autonomous driving systems (ADS) have garnered significant attention in recent years, driven by the promise of enhanced transportation safety, efficiency, and accessibility. At the core of these systems lies the critical task of object detection, which enables the accurate identification and localization of various elements in the surrounding environment, including vehicles, pedestrians, cyclists, traffic signs, and other road objects. Reliable object detection is paramount for safe navigation and decision-making in self-driving vehicles, as it provides the foundational awareness necessary to plan and execute appropriate actions.

The advent of deep learning has revolutionized the field of computer vision, propelling object detection capabilities to unprecedented levels. Modern deep learning-based object detectors have demonstrated remarkable accuracy and robustness, outperforming traditional computer vision techniques. However, the unique challenges posed by autonomous driving scenarios demand even higher standards of performance, reliability, and efficiency.

Autonomous vehicles must operate in dynamic and complex environments, where varying lighting conditions, adverse weather, occlusions, and diverse object appearances can significantly impact detection accuracy. Moreover, the real-time nature of autonomous driving necessitates object detectors capable of processing high-resolution video streams at high frame rates while maintaining low latency, ensuring timely decision-making and response.

By synthesizing the latest research efforts and insights, this paper aims to provide a comprehensive understanding of the state-of-the-art in deep learning-based object detection for autonomous driving and pave the way for future advancements in this critical domain.

Applications and Open Challenges.

Pedestrian detection. This component of autonomous driving systems is critical to ensure the safety of vulnerable road users. It enables self-driving vehicles to identify and track pedestrians in their surroundings, allowing them to make informed decisions and take appropriate actions to avoid collisions or other hazardous situations. Accurate pedestrian detection is essential for maintaining the trust and acceptance of autonomous vehicles by the public.

Despite significant advancements in computer vision and machine learning techniques, pedestrian detection for autonomous driving still faces several challenges. Occlusion, where pedestrians are partially obscured by other objects or vehicles, remains a significant hurdle. Varying lighting conditions, weather patterns, and diverse pedestrian appearances (clothing, posture, etc.) can also impact the reliability of detection algorithms. Additionally, distinguishing between static and moving pedestrians, predicting their intentions, and handling edge cases like crowded scenarios pose ongoing challenges.

(Lyssenko et al., 2024) address the safety-critical aspect of pedestrian detection in automated driving, where misdetections of critical pedestrians can endanger vulnerable road users. They introduce a safety-adapted loss function that leverages time-to-collision and distance information to quantify the criticality of pedestrians during training. Their approach aims to mitigate the misdetection of critical pedestrians without sacrificing overall performance.

Several papers discuss using the YOLO object detection algorithm for pedestrian detection in real-time scenarios. (Mishra & Jabin, 2023) and (Zuo et al., 2021) show that YOLO has proven effective at detecting and localizing objects in images with impressive speed. YOLO using transfer learning on pre-trained models is discussed by (Mishra & Jabin, 2023). (Zuo et al., 2021) evaluate different YOLO variants like YOLO-Tiny, YOLO, and YOLO-SPP, with YOLO and YOLO-SPP showing high average confidence, and YOLO-Tiny having fast detection speed suitable for real-time scenarios. The improved YOLO-R model proposed by (W. Lan et al., 2018) with added Passthrough layers can effectively improve pedestrian detection accuracy while reducing false and missed detections, achieving 25 FPS.

Other works focus on pedestrian detection using infrared images and multimodal data fusion techniques. (Wei et al., 2023) propose an approach using an improved UNet and YOLO network that shares visible light information from related datasets, achieving high detection accuracy on infrared pedestrian datasets, with a real-time speed of 25.6 FPS on edge devices. (Y. Zhang et al., 2022) introduce a lightweight vehicle-pedestrian detection algorithm based on YOLOv4 with a MobileNetv2 backbone, multi-scale feature fusion, and coordinate attention mechanism, improving accuracy and speed over the original YOLOv4. (Y. Chen et al., 2023) propose the TF-YOLO detector that uses a transformer-fusion module in a two-stream backbone to robustly integrate visible and infrared images, improving pedestrian detection performance under various illumination conditions compared to state-of-the-art approaches.

Pedestrian detection in crowded scenes poses challenges due to occlusion and scale variations. The recently proposed end-to-end detectors DETR and deformable DETR, based on transformer architectures, have shown promising results by avoiding hand-crafted components, but (Lin et al., 2021) found their performance surprisingly poor on crowd pedestrian detection compared to Faster-RCNN. Several works aim to address this issue. (Han et al., 2024) propose an improved deformable DETR (IDPD) with a dynamic neck and hybrid decoding loss to alleviate information loss and positive-negative imbalance. (Yuan et al., 2022) demonstrate the effectiveness of vision transformers for fast and accurate single-stage pedestrian detection by proposing a spatial and multi-scale feature enhancement module. (Deng & Li, 2024) introduce an efficient dense pedestrian detection algorithm using EfficientNet as the backbone, a cross-fertilization module for fusing multi-scale features, and noise removal training to improve detection of occluded and small-scale pedestrians while reducing model size and computation.

Moving object detection (MOD) is essential for identifying potential collision risks from dynamic objects in the environment. Several works have proposed deep learning approaches to tackle this problem by jointly modeling motion and appearance cues. (Siam et al., 2018) introduced MODNet, a two-stream architecture that combines object detection and motion segmentation for improved accuracy. (Yahiaoui et al., 2019) extended this idea to fisheye surround-view cameras with FisheyeMODNet. (Rashed, Essam, et al., 2021) explored end-to-end MOD in the bird's eye view (BEV) space using monocular images, demonstrating significant improvements over traditional inverse perspective mapping methods.

