Artificial Intelligence Review (2026) 59:19
https://doi.org/10.1007/s10462-025-11425-1

# Unifying ground and air: a comprehensive review of deep learning-enabled CAVs and UAVs

**Muhammad Umer Zia[1] · Wei Xiang[2] · Tao Huang[1] · Jameel Ahmad[3] · Jawwad Nasar Chattha[4] · Ijaz Haider Naqvi[5] · Faran Awais Butt[6]**

Received: 24 September 2024 / Accepted: 7 October 2025
© The Author(s) 2025

**Abstract**
The tremendous advancements in artificial intelligence (AI) techniques, particularly those pertinent to computer vision and image recognition, are revolutionizing the automotive industry towards the development of intelligent transportation systems for smart cities. Integrating AI techniques into connected autonomous vehicles (CAVs) and unmanned aerial vehicles (UAVs) and their data fusion, enables a new paradigm that allows for unparalleled real-time awareness of the surrounding environment. The potential of emerging wireless technologies can be fully exploited by establishing communication and cooperation among AI-augmented CAVs and UAVs. However, configuring appropriate deep learning (DL) models for connected vehicles is a complex task. Any errors can result in severe consequences, including loss of vehicles, infrastructure, and human lives. These systems are also susceptible to cyber attacks, necessitating a thorough and timely threat analysis and countermeasures to prevent catastrophic events. Our findings highlight the effectiveness of AI-driven data fusion in enhancing cooperative perception between CAVs and UAVs, identify security vulnerabilities in DL-based systems, and demonstrate how V2X-enabled UAVs can significantly improve situational awareness in corner cases.

**Keywords** Deep learning · Artificial intelligence · Connected and autonomous vehicles · Unmanned aerial vehicles · Cybersecurity

## 1 Introduction

The last decade has seen immense progress in making the dream of connected and autonomous vehicles (CAVs) a reality. Deep learning (DL) is undoubtedly the primary technology behind many breakthroughs in image recognition, and robotics (Bonsignorio et al. 2020). The success of DL techniques in the mentioned fields has led to widespread deployment of this technology with the aim of passenger safety, elimination of roadside accidents, and optimal path planning in self-driving cars (Grigorescu et al. 2020; Kuutti et al. 2020; Rao and Frtunikj 2018; Ni et al. 2020). The automotive industry has started testing CAVs on "controlled" roads with different capabilities termed "scales" graded from zero to five. The

Extended author information available on the last page of the article

lower scales feature basic driver assistance, while higher scales indicate a vehicle that needs no human intervention in driving (Yurtsever et al. 2020). A complete CAV system combines technologies, sensors, algorithms, and communication infrastructure. The involvement of DL blocks in a CAV system also depends on its scale. A fully automated car operating on scale five may have a distinct DL module attached to its key decision and control systems.

To complement CAVs in challenges like surveillance, acquiring aerial data, and combating emergencies, a promising solution is to adopt unmanned aerial vehicles (UAVs) (Hildmann and Kovacs 2019; Amer et al. 2020; Moukahal et al. 2020; Guillen-Perez and Cano 2018). A UAV is an unmanned autonomous or semi-autonomous machine that can be controlled remotely and allows us to monitor activities at different locations. UAVs can play a vital role in assisting a CAV's network in conjunction with Vehicle-to-Everything (V2X) communication technology and other advanced network technologies, such as software-defined networking, network function virtualization, mobile edge computing (MEC), and fog computing (Mishra and Natalizio 2020). Recently, UAVs' applications in the communications domain, along with their challenges and open problems, are investigated in Mozaffari et al. (2019). Due to UAV's versatile nature, automation, and low cost, it enjoys widespread use in civilian applications like surveillance, disaster rescue, parcel delivery, power line inspection, agriculture support, and mobile sensing platforms (Giordan et al. 2020; Menouar et al. 2017a). Integrating UAVs in CAV networks unveils many benefits and new use cases. Figure 1 depicts a typical scenario of a composite UAV-assisted CAV network, illustrating the various entities involved and their corresponding communication link types, thereby providing a clear and comprehensive representation of the considered operational environment. Besides surveillance and aerial information exchange, a UAV can take the roles of flying or emergency roadside unit (RSU), base station, or reconfigurable intelligent surface (RIS) (Hildmann and Kovacs 2019; Menouar et al. 2017a). These use cases can be extremely helpful in hardware malfunctions and disaster situations like fire or earthquake, thus guaranteeing an operational CAV system at all times.

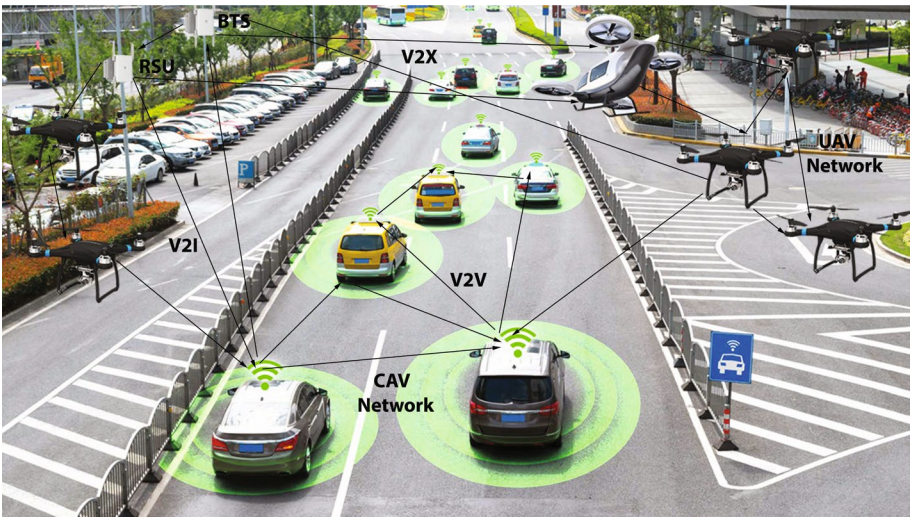


**Fig. 1** A pictorial view of UAV-assisted CAVs in a vehicular network

### 1.1 Research motivation

The motivation for this review stems from a gap identified in the existing literature regarding DL-assisted CAVs and UAVs. Previous surveys, as referenced in Table 1, have focused on aspects specific to either CAVs or UAVs, failing to provide an integrated view of how the unique capabilities of UAVs, particularly their aerial view (that presents a holistic view of traffic conditions), can address significant challenges in the deployment of the connected vehicular system. Numerous issues, such as handling corner cases, computer vision errors, adapting to diverse driving conditions, accurately predicting human behaviour and legal, ethical, and regulatory obstacles still need to be addressed for the real-world deployment of CAVs.

To this end, the key contributions of this article are highlighted as follows:

1. We provide a comprehensive system-level overview of UAVs-assisted CAV network architecture, integration efforts, CAV-UAV data fusion, CAVs and UAVs DL designs, along with highlighting the potential use cases of integrating UAVs with CAVs.
2. We compare two state-of-the-art deep learning frameworks applicable to CAVs, namely, the pipeline-based modular approach and the single block processing approach, also known as "End-to-End (E2E)" learning. Furthermore, we discuss the role of the latest Large Language Models (LLM) based DL designs in CAVs and UAVs, critically analyzing the strengths and limitations of these models.
3. We analyze the domains where deep learning can assist UAVs in enhancing their perception, path planning, navigation, and control. We also explore the state-of-the-art DL designs that enable UAVs to detect entities of vehicular networks and discuss their limitations.

**Table 1** Literature comparison

| Detailed research analysis | UAV-CAV Networking | Deep learning in CAVs | Deep learning in UAVs | Cybersecurity in CAVs and UAVs | Critical analysis | Challenges | Trends and future directions |
|---|---|---|---|---|---|---|---|
| Oubbati et al. (2021) | ✓ | × | ✓ | × | × | × | × |
| Shin et al. (2022) | ✓ | ✓ | ✓ | × | × | ✓ | × |
| Shi et al. (2018) | ✓ | × | × | × | × | ✓ | ✓ |
| Bouguettaya et al. (2021) | ✓ | ✓ | ✓ | × | × | ✓ | ✓ |
| Hu et al. (2021) | ✓ | ✓ | ✓ | × | × | ✓ | ✓ |
| Telikani et al. (2024) | × | ✓ | ✓ | × | × | ✓ | ✓ |
| Biswas et al. (2022) | ✓ | × | × | ✓ | × | ✓ | ✓ |
| Abir et al. (2023) | ✓ | ✓ | × | × | × | ✓ | ✓ |
| Ahmad et al. (2024) | × | ✓ | × | ✓ | × | ✓ | × |
| Telikani et al. (2025) | ✓ | × | ✓ | × | ✓ | ✓ | ✓ |
| Ours | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

4. We conduct a comprehensive review of AI-based cyberattacks of various types that can affect CAVs and UAVs. We also conduct a critical analysis of AI-based attacks, identifying the severity of each type. Furthermore, we explore the deep learning techniques that can be tailored as countermeasures against adversarial attacks.
5. We conclude our work by identifying current and prospective future challenges faced by UAV-assisted CAVs and propose future directions that can help overcome these challenges, leading to a successful deployment of CAVs in smart cities.

To clearly describe the scope of our investigation, the section breakdown of the paper is depicted in Fig. 2.

The research landscape on UAV-assisted CAVs has made notable strides, yet several critical gaps remain, requiring more focused efforts in the integration and enhancement of their coordination and vision capabilities. First, there is a clear need for extensive work on the integration of CAV-UAV systems, particularly in the joint testing of their coordination and vision capabilities through advanced deep learning constructs. The current literature predominantly addresses these domains separately, leaving an unexplored potential for their combined functionality in real-world scenarios. Second, the current body of work lacks an in-depth exploration of how these constructs can provide a more accurate and holistic

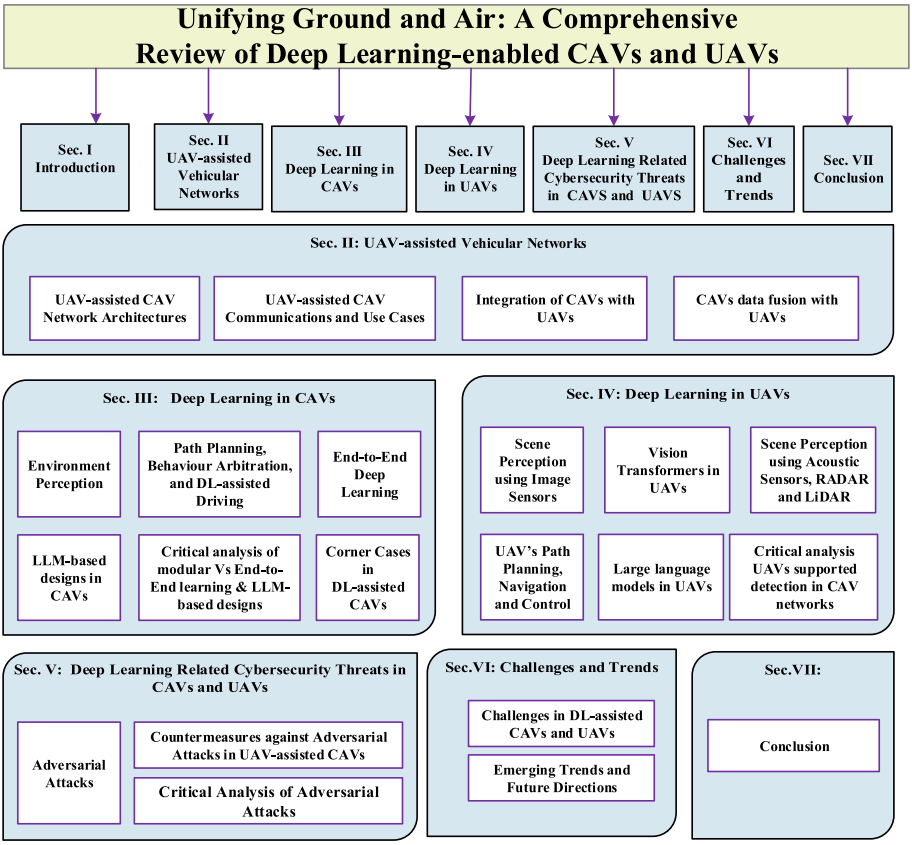


**Fig. 2** Sections breakdown of the paper

understanding of the CAV environment, essential for enabling reliable self-driving systems. At last, there is an urgent need to critically examine the cybersecurity threats that could compromise both autonomous driving and UAV-CAV coordination. Specifically, adversarial attacks that target the integrity of machine learning models present a significant risk to the safety and reliability of these systems. Future studies must address these vulnerabilities by proposing robust defense mechanisms that ensure resilience against cyber threats. Table 1 presents a comprehensive list of literature works on UAV-assisted CAVs, highlighting the contributions of each work.

### 1.2 Research methodology

This study was initiated to establish a foundational understanding of modular and end-to-end approaches in deep learning-enabled CAVs. The rapid advancement in both approaches necessitated a comprehensive review of their pros and cons to answer the crucial question of the most promising approach. Through comparative analysis and the identification of corner cases, we recognized the importance of incorporating aerial support to address unresolved scenarios. Consequently, we presented an integrated system that combines deep learning-enabled CAVs and UAVs to realize the vision of a successful driving system capable of overcoming challenges in diverse scenarios.

The literature selection process adhered to a rigorous and systematic methodology designed to ensure comprehensiveness and high relevance. To collate relevant studies, we have adopted a two-pronged search strategy. Initially, we combined key terms such as “CAVs or ITS”, “Perception”, “Path Planning”, “Motion Control” “Modular Vs End-to-End” and “Cameras”, “RADAR”, “LiDAR”. Later, we conducted an exhaustive search with terminologies such as “Deep Learning in UAVs”, “Deep Learning in CAVs-UAVs systems” “UAVs-CAVs integration” “UAV-CAV Sensor Fusion”, “UAV-CAV Cyber-threats and Countermeasures”. Finally, terms like ”UAV-CAV Vision Transformers”, “UAVs-CAVs LLM-based designs” and “UAV-CAV Challenges, Trends and Future Directions” were utilized to complete this comprehensive review.

A wide array of leading academic databases, including IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, and Google Scholar, was utilized to identify peer-reviewed studies published within the last decade. In addition, to identify industry trends and incorporate practical perspectives, we also referred to credible websites and news articles associated with the automotive and drone industry. Articles meeting inclusion criteria focused on experimental validation or theoretical innovations addressing DL-enabled UAV-CAV systems. The review excluded studies that lacked empirical grounding or DL designs unrelated to the vision capability of CAVs and UAVs.

Key insights from the selected literature were synthesized to provide a coherent and forward-looking perspective on the field. It also highlights the technical and operational challenges of system integration, such as the complexities of real-time sensor fusion and the mitigation of adversarial DL attacks, offering a critical analysis of proposed solutions. Furthermore, emerging trends like the incorporation of Vision Transformers, Large Language Models, and advancements in sensor fusion are explored for their potential to redefine UAV-CAV collaborations. By offering a structured and multidimensional perspective, this study not only provides a roadmap for addressing current research gaps but also serves as

a foundational resource for researchers and industry professionals aiming to advance this interdisciplinary domain.

## 2 UAV-assisted vehicular networks

UAV-assisted CAV networks comprise sophisticated systems that integrate advanced sensors, data fusion modules, and AI-driven functionalities. By leveraging aerial perspectives, UAVs significantly enhance vehicular networks, particularly in emergency scenarios and regions with limited infrastructure (Amponis et al. 2022). Their support enables adaptive and dynamic network topologies, thereby improving coverage and reliability for CAVs across diverse environments, including urban, rural, and highways. Furthermore, UAVs offer high mobility and flexibility, facilitating on-demand connectivity services such as data offloading, caching, and relaying. This collaboration not only augments network performance and efficiency but also addresses the limitations of traditional cellular networks with fixed or constrained resources.

### 2.1 UAV-assisted CAV network architectures

The architecture of UAV-assisted CAV networks can be categorized based on the role played by the drone within the communication network. These roles range from passive elements within the CAV system to active relay nodes or dynamic mobile RSUs. Multiple drones or swarms can also form independent network layers within the CAV ecosystem. Drones can function as regular vehicles in UAV-assisted CAV networks, transmitting cooperative awareness messages like other users (Valle et al. 2021). UAVs may also act as relay nodes to enhance inter-vehicle communication in V2V networks. By hovering above CAVs and observing network topology, UAVs can integrate themselves as relays to improve connectivity. As proposed by Lin et al. (2020), drone deployment can be optimized by predicting vehicle distributions, aiding in routing for isolated vehicles, non-line-of-sight communication, and network load balancing. UAVs can also serve as resource nodes, bridging coverage gaps and supporting V2I communication. By repositioning, they can establish reliable wireless links with infrastructure (Seliem et al. 2018). In Al-Hilo et al. (2020), a cooperative caching-based approach is proposed where UAVs assist RSUs in fetching, carrying, and forwarding content without accessing the backhaul. UAVs further enable real-time traffic monitoring by covering inaccessible areas without interrupting traffic flow (Zhang et al. 2023). Additionally, in jamming scenarios, UAV-assisted CAVs can provide direct, unobstructed communication links between vehicles and drones (Feng and Haykin 2019a). Multiple drones can also form a coordinated swarm to support critical communication. These swarms can act as relays, extending coverage and increasing data transmission rates in infrastructure-limited areas (Raza et al. 2021). Swarm networks offer flexibility and protocol diversity, acting as overlay networks that provide redundancy (Raza et al. 2021). In Jacob et al. (2020), intelligent swarm coordination is proposed to assist vehicular networks and maintain safe inter-vehicle distances. UAV-enabled CAVs, enhanced with AI, are increasingly contributing to smarter cities. UAV-assisted communication is expected to play a critical role in optimizing wireless connectivity during high-demand or emergency situations.

The block diagram in Fig. 3 illustrates the integration of UAVs with CAV networks and highlights their significance in future mobility systems.

## 2.2 UAV-assisted CAV communications and use cases

Communication between UAV and CAV networks can be established through various methods. Authors in Kavas-Torris et al. (2022a) discuss two of the most prominent communication protocols and evaluate them under real-world scenarios: dedicated short-range communication (DSRC) and fourth-generation (4 G) cellular communication. In Nazib and Moh (2020); Guillen-Perez et al. (2021, 2016), the authors comprehensively classified routing protocols utilized in UAV-aided vehicular networks. The study in Zanjie et al. (2014) explored bandwidth and energy allocation strategies to optimize sensing and data gathering in UAV-assisted CAV networks. The primary goal was to maximize the overall data rate while ensuring fairness among all connected users. Furthermore, Poudel and Moh (2019) reviewed various medium access protocols for UAV-aided networks. Now, we focus on the different use cases these communication technologies enable.

### 2.2.1 Use cases for UAV-assisted CAVs

Integrating UAVs into CAV systems unlocks diverse capabilities and services (Menouar et al. 2017b):

1. **Safety Message broadcast:** UAVs support rapid and reliable broadcasting of safety alerts and accident notifications using direct line-of-sight communication (Saputro et al. 2018).
2. **Dynamic Spectrum Provisioning:** Drones can augment network capacity by acting as mobile RSUs, dynamically allocating additional spectrum.

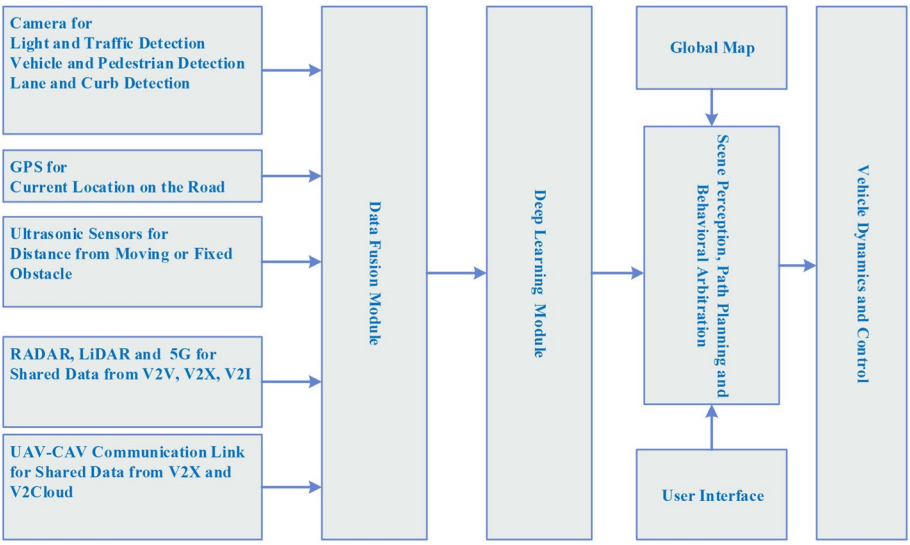


**Fig. 3** UAV's role in assisting CAVs

3. **Traffic Monitoring and Law Enforcement:** UAVs provide a 3D vantage point for real-time traffic surveillance, facilitating the detection of violations and criminal activities (Kang et al. 2020).
4. **Connectivity Enhancement:** Acting as relay nodes, UAVs improve network resilience by bridging coverage gaps, alleviating congestion, and balancing traffic load (Ahmed et al. 2021).
5. **Secure Communication:** UAVs enhance the robustness of V2V links by offering anti-jamming capabilities, thereby strengthening communication security (Feng and Haykin 2019b).
6. **Edge Computing Support:** Equipped with onboard compute resources, UAVs can function as mobile edge computing (MEC) servers, enabling task offloading from resource-constrained vehicles (He et al. 2021).

### 2.3 Integration of CAVs with UAVs

This subsection highlights key research regarding UAV-CAV integration, focusing on communication, optimization, and security aspects. In communication, Kavas-Torris et al. (2022b) implemented a V2X system based on a real-world use case, "Quick Clear," evaluating four communication protocols: DSRC, User Datagram Protocol, 4 G-based WebSocket, and Transmission Control Protocol. Su et al. (2023) investigated UAVs as relays to assist ground user equipment when RSUs are unavailable or provide poor coverage, analyzing both single- and multi-UAV deployments with user mobility. Similarly, Zou et al. (2022) addressed data distribution and offloading by proposing a UAV-assisted method that serves both stationary and mobile edge nodes, ensuring low latency and service reliability for vehicles. In the domain of optimizing UAV-aided CAV systems, the utilization of UAVs to enhance mobile edge computing for vehicles in a platoon was explored by Liu et al. (2022). Their model considered UAV-platoon interaction, ground-to-air communication, onboard computing, and energy harvesting. Extending this line of work, Liao et al. (2023) introduced 3D-UAV, an energy-aware deployment strategy designed for complex environments such as interchange bridges. Their approach addressed line-of-sight challenges, optimizing UAV altitude and vehicle clustering to maximize uplink rates while minimizing UAV usage in the Internet of Vehicles (IoV). In security domain Feng and Haykin (2019c) proposed a UAV-assisted secure communication framework resilient to hybrid attacks involving malicious CAVs and UAVs. They introduced a "cognitive dynamic system" utilizing cognitive risk control and intelligent jamming resistance. Likewise Khan et al. (2022) tackled secure data exchange in complex UAV-CAV hierarchies, presenting "B-UV2X," a blockchain-based modular V2X infrastructure enabling transparent and secure communication in distributed vehicle networks. Table 2 summarizes additional integration challenges and considerations. Parallel to these research efforts, standardization plays a pivotal role in enabling UAV-CAV interoperability. Notable standards include IEEE 802.11n (Zhou et al. 2015), IEEE 802.11ah (Adame et al. 2014), and IEEE 802.11p (Shilin et al. 2016). IEEE 802.11n (Wi-Fi 4) enhanced data rates, range, and reliability, supporting UAV applications such as video streaming and telemetry across the 2.4 and 5 GHz bands. IEEE 802.11ah (Wi-Fi HaLow) was developed for low-power, long-range IoT communication, treating UAVs as networked "Things." IEEE 802.11p, designed for vehicular ad-hoc networks (VANETs), operates in the 5.9 GHz band and supports low-latency communication between vehicles. While primar-

**Table 2** Prospects of integrating UAVs with CAVs

| Integration issues | Keypoints | Details |
|---|---|---|
| Air traffic management and collision avoidance (Kavas-Torris et al. 2021) | Communication and connectivity<br>Coordinating airspace<br>Sensing and perception<br>Regulatory and legal hurdles<br>Interoperability and standards<br>Safety and reliability | Requires robust communication and coordination systems.<br>Reliable links between UAVs, CAVs, and central control are essential.<br>Altitude, weather, and fusion challenges affect detection and response. Collaboration is needed for liability, licensing, and airspace management.<br>Common standards for communication, navigation, and control are vital. Robust mechanisms are needed as faulty UAVs are dangerous. |
| Energy Efficiency and Range (Oubbati et al. 2019) | Drones are usually battery powered | Optimizing energy usage for both platforms is complex. Ensuring sufficient range for integration of UAVs with CAVs is essential. |
| Privacy and Security (Khan et al. 2022) | Integration raises privacy concerns. | Ensuring security in the aerial dimension, considering the limited processing power of drones is a challenge. Securing communication channels is paramount as the human transportation is involved. |
| Infrastructure and Urban Planning (Zhu et al. 2019; Motlagh et al. 2016) | Variable building heights can obstruct drone route | Adapting urban infrastructure for UAVs/CAVs requires careful planning. |
| Collaboration and Testing (Khan et al. 2022) | Diverse UAV & CAV makers | Collaboration among stakeholders is essential. Comprehensive testing, simulation, and iterative development are necessary. |
| Technical (Khabbaz et al. 2019) | Communication<br>Sensing and Perception<br>Integration of Control Systems | Establishing reliable communication links between UAVs and CAVs.<br>Ensuring UAVs can accurately detect and respond to ground-based CAVs.<br>Coordinating control algorithms for UAVs and CAVs to avoid collisions. |
| Regulatory (Shrestha et al. 2021) | Airspace Regulations<br>Traffic Management | Complying with airspace regulations and obtaining necessary permissions. Integrating UAVs into existing traffic management systems. |
| Operational (Xu et al. 2020) | Scalability | Handling a large number of UAV-CAV interactions in urban environments. |
| Public acceptance and perception (Cawthorne and Juhl 2022) | Fear of spying<br>Drone noise disturbance | Skepticism due to noise, congestion, and perceived risks should be addressed. |
| Economics of integration (Motlagh et al. 2016) | Cost and Scalability | Managing costs while ensuring scalable technology. |

ily focused on ground vehicles, it is extensible to UAVs serving as communication relays within vehicular networks.

The investigation by Kavas-Torris et al. (2022a) focused on the empirical study of hardware implementation and real-life testing of a V2X communication framework between a CAV and a UAV. Investigators tried to establish reliable communication links using four methods: DSRC, User Datagram Protocol (UDP), 4 G-based WebSocket, and Transmission Control Protocol (TCP). The coordinated mission involved transmitting accident location data from the CAV to the UAV, which was further relayed to a Contingency Management Platform (CMP) and a web server for situational awareness. The study evaluated the per-

formance of these communication methods through latency and package drop percentage metrics. The experimental results of this study are summarized in the Table 3.

### 2.4 CAVs data fusion with UAVs

Data fusion in connected vehicles employs diverse methodologies to integrate multi-sensor data, enhancing perception accuracy and decision-making reliability. The overviews in Khezaz et al. (2022a); Ounoughi and Yahia (2023); Butt et al. (2022) cover both intra-CAV sensor fusion and fusion involving UAVs, although most studies concentrate on the former. This section elaborates on key techniques applicable to CAV-UAV data fusion.

1. **Probabilistic Methods** Probabilistic approaches effectively manage uncertainty and improve estimation accuracy, especially in dynamic environments. Following Kalman filter variants are widely adopted for real-time applications (Ounoughi and Yahia 2023; Montañez et al. 2023):

   (a) *Extended Kalman Filter (EKF)*: Linearizes nonlinear models using the current state estimate and covariance.
   (b) *Unscented Kalman Filter (UKF)*: Uses the unscented transform for more accurate nonlinear estimation.
   (c) *Sequential Kalman Filter (SKF)*: Processes data incrementally as it becomes available.
   (d) *Federated Kalman Filter (FKF)*: Aggregates outputs from multiple Kalman filters to produce a global estimate.
   (e) *Cubature Kalman Filter (CKF)*: Applies third-degree spherical-radial cubature rules for nonlinear filtering.

These filters iteratively refine state predictions using incoming sensor data, making them vital for real-time vehicular contexts.

2. **Evidence-Based Methods:** The Dempster-Shafer theory provides a robust alternative to traditional probabilistic methods by combining uncertain and imprecise information (Kusenbach et al. 2020; Cai et al. 2023; Xiang et al. 2023a). It is particularly suitable when sensor inputs are incomplete or noisy.

   (a) *Belief Functions*: Represent degrees of belief across hypotheses, capturing imprecision beyond what standard probabilities allow.
   (b) *Combination Rules*: Dempster's Rule of Combination merges evidence from different sources via Basic Probability Assignments (BPAs).

**Table 3** Package Drop Percentage

| Communication Method | Package Drop (%) |
|---|---|
| DSRC | 0.36 |
| UDP | 1.72 |
| WebSocket | 1.56 |
| TCP | 10.45 |

These features enable more flexible and reliable situational awareness in heterogeneous vehicular networks.

3. **Knowledge-Based Methods**: DL methods facilitate high-level feature extraction and data fusion from raw sensor streams (Butt et al. 2022; Harun et al. 2022). Traditional machine learning algorithms, including support vector machines, random forests, and Gaussian mixture models are also used for classification and sensor integration.
   Advanced DL techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs), and federated learning process large, nonlinear, and heterogeneous sensor data for object detection and predictive tasks (Wang et al. 2023b). However, their growing complexity and "black-box" nature pose challenges in interpretability and trust (Ounoughi and Yahia 2023).
4. **Statistical Methods**: Statistical models interpret and fuse sensor data through probabilistic relationships, enabling robust decision-making under uncertainty (Butt et al. 2022).

   (a) *Bayesian Networks*: Graphical models that capture conditional dependencies among variables to enhance environmental awareness (Lim et al. 2021a).
   (b) *Particle Filters*: Use sample-based approximations for non-linear, non-Gaussian state estimation (Tekeli et al. 2018).
   (c) *Expectation-Maximization (EM)*: Iteratively estimates parameters in latent variable models by alternating between expectation and maximization steps (Kim et al. 2021).

These methods are valued for accurately modeling uncertainty and interdependencies in real-world sensing environments.

5. **Hybrid Methods**: Hybrid approaches integrate multiple data fusion techniques to capitalize on their strengths while mitigating individual limitations (Malawade et al. 2022; Butt et al. 2022; Yeong et al. 2021). Common combinations include:

   (a) *Kalman Filter and Deep Learning*: Merges real-time uncertainty handling with deep neural feature extraction.
   (b) *Bayesian Networks and Dempster-Shafer*: Combines probabilistic and evidence-based reasoning for robust decision-making.
   (c) *Multi-Sensor Fusion Frameworks*: Organize fusion hierarchically–using probabilistic techniques at low levels and DL or Dempster-Shafer methods, for high-level reasoning.

Hybrid methods provide a flexible, adaptive, and comprehensive framework that significantly enhances the robustness and reliability of CAV-UAV systems.

### 2.4.1 Comparative analysis

Sensor fusion techniques for CAVs and UAVs offer distinct advantages and limitations. Probabilistic methods provide reliable state estimation by effectively handling uncertainty

but require significant adaptations for highly non-linear or complex data (Ounoughi and Yahia 2023; Montañez et al. 2023). In contrast, evidence-based methods, such as Dempster-Shafer theory, excel in uncertain and incomplete data environments, offering robust situational awareness through multi-source evidence aggregation, albeit at high computational cost (Kusenbach et al. 2020; Xiang et al. 2023a).

Knowledge-based methods leverage deep learning to extract high-level features from heterogeneous, non-linear sensor data, enabling advanced perception tasks such as object detection and lane recognition. However, their computational intensity and lack of interpretability due to the ”black-box” nature pose a challenge for safety-critical autonomous systems (Butt et al. 2022; Harun et al. 2022). Statistical methods, including Bayesian networks and particle filters, accurately model uncertainties and are effective in non-Gaussian, dynamic environments (Butt et al. 2022; Tekeli et al. 2018). Despite their strengths, these methods can be complex and resource-intensive. Hybrid methods integrate probabilistic, evidence-based, and deep learning techniques into a unified framework, offering a balanced and adaptive solution for data fusion (Malawade et al. 2022; Butt et al. 2022; Yeong et al. 2021). While their design and implementation can be complex, they achieve enhanced robustness, scalability, and accuracy.

Given the strengths and limitations of individual approaches, hybrid methods emerge as the most suitable strategy for CAV-UAV data fusion. Their ability to address diverse and dynamic real-world scenarios makes them well-aligned with the requirements of intelligent transportation systems.

## 3 Deep learning in CAVs

In order to respond to the environment, CAVs should be familiar with their surroundings. The primary task of environment perception is achieved using various sensing devices such as radio detection and ranging (RADAR), light detection and ranging (LiDAR), and cameras. Environmental perception is of paramount importance as it is directly related to the safety of the passengers. AI can play a crucial role in developing environment perception methods using novel machine learning algorithms. One important application of AI is path planning and behavior arbitration to accurately plan car routes and arbitrate different driving strategies. The information from sensors can be fed to an AI black box using modular or End-to-End learning approach as shown in Fig. 4. Once the AI module is trained on a certain dataset, it must be tested in different scenarios so that rare situations/corner cases do not deceive the model and fatal accidents can be avoided.

In addition to path planning and behavior arbitration, a vehicle is also required to avoid collisions, take safe turns, and make overtaking decisions with adherence to traffic laws. In this scenario, one important aspect involves interaction with other drivers and their behavior in certain conditions. An ideal vehicle tries to preempt the behavior of aggressive drivers and take action to keep itself (and others on the road) safe.

### 3.1 Environment perception

This section will review different deep-learning techniques applied to sensor data for environment perception. In connected and autonomous vehicles, perception systems understand

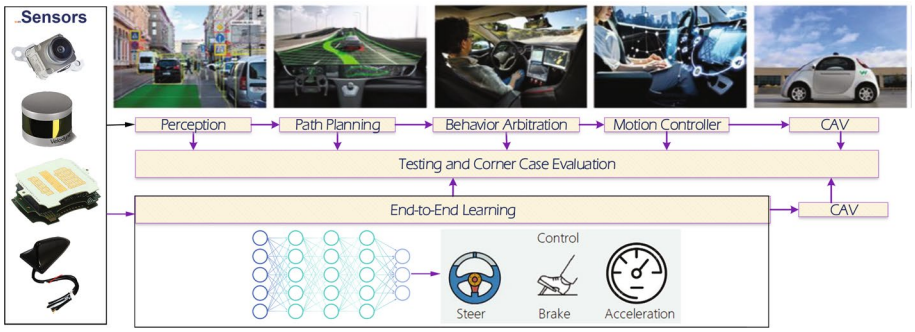


**Fig. 4** Deep learning aided CAV system. The modular approach and End-to-End learning approach are shown at the top and bottom, respectively

the surrounding environment utilizing the input from various sensors, such as camera, RADAR, LiDAR, and inertial sensors (Schoettle 2017). These sensors provide information regarding the fast-changing environment. The environment perception in CAVs includes detection of weather conditions such as fog, snow, and rain as well as a range of fixed (traffic signals and signs, buildings, road markings, etc.) and moving objects (cars, pedestrians, bicycles, etc.), in the surrounding along with their distance from the sensors (Yurtsever et al. 2020; Hafeez et al. 2020; Wang et al. 2017; Guillen-Perez and Cano 2019). The data provided by these sensors is utilized for driver assistance and vehicle control. In the detection process, a bounding box is drawn around important objects, and multiple bounding boxes capture multiple objects to accomplish environmental perception in real-time. Despite technological advancements, one sensor cannot satisfy all autonomous driving requirements in all weather conditions and ranges. Marti et al. (2019). For example, the advanced cameras that produce high-resolution 2D images suffer severe deterioration in their performance at low or high-intensity light and unclear weather. Similarly, the RADAR works very well in bad weather. However, the resolution of RADAR data is not enough for object identification (Dickmann et al. 2014). A detailed description regarding environment perception using sensors and their fusion is covered in Butt et al. (2022). As perceiving the surrounding environment and extracting information is critical for the operation and safety of a CAV system, we will review each sensor utilized in CAV along with the employed machine learning techniques.

### 3.1.1 Scene perception using cameras

The camera sensors do not transmit any signal and depend upon incoming rays for perceiving the environment. Several types of cameras sense the environment images are utilized in autonomous vehicles, including flash cameras, thermal cameras, and event cameras (Maqueda et al. 2018). Machine learning algorithms have achieved remarkable success in object detection and image classification and are considered state-of-the-art these days. Moreover, image processing using deep learning techniques is vital for detecting unusual objects. For example, the authors in Ramos et al. (2017) have presented a framework that employs appearance and contextual information to detect small unforeseen obstacles for

self-driving cars. These methods can be categorized based on their framework into the following types (Carranza-García et al. 2021; Jiao et al. 2019; Wang et al. 2019).

1. Single-stage detection: The architecture in this category uses a single network to detect an object while predicting its class simultaneously.
2. Region proposal detection: The architecture in this category utilizes two-stage designs where general regions of interest are identified, followed by their class identification by another network.

We will now briefly discuss and compare single-stage and regional proposal detection methods.

### 3.1.2 Single stage detection

The single-stage detectors comprise one feed-forward CNN network that produces the bounding boxes and classifies the object. Several investigations for one-stage-based object detection have been conducted for autonomous driving. The initial work concerning single unified architecture was SSD: Single Shot Multi-Box Detector (Liu et al. 2016a) and YOLO (You Only Look Once) (Redmon et al. 2016). YOLO offers real-time detection of images; moreover, feature extraction using this unified architecture is straightforward, utilizing input images to predict bounding boxes and class probabilities. YOLO is good for real-time processing and improved accuracy. Moreover, it can be trained end-to-end. With additional contextual information, YOLO exhibits fewer false positives in background areas. The shortcoming of YOLO includes reduced localization accuracy which is the main source of prediction error; moreover, there are only a few close-by objects that YOLO can predict. The YOLO design triggered a series of investigations that led to improved single-stage detection architectures.

YOLOv2 (Redmon and Farhadi 2017), the second version of YOLO, incorporates several design improvements such as batch normalization (BN), convolution with anchor boxes, multi-scale training, and addition of fine-tuning process to the classifier neural network. The YOLOv2 achieves 78.6% mAP (mean Average Precision) and 40fps (frame per second) in comparison to YOLO with 63.4% mAP and 45fps. A further improved version of YOLO2, YOLO3 was presented in Redmon and Farhadi (2018) which incorporates several enhancements, including multi-label classification, three different scale feature maps, and a deeper and more robust feature extractor. The multi-scale predictions lead to better detection of small objects at the expense of the detection performance of medium and large-sized objects. The work in Bochkovskiy et al. (2020) proposed YOLOv4, which outperforms the previous YOLO versions with more accurate results when tested on the MS COCO dataset. Multiple other incremental improvements on the YOLO construct are proposed by researchers such as YOLOv5 (Mahaur and Mishra 2023) which focused on improved detection of small objects along with better performance and speed. YOLOv5 is fully written in PyTorch contrary to using any form of the Darknet framework. In this context, another contribution (Benjumea et al. 2021) dubbed "YOLO-Z" focused on detection of small objects present in vehicular networks with higher speed and accuracy. There are several other detectors like SSD513 (Fu et al. 2017), RefineDet (Zhang et al. 2018), RetinaNet (Lin et al. 2017), and M2Det512 (Zhao et al. 2019a) with competing performance concerning speed and accuracy.