To further improve MOD performance, researchers have explored incorporating additional information into the models. VM-MODNet (Rashed, Sallab, et al., 2021) leverages vehicle motion information to enable ego-motion compensation, leading to substantial gains in accuracy. (Hernandez et al., 2020) fused semantic information from a deep learning detector with occupancy grid estimations to recognize moving objects. RST-MODNet (Ramzy et al., 2019) developed a real-time spatio-temporal architecture that exploits temporal motion information from sequential images and optical flow for increased robustness.

In addition to deep learning approaches, researchers have also explored alternative strategies for real-time MOD in autonomous driving systems. (Jha et al., 2021) introduced a system that combines object detection and tracking algorithms, adaptively controlling their execution cycles to ensure real-time performance in various edge computing environments. (Z. Zhou et al., 2023) proposed RENet, a novel RGB-Event fusion network that jointly exploits complementary modalities from conventional cameras and event cameras for more robust MOD under challenging scenarios. (D. Liu et al., 2020) presented MFCN, an end-to-end deep learning framework that leverages temporal coherence and motion patterns in video features for improved object detection accuracy while maintaining efficiency.

Traffic sign detection task faces challenges such as multi-scale targets and real-time performance requirements. Several studies have aimed to improve the detection accuracy and speed for multi-scale traffic signs by enhancing the feature pyramid network of object detectors like YOLOv5. (J. Wang et al., 2021) proposed an adaptive attention module and feature enhancement module in the feature pyramid to reduce information loss and enhance representation ability. The ETSR-YOLO algorithm (H. Liu et al., 2023) generated an additional high-resolution feature layer and introduced improved C3 modules to suppress background noise and enhance feature extraction. (T. Chen & Ren, 2023) designed a cross-level loss function to enable each level of the MFL-YOLO model to learn diverse features and improve fine-grained details for detecting damaged traffic signs.

Other approaches have focused on improving the performance of object detectors like SSD for traffic sign detection. (You et al., 2020) proposed a lightweight SSD network with 1x1 convolution kernels and color-based filtering to improve detection speed while maintaining accuracy. (Wu & Liao, 2022) combined SSD with a receptive field module and path aggregation network to improve small traffic sign detection and integrate multi-scale features. (Greer, Gopalkrishnan, Deo, et al., 2023; Greer, Gopalkrishnan, Landgren, et al., 2023) defined the concept of «salient» traffic signs/lights that influence driver decisions and trained Deformable DETR models with a salience-sensitive loss function to emphasize performance on these salient objects.

Some studies have explored alternative architectures like transformers for traffic sign recognition. (Farzipour et al., 2023) proposed a hybrid model combining convolutional and transformer-based blocks with a locality module to capture local and global features, achieving high accuracy on traffic sign datasets. The DSRA-DETR algorithm (Xia et al., 2023) introduced dilated spatial pyramid pooling and multi-scale feature residual aggregation to improve multi-scale traffic sign detection with DETR. (S. Chen et al., 2024) presented a semi-supervised learning framework combining CNN and multi-scale transformer with hierarchical sampling and local/global information aggregation for accurate traffic sign detection and recognition from vehicle panoramic images.

Semi-supervised object detection methods leverage both labeled and unlabeled data to improve performance, and they are widely used in autonomous driving systems where only a fraction of objects are labeled (Hu et al., 2022). These methods generate pseudo-labels for unlabeled objects, which can significantly improve performance but also introduce noise and errors, especially for video data (W. Chen et al., 2024; Hu et al., 2022). Approaches have been proposed to generate more robust pseudo-labels by leveraging motion continuity in video frames (Hu et al., 2022) and using semi-supervised co-training with unsupervised data augmentation to improve generalization and robustness under adversarial attacks (W. Chen et al., 2024). Additionally, methods have been developed to correct and refine pseudo-labels to reduce classification and localization noise (He et al., 2023), as well as to enhance the quality of object queries and selectively filter high-quality pseudo-labels in transformer-based object detection models (Shehzadi et al., 2024). These advancements in semi-supervised object detection aim to improve accuracy, consistency, and robustness, particularly for challenging scenarios involving small or occluded objects in autonomous driving applications.

Real-time object detection. Recent advances in real-time object detection for autonomous driving have focused on developing efficient and accurate models. (Miraliev et al., 2024) propose a real-time memory-efficient multitask learning model for joint object detection, drivable area segmentation, and lane detection, achieving high accuracy and a processing speed of 112.29 fps. (Mahaur et al., 2023) introduce architectural modifications to the

YOLOv5 model, including group depthwise separable convolutions and attention-based dilated blocks, improving small object detection accuracy by 8.35% on the BDD100K dataset while increasing speed by 3.1%. The DPNet algorithm (Q. Zhou et al., 2022) presents a dual-path network with a lightweight attention scheme, enabling parallel extraction of high-level semantic features and low-level object details, achieving state-of-the-art trade-off between detection accuracy and efficiency. (A. Wang et al., 2024) introduce YOLOv10, a real-time end-to-end object detector with consistent dual assignments for NMS-free training and holistic efficiency-accuracy driven model design, significantly reducing computational overhead and enhancing capability compared to previous YOLOs. (Zhao et al., 2024) propose RT-DETR, the first real-time end-to-end object detector, featuring an efficient hybrid encoder and uncertainty-minimal query selection, outperforming advanced YOLOs in both speed and accuracy on the COCO dataset.