### 3.1.3 Region proposal detection

Region proposal methods provide higher accuracy than single-stage detectors but at expense of higher computational complexity. Region Proposal Networks (RPN) use a cumbersome two-stage detection framework to train and calibrate classifiers. However, their object recognition and localization accuracy are higher. The region proposal method is a detection process that comprises the region proposals and the classifier. Several object candidates, known as Regions of Interest (RoI) are proposed first using reference boxes (anchors) and in the next step, these proposals are classified. The pioneering deep learning-based work in this context is R-CNN (Girshick et al. 2014). in (Girshick et al. 2014), authors used an external selective search to generate proposals fed to a CNN to perform classification and bounding box regression. A year later, Girshick (2015) proposed an improved version, namely Fast R-CNN. The Fast R-CNN is composed of a fully convolutional neural network (CNN) whose inputs are multiple RoIs and an entire image. The proposed network produces two output vectors which are softmax probabilities and per-class bounding-box regression offsets. Shortly after Fast R-CNN, a further improved version dubbed Faster R-CNN (Ren et al. 2015), was proposed. Faster R-CNN consisted of two modules, namely a deep CNN that proposes regions and a Fast R-CNN detector that uses the proposed regions. Faster R-CNN replaces selective RoI search with a novel region proposal network. The region proposal network accelerates the formation of proposals because it contributes full-image convolutional features and a common set of convolutional layers with the detection network. Moreover, this research proposes a novel method in which multi-scale anchors are employed as a reference for different-sized object detection. There is no straight answer regarding which model is the best, as they have different performances on different datasets and objects. For real-world applications, there is a trade-off between accuracy and speed. Table 4 provides the performance comparison of different models' scene perception capabilities.

### 3.1.4 Vision transformer in CAVs

Vision Transformers (ViTs) are a type of neural network architecture designed for processing images, based on the Transformer model originally developed for natural language processing. State-of-the-art vision transformers are revolutionizing critical tasks such as object detection, lane detection, and segmentation, and can be integrated with reinforcement learning for complex pathfinding (Lai-Dang 2024). They excel at processing spatial and temporal data, surpassing traditional CNNs and RNNs in functions like scene graph generation and tracking. The self-attention mechanism of Transformers offers a deeper understanding of dynamic driving environments, which is crucial for the safe navigation of autonomous vehicles. This comprehensive approach makes Transformers particularly effective in enhancing the performance and safety of CAV systems (Zhu et al. 2024).

In the domain of transformer-based designs, Deshmukh et al. (2023) addressed the vehicle detection challenge in traffic environments with mixed vehicle types and non-standard traffic behaviour. To tackle the shortcomings of conventional CNNs, a new Swin transformer-based vehicle detection (STVD) framework is proposed. This framework enhanced feature extraction by facilitating thorough information exchange within and between image patches while incorporating a bi-directional feature pyramid network (BiFPN). The results demonstrated the framework's superiority, achieving a 91.32 percent accuracy on DTLD,

**Table 4** Performance of several models on different image datasets

| Method | FPS (M) | AP % | Dataset | Critical Analysis |
|---|---|---|---|---|
| YOLOv9 (Wang et al. 2024) | N/A | 72.8 | MS COCO | The performance of machine learning models in object detection varies significantly across datasets such as MS COCO, KITTI, and VOC 07 due to differences in object types, environmental conditions, and other dataset-specific characteristics. This inconsistency makes it difficult to fairly compare models, as a model that excels on one dataset may underperform on another. To address this challenge, there is a need for a universal dataset or standardized benchmark that enables consistent and objective evaluation. Such a benchmark would not only facilitate fair comparisons but also drive the development of more robust models capable of performing well across diverse real-world scenarios. |
| YOLOv7 (Wang et al. 2023a) | N/A | 74.4 | MS COCO | |
| YOLO-Z (Benjumea et al. 2021) | 30.6 | 96.5 | MS COCO | |
| YOLOv4 (Bochkovskiy et al. 2020) | 31 | 43.0 | MS COCO | |
| YOLOv3 (Bochkovskiy et al. 2020) | 35 | 31.0 | MS COCO | |
| YOLOv2 (Lin et al. 2017) | N/A | 21.6 | MS COCO | |
| SSD (Bochkovskiy et al. 2020) | 22 | 28.8 | MS COCO | |
| RefineDet (Bochkovskiy et al. 2020) | 22.3 | 33.0 | MS COCO | |
| RetinaNet (Bochkovskiy et al. 2020) | 13.9 | 32.5 | MS COCO | |
| M2det (Bochkovskiy et al. 2020) | 33.4 | 33.5 | MS COCO | |
| Faster R-CNN (Bochkovskiy et al. 2020) | 9.4 | 39.8 | MS COCO | |
| Fast R-CNN (Jiao et al. 2019) | N/A | 19.7 | MS COCO | |
| YOLOv2 (Chun et al. 2019) | 85.5 | 64.8 | KITTI | |
| YOLOv3 (Chun et al. 2019) | 43.6 | 80.5 | KITTI | |
| RefineDet (Chun et al. 2019) | 27.8 | 84.4 | KITTI | |
| SSD (Chun et al. 2019) | 28.9 | 14.1 | KITTI | |
| YOLO (Zhao et al. 2019c) | 45 | 63.4 | VOC 07 | |
| YOLOv2 (Zhao et al. 2019c) | 40 | 78.6 | VOC 07 | |
| SSD (Zhao et al. 2019c) | 19 | 76.8 | VOC 07 | |
| Fast R-CNN (Zhang et al. 2018) | 0.5 | 70 | VOC 07 | |
| Faster R-CNN (Zhang et al. 2018) | 7 | 73.2 | VOC 07 | |
| RefineDet (Zhang et al. 2018) | 40.3 | 80.0 | VOC 07 | |

87.4 percent on IITM-here, and 88.45 percent on KITTI datasets, outperforming existing methods. Working on similar lines but considering the safety of autonomous vehicles in mixed traffic and connected environments, the authors in Ji et al. (2024) focused on interpreting the intentions of human-driven vehicles when they change lanes. The authors of this investigation presented an innovative method for identifying lane-changing intentions by analyzing the driving state and relative motion of the target vehicle and its neighbouring vehicles. This method utilizes various techniques such as short-time Fourier transform, Gramian angular summation field, and Gramian angular difference field to convert time-series data into grayscale images, which are then combined into an information fusion image (IFI). These IFIs are then categorized into lane-keeping, lane-changing left, and lane-changing right using a Vision Transformer model with transfer learning. This approach has demonstrated superior performance compared to traditional methods, achieving 95.65 percent accuracy in recognizing lane-changing intentions 3 s prior to the lane change. Ramana et al. (2023) focused on predicting urban traffic patterns due to severely deteriorating urban conditions such as population growth, congestion, air pollution, fuel consumption, traffic violations, noise, accidents, and time loss. The authors proposed a method for accurate traffic prediction using ViTs in combination with CNNs. In this approach, CNNs process traffic images to generate feature maps, which are then tokenized and projected by ViTs before being analyzed by LSTM. The results demonstrate that this ViT-based method is especially effective in predicting traffic flow, even under unusual traffic conditions.

Considering the importance of Bird's Eye View (BEV) perception and the reliance of its accuracy on large data sets, Song et al. (2023) introduced FedBEVT, a federated transformer learning approach for BEV perception. They addressed data heterogeneity issues, such as diverse sensor poses and varying sensor numbers, by using Federated Learning with Camera-Attentive Personalization and Adaptive Multi-Camera Masking. Their method outperformed baseline approaches in four typical federated use cases, showing promise for enhancing BEV perception in autonomous driving.

To address the "black box" nature and lack of interpretability in deep learning approaches, Dong et al. (2021) proposed an explainable end-to-end autonomous driving system using a state-of-the-art self-attention-based Transformer. This system maps visual features from images collected by onboard cameras to guide driving actions, while providing corresponding explanations. The results show that their model significantly outperforms the benchmark model in both action and explanation prediction while reducing computational costs. In Li et al. (2022), proposed a lightweight transformer-based end-to-end model with built-in risk awareness to reduce the high computational burden in autonomous vehicle decision-making. The model utilizes a lightweight network combining depth-wise separable convolution and transformer modules for efficient image semantic extraction from trajectory data sequences. Driving risk is then assessed using a probabilistic model that accounts for position uncertainty, which is integrated into deep reinforcement learning to identify strategies with minimal expected risk. The method was validated in three lane change scenarios, demonstrating its effectiveness and superiority. To enhance driver assistance systems, Gao et al. (2022) proposed a novel hybrid deep learning framework called Multi-Modal CNN-Transformer (M2-Conformer) for detecting driving behavior using video frames and multivariate vehicle signals. The M2-Conformer integrates both Transformer and CNN architectures in parallel branches to extract features from driving scenes and vehicle dynamics. It employs dynamic token sparsification in the Transformer branch to prune redundant tokens, improving processing speed. Additionally, a custom Feature Aggregation Module (FAM) is designed to combine high-quality features from different branches. Experiments on a naturalistic driving dataset show that M2-Conformer offers a superior balance between complexity and accuracy compared to other state-of-the-art methods for driving behavior detection. Kang et al. (2022) developed ViT-TA, a customized Vision Transformer, to enhance autonomous vehicles' safety by accurately classifying critical traffic accident situations and identifying probable causes. ViT-TA outperformed existing methods in detecting critical moments and helped systematize the creation of functional scenarios for improvements in CAV safety. This framework offers a scalable and reliable approach to generating safety plans for CAVs.

In Islam et al. (2023), investigators proposed a novel ensemble framework integrating transformer and conformer models for crash prediction utilizing connected vehicle trajectory data. Their prominent contribution lies in the synergistic combination of the models, capitalizing on their complementary strengths to enhance predictive accuracy. Empirical evaluations demonstrate promising results, underscoring the potential of this approach for proactive safety interventions. In Tian et al. (2023), authors presented a cutting-edge approach dubbed "VistaGPT", leveraging generative parallel transformers. A notable research strength was enabling efficient processing of multimodal sensor data. The paper's experimental results demonstrate impressive performance in complex driving scenarios, showcasing VistaGPT's potential for transport automation. While the paper provides a rea-

sonable basis, further examination is needed to establish the computational efficiency and scalability of VistaGPT for real-time autonomous driving.

The investigators in Dalwai et al. (2023) focused on sub-optimal performance of conventional DL designs in dynamic scenes and lighting variations and proposed using Vision Transformers' self-attention mechanisms to capture spatial relationships and contextual information in video frames. This solution offered an innovative approach for real-time vehicle collision detection in CCTV footage. The results showed that this method improves accuracy, with insights into future research directions highlighting the potential impact of ViT-based systems.

HM-ViT, a novel hetero-modal vehicle-to-vehicle cooperative perception framework leveraging Vision Transformers, is proposed in Xiang et al. (2023b). A key strength of this research was its ability to fuse multi-modal data for enhanced perception accuracy effectively. While HM-ViT demonstrated promising results in cooperative perception tasks, its applicability and scalability due to reliance on high-quality sensor calibration require further investigation.

#### 3.1.5 Scene perception using LiDAR

To ensure secure driving of the CAVs, the investigators in Feng et al. (2018) model the uncertainties in vehicle identification and 3D bounding box regression. Moreover, it is also shown that the uncertainty model can be applied to enhance tracking and detection accuracy. The researchers in Velas et al. (2018) segmented the sparse point cloud into the ground and non-ground points using CNN to LiDAR expressed by multi-channel range images. The proposed design was shown to significantly improve over the state-of-the-art method in terms of speed and minor improvements in terms of accuracy. In Yang et al. (2018), the detection of autonomous vehicles by a CNN-based proposal-free single-stage detector in a bird's eye view representation of LiDAR points is suggested. A more complicated neural network in which the sparse 3D point cloud was encoded with a short multi-view design description is proposed in Chen et al. (2017). In Capellier et al. (2019b), authors proposed the processing of LiDAR rings instead of a full LiDAR point cloud for road segmentation and mapping. In Milioto et al. (2019) RangeNet was proposed, which utilizes range images as an intermediate representation and a CNN utilizing the rotating LiDAR sensor model. In contrast, Wu et al. (2018) attained real-time segmentation by using a CNN in-range view of LiDAR points.

Instead of splitting point clouds into clusters, an evidential end-to-end deep neural network (DNN) for classifying LiDAR objects is proposed in Capellier et al. (2019a), and their suggested design was able to classify the known objects and identify unknown objects correctly. The investigators in Asvadi et al. (2017) use point cloud segmentation and segmented obstacles projected onto a dense-depth map followed by bounding boxes fitting to the segmented objects as vehicle hypotheses. Lastly, the classification objective is achieved using the bounding boxes as inputs to a Deep CNN. The authors in Lu et al. (2019) present a new deep learning architecture termed $L_3$ for high localization accuracy using LiDAR. The proposed framework learns features by PointNet and utilizes convolutional neural networks and RNNs to predict the optimal pose.

### 3.1.6 Scene perception using DL-assisted RADAR

RADAR sensors are useful for scene perception in adverse conditions for connected vehicles as their performance is not influenced by brilliance. RADARs are active sensors that emit radio waves that aid localization and speed estimation when they bounce back from objects. The time of arrival of the reflected signal allows range and localization of the objects in the environment. In contrast to passive sensors, active sensors are prone to interference from other systems. RADAR has been used for centuries, and thus it has the advantage due to its reduced weight and cost-effectiveness.

In the context of CAVs, RADARs can be installed inside the vehicle's side mirrors. RADARs can detect an object at a high range and estimate its velocity, but they are not as accurate as LiDAR. This accuracy deficiency in estimating the shape of objects is a major flaw concerning its deployment in perception systems. The importance of RADARs, however, lies in their complementary role in poor weather conditions. The RADAR also faces a challenge of a very limited field of view, which is generally sorted using a complex array of RADAR sensors to cater full field of view. Due to these reasons, the use of RADAR is widespread in CAVs for its utilization in issuing proximity warnings and adaptive cruise control. However, Deep learning research using RADAR data for object detection is limited compared to LiDAR.

In the domain of DL-aided scene perception using radar, investigators in Major et al. (2019) proposed two ways to process the RADAR tensor. The first technique eliminates the Doppler dimension by adding the signal power over that dimension providing range-azimuth tensor. In contrast, the second strategy provides range-Doppler and azimuth-Doppler tensors as input. This leads to three model inputs being combined after primary processing in a range-azimuth-doppler model. In the end, the authors illustrated the model's viability by comparing its characteristics with LiDAR-based techniques. Similarly, authors in Patel et al. (2019) proposed a new design for RADAR-based classification that employs RADAR spectra generated by multi-dimensional Fast Fourier Transform (FFT). Their proposed technique applies deep CNNs directly to ROIs in the RADAR spectrum and thus achieves precise classification of various objects. The researchers claim that their proposed technique is a suitable substitute for classical RADAR signal processing techniques and performs better than other DL strategies. The investigators in Sligar (2020) used a physics-assisted electromagnetic simulation of a multiplex scattering environment to produce a virtual dataset. The data regarding object's distance and speed are determined from EM fields and converted into a range-doppler map. These range-doppler maps are used as input to train popular DL models based on the YOLOv3 backbone. The researchers emphasize the usability of the model for various scenarios and environments. In Engelhardt et al. (2019), researchers utilized raw RADAR data as input to deep neural networks and generated occupancy grids in the RADAR's field-of-view. The authors also validated the idea of deep learning-aided object detection by applying frustum representation. Furthermore, the authors developed a semi-automatic labeling tool using raw RADAR data collected using a test CAV. In Ristea et al. (2020), RADAR interference mitigation that relies on CNN is investigated. The proposed neural network predicts range profile magnitude with compensated noise and interference using spectrograms of noisy beat signals along with interference.

In Scheiner et al. (2019), the authors presented an architecture in which data is first converted to a common coordinate system and subsequently clustered and labeled before fea-

ture extraction. It is also claimed that the proposed method enhances overall classification performance. The research work in Lombacher et al. (2016) investigated the potential for which static object classes can be recognized in RADAR grids using deep learning methods. Several static objects which occur near roads, such as buildings, cars, fences, poles, shrubs, trees, traffic signs, and fields, were successfully classified using this method. Investigators in Scheiner et al. (2019) worked on similar grounds and evaluated the power of the deep learning method to detect the different road users. Investigators in Kim et al. (2019) proposed a recurrent convolutional neural network for V2X communications, which classifies moving targets in an automotive RADAR system.

### 3.2 Path planning, behaviour arbitration, and DL-assisted driving

An autonomous vehicle's capability to figure out a route between the starting position and destination is termed path planning. The path determination process includes evaluating all likely obstacles in the surroundings and discovering a track along a collision-free route (El Khatib et al. 2019; Grigorescu et al. 2020). Autonomous driving involves interaction with all the parties on the road while overtaking, changing lanes, giving proper way to vehicles, and taking turns on roadways that can lead to speedy arrivals at destinations. Research on connected vehicles' decision-making, safety, security, control, and standardization of rules has increased exponentially in recent years. A wide range of techniques in the deep learning domain has also been developed for these tasks.

The two main deep learning techniques regarding path planning are Imitation Learning (IL) (Rehder et al. 2017; Sun et al. 2018; Grigorescu et al. 2019), and Deep Reinforcement Learning (DRL) (Yu et al. 2018; Paxton et al. 2017). Imitation learning means learning to plan vehicle motion by imitating the observed behavior of humans. In the domain of IL, the researchers in Rehder et al. (2017) trained a network from previously observed paths. They proposed to model motion planning of an intelligent vehicle as a value iteration network. Moreover, the network performance was demonstrated by training a cost function from aerial images to resemble human driving behavior. Investigators in Sun et al. (2018) emphasized the reduction in computational complexity by proposing a two-layer architecture in which the layers perform driving policy generation and its execution subsequently. The authors in Grigorescu et al. (2019) proposed a DNN for perception planning that acquires the desired state trajectory of the vehicle under test over a finite prediction horizon. In a similar framework to IL, Inverse Reinforcement Learning is utilized in Gu et al. (2016) to learn the reward function from an individual driver and subsequently generate human-like driving trajectories.

On the other hand, DRL-based planning was proposed in Yu et al. (2018), where the environmental model was condensed into a simple virtual environment model first. Then DRL training was applied to obtain the optimal control-trajectory sequence. In Paxton et al. (2017), authors discussed a methodology based on reinforcement learning to learn both linear temporal logic constraints and control policies to generate task and motion plans, whereas in Panov et al. (2018), the usability of the DRL approach is evaluated for path planning on square grids. Driver inattention and vehicle automation interact in a complex way depending on the level of vehicle autonomy. Instead of using the electrocardiography or photoplethysmography signal for driver alertness, in Trenta et al. (2019), the researchers examined the skin micro-movements and variations in face color due to blood flow to

extract facial landmarks. The researchers in Zyner et al. (2018) presented an RNN-aided prediction method that uses LiDAR input and estimates driver plans at an un-signalized roundabout, whereas Baheti et al. (2018); Kim et al. (2017); Xing et al. (2019) and Le et al. (2016) focus on using CNN and Faster R-CNN respectively to detect driver alertness. AI assistance in driving is leading towards a massive increase in different applications for ease of driving such as blind-spot reduction (Virgilio et al. 2020; Zhao et al. 2019b; Shen and Yan 2018), traffic signal and signs detection (Alghmgham et al. 2019; Tabernik and Skočaj 2020; Nagpal et al. 2019; Kukreja et al. 2020), lane deviation detection (Wei et al. 2019; Satti et al. 2021; Du et al. 2020a) and vehicle make and model classification (Satar and Dirik 2018; Nazemi et al. 2020; Manzoor et al. 2019; Artan et al. 2019).