Low light condition. Operate seamlessly 24/7 without being limited by nighttime or poor visibility scenarios is crucial for enabling safe and reliable autonomous driving in all environments and conditions. (X. Wang et al., 2022) and Pham et al. (Pham et al., 2020) propose methods to enhance low-light images and improve object detection accuracy, which is crucial for applications like traffic monitoring and autonomous driving. (X. Wang et al., 2022) utilize dark channel prior and adaptive gamma transformation to restore scene radiance and develop the LL4PH-Net framework for low-light traffic object detection. (Pham et al., 2020) introduce DriveRetinex-Net, a deep retinex neural network trained on a low-light driving dataset (LOL-Drive), which decomposes images into reflectance and illumination maps, enhancing the latter for improved object detection.

Several other approaches tackle the challenge of object detection in adverse weather and low-light conditions for autonomous driving. The IDOD-YOLOV7 algorithm (Qiu et al., 2023) jointly optimizes image defogging (AOD) and enhancement (SAIP) modules with YOLOV7 detection, improving perception in low-light foggy environments. (W. Liu et al., 2022) propose Image-Adaptive YOLO (IA-YOLO), where a differentiable image processing module adaptively enhances images for better object detection. (Guo et al., 2024) introduce HawkDrive, a transformer-based visual perception system with stereo vision and edge computing for depth estimation and semantic segmentation in night scenes. (Ye et al., 2024) propose VELIE, a vehicle-based efficient low-light image enhancement method using Swin Vision Transformer and U-Net for real-time inference and edge deployment in intelligent vehicles.

Overview of Modern Techniques.

Modern deep learning-based object detection systems for autonomous driving must possess exceptional accuracy in detecting and precisely localizing a wide range of objects, including vehicles, pedestrians, cyclists, traffic signs, and other road elements. They should demonstrate this high level of accuracy under diverse lighting conditions, varying weather scenarios, and situations with partial occlusion. Additionally, these systems need to deliver real-time performance, capable of processing high-resolution video streams at high frame rates while maintaining low latency to enable timely decision-making and response. Furthermore, robustness and reliability are crucial, ensuring safe and consistent operation even in challenging scenarios. There are some approaches that may help to meet these requirements.

Incorporating attention mechanisms into object detectors can help them focus on relevant regions and better handle occlusions. (S. Zhang et al., 2021) propose an attention mechanism across CNN channels to represent various occlusion patterns for pedestrian detection and re-identification, employing attention guided self-paced learning to balance optimization across different occlusion levels. (Zou et al., 2020) proposed an attention guided neural network model (AGNN) that uses an attention mechanism to selectively weight and integrate features from sub-images representing body parts of occluded pedestrians for improved detection performance.

Leveraging temporal information from video sequences can help in tracking and predicting the movement of occluded objects. It was proposed a novel spatio-temporal fusion Transformer (STFT) (Qi et al., 2024) model that incorporates a dynamic template update strategy based on salient points feature representation, an IoU-Aware target state estimation head, and an IoU-Aware criterion for robust thermal infrared object tracking to address the limitations of existing approaches in handling scale variations, appearance changes, and occlusions.

Incorporating FPNs or similar multi-scale architectures can help object detectors capture both fine-grained and contextual information, improving small object detection. (Huang et al., 2023) propose the Multiple Link Feature Pyramid Networks (MLFPN) with novel information transfer pathways and Poly-Scale Convolution (PSconv) to reduce feature information loss and improve pedestrian detection. (Yahya et al., 2023) introduce the Fast Region-Convolutional Neural Network (R-CNN) with the Attention-guided Context Feature Pyramid Network (ACFPN) for object detection in autonomous vehicles, achieving better mean Average Precision (mAP). (Dang et al., 2023) propose the hierarchical attention feature pyramid network (HA-FPN), comprising transformer feature pyramid networks (TFPNs) that apply self-attention across scales to capture contextual information, and channel attention modules (CAMs) that select channels with rich information to improve bounding box detection accuracy and object localization, while minimizing computational overhead. (Xie et al., 2022) proposed FocusTR (Focusing on the valuable features by multiple Transformers) architecture that presents novel self-attention mechanisms, including spatial-wise boxAlign attention, context-wise affinity attention, and level-wise attention, along with low and high-level fusion and Pre-Ln, to effectively fuse multi-level and multi-sensor feature pyramids for object detection in autonomous driving.

Model Compression and Quantization techniques like model pruning, quantization, and knowledge distillation can significantly reduce the computational complexity and memory footprint of object detectors. (H. Liu et al., 2021) and (Youn et al., 2023) explore knowledge distillation as a technique for compressing large neural networks like vision transformers and convolutional neural networks into smaller and more efficient models suitable for resource-constrained devices such as autonomous vehicles. The core idea involves training a smaller «student» model to mimic the behavior of a larger and more accurate «teacher» model, transferring knowledge from the teacher to the student. (Q. Lan & Tian, 2023) and (Agand, 2024) propose enhancing this process with techniques like model pruning, quantization, and adaptive instance/scale-wise distillation.

In autonomous driving contexts, Flexi-Compression (H. Liu et al., 2021), Quasar-ViT (Li et al., 2024), and the approach by (Youn et al., 2023) focus on compressing vision transformers and convolutional neural networks used for object detection, semantic segmentation, and navigation by employing architectural modifications, hardware-aware neural architecture search, and combining distillation with pruning and quantization. Additionally, (Q. Lan & Tian, 2023) introduce multi-teacher adaptive instance distillation, while (Agand, 2024) explores an end-to-end transformer-based sensor fusion method using knowledge distillation to improve performance and handle challenging scenarios in autonomous driving.

Incorporating meta-learning techniques, which learn to adapt to new domains quickly, can enhance the generalization capabilities of object detectors. (Sun et al., 2024) propose a transformer-based few-shot object detection approach for traffic scenarios. It employs class-agnostic training to extend the detector to novel classes and combines visual prompts with pseudo-class embeddings to improve query generation. This approach does not require retraining during inference and accurately localizes novel objects through an improved query generation mechanism.

Integrating explainable AI techniques can provide local explanations for individual predictions made by object detectors and improve transparency and interpretability. (Dong et al., 2023) propose a novel approach to enhance trustworthiness in autonomous driving systems through explainable deep learning models. Instead of treating decision-making as a classification task, it frames it as an image-based language generation (image captioning) task, where the model generates textual descriptions of driving scenarios to serve as explanations for its decisions. (Cultrera, Luca., 2023) presents an approach to autonomous driving based on imitation learning using visual attention mechanisms. By selectively weighting and prioritizing relevant regions in the image, this method aims to mimic the human approach to driving and enhance interpretability and explainability. (Adom & Mahmoud, 2024) introduce RB-XAI, a Relevance-Based Explainable AI algorithm that uses Concept Relevance Propagation (CRP) to provide transparent concept-level explanations for the behavior of object detection models used in autonomous vehicles. CRP generates explanations that automatically identify and visualize relevant concepts within the input space, shedding light on the crucial areas responsible for the models' decisions. (Kolekar et al., 2022) propose an explainable inception-based U-Net model with Grad-CAM visualization for semantic segmentation in unstructured traffic environments on Indian roads. The inception U-Net model combines an inception-based module as an encoder for automatic feature extraction and a decoder for reconstructing the segmentation feature map. Grad-CAM is used to interpret the deep learning model, increasing consumer trust by providing visual explanations.

Conclusions. The field of object detection for autonomous driving has seen remarkable progress, with innovative architectures demonstrating state-of-the-art performance. However, several challenges remain, including handling occlusions, varying lighting conditions, diverse object appearances, and achieving real-time performance with high accuracy and robustness.

To tackle these challenges, future research should prioritize integrating attention mechanisms to focus on relevant regions, leveraging temporal information from video for tracking occluded objects, and incorporating multi-scale architectures like FPNs to capture fine-grained and contextual information. Additionally, model compression techniques like knowledge distillation, quantization, and pruning can reduce computational complexity for efficient inference. Incorporating meta-learning can enhance generalization to new domains, while explainable AI techniques can provide local explanations, improving transparency and interpretability. Other promising directions include multi-task learning for joint perception tasks, sensor fusion approaches leveraging complementary modalities, and semi-supervised methods to leverage unlabeled data. Ultimately, continued research efforts in these areas, coupled with hardware advancements, will pave the way for safe and reliable autonomous driving by enabling efficient, accurate, and robust object detection systems capable of operating in diverse real-world environment.

BIBLIOGRAPHY:

1. Lyssenko M. A safety-adapted loss for pedestrian detection in automated driving / M. Lyssenko, P. Pimplikar, M. Bieshaar, [et al.]. arXiv, 2024.
2. Mishra S. Real-time pedestrian detection using yolo / S. Mishra, S. Jabin. 2023.
3. Zuo X. Pedestrian detection based on one-stage yolo algorithm / X. Zuo, J. Li, J. Huang, [et al.]. *Journal of Physics: Conference Series.* 2021. Vol. 1871, No. 1. P. 012131.
4. Lan W. Pedestrian detection based on yolo network model. W. Lan, J. Dang, Y. Wang, S. Wang. 2018.
5. Wei J. Infrared pedestrian detection using improved unet and yolo through sharing visible light domain information. J. Wei, S. Su, Z. Zhao, [et al.]. Measurement. 2023. Vol. 221. P. 113442.
6. Zhang Y. A lightweight vehicle-pedestrian detection algorithm based on attention mechanism in traffic scenarios / Y. Zhang, A. Zhou, F. Zhao, H. Wu. Sensors (Basel, Switzerland). 2022. Vol. 22, No. 21. P. 8480.
7. Chen Y. TF-yolo: a transformer–fusion-based yolo detector for multimodal pedestrian detection in autonomous driving scenes / Y. Chen, J. Ye, X. Wan. World Electric Vehicle Journal. 2023. Vol. 14, No. 12. P. 352.
8. Lin M. DETR for crowd pedestrian detection / M. Lin, C. Li, X. Bu, [et al.]. arXiv, 2021.
9. Han W. IDPD: improved deformable-detr for crowd pedestrian detection / W. Han, N. He, X. Wang, [et al.]. Signal, Image and Video Processing. 2024. Vol. 18, No. 3. P. 2243–2253.
10. Yuan J. Effectiveness of vision transformer for fast and accurate single-stage pedestrian detection / J. Yuan, P. Barmpoutis, T. Stathaki. Advances in Neural Information Processing Systems. 2022. Vol. 35. P. 27427–27440.
11. Deng S. Efficient dense pedestrian detection based on transformer / S. Deng, J. Li. 2024.
12. Siam M. MODNet: motion and appearance based moving object detection network for autonomous driving / M. Siam, H. Mahgoub, M. Zahran, [et al.]. 2018.
13. Yahiaoui M. FisheyeMODNet: moving object detection on surround-view cameras for autonomous driving / M. Yahiaoui, H. Rashed, L. Mariotti, [et al.]. arXiv, 2019.
14. Rashed H. BEV-modnet: monocular camera based bird's eye view moving object detection for autonomous driving / H. Rashed, M. Essam, M. Mohamed, [et al.]. 2021.
15. Rashed H. VM-modnet: vehicle motion aware moving object detection for autonomous driving / H. Rashed, A. E. Sallab, S. Yogamani. 2021.
16. Hernandez A. E. G. Recognize moving objects around an autonomous vehicle considering a deep-learning detector model and dynamic bayesian occupancy / A. E. G. Hernandez, O. Erkent, C. Laugier. Shenzhen, China: IEEE, 2020.
17. Ramzy M. RST-modnet: real-time spatio-temporal moving object detection for autonomous driving / M. Ramzy, H. Rashed, A. El Sallab, S. Yogamani. 2019.
18. Jha S. Real time object detection and trackingsystem for video surveillance system. S. Jha, C. Seo, E. Yang, G. P. Joshi. Multimedia Tools and Applications. 2021. Vol. 80, No. 3. P. 3981–3996.
19. Zhou Z. RGB-event fusion for moving object detection in autonomous driving / Z. Zhou, Z. Wu, R. Boutteau, [et al.]. arXiv, 2023.
20. Liu D. Video object detection for autonomous driving: motion-aid feature calibration / D. Liu, Y. Cui, Y. Chen, [et al.]. Neurocomputing. 2020. Vol. 409. P. 1–11.
21. Wang J. Improved yolov5 network for real-time multi-scale traffic sign detection / J. Wang, Y. Chen, M. Gao, Z. Dong. arXiv, 2021.