### 3.3 End-to-end deep learning

End-to-End deep learning in CAV can be defined as "direct mapping by a neural network from sensor data to vehicular control instructions". The inputs to a DNN can be high-dimensional sensor data like images or point clouds interpreted to control commands by End-to-End networks.

One of the first works on End-to-End learning was introduced in the 1990 s when a 3-layer back-propagation network called Autonomous Land Vehicle In a Neural Network (ALVINN) (Pomerleau 1988). The ALVINN was devised for following the road and steering as per the perceived road curvature. The training of ALVINN was carried out using simulated road images and test results showed that it can efficiently follow actual roads. In Bojarski et al. (2016), authors illustrated that CNNs are able to learn tasks such as lane or road following using crude pixel information from a single front-facing camera and can map directly to steering commands. Their scheme (dubbed DAVE-2) for End-to-End deep learning can be visualized in Fig. 5 demonstrating that CNNs can learn the entire task of the road following without any manual breakdown into subtasks.

In Xu et al. (2017), researchers proposed combining a convolutional network and a Long Short-Term Memory (LSTM) network to learn a general model of vehicle movement from large-scale video data. In Eraqi et al. (2017), it is taken into account the combination of visual and dynamic temporal dependencies of the input data where the convolutional long–short-term memory (C-LSTM) network has been utilized for steering control. In Hecker et al. (2018), the $360^o$ view of the surrounding area is captured via sensors. All the information around the vehicle was united into the network model to produce an appropriate control command. In Rausch et al. (2017), investigators designed a CNN to map pixel data taken from a frontal camera to steering commands without involving other sensors and compared their designed system performance with the human steering behavior.

The researchers presented DeepPicar in Bechtel et al. (2018), a deep convolutional neural network and a low-cost mini-model of DAVE-2 (a self-driving car by NVIDIA). The vehicles having DeepPicar can estimate the steering angles of a CAV in real-time utilizing a webcam in conjunction with a Raspberry Pi 3 quad-core platform. In Yang et al. (2017), the research team used "The open racing car simulator" for data collection and classified the image features into sky-related, roadside-related, and road-related categories. Moreover, multiple experimental evaluations are employed to investigate the influence of every feature for training a CNN controller. The investigators in Sallab et al. (2017), incorporated RNN for information synthesis, equipping the car to handle partly visible situations. The

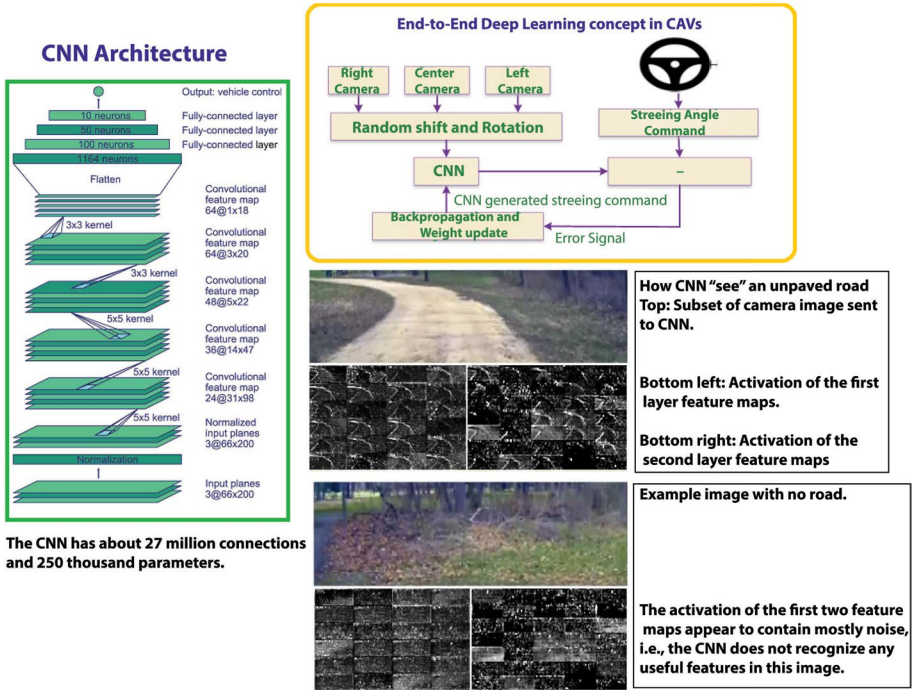


**Fig. 5** NVIDIA End to End Learning for Self-Driving Cars (Bojarski et al. 2016). NVIDIA DevBox and Torch 7 for training and an NVIDIA DRIVE-PX self-driving car computer also running Torch 7 for determining where to drive. The system was trained using three cameras, a single camera in operation, and processing 30 frames per second

investigators have also shown its ability to learn complex road curvatures in the racing car simulator. The researchers in Pan et al. (2018) trained a CNN control policy to map raw observations to steering and throttle commands that enable a speedy off-road vehicle movement. An asynchronous advantage actorcritic framework had been adopted in Jaritz et al. (2018) to learn the car control in a realistic rally game in which the agents can emerge concurrently on different tracks. In Codevilla et al. (2018), command-conditional imitation learning is proposed based on learning basic controls and advanced-level commands from the presentations of an expert. The researchers in Hecker et al. (2018) extended the driving data set by utilizing eight cameras to capture videos while driving, furthermore presented a new DNN that can map the sensor inputs to future driving maneuvers. A blend of modular designs and End-to-End deep learning approaches were suggested in Müller et al. (2018), so that driving policy is not revealed to raw input or basic vehicle dynamics. The investigators in Sauer et al. (2018) proposed a perception method that can map video input to intermediate forms fit for autonomous navigation in complicated urban environments with an improvement claim of up to 68 percent compared to the latest reinforcement and conditional imitation learning designs.

In the deep reinforcement learning domain, Kendall et al. (2019) applied the DRL technique to learn a policy for lane following using a single monocular image as input and reward as the distance traveled by the vehicle without the driver taking control. In Liang et al. (2018), the authors presented a novel model termed Controllable Imitative Reinforce-

ment Learning (CIRL) for challenging vision-aided autonomous driving. The proposed method takes into account the controllable imitation learning with Deep Deterministic Policy Gradient (DDPG) policy learning to fix the reinforcement learning poor efficiency issues. In Amini et al. (2019), authors utilized raw camera data and higher-level road-maps in a novel variational network to estimate probability distribution over the possible and deterministic control command for navigation. Authors in Bansal et al. (2018) trained a policy for driving CAVs through imitation learning where mid-level input and output representations that exploit perception and control components are selected to diminish complexity. The mid-level input is fed to an RNN, dubbed ChauffeurNet, whose output drives trajectory rendered into steering and acceleration by the controller. In Bewley et al. (2019), a system that uses simulation to learn an End-to-End driving policy readily transferable to real-world scenarios is presented and validated against several baselines. The researchers in Codevilla et al. (2018) focused on the issue of the unrealistic approach of modeling a wide variety of complex environmental conditions and proposed behavior cloning to achieve state-of-the-art results.

The investigation in Xiao et al. (2020) explored the combination of RGB and depth modalities producing better end-to-end AI drivers, whereas an End-to-End conditional imitation learning by linking lateral and longitudinal control on vehicles is explored in Hawke et al. (2020). The research in Maanpää et al. (2021) is an effort to extend End-to-End learning using multi-modal data collected by 28 h of driving on several roads in adverse weather conditions. The work in Chi and Mu (2017) focuses on a vision-based model that autonomously drives a car solely from its camera's visual observation by mapping it to steering angles. The novelty is claimed based on learning from real human driving videos instead of being trained from synthetic data, taking informative historical states of a vehicle into account, and using the visual back-propagation scheme for visualizing image regions. In Gurghian et al. (2016), the images from laterally mounted down-facing cameras are used in a convolutional neural network for lane detection. The research claims to achieve high-accuracy lane position for keeping vehicle lane alignment to center and real-time navigation. In Kocić et al. (2019), the investigators proposed a very light neural network that leads to lower latency and successful autonomous driving with similar effectiveness compared to the state-of-the-art models in autonomous driving.

### 3.4 LLM-based designs in CAVs

Large Language Models are advanced deep learning models trained on massive amounts of text data to understand, generate, and manipulate human language. While traditionally used in natural language processing (NLP) tasks such as translation, summarization, and conversation, LLMs are now being explored for their potential in designing and operating CAVs (Cui et al. 2024). LLMs are capable of sophisticated human–machine interactions, allowing vehicles to understand and respond to complex voice commands and predict passenger needs based on conversational cues. LLMs can also be adapted to process CAVs' multi-modal data, providing a unified understanding of the vehicle's environment (Cui et al. 2024; Tong and Solmaz 2024). LLMs can also generate realistic driving scenarios and dialogues for training and testing autonomous systems. This trait enables CAVs to handle a wide range of situations, including rare and complex events and associated policy-making.

In Cui et al. (2024), authors presented a survey on multimodal large language models (LLMs) for autonomous driving, focusing on heterogeneous modalities and cross-modal learning. This investigation introduced the topic well, however, without an in-depth analysis. In the domain of end-to-end autonomous driving systems based on LLMs, Xu et al. (2024b) introduced DriveGPT4, as an innovative, interpretable solution. DriveGPT4 could process multi-frame video inputs and textual queries, enabling it to interpret vehicle actions, provide reasoning, and predict low-level control signals. This pioneering effort employed LLMs for autonomous driving, using a custom visual instruction tuning dataset and a mix-finetuning strategy to achieve driving capabilities. Tong and Solmaz (2024) explored the integration of LLM-based DL with CAVs focusing on improving the traffic conditions. The authors proposed "ConnectGPT," a pipeline that connects LLMs with CAVs, using GPT-4 to monitor traffic, identify hazards, and automatically generate standardized safety messages for smooth CAVs operations.

Sha et al. (2023) employed LLMs to address challenges of complex autonomous driving scenarios. They created cognitive pathways for comprehensive reasoning with LLMs and developed algorithms to translate their decisions into driving commands. By integrating LLM decisions with low-level controllers using guided parameter matrix adaptation, their approach outperformed baseline methods in both single-vehicle tasks and multi-vehicle coordination. Cui et al. (2023b) presented a pioneering approach, Drivellm, that harnesses the power of large language models for autonomous driving. The authors demonstrate a promising direction, leveraging LLMs' capabilities in processing complex scenarios. While this approach pioneers the application of large language models in autonomous driving, its methodology raises concerns. The authors' reliance on pre-trained LLMs without thorough fine-tuning may lead to biased or inaccurate decision-making in complex driving scenarios. Chen et al. (2024b) presents a novel approach leveraging LLMs for autonomous driving. A significant strength is the authors' innovative fusion of object-level vector modality, enabling explainable decision-making. The paper's experimental results demonstrate improved performance in various driving scenarios, showcasing the potential of LLMs in autonomous driving. Yildirim et al. (2024) presented an innovative approach, called HighwayLLM, combining reinforcement learning and language models for highway driving. A notable strength is the authors' attempt to leverage language models for decision-making, showcasing promising results in simulated environments. Additionally, the paper's discussion on integrating reinforcement learning and language models highlights potential benefits for autonomous driving.

### 3.5 Critical analysis of modular versus end-to-end learning & LLM-based designs

The conventional method for controlling autonomous vehicles is the modular CAV design. This approach breaks down the driving task into smaller sub-problems, where distinct DL modules can be trained for environmental perception, path planning, and motion control. The modular design is favored for its transparency and interpretability compared to the End-to-End (E2E) design. However, it is susceptible to error propagation, where inaccuracies in one module can lead to compounded errors in subsequent modules. Conversely, the E2E approach integrates all driving tasks into a single model, facilitating easier error detection and correction. This design enables the learning of advanced driving strategies and generally performs well on simple datasets. Despite these advantages, the E2E approach

has significant limitations, notably a lack of interpretability crucial for safety assurance. Additionally, its single-step mapping can result in less efficient learning processes. furthermore, E2E models require vast amounts of training data to learn to drive safely in various conditions. Collecting, annotating, and curating such data can be resource-intensive. E2E models are often considered "black boxes," making it challenging to understand why a particular decision was made (Chib and Singh 2023). This lack of interpretability is a significant concern for safety and regulatory approval. The learned behavior in E2E models might not be easily transferable to different geographic locations, weather conditions, or vehicle types, while modular approaches can be more adaptable. Modular systems often struggle to handle corner cases or rare, unanticipated scenarios because each module may not have been explicitly designed for them. E2E systems, on the other hand, exhibit better generalization across diverse scenarios due to their learning from extensive and varied datasets. Furthermore, modular systems can experience delayed decision-making due to the sequential processing stages. In contrast, E2E systems potentially reduce latency by processing all relevant information in a single step. Another advantage of E2E systems is their continuous trainability on new data, allowing adaptation to evolving road conditions, traffic patterns, and regulations. Modular systems, in comparison, might require manual updates and adjustments to individual modules (Tampuu et al. 2020).

There has been a growing trend in recent years of testing DL-enabled autonomous driving models using open-source simulators. Figure 6 illustrates a generic simulator architecture.

Several simulators are available for testing DL-enabled autonomous driving models, including CarSim (Johansson et al. 2004), PreScan (Ortega et al. 2020), Gazebo (Ahamed et al. 2018), LGSVL (LG–Autonomous 2023) and CARLA (Dosovitskiy et al. 2017). Notably, CARLA (CAR Learning to Act) (Dosovitskiy et al. 2017) supports developing, training, and validating autonomous urban driving systems. CARLA's versatility lies in its capacity to allow flexible specification of sensor suites and environmental conditions. This simulation platform is particularly effective in evaluating three distinct approaches to autonomous driving: a traditional modular pipeline, an E2E model trained through imitation learning, and an E2E model trained via reinforcement learning. Authors in Niranjan et al. (2021) highlighted the usability of the CARLA simulator for testing object detection algorithms and getting meaningful results. They utilized CARLA to generate a dataset for training an object detection model, which was subsequently evaluated on test images to assess its performance within the CARLA environment.

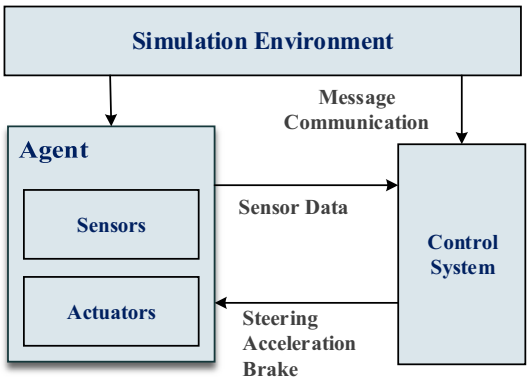


**Fig. 6** Simulator architecture

The CARLA leaderboard serves as an evaluative tool for gauging the proficiency of autonomous driving systems in uncertain environments. This tool presents vehicles with predetermined routes that encompass challenging scenarios, including sudden lane changes and unforeseen pedestrian crossings. The evaluation criteria include measuring the distance successfully navigated by the vehicle within a specified time frame on a designated town route and the tally of infractions incurred during the journey. The assessment employs multiple metrics, offering a thorough analysis of the driving system's performance. This comprehensive evaluation framework enables a detailed understanding of how autonomous vehicles respond to and manage unpredictable driving situations.

Average driving proficiency score or Driving Score (DS) is a metric that reflects the average route completion percentage with average infraction penalty. Similarly, route completion depicts the average percentage of Routes Completed (RS) by the model. Based on the submission in the CARLA leaderboard till August 2024, we can deduce that E2E models, including ResonNet and InterFuser (Shao et al. 2023b, a), are leading the leaderboard with DS value of 79.95 and 76.18 respectively. Corresponding RS values for the two leading models are 89.89 and 88.23. Modular approaches, such as Rosero et al. (2022) and Rosero et al. (2020) with DS of 15.40 and 4.56 as well as RS values of 50.05 and 23.80, respectively, were lagging in comparison with E2E designs. These values show that the modular approaches perform significantly less when benchmarked in CARLA leaderboard KPIs.

In the autonomous vehicle industry, companies are adopting different approaches based on the requirements of their systems, safety protocols, and operational challenges. Tesla, a leader in autonomous driving, is shifting from the modular approach towards end-to-end learning Cohen (2023). As this approach relies heavily on DL models that process raw sensory inputs directly into control outputs, it gives the liberty to jointly optimize perception, planning, and control, potentially resulting in better overall performance. This approach is beneficial in optimizing real-time decision-making, such as lane changes or complex maneuvers. However, the data dependence of Tesla's approach is a significant concern. Tesla requires massive amounts of driving data to train its neural networks, and while its fleet provides real-world data, the model's lack of transparency poses a risk when things go wrong. The system operates as a "black box," which makes debugging and understanding the decision-making process difficult. This raises safety concerns, especially in corner cases where the system's behavior might not be fully understood or predictable.

Waymo, another leader in autonomous driving, employs a modular learning architecture, which divides its learning module into distinct components such as perception, planning, and control (Cortese 2025). This approach ensures reliability, safety, and transparency, all of which are crucial aspects for self-driving vehicles operating in real-world environments. Modular systems allow each module to be specialized and optimized independently, making it easier to test and validate the system in varied scenarios. However, this separation often leads to challenges in integrating these modules efficiently, resulting in performance losses and slower adaptation to dynamic conditions.

In conclusion, there is a growing shift towards end-to-end driving systems, as demonstrated by companies like Motional and other autonomous vehicle developers, is consistent with the findings of deep learning models tested on simulation platforms such as CARLA (Shao et al. 2023b). While modular approaches have shown considerable advantages in terms of reliability and transparency, which are essential for ensuring the safety of both vehicle occupants and other road users, they also face certain specific challenges (Hussain

et al. 2025). In particular, the integration of separate modules may result in suboptimal performance and a delayed response to dynamic environmental changes. The ongoing research into end-to-end systems aims to address these issues by improving adaptability and performance across complex, real-world scenarios, suggesting a potential path forward for the future of autonomous driving (Coelho and Oliveira 2022).

The integration of Large Language Models (LLMs) into Connected and Autonomous Vehicles represents a promising frontier in the advancement of autonomous driving technologies. However, despite the innovative potential of LLM-based designs, several critical shortcomings must be addressed to ensure their practical applicability and safety in real-world scenarios.

One of the primary concerns with LLM-based designs, such as those proposed in Drivellm (Cui et al. 2023b), is the lack of rigorous experimental evaluation. These designs often fail to rigorously test critical corner cases, omitting essential comparisons with established autonomous driving approaches. This oversight can lead to significant gaps in understanding how these models perform under challenging and unpredictable conditions, which are common in real-world driving environments. The experimental setups in these studies often rely heavily on simulated scenarios, which, while useful, may not fully capture the complexities and uncertainties of real-world driving. Consequently, the applicability of these LLM-based systems in actual driving situations remains uncertain, raising concerns about their reliability and robustness.

The investigation by Chen et al. (2024b) also has some room for improvement. The reliance on pre-trained LLMs may lead to biases and limitations in handling corner cases. Additionally, the explainability aspects, while promising, require further development to provide more insightful interpretations. The paper's evaluation could benefit from more comprehensive metrics, including safety and robustness assessments. Nevertheless, investigation by Chen et al. (2024b) contributes meaningfully to the emerging field of LLM-based autonomous driving, offering a promising direction for future research.

In the similar manner, the investigation by Yildirim et al. (2024) relies heavily on simulated scenarios, which may not fully capture real-world complexities. Furthermore, the evaluation metrics could be more comprehensive, incorporating safety, robustness, and computational efficiency assessments. Nevertheless, HighwayLLM (Yildirim et al. 2024) contributes to the growing body of research exploring AI-driven autonomous driving solutions, and its ideas warrant further exploration and refinement.