22. Liu H. ETSR-yolo: an improved multi-scale traffic sign detection algorithm based on yolov5 / H. Liu, K. Zhou, Y. Zhang, Y. Zhang. PLOS ONE. 2023. Vol. 18, No. 12. P. e0295807.

23. Chen T. MFL-yolo: an object detection model for damaged traffic signs / T. Chen, J. Ren. arXiv, 2023.

24. You S. Traffic sign detection method based on improved ssd / S. You, Q. Bi, Y. Ji, [et al.]. Information. 2020. Vol. 11, No. 10. P. 475.

25. Wu J. Traffic sign detection based on ssd combined with receptive field module and path aggregation network / J. Wu, S. Liao. Computational Intelligence and Neuroscience. 2022. Vol. 2022. P. 4285436.

26. Greer R. Salient sign detection in safe autonomous driving: ai which reasons over full visual context / R. Greer, A. Gopalkrishnan, N. Deo, [et al.]. arXiv, 2023.

27. Greer R. Robust traffic light detection using salience-sensitive loss: computational framework and evaluations / R. Greer, A. Gopalkrishnan, J. Landgren, [et al.]. arXiv, 2023.

28. Farzipour A. Traffic sign recognition using local vision transformer / A. Farzipour, O. N. Manzari, S. B. Shokouhi. arXiv, 2023.

29. Xia J. DSRA-detr: an improved detr for multiscale traffic sign detection / J. Xia, M. Li, W. Liu, X. Chen. Sustainability. 2023. Vol. 15, No. 14. P. 10862.

30. Chen S. A semi-supervised learning framework combining cnn and multiscale transformer for traffic sign detection and recognition / S. Chen, Z. Zhang, L. Zhang, [et al.]. *IEEE Internet of Things Journal*. 2024. Vol. 11, No. 11. P. 19500–19519.

31. Hu S. PseudoProp: robust pseudo-label generation for semi-supervised object detection in autonomous driving systems / S. Hu, C.-H. Liu, J. Dutta, [et al.]. 2022.

32. Chen W. Robust object detection for autonomous driving based on semi-supervised learning / W. Chen, J. Yan, W. Huang, [et al.]. Security and Safety. 2024. Vol. 3. P. 2024002.

33. He Y. Pseudo-label correction and learning for semi-supervised object detection / Y. He, W. Chen, K. Liang, [et al.]. arXiv, 2023.

34. Shehzadi T. Sparse semi-detr: sparse learnable queries for semi-supervised object detection / T. Shehzadi, K. A. Hashmi, D. Stricker, M. Z. Afzal. arXiv, 2024.

35. Miraliev S. Real-time memory efficient multitask learning model for autonomous driving / S. Miraliev, S. Abdigapporov, V. Kakani, H. Kim. IEEE Transactions on Intelligent Vehicles. 2024. Vol. 9, No. 1. P. 247–258.

36. Mahaur B. An improved lightweight small object detection framework applied to real-time autonomous driving / B. Mahaur, K. K. Mishra, A. Kumar. Expert Systems with Applications. 2023. Vol. 234. P. 121036.

37. Zhou Q. DPNet: dual-path network for real-time object detection with lightweight attention / Q. Zhou, H. Shi, W. Xiang, [et al.]. arXiv, 2022.

38. Wang A. YOLOv10: real-time end-to-end object detection / A. Wang, H. Chen, L. Liu, [et al.]. arXiv, 2024.

39. Zhao Y. DETRs beat yolos on real-time object detection / Y. Zhao, W. Lv, S. Xu, [et al.]. arXiv, 2024.

40. Wang X. Low-light traffic objects detection for automated vehicles / X. Wang, D. Wang, S. Li, [et al.]. 2022.

41. Pham L. H. Low-light image enhancement for autonomous driving systems using driveretinex-net / L. H. Pham, D. N.-N. Tran, J. W. Jeon. 2020.

42. Qiu Y. IDOD-yolov7: image-dehazing yolov7 for object detection in low-light foggy traffic environments / Y. Qiu, Y. Lu, Y. Wang, H. Jiang. Sensors. 2023. Vol. 23, No. 3. P. 1347.

43. Liu W. Image-adaptive yolo for object detection in adverse weather conditions / W. Liu, G. Ren, R. Yu, [et al.]. arXiv, 2022.

44. Guo Z. HawkDrive: a transformer-driven visual perception system for autonomous driving in night scene / Z. Guo, S. Perminov, M. Konenkov, D. Tsetserukou. arXiv, 2024.

45. Ye L. VELIE: a vehicle-based efficient low-light image enhancement method for intelligent vehicles / L. Ye, D. Wang, D. Yang, [et al.]. Sensors. 2024. Vol. 24, No. 4. P. 1345.

46. Zhang S. Guided attention in cnns for occluded pedestrian detection and re-identification / S. Zhang, D. Chen, J. Yang, B. Schiele. *International Journal of Computer Vision.* 2021. Vol. 129, No. 6. P. 1875–1892.

47. Zou T. Attention guided neural network models for occluded pedestrian detection / T. Zou, S. Yang, Y. Zhang, M. Ye. Pattern Recognition Letters. 2020. Vol. 131. P. 91–97.

48. Qi M. Exploring reliable infrared object tracking with spatio-temporal fusion transformer / M. Qi, Q. Wang, S. Zhuang, [et al.]. Knowledge-Based Systems. 2024. Vol. 284. P. 111234.

49. Huang T. Dense pedestrian detection based on multiple link feature pyramid networks / T. Huang, S. Yang, T. Yu, X. Fu. Hangzhou, China: SPIE, 2023.

50. Yahya M. Object detection and recognition in autonomous vehicles using fast region-convolutional neural network / M. Yahya, R. A. Reddy, K. Al-Attabi, [et al.]. 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS). 2023. P. 1–5.

51. Dang J. HA-fpn: hierarchical attention feature pyramid network for object detection / J. Dang, X. Tang, S. Li. Sensors. 2023. Vol. 23, No. 9. P. 4508.

52. Xie B. FocusTR: focusing on valuable feature by multiple transformers for fusing feature pyramid on object detection / B. Xie, L. Yang, Z. Yang, [et al.]. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2022. P. 518–525.

53. Liu H. Flexi-compression: a flexible model compression method for autonomous driving / H. Liu, Y. He, F. R. Yu, J. James. Proceedings of the 11th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications. 2021. P. 19–26.

54. Youn E. Compressing vision transformers for low-resource visual learning / E. Youn, S. M. J, S. Prabhu, S. Chen. 2023.

55. Lan Q. Instance, scale, and teacher adaptive knowledge distillation for visual detection in autonomous driving / Q. Lan, Q. Tian. IEEE Transactions on Intelligent Vehicles. 2023. Vol. 8, No. 3. P. 2358–2370.

56. Agand P. Knowledge distillation from single-task teachers to multi-task student for end-to-end autonomous driving / P. Agand. Proceedings of the AAAI Conference on Artificial Intelligence. 2024. Vol. 38, No. 21. P. 23375–23376.

57. Li Z. Quasar-vit: hardware-oriented quantization-aware architecture search for vision transformers / Z. Li, A. Lu, Y. Xie, [et al.]. 2024.

58. Sun E. Transformer-based few-shot object detection in traffic scenarios / E. Sun, D. Zhou, Y. Tian, [et al.]. Applied Intelligence. 2024. Vol. 54, No. 1. P. 947–958.

59. Dong J. Why did the ai make that decision? towards an explainable artificial intelligence (xai) for autonomous driving systems / J. Dong, S. Chen, M. Miralinaghi, [et al.]. Transportation Research Part C: Emerging Technologies. 2023. Vol. 156. P. 104358.

60. Cultrera, Luca. Visual attention and explainability in end-to-end autonomous driving / Cultrera, Luca. 2023.

61. Adom I. RB-xai: relevance-based explainable ai for traffic detection in autonomous systems / I. Adom, M. N. Mahmoud. SoutheastCon 2024. P. 1358–1367.

62. Kolekar S. Explainable ai in scene understanding for autonomous vehicles in unstructured traffic environments on indian roads using the inception u-net model with grad-cam visualization / S. Kolekar, S. Gite, B. Pradhan, A. Alamri. Sensors. 2022. Vol. 22, No. 24. P. 9677.

REFERENCES:

1. Lyssenko, M. A. (2024). safety-adapted loss for pedestrian detection in automated driving / M. Lyssenko, P. Pimplikar, M. Bieshaar, [et al.]. arXiv

2. Mishra, S. (2023). Real-time pedestrian detection using yolo / S. Mishra, S. Jabin.

3. Zuo, X. (2021). Pedestrian detection based on one-stage yolo algorithm / X. Zuo, J. Li, J. Huang, [et al.]. *Journal of Physics: Conference Series*. Vol. 1871, No. 1. P. 012131.

4. Lan, W. (2018). Pedestrian detection based on yolo network model / W. Lan, J. Dang, Y. Wang, S. Wang.

5. Wei, J. (2023). Infrared pedestrian detection using improved unet and yolo through sharing visible light domain information / J. Wei, S. Su, Z. Zhao, [et al.]. Measurement. Vol. 221. P. 113442.

6. Zhang, Y. A. (2022). lightweight vehicle-pedestrian detection algorithm based on attention mechanism in traffic scenarios / Y. Zhang, A. Zhou, F. Zhao, H. Wu. Sensors (Basel, Switzerland). Vol. 22, No. 21. P. 8480.

7. Chen, Y. (2023). TF-yolo: a transformer–fusion-based yolo detector for multimodal pedestrian detection in autonomous driving scenes / Y. Chen, J. Ye, X. Wan. World Electric Vehicle Journal. Vol. 14, No. 12. P. 352.