### 3.6 Corner cases in DL-assisted CAVs

The recent developments in autonomous driving and deep learning techniques rely upon the availability of huge amounts of training data for training purposes. The use of deep learning systems in CAVs is a black-box approach that furnishes a quick mapping solution. However, it also poses a risk. Recently, real-world accidents related to CAVs confirm the lack of robustness in these systems (Bolte et al. 2019). One cause of such accidents is the deficiency in training data concerning capturing all critical situations (Tian et al. 2018). A classic example of a corner case is the crash between a Tesla car and a trailer due to unsuccessful differentiation of "white color against a brightly lit sky" and the "high ride height" by Tesla's DL system (Ouyang et al. 2021). At the conceptual level, the corner-case role in DNN is just like conventional software logic bugs. However, these cases lead to potentially

fatal collisions. To avoid such situations, testing, evaluation and update are crucial steps in developing a safe CAV system. In the testing phase, the CAV is evaluated in safety-critical situations, which infrequently happen in a common driving environment. How to systematically generate these corner cases aiming for training is a challenging task. Several investigations aim to develop a corner case. detection system to identify unusual scenarios. In Bolte et al. (2019), a formal definition for a corner case for driving a CAV, along with a system framework that provides both the online and the offline use case for cameras and subsequently outputs a corner case score, is offered. In Sun et al. (2021), a single framework is introduced to generate corner cases for the decision-making systems where the high dimensionality issue is addressed using the Markov decision process. In Tian et al. (2018), a systematic testing tool dubbed "DeepTest" was designed and evaluated to automatically detect erroneous behaviors of DNN-driven vehicles to avoid fatal crashes. "DeepTest" was found capable of detecting numerous erroneous behaviors under different practical driving conditions like rain, lightning, fog, and blurring, leading to lethal accidents in three high-performing DNNs of the Udacity self-driving car challenge. In Yu et al. (2021), the Multi-Relation Graph Convolution Network (MR-GCN) and attention layers are introduced to model the risk of driving manoeuvres. In Zhao et al. (2017a, 2017b), authors proposed important sampling techniques to generate test cases regarding lane-changing manoeuvres and car-following scenarios. The overvalue problem of most serious cases is addressed in Feng et al. (2021, 2020a, 2020b, 2020c) by defining the manoeuvre challenge and exposure frequency and by forming cases on several environment settings such as car-following scenarios, cut-in scenarios, and highway driving environment. Some investigators offered the risky index and the probabilistic model of the environment to assist in creating critical cases, such as Akagi et al. (2019) used a self-defined risky index and naturalistic driving data to sample critical cut-in scenarios. Adding to this effort, O'Kelly et al. (2018) utilized a deep learning framework to calibrate a naturalistic driving model. Investigators in Ding et al. (2020) modeled the environment as the union of blocks and used REINFORCE algorithm to create corner cases in limited traffic load. The adaptive stress testing method is proposed by Koren et al. (2018) that suggested Monte Carlo tree search and deep reinforcement learning to resolve the pedestrian-crossing problem. However, this study also considered limited traffic and pedestrians. The study by Karunakaran et al. (2020) used the Deep Q Network (DQN) to generate corner cases. Despite all these efforts, due to difficulties in modeling complex scenarios, the generation and identification of corner cases that are a true representation of actual situations with high coverage and variability remain an open challenge.

A number of potential solutions are also suggested by researchers concerning the corner cases problem. The lack of robustness in deep learning systems due to deficiency in training data regarding uncommon situations dubbed "corner cases" must be addressed by the automotive industry to ensure safety in CAVs (Sun et al. 2021). The first straightforward solution concerning the testing and evaluation of CAVs is to purposely and systematically generate these corner cases. The corner cases are unpredictable, and engineers can't figure out and cover all the scenarios without road tests in the real world. Consequently, the CAV industry is working on virtual simulation to simulate the unusual scenarios that seldom happen in the real world and subsequently train CAVs. The other dimension in resolving this issue is multi-sensor fusion. Because CAV's sensors perceive the environment differently, a corner case for one sensor might be a common scenario for others, such as in dim light or

night. A camera might not detect a black car, whereas, for a RADAR, this scenario does not fall in a corner case category.

The CAV's ability to autonomously find the most similar corner case relevant to the current situation can further improve the performance of well-coordinated sensor fusion algorithms. The sensor fusion algorithms combine data from several sensors to determine the most accurate object information. For example, if LiDAR provided the most accurate data in a certain "corner case" previously, the sensor fusion algorithm will assign more weight to the LiDAR sensor considering its better performance in a similar situation.

## 4 Deep learning in UAVs

DL constructs for UAVs differ significantly from those used in CAVs, largely due to variations in data characteristics. UAV datasets benefit from flexible data collection without geographic constraints but also face challenges such as varying sensor-object distances, wide viewing angles, and substantial illumination changes. These factors contribute to low spatial resolution, diverse object sizes, complex backgrounds, and a higher object count per image (Wu et al. 2022; Jain et al. 2021).

Aerial detection is further complicated by occlusions caused by buildings and trees. Additionally, UAV-specific limitations such as power consumption, payload capacity, flight time, and operational range must be considered when designing DL algorithms (Carrio et al. 2017). These distinctions necessitate specialized DL approaches tailored to aerial datasets rather than directly applying models developed for connected vehicles. The following subsections detail various DL architectures explored for object detection using UAV sensor data.

### 4.1 Scene perception using image sensors

Deep learning methods for UAV-based scene perception rely mainly on CNNs for feature extraction (Karim Amer et al. 2019). Scene perception tasks include object identification and scene classification. In Gangopadhyay et al. (2015), a statistical aggregation approach using CNNs was proposed to classify videos of natural dynamic scenes. For object detection, Lee et al. (2017) applied Faster R-CNNs, proposing a computational split between low-level object detection and short-term navigation for online processing. In Wang et al. (2018), the performance of SSD, Faster R-CNN, and RetinaNet was evaluated on the Stanford Drone Dataset. Similarly, Ammour et al. (2017) used a pre-trained CNN with a linear SVM to detect car regions, while Radovic et al. (2017) demonstrated that parameter tuning significantly improved CNN classification accuracy on aerial images.

DL techniques trained on UAV aerial datasets have attracted interest, especially for surveillance and rescue. Mittal et al. (2020) provided a comprehensive review of state-of-the-art DL algorithms implemented on low-altitude UAV datasets. In Kim and Chervonenkis (2015), DL-based image segmentation detected accidents and abnormal traffic using UAV vision systems. Antonio and Maria-Dolores (2022); Guillen-Perez and Cano (2019) proposed an end-to-end Multi-Agent Deep Reinforcement Learning framework for collaborative CAV control at intersections, capturing complex traffic dynamics. Despite their

effectiveness, UAV constraints such as flight time, energy, and payload necessitate low-complexity DL algorithms tailored for onboard deployment (Carrio et al. 2017).

Addressing these constraints, an empirical study by Purdue and CCAT (Zong et al. 2023) focused on intersection monitoring for crash risk assessment under rainy weather using UAVs. The study leveraged the VisDrone dataset comprising 400 videos (265,228 frames) captured via drone-mounted cameras in diverse scenes. They developed a monitoring framework evaluated using Multiple Object Tracking Accuracy (MOTA), reporting $64.89\%$ on the training set and $63.12\%$ on testing. The study also introduced a denoising framework evaluated through Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). However, a key limitation was the use of synthesized rain images, with minimal validation of their real-world fidelity.

In another empirical study, Barmpounakis and Geroliminis (2020) conducted a big data-based traffic congestion analysis by creating an extensive UAV-based traffic stream data repository. The work offered an opportunity to test traffic models developed from diverse disciplines. A swarm of 10 drones hovered over the central business district of Athens over multiple days. This swarm of drones recorded traffic streams in a congested area of a $1.3km^2$ area with more than 100 km lanes of road network, around 100 busy intersections (signalized or not), including many bus stops, and close to half a million trajectories. The experiment aimed to record traffic streams in a multimodal congested environment. Analysis of the dataset revealed that taxis, stopping randomly for $5-15$ seconds, and buses, stopping at fixed locations for $30-40$ seconds, frequently created static and moving bottlenecks. For instance, a taxi stop caused a queue behind it, with a waiting time of more than 10 seconds for affected vehicles before a lane change was possible. Similarly, a bus stop near a traffic light resulted in a queue where no vehicles passed the stop line for a 20-second interval, underutilizing the green phase. In contrast, another bus stop of similar duration had no capacity loss due to better traffic flow management, emphasizing the variability of impacts. These findings underline the significant role of such stops in affecting lane capacity and multimodal traffic interactions.

#### 4.1.1 Vision transformers in UAVs

This subsection reviews the application of ViTs in UAV-based computer vision tasks such as image classification, object detection, segmentation, and tracking. ViTs have gained significant focus in UAV imagery, particularly for object detection, due to their potential to enhance autonomy, accuracy, and efficiency. However, challenges like data quality, class imbalance, real-time processing, and object scale variation caused by varying altitudes and motion blur during low-altitude flights complicate deployment. Various transformer architectures have been proposed to address these issues.

Several ViT variants, ViT-Base, ViT-Large, ViT-Hybrid, and Swin Transformer, have been tailored for UAV tasks using datasets like UAVid, VisDrone, Campus, UAV123, UAVDT, MDOT, AU-AIR, and SynDrone (Rizzoli et al. 2023). Zhao et al. (2023) introduced TPH-YOLOv5, an enhanced YOLOv5 with an added tiny-object prediction head and transformer-based heads. Its successor, TPH-YOLOv5++, integrates a cross-layer asymmetric transformer module to reduce computational cost while maintaining performance.

Tahir et al. (2024) proposed PVswin-YOLOv8s for pedestrian and vehicle detection, combining Swin Transformer based global feature extraction with channel and spatial atten-

tion modules. Soft-NMS was employed to improve occlusion handling. A comprehensive evaluation of the PVswin-YOLOv8s model benchmarked against various YOLO versions (YOLOv3, YOLOv5, YOLOv6, and YOLOv7) and classical object detectors (Faster-RCNN, Cascade R-CNN, RetinaNet, and CenterNet), revealed a significant improvement in average detection accuracy (mAP) of 4.8% over YOLOv8s on the VisDrone2019 dataset, thereby validating its efficacy in detecting small objects and enhancing overall detection performance.

Tran et al. (2024) developed an unsupervised transformer-based framework for anomaly detection in aerial surveillance. By predicting future frames and analyzing reconstruction errors, their model outperformed state-of-the-art methods on the UIT-ADrone and Drone-Anomaly datasets. Chen et al. (2024a) addressed decentralized, scalable UAV navigation with a transformer-based multi-agent reinforcement learning (T-MARL) algorithm. T-MARL integrates the Transformer's adaptability and attention mechanism with deep RL to optimize cooperative UAV trajectories for area coverage. Xu et al. (2022a) tackled dense object distribution using Foreground Enhancement Attention Swin Transformer (FEA-Swin), which enriches Swin Transformer with contextual information and integrates an improved BiFPN to retain small object details. The model demonstrated a balanced trade-off between accuracy and efficiency.

In visual tracking, Xu et al. (2022b) proposed STN-Track, combining STN-YOLOX detection and G-Byte tracking to improve accuracy and identity retention on UAVDT and VisDrone MOT datasets. Similarly, Ye et al. (2023) introduced RTD-Net for real-time object detection. The model integrates a Feature Fusion Module (FFM) for small object detection, a Lightweight Extraction Module (LEM) for real-time efficiency, and a Convolutional Multihead Self-Attention (CMHSA) block to enhance occluded object recognition, achieving $86.4\%$ mAP on a UAV dataset.

In summary, ViT-based models significantly advance UAV perception by improving object detection and tracking, particularly for small or occluded objects. Despite these gains, challenges such as data variability and real-time constraints necessitate continued innovation in ViT architectures for UAV deployment.

### 4.2 Scene perception using acoustic sensors, RADAR and LiDAR

DL has demonstrated superior performance over traditional computer vision methods in processing UAV aerial data. While CNNs are the most commonly employed architectures for aerial image analysis, other DL models have also been applied across various UAV sensing modalities.

In acoustic sensing, a partially shared deep neural network was used in Morito et al. (2016) to extract human voices from noise-suppressed signals for detecting help requests in disaster scenarios. Similarly, Jeon et al. (2017) explored acoustic classification using Gaussian Mixture Models and CNNs. These studies highlight the growing use of machine learning in UAV-based emergency response.

The recent advances in RADAR research point toward the increasing use of machine learning techniques in advanced target classification (Mendis et al. 2016; Huizing et al. 2019; Park et al. 2021; Samaras et al. 2019). Mendis et al. (2016) proposed using deep belief networks to classify spectral correlation function signatures of micro UAV systems. Huizing et al. (2019) utilized CNN and LSTM-RNN architectures to classify targets based on micro-

Doppler signatures. Park et al. (2021) introduced a ResNet-SP model, an enhancement over ResNet-18, trained on RADAR spectrogram images. Their model achieved higher accuracy with reduced computational complexity.

For LiDAR data processing, Maturana and Scherer (2015) presented a 3D-CNN framework coupled with a volumetric occupancy map for identifying safe UAV landing zones. UAV-based LiDAR has also been explored in infrastructure monitoring. For example, Liu et al. (2019b) applied a random forest model to classify pavement distresses such as cracks, potholes, and rutting using low-altitude UAV-generated point clouds. A broader review of UAV LiDAR applications in road safety, traffic surveillance, and infrastructure management is provided in Outay et al. (2020).

These works collectively underscore the versatility of DL in fusing data from acoustic, RADAR, and LiDAR sensors to enhance UAV-based perception across safety, surveillance, and control tasks.

### 4.3 UAV's path planning, navigation, and control

DL has significantly advanced UAV path planning, navigation, and control, particularly in unstructured and dynamic environments. Situational awareness, i-e, UAVs' understanding of their state and environment is crucial for selecting optimal routes. Chang et al. (2019a) provided a comprehensive review of DL methods for UAV path planning. For localization, Lin et al. (2015) proposed matching ground-level query images with aerial views using CNNs. Padhy et al. (2021) presented a deep neural network framework that utilizes RGB images from a UAV's front camera to enable corridor navigation.

In autonomous landing, LI and HU (2021) developed a model integrating: i) a DNN-based bounding box detector, ii) an extended Kalman filter-based coordinate combiner, and iii) PointRefine-Net for improving detection accuracy. For adaptive navigation, Theile et al. (2020) proposed a double deep Q-network to handle diverse mission scenarios, while Luo et al. (2018) introduced "Deep-Sarsa," an on-policy reinforcement learning algorithm that facilitates path planning and obstacle avoidance via environmental feedback.

The capability of neural networks to handle high-dimensional data has enabled their application in complex control problems, where classical control theory falls short under model variations or disturbances (e.g., damaged propellers, wind, or rain). Shah et al. (2016) introduced "DeepFly," an autonomous flight system using a monocular camera and disparity images to select obstacle-free waypoints. Lin et al. (2014) proposed a recurrent wavelet neural network (RWNN)-based control system for robust motion tracking under crosswind and control disturbances. Punjani and Abbeel (2015) designed a hierarchical ReLU-based network for executing complex helicopter maneuvers.

Recent studies focus on deep reinforcement learning (DRL) for tasks like target tracking, attitude control, and landing on static and mobile platforms. Li et al. (2017) proposed a hierarchical control scheme that combines model-free policy gradient methods with PID controllers for safe target tracking. Koch et al. (2019) addressed control issues in unpredictable environments by developing intelligent DRL-based flight controllers. Polvara et al. (2018) utilized low-resolution, earth-oriented camera images in a DRL-based framework for autonomous landing. Qing et al. (2018) employed an adaptive radial basis function neural network and backstepping control to manage unknown disturbances during UAV landing.

In Ma et al. (2024), authors introduced GN-Trans, a hybrid Graph Neural Network (GNN) and Transformer architecture for mission planning in UAV-CAV systems. The model combines a global Transformer for high-level behavior modeling and a local Transformer for region-specific task allocation and path planning. Evaluations on the Stanford Drone and CityScapes datasets showed GN-Trans achieved $92\%$ task allocation accuracy and $88\%$ resource utilization–outperforming Dijkstra's algorithm ($70\%$) and RL-based models ($82 - -89\%$). Ablation studies demonstrated complementary benefits of GNNs ($87\%$) and Transformers ($89\%$), while GN-Trans yielded $12 - -15\%$ improvements in dynamic scenarios. The model scaled to 50 UAVs and 30 CAVs, achieving robust performance across varied environments ($97.5\%$ UAV accuracy in urban settings). GN-Trans effectively bridges relational and contextual AI, setting a new benchmark in autonomous IoT mission coordination.

### 4.4 Large language models in UAVs

The integration of Large Language Models (LLMs) into UAV systems represents a significant advancement, enabling enhanced decision-making, natural language interaction, and autonomous mission planning. By employing their predictive and generative capabilities, LLMs support real-time adaptability, efficient communication, and greater autonomy, particularly valuable in domains such as search and rescue, environmental monitoring, remote sensing, and military operations. Furthermore, LLMs extended to the vision domain have demonstrated strong multi-modal reasoning, opening new avenues in UAV-based applications.

Several notable contributions illustrate the application of LLMs in UAVs. Zhan et al. (2024) proposed SkyEyeGPT, a multi-modal LLM (MLLM) designed for remote sensing. It employs a two-stage tuning strategy to improve instruction-following and multi-turn dialogue across different granularities, achieving superior performance on eight remote sensing vision-language datasets. Similarly, Xu et al. (2024a) introduced RS-Agent, an autonomous remote sensing agent that combines LLMs with advanced remote sensing image processing tools. This RS-Agent excels in tasks such as scene classification, visual question answering, and object counting across multiple benchmarks.

Beyond remote sensing, LLM integration into future wireless networks has been explored. Javaid et al. (2024b) emphasized LLMs' potential to reduce latency, optimize data flow, enhance signal processing, and manage network traffic through advanced prediction and real-time decision-making. Extending this idea, Jiang et al. (2024) demonstrated how collaborative, self-improving LLM-enhanced agents can address complex problems in 6 G communication. A broader survey by Javaid et al. (2024a) reviewed LLM architectures suited for UAV deployment, summarizing current trends, design frameworks, and potential integration pathways for future LLM-based UAV systems.

In Wang et al. (2023c), authors explore the integration of Large Language Models (LLMs) into autonomous driving systems, applying LLMs to behavior planning and safety enhancement. From a UAV perspective, De Curtò et al. (2023) combined LLMs and Vision-Language Models (VLMs) to enable zero-shot scene-to-text descriptions using UAV imagery via a state-of-the-art detection pipeline. Lastly, Abu Tami et al. (2024) proposed an MLLM framework that utilizes object-level question-answering prompts to enhance safety-critical event detection, offering actionable insights through robust logical and visual reasoning.

### 4.5 Critical analysis of UAVs-supported detection in CAV networks

This section discusses the use of UAVs for real-time detection and tracking of objects such as vehicles, pedestrians, and obstacles within vehicular networks. Despite their unobstructed aerial view advantages, UAVs face several challenges in object detection:

- UAV imagery is typically captured from altitudes far exceeding inter-vehicular distances. This results in relatively small object representations, distorted views due to oblique angles, and motion blur from UAV-object relative movement, increasing false detection rates (Zhou et al. 2022; Li et al. 2018).
- Occlusion is caused by environmental elements or poor illumination, impairs visibility, and leads to missed or false detections (Scott et al. 2016). Researchers have proposed various methods to address occlusion under diverse conditions.
- Further issues in deep learning-based detection include scale variation, object similarity, and real-time processing demands (Bouguettaya et al. 2022).

To tackle these challenges, numerous DL-based vehicle detectors have been developed. Single-stage detectors, especially the YOLO series, have evolved to enhance real-time performance and improve detection of small-scale targets in complex scenes.

YOLOv3 (Bochkovskiy et al. 2020) significantly improved computational efficiency and resource utilization. Innovations like YOLO-GCC and Traffic-DQN presented in Li et al. (2021a; 2021b) further refined small-object detection in UAV imagery. Enhanced bounding box accuracy was achieved through Soft-NMS and K-means++ algorithms, aiding occlusion management and complex background scenarios. YOLOv4 introduced additional data augmentation techniques and clustering improvements (Iftikhar et al. 2023). Its Drone-specific model incorporated receptive field block (RFB) and ultra-lightweight subspace attention mechanism (ULSAM) modules for better precision (Koay et al. 2021).

YOLOv6 (Norkobil Saydirasulovich et al. 2023) and YOLOv7 (Wang et al. 2023a) achieved gains in detection speed and accuracy. YOLOv8 introduced an anchor-free architecture, simplifying detection and achieving higher precision than YOLOv5 on MS-COCO (Sirisha et al. 2023). Aerial image detection enhancements in Li et al. (2023) utilized YOLOv8-s for real-time detection, integrating bidirectional path aggregation network-feature pyramid networks (Bi-PAN-FPN) into the network neck to strengthen multiscale feature fusion. This approach addresses common aerial image issues such as small object size, variable lighting, and diverse backgrounds, while maintaining edge-device compatibility and reducing parameter costs.

In the domain of two-stage detectors, the low detection accuracy issue of vehicular objects from aerial images is addressed in Wang et al. (2020c) utilizing a modified version of Faster R-CNN. Working on similar lines, authors in Benjdira et al. (2019) showed the competitive performance between YOLOv3 and Faster R-CNN. Their results indicated YOLOv3 and Faster R-CNN are comparable regarding precision, while YOLOv3 outperformed Faster R-CNN in terms of sensitivity and processing time. Authors in Avola et al. (2021) proposed a model dubbed "Multi-Stream Faster R-CNN" which employed different kernel sizes for each captured stream to simulate a multi-scale image analysis, thus efficiently detecting objects at different heights. Author in Khezaz et al. (2022b) highlighted

the limited role of RADAR and LiDAR sensors, i.e., aiding the vision sensors to enhance perception accuracy, for object detection in vehicular networks.

## 5 Deep learning related cybersecurity threats in CAVs and UAVs

This section covers a comprehensive study of ML techniques that can be utilised to attack CAVs and UAVs. Deep learning has been successfully applied to several applications in recent years, and among them, many applications are critical as human safety depends upon their error-free operations. The deep learning-aided CAVs is one such application (Sarker et al. 2020). The crucial association of human safety with AI is a big concern in the cybersecurity domain. The immense power of deep learning-aided designs necessitates a high level of responsibility (Yuan et al. 2019). Conventional cyberattacks typically involve exploiting vulnerabilities in the CAV's software or its communication capabilities. In contrast, adversarial attacks target loopholes in perception systems, i.e., cameras, lidar, RADAR, and their software counterpart that identify vehicular entities. The pioneer investigation by authors in Papernot et al. (2016b) explored the weakness of DL constructs that make them vulnerable against carefully altered input samples, termed "adversarial examples." These precisely crafted samples can easily deceive a nicely working deep learning systems with little changes dubbed "perturbations," which are generally undetectable by humans.

Current investigations also reveal that adversarial examples can be applied to deceive autonomous systems by altering input segments in an object detection system (Xie et al. 2017). As a cyber-physical system (Rong-xiao et al. 2020), UAVs are part of the distributed flying ad-hoc network (FANET) deployed in smart cities to assist CAVs on the ground and are also vulnerable to various deep learning-supported malicious attacks by hostile nodes. An adversarial example can be formally defined as "inputs to machine learning models that have been intentionally modified in a subtle way to cause the model to make a mistake." The adversarial example can be expressed as

$$\bar{\mathrm{x}} = \mathrm{x} + \arg\min_{\boldsymbol{\eta}} \{\| \boldsymbol{\eta} \| \ f(x + \eta = t)\} \tag{1}$$

where $\bar{\mathrm{x}}$ is an adversarial sample, x is the correctly classified sample, $\boldsymbol{\eta}$ is perturbation, $f()$ is the ML classifier, and $t$ is the targeted class. The adversarial attacks target the input of a deep learning module by adding adversarial perturbations, so they can be discussed in an integrated fashion without differentiating whether UAVs or CAVs captured these images.

### 5.1 Adversarial attacks

As discussed earlier in this section, adversarial attacks can be conducted using an input crafted by a distinct method to obtain incorrect results from the model. In the literature, adversarial attacks that impact the training process of machine learning are designated as poisoning attacks, whereas the adversarial attacks that affect the inference stage of machine learning are called evasion attacks (Jiang et al. 2020). The evasion attacks occur if test samples or live inputs of a model are manipulated in order to generate an inaccurate outcome. Adversarial attacks can also be divided into three main categories, namely white-

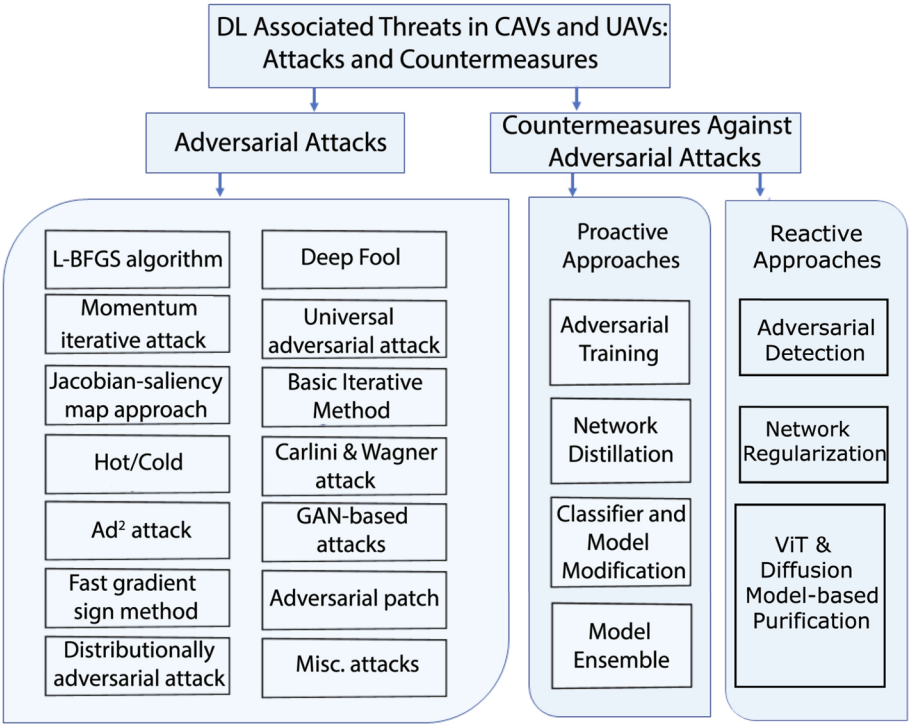


**Fig. 7** Deep Learning related cybersecurity threats and defence in CAVs and UAVs

box, grey-box, and black-box attacks, depending on the availability of information needed for execution. In order to execute white-box attacks, a deep understanding of the target is needed, including its training data, neural network structure, parameters, hyperparameters, access to gradients, and prediction results. Gray-box attacks assume a partial knowledge about the targeted model, whereas black-box attacks only necessitate the ability to query the model using arbitrary input and obtain the corresponding prediction. In the case of black-box attacks, attackers can construct a substitute model based on interactions with the target model, utilizing input–output pairs. Subsequently, they can transform this substitute model into their own white-box model, enabling them to generate adversarial examples. These adversarial examples can then be employed to launch attacks on the original black-box target model, a phenomenon referred to as the transferability of adversarial examples. Numerous other classifications of adversarial attacks depending upon their perturbations generation method and attack recurrence exist in literature (Qayyum et al. 2020); however, in order to present a holistic overview of threats and defences to the reader, we will concentrate on adversarial attack generation and their combat methods. Figure 7 represents different categories of adversarial cybersecurity threats and defense mechanisms within the context of CAVs and UAVs. These categories visually demonstrate the diversity of these threats and highlight the proposed defense strategies, offering a clear representation of ways through which the DL models can be compromised and the countermeasures in that domain. Before discussing the attack types, we explain the terminology useful for understanding adversarial attacks.

A dataset comprising $N$ samples is expressed as $\{\mathrm{x}_i y_i\}_{i=1}^{N}$, where $\mathrm{x}_i$ refers to the input data and $y_i$ represents the corresponding labels. The neural network is modelled by the function $f(.)$, which makes predictions $f(x)$ based on the input x. The adversarial loss function is written as $J(\boldsymbol{\theta}, \mathrm{x}, y)$, with $\boldsymbol{\theta}$ denoting the parameters of the model. For classification problems, the cross-entropy loss, represented by $J(f(x), y)$, is commonly used. Additionally, the adversarial version of the input x is represented by $\bar{\mathrm{x}}$ and formulated as

$$\bar{\mathrm{x}} \; : \; D(\mathrm{x}, \bar{\mathrm{x}}) \; < \boldsymbol{\eta}, \; f(\bar{x}) \neq y. \tag{2}$$

Here, $D(x)$ represents the distance metric, and $\boldsymbol{\eta}$ is the permissible perturbation, typically chosen to be minimal in order to ensure that x and $\bar{\mathrm{x}}$ remain similar.

Adversarial attacks are a topic of extensive discussion (Sadeghi et al. 2020; Hafeez et al. 2019; Sharma et al. 2019; Assion et al. 2019; Qayyum et al. 2020; Ren et al. 2020) and these can be generated and launched employing several methods. Here we will take a brief but holistic survey of the attacks and perturbations that can severely impact the output of a deep learning model.

### 5.1.1 L-BFGS algorithm

The vulnerability of deep neural networks regarding adversarial examples was first discussed in Papernot et al. (2016b). In this investigation, Papernot et al. crafted adversarial examples using the Limited Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method. The L-BFGS method finds the adversarial perturbations with the minimum $L_p$ norm, which is expressed as

$$\min_{\mathrm{x}} = \| \mathrm{x} - \bar{\mathrm{x}} \|_p \;\; \text{subject to} \, f(\bar{x}) \neq \bar{y}. \tag{3}$$

where $\bar{y}$ is the adversarial target label, ($\bar{y} \neq y$). The L-BFGS method introduced perturbations that were almost imperceptible, resulting in misclassified DNN output. The work in Papernot et al. (2016b) also explored that these adversarial examples can be generalized across various models and datasets. noticed that the generated adversarial examples could be generalized to different models and datasets. Another investigation by Papernot et al. (2016c) demonstrated the potential of binary search to get the optimal perturbation for executing an L-BFGS attack.

### 5.1.2 Fast gradient sign method

The issue of extended search time for obtaining the optimal value to launch L-BFGS attack was investigated in Goodfellow et al. (2015). Goodfellow et al. (2015) suggested "Fast Gradient Sign Method" (FGSM) to generate perturbations. FGSM was speedy, as it follows the steepest direction toward the optimal value and could execute the one-step update towards the direction of the gradient of the adversarial loss $J(\boldsymbol{\theta}, \mathrm{x}, y)$. The FGSM-generated adversarial sample can be mathematically written as

$$\bar{\mathrm{x}} \; = \mathrm{x} + \epsilon . \, \mathrm{sgn}[\nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \mathrm{x}, y)], \tag{4}$$

where $\epsilon$ denotes the size of the perturbation, FGSM can be adapted to perform an attack by moving along the gradient's direction with respect to the loss function $J(\boldsymbol{\theta}, \mathrm{x}, \bar{y})$, where $\bar{y}$ is the target label. The corresponding update rule is given by

$$\bar{\mathrm{x}} = \mathrm{x} + \epsilon.\,\mathrm{sgn}[\nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \mathrm{x}, \bar{y})]. \tag{5}$$

Another variant, suggested by Rozsa et al. (2018), dubbed "Fast Gradient Value Method (FGVM)" changed the gradient sign with the raw gradient, i.e., $\boldsymbol{\eta} = \nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \mathrm{x}, y)$. The FGVM can generate images with much higher local differences and without pixel restrictions.

### 5.1.3 Basic iterative method (BIM)

The BIM method explored in Kurakin et al. (2017), refined FGSM by applying it repeatedly with a small step size. In each iteration, pixel values are clipped to limit significant modifications, ensuring minimal changes per pixel. This iterative addition of perturbations generated adversarial examples that closely resemble the original input, leading to higher chances of misleading the network. The update rule for the *t*-th iteration is given by

$$\bar{\mathrm{x}}_{t+1} = \mathrm{Clip}\left[\bar{\mathrm{x}}_t + \alpha.\mathrm{sgn}\left\{\nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \bar{\mathrm{x}}_t, y)\right\}\right]. \tag{6}$$

This method utilized three hyper-parameters, namely, the step size $\alpha$, the maximum allowable perturbation, and the number of iterations. Another variant of BIM dubbed "Projected Gradient Descent (PGD)" offers a version free of the constraint $\alpha T = \epsilon$. PGD applies smaller adversarial perturbations using the update rule given below

$$\bar{\mathrm{x}}_{t+1} = \mathrm{proj}\{\bar{\mathrm{x}}_t + \alpha.\mathrm{sign}\left[\nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \bar{\mathrm{x}}_t, y)\right]\} \tag{7}$$

where $\mathrm{proj}$ denotes the projection operation.

### 5.1.4 Momentum iterative attack

Authors in Dong et al. (2018) proposed a momentum iterative FGSM attack based on the findings that one-step attacks are easily transferable as well as relatively simple to defend. Momentum iterative FGSM boosts FGSM with momentum to produce adversarial examples with additional iterations. Mathematically, momentum iterative FGSM updates the adversarial sample iteratively as given below

$$\bar{\mathrm{x}}_{t+1} = \mathrm{Clip}\left[\bar{\mathrm{x}}_t + \alpha.sgn\left\{\mathrm{g}_{t+1}\right\}\right] \tag{8}$$

where gradient $\mathrm{g}$ is updated according to

$$\mathrm{g}_{t+1} = \xi \mathrm{g}_t + \frac{\nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \bar{\mathrm{x}}_t, y)}{\| \nabla_{\mathrm{x}} J(\boldsymbol{\theta}, \bar{\mathrm{x}}_t, y) \|}. \tag{9}$$

Additionally, Dong et al. (2018) suggested incorporating gradients from multiple models with respect to input and identifying a gradient direction that transfers more effectively to other models.

#### 5.1.5 Distributionally adversarial attack

Authors in Zheng et al. (2020) explored the incorporation of probability space to form a novel attack type dubbed "Distributionally Adversarial Attack (DAA)". In contrast to the Projected Gradient Descent's loss function-dependent generation of adversarial samples, DAA focuses on optimizing over possible adversarial distributions. The basic idea was to incorporate the Kraft–McMillan (KL) divergence between the adversarial and benign data distributions to evaluate the loss function. The optimization function can be stated as

$$\max_{\mu} \int_{\mu} J(\boldsymbol{\theta}, \bar{\mathrm{x}}, y) d\mu + \mathrm{KL}(\mu\bar{\mathrm{x}} \parallel \pi(x)) \tag{10}$$

where $\mu$ and $\pi(x)$ represents adversarial and non-adversarial data distributions, respectively. DAA discovers new adversarial patterns and is recognized as one of the most effective attacks against multiple defence models.

#### 5.1.6 Carlini and Wagner (C&W) attack

Carlini and Wagner (2017) introduced a new set of attack algorithms and demonstrated that defensive distillation offers limited improvement in neural network's robustness. C&W proposed a series of optimization-based adversarial attacks, capable of generating adversarial examples measured by different norms, known as $\mathrm{CW}_0$, $\mathrm{CW}_2$, and $\mathrm{CW}_\infty$. The objective function for these attacks is given below

$$\min_{\boldsymbol{\delta}} \mathrm{D}(\mathrm{x}, \mathrm{x} + \boldsymbol{\delta}) + c.\, f(x + \delta), \text{ where } \mathrm{x} + \boldsymbol{\delta} \in [0, 1] \tag{11}$$

where $\boldsymbol{\delta}$ represents the perturbation, D is the distance metric, and $f(x + \delta)$ is the adversarial loss, which holds true under the condition $f(x + \delta) \leq 0$, indicating that the attack successfully causes the DNN to misclassify the target. Subsequent work by authors of Sharma et al. (2020); Alsheikh and Mahmoud (2020) showed that C&W's attack is effective against several adversarial defences.

Concerning image classification attacks of UAVs, optimization base attacks are considered relatively more time-consuming, however, can achieve the objective of targeted wrong classification (Wu et al. 2019).

#### 5.1.7 Jacobian-based saliency map approach

The investigations in Papernot et al. (2016a), resulted in an effective target attack dubbed "Jacobian-based saliency map approach" (JSMA) that can fool neural networks with minor perturbations. Initially, this technique evaluates the Jacobian matrix of the logit outputs. The Jacobian matrix of the sample $x$ is given by

$$\nabla \mathrm{l}(\mathrm{x}) = \frac{\partial \mathrm{l}(\mathrm{x})}{\partial \mathrm{x}} = \left[ \frac{\partial \mathrm{l}_j(\mathrm{x})}{\partial \mathrm{x}_\gamma} \right]_{\gamma \in 1\ldots, M_{in}, j \in 1\ldots, M_{out}} \tag{12}$$

where $M_{in}$ and $M_{out}$ represent the number of neurons in the input and output layers, respectively, and $\gamma$ and $j$ are the indices of the input x and output components. The Jacobian matrix addresses the question of how elements in the input x influence the logit outputs that are ready for classification. Specifically, the adversarial saliency map, derived from the Jacobian matrix, identifies the pixels that can be perturbed to achieve a desired change in the logit outputs. By altering a small subset of input elements, the network can be easily fooled into misclassifying the data. Recently, authors in Tian et al. (2022a) exposed the vulnerability of WiFi-supported UAVs against such attacks. The proposed approach dubbed "Forward Derivative-Based Attack" is claimed as an efficient non-targeted attack regarding image classification tasks.

### 5.1.8 DeepFool

In Moosavi-Dezfooli et al. (2016), the Deep Fool algorithm was proposed in order to find the smallest distance from an original input to the decision boundary of an adversarial example. This method involves the affine binary classifier and a general binary differentiable classifier. Initially, the authors showed that in the case of an affine classifier, the minimal perturbation is the same as the distance to the separating affine hyperplane

$$F = \left\{ \mathrm{x} : \mathrm{w}^T \mathrm{x} + b = 0 \right\}. \tag{13}$$

The perturbation for an affine classifier $f$ can be denoted as $-\frac{f(\mathrm{x})}{\|\mathrm{w}\|^2}\mathrm{w}$. For a general differentiable classifier, DeepFool assumes $F$ as linear around $\bar{x}t$ and iteratively computes the perturbation $\boldsymbol{\delta} t$ as

$$\underset{\boldsymbol{\delta}_t}{\operatorname{argmin}} \parallel \boldsymbol{\delta} \parallel_2 \text{ subject to } f(\bar{x}_t) + \nabla f(\bar{x}_t)^T \boldsymbol{\delta}_t = 0. \tag{14}$$

This result can be extended to multi-class classifiers by hunting the nearest hyperplanes and identifying more general $l_p$ norms. Studies on the DeepFool algorithm have shown that the perturbations generated by DeepFool are relatively small compared to those produced by FGSM and JSMA on several datasets.