8. Lin, M. (2021). DETR for crowd pedestrian detection / M. Lin, C. Li, X. Bu, [et al.]. arXiv,

9. Han, W. (2024). IDPD: improved deformable-detr for crowd pedestrian detection / W. Han, N. He, X. Wang, [et al.]. Signal, Image and Video Processing. Vol. 18, No. 3. P. 2243–2253.

10. Yuan, J. (2022). Effectiveness of vision transformer for fast and accurate single-stage pedestrian detection / J. Yuan, P. Barmpoutis, T. Stathaki. Advances in Neural Information Processing Systems. Vol. 35. P. 27427–27440.

11. Deng, S. (2024). Efficient dense pedestrian detection based on transformer / S. Deng, J. Li.

12. Siam, M. (2018). MODNet: motion and appearance based moving object detection network for autonomous driving / M. Siam, H. Mahgoub, M. Zahran, [et al.].

13. Yahiaoui, M. (2019). FisheyeMODNet: moving object detection on surround-view cameras for autonomous driving / M. Yahiaoui, H. Rashed, L. Mariotti, [et al.]. arXiv,

14. Rashed, H. (2021). BEV-modnet: monocular camera based bird's eye view moving object detection for autonomous driving / H. Rashed, M. Essam, M. Mohamed, [et al.].

15. Rashed, H. (2021). VM-modnet: vehicle motion aware moving object detection for autonomous driving / H. Rashed, A. E. Sallab, S. Yogamani.

16. Hernandez, A. E. G. (2020). Recognize moving objects around an autonomous vehicle considering a deep-learning detector model and dynamic bayesian occupancy / A. E. G. Hernandez, O. Erkent, C. Laugier. Shenzhen, China: IEEE.

17. Ramzy, M. (2019). RST-modnet: real-time spatio-temporal moving object detection for autonomous driving / M. Ramzy, H. Rashed, A. El Sallab, S. Yogamani.

18. Jha, S. (2021). Real time object detection and trackingsystem for video surveillance system / S. Jha, C. Seo, E. Yang, G. P. Joshi. Multimedia Tools and Applications. Vol. 80, No. 3. P. 3981–3996.

19. Zhou, Z. (2023). RGB-event fusion for moving object detection in autonomous driving / Z. Zhou, Z. Wu, R. Boutteau, [et al.]. arXiv,

20. Liu, D. (2020). Video object detection for autonomous driving: motion-aid feature calibration / D. Liu, Y. Cui, Y. Chen, [et al.]. Neurocomputing. Vol. 409. P. 1–11.

21. Wang, J. (2021). Improved yolov5 network for real-time multi-scale traffic sign detection / J. Wang, Y. Chen, M. Gao, Z. Dong. arXiv,

22. Liu, H. (2023). ETSR-yolo: an improved multi-scale traffic sign detection algorithm based on yolov5 / H. Liu, K. Zhou, Y. Zhang, Y. Zhang. PLOS ONE. Vol. 18, No. 12. P. e0295807.

23. Chen, T. (2023). MFL-yolo: an object detection model for damaged traffic signs / T. Chen, J. Ren. arXiv,

24. You, S. (2020). Traffic sign detection method based on improved ssd / S. You, Q. Bi, Y. Ji, [et al.]. Information. Vol. 11, No. 10. P. 475.

25. Wu, J. (2022). Traffic sign detection based on ssd combined with receptive field module and path aggregation network / J. Wu, S. Liao. Computational Intelligence and Neuroscience. Vol. 2022. P. 4285436.

26. Greer, R. (2023). Salient sign detection in safe autonomous driving: ai which reasons over full visual context. R. Greer, A. Gopalkrishnan, N. Deo, [et al.]. arXiv,

27. Greer, R. (2023). Robust traffic light detection using salience-sensitive loss: computational framework and evaluations. R. Greer, A. Gopalkrishnan, J. Landgren, [et al.]. arXiv,

28. Farzipour, A. (2023). Traffic sign recognition using local vision transformer / A. Farzipour, O. N. Manzari, S. B. Shokouhi. arXiv.

29. Xia, J. (2023). DSRA-detr: an improved detr for multiscale traffic sign detection / J. Xia, M. Li, W. Liu, X. Chen. Sustainability. Vol. 15, No. 14. P. 10862.

30. Chen, S. (2024). A semi-supervised learning framework combining cnn and multiscale transformer for traffic sign detection and recognition / S. Chen, Z. Zhang, L. Zhang, [et al.]. IEEE Internet of Things Journal. Vol. 11, No. 11. P. 19500–19519.

31. Hu, S. (2022). PseudoProp: robust pseudo-label generation for semi-supervised object detection in autonomous driving systems. S. Hu, C.-H. Liu, J. Dutta, [et al.].

32. Chen, W. (2024). Robust object detection for autonomous driving based on semi-supervised learning / W. Chen, J. Yan, W. Huang, [et al.]. Security and Safety. Vol. 3. P. 2024002.

33. He, Y. (2023). Pseudo-label correction and learning for semi-supervised object detection / Y. He, W. Chen, K. Liang, [et al.]. arXiv,

34. Shehzadi, T. (2024). Sparse semi-detr: sparse learnable queries for semi-supervised object detection / T. Shehzadi, K. A. Hashmi, D. Stricker, M. Z. Afzal. arXiv,

35. Miraliev, S. (2024). Real-time memory efficient multitask learning model for autonomous driving / S. Miraliev, S. Abdigapporov, V. Kakani, H. Kim. IEEE Transactions on Intelligent Vehicles. Vol. 9, No. 1. P. 247–258.