### 5.1.9 GAN-based attacks

The investigation in Xiao et al. (2018) first discussed the generation of adversarial samples with the GAN. GAN will be elaborated here for clarity before describing the loss model and other details. On the basis of a huge dataset, GAN can form a totally new dataset that closely resembles the original dataset. The two essential parts of typical generative adversarial networks are 1) a Generator and 2) a Discriminator. The functionality of the generator is to make new instances of an object, while the Discriminator's task is to determine whether these instances are part of the original dataset. The Generator gets feedback from the Discriminator and employs it to compose more "real" images. Let the neural networks

formed by GAN be represented by a Generator network $G$ and a Discriminator network as $D$. Furthermore, let the real data distribution be $P_{data}$, the noise vector input to the generator be z that is formed utilizing the distribution $P_z$, whereas the generated samples are referred to as $G(\mathrm{z})$. The Discriminator acts as a binary classifier, taking real and generated samples as input and estimating the probability of a sample being real. Training a GAN involves solving the optimization problem formulated in Goodfellow et al. (2020).

$$\min_G \max_D V(D,G) = \underset{\mathrm{x}\sim P_{data}}{E}[log(D(\mathrm{x}))] + \underset{\mathrm{z}\sim P_z}{E}[log(1-D(G(\mathrm{z})))] \tag{15}$$

where, $V(D, G)$ represents objective function, $D(\mathrm{x})$ is the probability that $D$ discriminates x as real data, $G(\mathrm{z})$ denotes sample generated by the generator, and $D(G(\mathrm{z}))$ represents the probability that $D$ identifies as the sample formed by generator $G(\mathrm{z})$. A number of GAN variants emerged after the pioneering work, such as the Conditional Generative Adversarial Net (Mirza and Osindero 2014), Auxiliary Classifier GAN (Odena et al. 2017), while with further enhancement in training performance introduced by Arjovsky et al. (2017) and (Gulrajani et al. 2017).

#### 5.1.10 Hot/Cold

In Rozsa et al. (2016), authors investigated Hot/Cold method to discover several adversarial examples for every single image. Their idea was to permit small translations and rotations if they are imperceptible. The judge the identifiable similarity to humans, a new metric, "Psychometric Perceptual Adversarial Similarity Score" was defined. The proposed Hot/Cold method was designed to ignore the unnoticeable difference based on pixels and to use PASS instead of commonly used $l_p$ distance. A two-step procedure adopted by PASS was to a) align the modified image with the original image; 2) measure the similarity between the aligned image and the original one. Let $\phi(\bar{x}, x)$ be a homography transform from the adversarial example $\bar{x}$ to the original example $x$. $H$ is the tomography matrix, with size $3 \times 3$, $H$ is solved by maximizing the enhanced correlation coefficient between $\bar{\mathrm{x}}$ and x. The optimization function can be written as

$$\arg_H \min \left\| \frac{\bar{\mathrm{x}}}{\|\bar{\mathrm{x}}\|} - \frac{\overline{\phi(\bar{\mathrm{x}},\mathrm{x})}}{\|\overline{\phi(\bar{\mathrm{x}},\mathrm{x})}\|} \right\| \tag{16}$$

where $\overline{[.]}$ represents image normalization.

#### 5.1.11 Universal adversarial attack

The adversarial attacks are usually designed to target specific benign samples. Due to this case, adversarial perturbations generally do not transfer across benign samples. Investigators in this domain are enthusiastic about discovering a universal perturbation that can deceive the network across a wide range of benign samples. The study in Moosavi-Dezfooli et al. (2017) represents an effort to identify the minimum additional perturbation required to compromise samples. In a further step, the minimum additional perturbation is then aug-

mented to the current perturbation. Over time, this iterative process identifies a perturbation that can fool the network on the majority of benign samples.

Several universal adversarial perturbation (UAP) techniques have been proposed, including the Vanilla Universal Attack (Moosavi-Dezfooli et al. 2017), SV-UAP (Khrulkov and Oseledets 2018), F-UAP (Zhang et al. 2020), and the Network for Adversary Generation (NAG) (Mopuri et al. 2018). These approaches highlight the potential of universal attacks to broadly compromise network robustness.

#### 5.1.12 Adversarial patch

Adversarial patches are perturbations in a specific region of the benign samples. Precisely crafted adversarial patches can easily deceive a deep-learning model. In this domain, Sharif et al. (2016) revealed that cutting-edge face recognition systems can be deceived by forming some accessories, e.g., eyeglass frames. Work in Parkhi et al. (2015) extended the investigation in this context by demonstrating the venerability of commonly employed adversarial loss, e.g., cross-entropy in the case where a locally generated perturbation is employed to trick the VGG-Face convolutional neural network. In Brown et al. (2017), it was revealed that a DNN could be fooled by totally replacing a portion of an image with their carefully crafted patch. In Liu et al. (2021), investigators offered a black-box adversarial patch dubbed "D-PATCH" capable of simultaneously targeting both the object classification and bounding box regression of models. Authors in Athalye et al. (2018b), showed a general-purpose algorithm "expectation over transformation (EOT)", can create robust adversarial examples, and effectively fabricate three-dimensional adversarial objects. In Liu et al. (2018b), authors proposed using trojan patches attached to benign samples to generate adversarial examples. Regarding UAVs remote sensing of images, an adversarial patch attack for multi-scale objects along with a novel optimization technique was proposed by Zhang et al. (2021).

#### 5.1.13 $Ad^2$ Attack

This work in Fu et al. (2022) proposed an attack against UAV object tracking dubbed "$Ad^2$ Attack". The attack theme utilizes the image resampling technique instead of crafting adversarial using perturbations. The proposed scheme adaptively attains a complex adversarial mapping from low-resolution image to higher resolution image by first directly downsampled in order to lose pixel features and, subsequently, resampling of a lower-resolution image utilizing super-resolution upsampling network to generate adversarial examples and mislead UAV tracking capability. According to Fu et al. (2022), the proposed method can successfully deceive advanced siamese trackers, and the approach can assist in exposing the drawbacks of UAV trackers.

#### 5.1.14 Miscellaneous attacks

This subsection briefly overviews several commonly referenced attacks to save space. Researchers have investigated several variations of the attack-generating methods such as Obfuscated-gradient circumvention attacks (Athalye et al. 2018a), Elastic-net attack (Chen et al. 2018), CPPN EA Fool (Nguyen et al. 2015), and Model-based Ensembling Attack (Liu et al. 2016b). Additionally, various concerns have been raised regarding the practical

applications of these attacks, such as the potential for adversarial perturbations to be neutralized by environmental noise and natural transformations, as well as challenges in applying perturbations to the background of images.

### 5.2 Countermeasures against adversarial attacks in UAV assisted CAVs

The research concerning practical applications of adversarial attacks on CAVs and their countermeasures is huge. Various machine learning approaches such as long short-term memory solutions, bi-LSTM, convolution neural networks, recurrent neural networks, and deep reinforcement learning techniques are also proposed in the literature to protect UAVs (Challita et al. 2019). Other futuristic technologies like Blockchain are investigated in Aloqaily et al. (2022). Data sharing between UAVs are subject to adversarial attacks, in which case an attacker can mingle with the swarm and personify a UAV, thus modifying shared data. Federated learning schemes (Wang et al. 2021; Nie et al. 2021; Do et al. 2021; Song et al. 2021; Pham et al. 2021; Ng et al. 2021; Brik et al. 2020; Shiri et al. 2020; Ng et al. 2020; Zhang and Hanzo 2020; Lim et al. 2021b, c) are proposed as countermeasures.

Recent investigations also explored attacks against the federated learning model. For example, Almutairi and Barnawi (2024) benchmarked Byzantine-robust aggregation methods against model poisoning attacks in federated learning-enabled CAVs. The investigators evaluated performance under various data distributions and adversarial scenarios. Their results challenge existing assumptions about data security and highlight the effectiveness of client-selection strategies. Federated learning is also vulnerable to Advanced Persistent Threats (APTs), which are stealthy, prolonged cyberattacks designed to infiltrate systems and exfiltrate sensitive data. To counter such threats, GK et al. (2025) investigated a Federated Deep Neural Network (FDNN) with a privacy-preserving technique to detect APTs in IoT-enabled vehicular networks. The framework is evaluated on three benchmark datasets, achieving high detection accuracy while maintaining data privacy. Furthermore, the interpretability of the model is improved using Shapley Additive Explanations (SHAP), which quantifies the contribution of each input feature to the prediction of the model, thus identifying the most influential indicators of APT activity. In the investigation conducted by Cui et al. (2023a), two novel optimization-based data poisoning attacks are explored, namely, “black-box” and “clean-label” targeting federated learning in CAVs. Their investigated attacks use hybrid methods combining particle swarm optimization with simulated annealing and genetic algorithms. Experiments conducted by them on traffic sign recognition demonstrate significant model degradation from minimal poisoned data, revealing critical FL vulnerabilities.

The existing defence strategies against adversarial cyber-threats are primarily founded on two key methods (Moosavi-Dezfooli et al. 2018; Wang et al. 2020b). These are the proactive approach, where the cyber-physical system is prepped for potential threats and attacks before they occur, and the reactive approach, where defensive measures are implemented after an attack has taken place. The majority of defence techniques rely on the proactive approach to minimize potential damage. A detailed description of these techniques and their associated research efforts is given below.

- Proactive defences rely on the enhanced robustness of the model during the training phase, making the model inherently more resistant to adversarial perturbations. These

techniques can be integrated into the model development process utilizing the following methods (Bai et al. 2017; Tramèr et al. 2017; Xie et al. 2019; Carlini et al. 2018; Goodfellow et al. 2015; Kurakin et al. 2016; Kannan et al. 2018; Zheng et al. 2016; Engstrom et al. 2018).

1. Adversarial training is one of the most widely used proactive defence strategies. This method involves incorporating adversarial examples directly into the training dataset, enabling the model to learn from these challenging inputs. By doing so, the model becomes more adept at identifying and mitigating adversarial attacks during inference. Various techniques within adversarial training include:

   (a) PGD Adversarial Training
   (b) Ensemble Adversarial Training
   (c) Adversarial Logic Pairing
   (d) Generative Adversarial Training

These methods collectively contribute to the model's ability to generalize better and withstand adversarial manipulations.

2. Network distillation method is primarily known for DNN size reduction by transferring knowledge from a large, complex model (teacher network) to a smaller, simpler model (student network). Apart from this role, this technique can be tailored to defend against adversarial attacks. Smoother decision boundaries of the distilled model make it less susceptible to adversarial perturbations. This approach employs the knowledge transfer process to enhance the model's ability to resist adversarial examples (Papernot et al. 2016d; Hinton et al. 2015; Soll et al. 2019; Papernot and McDaniel 2017).
3. Classifier and model modification defence strategies involve modifying the classifier or model architecture to make it more robust against adversarial attacks. The methods under this category investigated by researchers in Bradshaw et al. (2017); Abbasi and Gagné (2017); Alabdulmohsin et al. (2014); Biggio et al. (2010, 2015); Papernot and McDaniel (2018); Srisakaokul et al. (2018); Lecuyer et al. (2019); Raghunathan et al. (2018); Wong and Kolter (2018), and are listed below

   (a) Creating classifiers specifically designed to be resilient to adversarial inputs.
   (b) At inference time, randomly selecting a classifier from a pool of classifiers to prevent adversaries from predicting the model's behaviour.
   (c) Aggregating outputs from multiple classifiers to improve robustness.
   (d) Integrating k-Nearest Neighbors with DNNs to leverage the strengths of both methods.
   (e) Constructing a family of classifiers from the target classifier, with random selection at test time to increase unpredictability.
   (f) Altering the architecture to create provably robust models against certain types of adversarial attacks.

4. Model ensemble techniques combine multiple models' predictions to arrive at a final decision. This approach is particularly effective in enhancing robustness. Even if one model is vulnerable to an adversarial attack, the other models in the ensemble can provide a corrective influence, reducing the overall risk of a successful attack. By aggregating the strengths of multiple models, the ensemble approach offers a more resilient defense against adversarial manipulations. Kurakin et al. (2018); Liu et al. (2018a); Pang et al. (2019).
5. Network regularization techniques aim to improve model robustness by introducing regularization terms into the training objective function. These regularizers are designed to penalize large perturbations in the input space, thereby discouraging the model from making drastic changes in its predictions due to small input variations. Perturbation-based regularization has been shown to significantly enhance the robustness of models against adversarial attacks (Yan et al. 2018; Gu and Rigazio 2014; Cisse et al. 2017).

- Reactive defences, in contrast to proactive defences, are deployed during the model's inference phase. These techniques focus on detecting and mitigating adversarial attacks after they have been attempted.

1. Adversarial detection involves using specialized detectors to identify adversarial examples before they can impact the model's decision-making process. Zheng and Hong (2018); Gu et al. (2019) Specialized detectors analyze input features, check for inconsistencies, and verify feature representations within the model. This approach also helps trace and identify compromised images. The effectiveness of adversarial detection lies in accurately distinguishing between benign and adversarial inputs, making it a critical part of a robust defense strategy (Zheng and Hong 2018; Gu et al. 2019; Gao et al. 2019; Chen et al. 2021; Wang et al. 2019).
2. Adversarial transformation techniques are designed to reverse the effects of adversarial perturbations by converting the adversarial examples back into their original clean versions (Guo et al. 2017). These methods usually involve preprocessing steps that filter or modify the input before it is fed into the model. By removing the adversarial noise, these transformations help to restore the input to a state that the model can correctly interpret, reducing the risk of incorrect predictions caused by adversarial attacks (Guo et al. 2017; Samangouei et al. 2018; Jin et al. 2019; Liao et al. 2018).

### 5.2.1 ViT and diffusion model-based adversarial purification

In line with our previous discussion, adversarial perturbations can compromise image classification and object detection systems, jeopardizing the safety and reliability of vehicular networks. Although Vision Transformers have shown considerable promise in image recognition tasks due to their attention mechanisms and ability to model spatial relationships, they remain susceptible to adversarial examples (Sun et al. 2024). Furthermore, current adversarial defence solutions are designed primarily for traditional CNN-based constructs and display limited effectiveness when applied to ViT-based models (Wu et al. 2024). Current investigations in this domain include the research effort by Sun et al. (2024) in which a

novel detection method dubbed "ViTGuard" was introduced. This method utilized Masked Auto-encoders and Vision Transformer features to defend against adversarial attacks, including patch-based threats, without requiring adversarial training. The approach outperforms seven existing methods across multiple datasets and demonstrates robustness against adaptive attacks. The investigation by Song et al. (2024) enhanced the under-display Camera's image restoration by introducing a defence framework that combines diffusion-based adversarial purification with fine-tuning to neutralize adversarial attacks while maintaining image quality. Wu et al. (2024) introduced "CeTaD", a novel Rapid Plug-in Defender that fine-tunes normalization layers of pre-trained transformer models to efficiently counter adversarial perturbations without altering the target model or clean data. CeTaD demonstrates adaptability to various attacks and scenarios, showcasing effectiveness, transferability, and potential for continuous learning.

The diffusion models are known for their robust generative capabilities. There are limited research efforts in this domain; however, investigations have demonstrated effectiveness in mitigating noise and perturbations in image restoration tasks (Nie et al. 2022). This capability aligns well with the requirements for adversarial purification, where the goal is to neutralize adversarial perturbations while preserving image quality. In Nie et al. (2022), proposed "DiffPure", a novel adversarial purification method utilizing diffusion models to remove adversarial perturbations and recover clean images through a reverse generative process. Their approach achieves state-of-the-art performance, outperforming existing adversarial training and purification methods across multiple datasets and architectures.

As diffusion models operate by iteratively refining an image through a denoising process, their iterative nature can provide the following advantages.

- Adversarial perturbations can be considered as high-frequency noise embedded in the image. Diffusion models, by their design, iteratively reverse noise processes, offering a natural mechanism for purifying adversarial perturbations.
- Unlike conventional denoising methods, diffusion models preserve the semantic integrity of images while removing adversarial noise, making them ideal for applications requiring high-accuracy image recognition.
- The iterative nature of diffusion models enables them to adapt to a wide range of adversarial attacks, including both global and localized (patch-based) perturbations.

While diffusion models have been explored in image restoration, their integration with ViTs for adversarial purification is uncharted territory. Building on insights from previous work on vision transformers (ViTs) and diffusion models, a hybrid ViT diffusion-based purification method can be effectively integrated into UAV-supported vehicular networks for robust image defence. The novel framework proposed here is composed of a dual-stage process:

- Stage 1: Diffusion-based purification to neutralize adversarial perturbations at the pixel level.
- Stage 2: Attention-aware refinement using ViT attention maps to ensure no semantic distortions remain after purification.

The method can be specifically tailored for UAV-supported vehicular networks, where adversarial perturbations can significantly impact navigation and situational awareness.

The lightweight and modular nature of diffusion models makes them suitable for real-time deployment in resource-constrained UAV systems. Furthermore, Diffusion models, combined with adversarial training or fine-tuning, can be iteratively updated to counter adaptive attacks. This dynamic capability aligns with the evolving threat landscape in adversarial machine learning. By employing diffusion models for adversarial purification, the proposed method not only enhances the robustness of UAV-supported vehicular networks but also establishes a novel integration of generative modelling with ViTs for adversarial defence.

### 5.3 Critical analysis of adversarial attacks

The rise of adversarial attacks on CAVs is becoming a big issue in the cybersecurity domain. It is crucial to thoroughly examine these attacks in order to safeguard CAVs and to properly train deep learning modules to recognize and counteract such threats. This analysis is vital for preventing potential damage to CAV systems. Qayyum et al. (2020). The continuous evolution of CAVs leads to the emergence of new vulnerabilities, making it challenging to ensure foolproof security (Girdhar et al. 2023). Researchers are continuously striving to develop comprehensive frameworks to counter potential attacks. Here, we critically analyze adversarial threats and state-of-the-art defense mechanisms.

Firstly, we consider the attacks based on the perturbation generation method that directly generates adversarial examples by adding the sign of the loss gradient with respect to each pixel in original images, such as FGSM (Goodfellow et al. 2015), BIM (Kurakin et al. 2017), MI-FGSM (Dong et al. 2018). The success of these attacks depends upon the adversarial knowledge (white or gray box) and the loopholes present in adversarial defense strategies. The work in Deng et al. (2020) concluded that these attacks are moderately potent and thus require a compound defense.

Concerning the second class of attacks, the adversarial example can be formulated as an optimization problem such as DAA (Zheng et al. 2020), C &W (Carlini and Wagner 2017). As discussed earlier, C &W attack attacks have not only evaded the DNN classifiers but also evaded the defensive distillation successfully. The C &W (Carlini and Wagner 2017) and DeepFool (Moosavi-Dezfooli et al. 2016) attacks rely on the attributes of classification models and hence are more deadly in crafting a targeted attack against a particular model. Lastly, the methods harnessing the power of generative models such as Generative Adversarial Network (Xiao et al. 2018) to create adversarial examples also pose a serious threat. These generative adversarial network variants are formidable due to their capacity to generate subtle, realistic perturbations that can deceive machine learning models in CAVs. These methods often require more computational resources and time to execute but can yield more convincing and resilient adversarial examples.

It's important to consider the level of sophistication of adversarial attacks. Studies have shown that even minor changes to input data, such as adding imperceptible alterations to images or modifying road signs, can trick the perception systems of CAVs. Defending against these attacks is challenging and often involves developing robust perception algorithms to detect and minimize the impact of such deceptive examples. Techniques like adversarial training, anomaly detection, and sensor fusion are crucial for enhancing the resilience of these systems. Although adversarial attacks on autonomous vehicles have been demonstrated in research environments, there have been no significant real-world incidents reported to date. However, the research community is aware of the importance of actively

addressing this vulnerability and is exploring all possible measures to counter adversarial alterations (Tian et al. 2022b; Qayyum et al. 2020; Girdhar et al. 2023; Sharma et al. 2019). In this study, we have identified key methodologies of crafting adversarial perturbations, recognizing their equally potent nature concerning UAV-assisted CAV networks. Here, we summarize the lesson learned from this comprehensive state-of-the-art review.