36. Mahaur, B. (2023). An improved lightweight small object detection framework applied to real-time autonomous driving / B. Mahaur, K. K. Mishra, A. Kumar. Expert Systems with Applications. Vol. 234. P. 121036.

37. Zhou, Q. (2022). DPNet: dual-path network for real-time object detection with lightweight attention / Q. Zhou, H. Shi, W. Xiang, [et al.]. arXiv,

38. Wang, A. (2024). YOLOv10: real-time end-to-end object detection / A. Wang, H. Chen, L. Liu, [et al.]. arXiv,

39. Zhao, Y. (2024). DETRs beat yolos on real-time object detection / Y. Zhao, W. Lv, S. Xu, [et al.]. arXiv,

40. Wang, X. (2022). Low-light traffic objects detection for automated vehicles / X. Wang, D. Wang, S. Li, [et al.].

41. Pham, L. H. (2020). Low-light image enhancement for autonomous driving systems using driveretinex-net / L. H. Pham, D. N.-N. Tran, J. W. Jeon.

42. Qiu, Y. (2023). IDOD-yolov7: image-dehazing yolov7 for object detection in low-light foggy traffic environments / Y. Qiu, Y. Lu, Y. Wang, H. Jiang. Sensors. Vol. 23, No. 3. P. 1347.

43. Liu, W. (2022). Image-adaptive yolo for object detection in adverse weather conditions / W. Liu, G. Ren, R. Yu, [et al.]. arXiv.

44. Guo, Z. (2024). HawkDrive: a transformer-driven visual perception system for autonomous driving in night scene / Z. Guo, S. Perminov, M. Konenkov, D. Tsetserukou. arXiv.

45. Ye, L. (2024). VELIE: a vehicle-based efficient low-light image enhancement method for intelligent vehicles / L. Ye, D. Wang, D. Yang, [et al.]. Sensors. Vol. 24, No. 4. P. 1345.

46. Zhang, S. (2021). Guided attention in cnns for occluded pedestrian detection and re-identification / S. Zhang, D. Chen, J. Yang, B. Schiele. *International Journal of Computer Vision*. Vol. 129, No. 6. P. 1875–1892.

47. Zou, T. (2020). Attention guided neural network models for occluded pedestrian detection / T. Zou, S. Yang, Y. Zhang, M. Ye. Pattern Recognition Letters. Vol. 131. P. 91–97.

48. Qi, M. (2024). Exploring reliable infrared object tracking with spatio-temporal fusion transformer / M. Qi, Q. Wang, S. Zhuang, [et al.]. Knowledge-Based Systems. Vol. 284. P. 111234.

49. Huang, T. (2023). Dense pedestrian detection based on multiple link feature pyramid networks / T. Huang, S. Yang, T. Yu, X. Fu. Hangzhou, China: SPIE.

50. Yahya, M. (2023). Object detection and recognition in autonomous vehicles using fast region-convolutional neural network / M. Yahya, R. A. Reddy, K. Al-Attabi, [et al.]. 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS). P. 1–5.

51. Dang, J. (2023). HA-fpn: hierarchical attention feature pyramid network for object detection / J. Dang, X. Tang, S. Li. Sensors. Vol. 23, No. 9. P. 4508.

52. Xie, B. (2022). FocusTR: focusing on valuable feature by multiple transformers for fusing feature pyramid on object detection / B. Xie, L. Yang, Z. Yang, [et al.]. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). P. 518–525.

53. Liu, H. (2021). Flexi-compression: a flexible model compression method for autonomous driving / H. Liu, Y. He, F. R. Yu, J. James. Proceedings of the 11th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications. P. 19–26.

54. Youn, E. (2023). Compressing vision transformers for low-resource visual learning / E. Youn, S. M. J, S. Prabhu, S. Chen.

55. Lan, Q. (2023). Instance, scale, and teacher adaptive knowledge distillation for visual detection in autonomous driving / Q. Lan, Q. Tian. IEEE Transactions on Intelligent Vehicles. Vol. 8, No. 3. P. 2358–2370.

56. Agand, P. (2024). Knowledge distillation from single-task teachers to multi-task student for end-to-end autonomous driving / P. Agand Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38, No. 21. P. 23375–23376.

57. Li, Z. (2024). Quasar-vit: hardware-oriented quantization-aware architecture search for vision transformers / Z. Li, A. Lu, Y. Xie, [et al.].

58. Sun, E. (2024). Transformer-based few-shot object detection in traffic scenarios / E. Sun, D. Zhou, Y. Tian, [et al.]. Applied Intelligence. Vol. 54, No. 1. P. 947–958.

59. Dong, J. (2023). Why did the ai make that decision? towards an explainable artificial intelligence (xai) for autonomous driving systems / J. Dong, S. Chen, M. Miralinaghi, [et al.]. Transportation Research Part C: Emerging Technologies. Vol. 156. P. 104358.

60. Cultrera, Luca. (2023). Visual attention and explainability in end-to-end autonomous driving / Cultrera, Luca.

61. Adom, I. (2024). RB-xai: relevance-based explainable ai for traffic detection in autonomous systems / I. Adom, M. N. Mahmoud. SoutheastCon. P. 1358–1367.

62. Kolekar, S. (2022). Explainable ai in scene understanding for autonomous vehicles in unstructured traffic environments on indian roads using the inception u-net model with grad-cam visualization / S. Kolekar, S. Gite, B. Pradhan, A. Alamri. Sensors. Vol. 22, No. 24. P. 9677.