1. No single "one-size-fits-all" defense technique can completely eliminate adversarial attacks in connected vehicles. Adversarial attacks are a complex and evolving threat, and effective defense often requires a combination of defense techniques.
2. White-box attacks are significantly more impactful than black-box attacks; this fact highlights the significance of safeguarding model particulars (such as model architecture and hyperparameters) through model obfuscation.
3. If computational resources allow, opting for driving models possessing intricate architectures is preferable, as they exhibit greater resilience against adversarial attacks compared to simpler models.
4. In contrast to the method of random testing, employing worst-case analysis has emerged as a potent technique to differentiate between a system that might fail once in a billion trials and a system that boasts flawless reliability. When an adversary aiming to induce deliberate malfunctions within a system cannot succeed, it bolsters the assurance that the system will uphold its proper functioning even in the face of unanticipated variables.
5. To foster the progress of robust machine learning techniques, comprehending the reasons behind the failures of ML algorithms within specific contexts holds paramount importance.
6. The assessment of adversarial robustness should encompass both targeted and untargeted attacks. In any scenario, it is crucial to clearly indicate the types of attacks taken into account during the evaluation process. While theoretically, an untargeted attack is deemed inherently less challenging than a targeted attack, in practice, executing an untargeted attack could yield more favorable outcomes than attempting to target multiple classes.
7. Conducting ablation analysis involves systematically eliminating a set of defense components and confirming whether the attack prevails against a comparable yet unprotected model. This practice proves highly valuable, as it aids in a straightforward understanding of the goals and gauging the efficacy of combining multiple defense strategies.
8. The assessment should be carried out across a spectrum of scenarios, encompassing testing against random noise, validating against more comprehensive threat models, and carefully evaluating the attack hyperparameters to determine the optimal settings yielding maximum robustness.
9. To ensure the effectiveness of a defense strategy, it is crucial to evaluate the proposed method in broader contexts. Important vectors regarding this are given below.

    - Evaluating the defense across multiple databases.
    - Create adversarial examples by ensembling over the randomness.
    - Check the transferability of the defense to other models
    - Establish robustness bounds by testing the model against all types of attacks.

- Implementing the input processing mechanisms offered by researchers to filter out adversarial perturbations.
- Regularly updating the defense model to counter new attacks.

It is vital for researchers, industry experts, and regulatory bodies to collaborate in order to establish best practices, standards, and regulations that can effectively reduce the risks associated with adversarial attacks.

## 6 Challenges and trends

Various deep-learning designs applied to CAV sensors are already discussed in previous sections. The massive amount of literature indicates significant interest in the research of such systems. However, these systems are still far from ready for widespread commercial deployment due to certain challenges. This section will discuss how the research community works to solve these issues.

### 6.1 Challenges in DL-assisted CAVs and UAVs

1. CAV's modular design dependency:
   The modular AI-aided CAV system consists of a series of AI black boxes that aid in a certain problem, and the solution of one problem is the input to another one, thus forming a multilevel decision-making system (Furda and Vlacic 2011). Researchers have noted significantly good performance in certain parts. However, the dependency of individual parts on the overall performance of a CAV system calls for joint optimization, which is very challenging.
2. Adaptability in CAVs:
   In Muhammad et al. (2021); Gupta et al. (2018), researchers suggest that the adaptability of a designed AI system is a big challenge. The mainstream AI techniques for CAV trained on data collected in a certain environment (weather conditions, surrounding objects, and vehicles in urban and rural environments) are found unreliable in cross-environments.
3. Gigantic data in CAVs:
   The massive variations in vehicle type, road structure, and objects worldwide demand a considerable amount of data that should be collected worldwide for high accuracy in vehicle, and object detection (Mahmood et al. 2018; Kumari et al. 2017). Currently, no such data is available, which poses a key hurdle in generic AI-aided system design. Adding to this complexity, the data taken from different environments gets multiplied by the number of available sensors. Researchers have shown that it is infeasible for an AI system to process all captured data as it has immense redundancies, and thus data prioritization mechanisms are needed (Muhammad et al. 2019; Hussain et al. 2020).
4. Adversarial resilience ML:
   Although, the high caliber of DL techniques in scene perception and object identification is an edge, however, these algorithms are also vulnerable to well-crafted adversarial attacks as discussed in Section V. These carefully crafted adversarial perturbations

can cause havoc in UAV-assisted CAV systems by attacking either CAV or UAV sensors. The imminent threat of adversarial perturbations demands novel deep-learning approaches that are robust against these attacks. Till now, the defense strategies concerning ML attacks are focused on implementing novel attacks and better training of ML models against those attacks, whereas comparatively limited attention was devoted to defensive frameworks as well as more robust ML models. In a study by Gürel et al. (2021), the investigators explored a comprehensive defence strategy to reduce the susceptibility of ML/DL models to adversarial attacks. Additionally, the distributed storage of CAV sensor data pose security vulnerabilities that need to be addressed. Authors of Gürel et al. (2021) emphasised that robust security mechanisms for training and testing data should be ensured, as connected vehicles' control and decision-making processes rely on accurate and error-free datasets. The standardisation of defence techniques for safeguarding UAV-assisted CAV systems is imperative to guarantee the safety of passengers and pedestrians.

5. Safety concerns in DL-assisted CAVs:
   The automotive safety standards have not fully evolved to address the challenges of deep learning safety, such as verification and performance limitations. The issue with deep learning methods is their optimization for average cost function, and they do not guarantee safety for all cases. There is a need to develop strategies to keep the vehicle on the road safely at the time of partial or full-scale vehicle malfunction. Moreover, safety margins are needed to be clearly defined, i.e., the difference between the model's performance on the training set and operational performance in the real world (Mohseni et al. 2019). The performance of a deep learning module should be investigated in rare and unseen situations dubbed "corner cases" in literature.
6. CAV's conceptual model for Accountability:
   One of the challenges in dealing with DL-aided systems is that while using such neural networks, it is tough for humans to understand the rules learned by simply examining their weights. Researchers are working day and night to investigate different ways to visualize and understand the logic and decision provided by AI models in scenarios where such decisions ultimately impact humans' safety (Arrieta et al. 2020). Authors in Adadi and Berrada (2018) proposed the eXplainable AI (XAI) technique, which reveals the internals and knowledge learned by DL models and assists in tracking and post-mortem analysis of wrong decisions, thus providing accountability and a means for model refinement. More efforts in this domain can be found in Samek et al. (2017); Montavon et al. (2018).
7. CAV's human–machine control divide:
   Currently, there are no guidelines or hard and fast rules regarding
   a) How will the human driver respond to various system alerts, and how much should he/she trust DL-assisted system warnings compared to his visual analysis and experience?
   b) In what scenarios a human driver should prefer having control in his hand instead of a DL-assisted system?
   c) What type of alerts can indicate the driver to assume control of the vehicle, and how much control can be shared in specific scenarios?
   All these questions are challenging and are currently being investigated by the research community.
8. Crash rescue system for CAVs:

Considering the possibility of AI-assisted CAV being involved in traffic accidents due to unanticipated vehicle malfunction, a cash rescue system is necessary, especially in sparsely populated areas or in the case of a sole driver in a CAV who is unable to call for help. Researchers are investigating deep learning-aided systems that can detect such scenarios and send a distress message to the proper authorities promptly (Chang et al. 2019b; Rahim and Hassan 2021; Wang et al. 2020a).

9. UAV's model uncertainty:
   Several challenges arise concerning the employment of DL techniques in UAVs, starting with their intellectual understanding. The troubleshooting and update of a system according to needs and changing environment constitutes a significant part of system design and analysis (Khan and Al-Mulla 2019). Regarding this essential demand, the lack of knowledge about the relation between the neural network optimized weights, and system dynamics and unawareness of the reasons behind specific architectures outperforming others pose big challenges (Osco et al. 2021).
10. UAV's data dimensions and labels:
    Nowadays, collecting unlabeled data is feasible and technologically easy compared to labeled data. Success in acquiring such databases leads to the massive use of unsupervised learning algorithms. Unsupervised Learning mimics human behavior to learn the systems by simply observing them. In addition, the practical scenarios commonly involve high-dimensional state spaces (possible actions) that severely diminish the tractability with modern techniques (Zeggada et al. 2017).
    The acquisition of UAV data using comprehensive measurements in diverse areas, such as rural, urban, and areas having high mobility or loaded with sky-skippers, to test the accuracy of the DL algorithms is minuscule. These DL algorithms' performance is highly volatile in scenarios with actively changing environments, which complicates the realization of the UAV-assisted CAV system.
11. UAV's DL resources:
    The feature extraction constitutes the core application of a DL-aided UAV system due to its gifted capability to learn and interpret raw sensor data. In contrast to feature extraction, the DL-aided UAV supervision/planning system that translates the learned features to implement different functions is far more complex. Despite having an edge of being straightforward, the feature extraction systems still require high computational resources (Carrio et al. 2017). The UAV's limited resources make it challenging to integrate all resources needed for an autonomous online UAV that acts according to changing situations and environments.
    Even with advances in energy-efficient hardware, the high-end communication and computational resources requirements pose the significant challenge of developing energy-efficient low computation demanding deep learning architectures to researchers (Carrio et al. 2017).
12. CAV-UAV testing platforms:
    Like CARLA for testing the performance of CAV's DL designs, a comprehensive software testing platform is essential to validate and ensure the reliability of deep learning-supported connected vehicles that work in conjunction with UAVs. Currently, there are no such platforms that can perform a rigorous assessment of enhanced perception capabilities after integrating DL-enabled CAVs and UAVs' sensor data. Additionally, by providing a controlled environment to test and fine-tune DL algorithms, the platform

helps optimize the performance and safety of integrated CAV-UAV systems before deployment in real-world scenarios.

13. DL-aided UAVs in cross environments:
The DL-aided UAV's performance trained on sensor data in static weather is still oblivious in diverse weather conditions. The design of a deep learning UAV system that works with robustness and reliability in all cross-weather environments and counter uncertainty and data deficiency pose a serious challenge to machine learning experts (Azar et al. 2021).

### 6.2 Emerging trends and future directions

This subsection will discuss directions that academia and industry should follow for the deployment of up-and-running UAV-assisted CAV system.

1. Online learning:
Researchers are trying to tackle the problem of varying environments with online learning (also known as incremental or out-of-core learning) strategy that updates the model with new data. Investigators have recently applied online learning strategies in many domains, such as surveillance, where the deep model iteratively fine-tunes itself. Stochastic Online Learning (Cui et al. 2019) and Deep Online Learning via meta-learning (Nagabandi et al. 2018) are new trends in this domain.
2. Edge computing:
The traditional deep learning training process is performed on devices with high computational capabilities, and then the trained models are applied on the edge devices. This scheme is not efficient concerning its future deployment using the Deep Online Learning construct, where there is a need for updating the knowledge captured by the model. Edge Computing can contribute to this scenario as proposed by Liu et al. (2019a).
3. Federated learning:
Federated learning framework has been recently proposed as an effective tool to reduce the transmission overhead while achieving privacy by transmitting only model updates of the learnable parameters rather than the complete dataset. Several researchers, such as Zeng et al. (2022); Pokhrel and Choi (2020); Savazzi et al. (2021); Du et al. (2020b) are investigating significant challenges in its implementation from the machine learning and communication perspective.
4. Energy efficiency:
Several investigations showed that CNN have obtained unprecedented success in various object detection tasks associated with CAVs, however, their immense memory and computational requirements diminish their usefulness (Muhammad et al. 2021). Thus, energy-friendly and efficient CNN models are under investigation to improve the driving safety of CAVs.
5. Industrial standardization:
The lack of large-scale industrialization standards is a critical hurdle in developing a universally accepted CAV system. Several companies like Google and NVIDIA are investing massive resources in building powerful AI-based self-driving cars, neglecting

the integration and generalization of the CAV system. Many researchers are pointing towards this gap that could be a big issue when integration is needed in the future.

6. Benchmark dataset:
   Many researchers are focusing on the need for a universal benchmark dataset. Although the availability of several publicly accessible datasets such as MS-COCO, KITTI, VOC 07, and VOC 12 aid in evaluating different aspects of CAV systems. However, for standardization and evaluation of the overall performance of CAV systems, the need for a universal benchmark dataset is eminent.
7. Fully autonomous UAVs:
   UAV's autonomous working with least or best possible no human guidance is currently an essential research domain that will likely remain hot in the near future (Lee et al. 2021). The fully autonomous UAV research encompasses environmental perception, decision-making, navigation, control, data transfer, and UAV's emergency response (Bithas et al. 2019; Azar et al. 2021; Lee et al. 2021). Indeed, we will observe the progress concerning energy-efficient deep learning designs for these tasks in the forthcoming years.
8. Quantum neural networks:
   In general the complexity of bit-based conventional neural networks increases with the increase of hidden layers and neurons, quantum neural networks (QNN) can be employed due to their higher performance and low complexity for 6 G cell-free MIMO networks to optimize their performance (Narottama and Duong 2022). QNN has recently been proposed for optimal resource allocation in future wireless systems (Narottama and Shin 2022). Designing quantum gates, the required number of Qubits and integration of the quantum processing unit in the UAV-CAV network could be a challenging problem. Qubits are used to speed up the process in the network. It is also reliable and more accurate. The addition of a quantum module in the vehicular network needs the training to make QNN self-capable when handling large datasets with enhanced prediction accuracy in less time.
9. Emerging technologies:
   Some of the emerging technologies from which UAV-CAV network can benefit are 6 G-V2X, quantum computing-assisted V2X, satellite-assisted V2X, hybrid radio frequency-visible light communication, and intelligent reflecting surfaces-assisted V2X. To attain enhanced cybersecurity and data privacy, blockchain-assisted V2X and quantum federated learning (QFL) (Chehimi and Saad 2022; Huang et al. 2022) are currently under investigation by academia on a massive scale (Noor-A-Rahim et al. 2022).

## 7 Conclusion

In this paper, we have thoroughly investigated the emergence of next-generation CAVs assisted by UAVs in the context of artificial intelligence. We have pinpointed the challenges faced by CAVs that UAVs can address, leveraging the aerial perspective for traffic analysis and pattern recognition. We delved into deep learning constructs in connected vehicles, offering a detailed overview of modular and end-to-end DL approaches, followed by a critical assessment of their advantages and disadvantages. Notably, end-to-end models have

demonstrated promising outcomes in driving simulators like CARLA, with the potential for real-world application being a subject of keen interest. We also explored Vision Transformers and LLM-based designs. LLM-based approaches have demonstrated their utility in interpreting complex instructions, enabling more nuanced human-vehicle interactions and, enhancing system autonomy.

In addition to exploring the DL designs adopted by UAVs utilizing sensors such as cameras, RADAR, and LiDAR, we scrutinized DL architectures employed by UAVs for object detection in vehicular networks. In the realm of UAV-supported detection of autonomous vehicles, the choice between single-stage and two-stage designs remains pivotal. Single-stage approaches, such as YOLO and SSD, emphasize real-time processing crucial for dynamic road scenarios, while two-stage methods like Faster R-CNN prioritize accuracy for precise localization and vehicle recognition. The trade-off between speed, accuracy, and computational efficiency must be tailored to autonomy requirements. Furthermore, in examining DL-associated cybersecurity threats in CAVs and UAVs, we analyzed adversarial attack strategies and their corresponding countermeasures, underscoring the severity of these threats and the necessity of a holistic and resilient defense strategy.

The paper concludes with a discussion of open challenges and future research directions. Traditional DL models are insufficient for the evolving demands of CAV-UAV ecosystems. Therefore, we recommend adopting online or incremental learning methods that enable real-time adaptation to changing environments. Meta-learning strategies should be pursued for continual model refinement across diverse conditions. Federated learning emerges as a critical avenue to reduce data transmission overhead and enhance privacy by exchanging model parameters instead of raw data. Moreover, promoting the development of unified industry-wide standards is essential to ensure the interoperability and scalability of CAV systems. For UAVs, there is a pressing need to research and develop systems capable of full autonomy with minimal human oversight especially in navigation, perception, and emergency handling. There is a pressing need for a universal benchmark dataset that helps in standardized evaluation of the overall performance of CAV systems.

Finally, we emphasize that no singular or universal defense mechanism can fully mitigate adversarial attacks within connected vehicle environments. Given the dynamic and multi-faceted nature of these threats, robust security demands an adaptive, multi-layered strategy. This comprehensive survey thus provides valuable insights into cutting-edge AI practices and technological trajectories in UAV-assisted CAV networks, offering a foundation for both future academic research and industrial innovation.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References


Abbasi M, Gagné C (2017) Robustness to adversarial examples through an ensemble of specialists. arXiv preprint arXiv:1702.06856

Abir MABS, Chowdhury MZ, Jang YM (2023) Software-defined UAV networks for 6G systems: requirements, opportunities, emerging techniques, challenges, and research directions. IEEE Open J Commun Soc

Abu Tami M, Ashqar HI, Elhenawy M et al. (2024) Using multimodal large language models (MLLMs) for automated detection of traffic safety-critical events. Vehicles 6(3):1571–1590

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adame T, Bel A, Bellalta B et al. (2014) IEEE 802.11 AH: the wifi approach for M2M communications. IEEE Wireless Commun 21(6):144–152

Ahamed MFS, Tewolde G, Kwon J (2018) Software-in-the-loop modeling and simulation framework for autonomous vehicles. In: 2018 IEEE international conference on electro/information technology (EIT), IEEE, pp 0305–0310

Ahmed GA, Sheltami TR, Mahmoud AS et al. (2021) A novel collaborative IoD-assisted VANET approach for coverage area maximization. IEEE Access 9:61211–61223

Ahmad U, Han M, Jolfaei A et al. (2024) A comprehensive survey and tutorial on smart vehicles: Emerging technologies, security issues, and solutions using machine learning. IEEE transactions on intelligent transportation systems

Akagi Y, Kato R, Kitajima S et al. (2019) A risk-index based sampling method to generate scenarios for the evaluation of automated driving vehicle safety. In: 2019 IEEE intelligent transportation systems conference (ITSC), IEEE, pp 667–672

Alabdulmohsin IM, Gao X, Zhang X (2014) Adding robustness to support vector machines against adversarial reverse engineering. In: Proceedings of the 23rd ACM International conference on conference on information and knowledge management, pp 231–240

Alghmgham DA, Latif G, Alghazo J et al. (2019) Autonomous traffic sign (ATSR) detection and recognition using deep CNN. Proc Comput Sci 163:266–274. https://doi.org/10.1016/j.procs.2019.12.108. (**16th Learning and Technology Conference 2019Artificial Intelligence and Machine Learning: Embedding the Intelligence**)

Al-Hilo A, Samir M, Assi C et al. (2020) Cooperative content delivery in UAV-RSU assisted vehicular networks. In: Proceedings of the 2nd ACM MobiCom workshop on drone assisted wireless communications for 5G and beyond, pp 73–78

Almutairi S, Barnawi A (2024) A comprehensive analysis of model poisoning attacks in federated learning for autonomous vehicles: a benchmark study. Results Eng 24:103295

Aloqaily M, Hussain R, Khalaf D et al. (2022) On the role of futuristic technologies in securing UAV-supported autonomous vehicles. IEEE Consum Electron Magazine 11(6):93–105. https://doi.org/10.1109/MCE.2022.3141065

Alsheikh M, Mahmoud A (2020) Adversarial attacks and defenses in deep learning: a survey. Mach Learn Knowledge Extract 2(2):456–482. https://doi.org/10.3390/make2020025

Amer R, Saad W, Marchetti N (2020) Mobility in the sky: performance and mobility analysis for cellular-connected UAVs. IEEE Trans Commun 68(5):3229–3246. https://doi.org/10.1109/TCOMM.2020.2973629

Amini A, Rosman G, Karaman S et al. (2019) Variational end-to-end navigation and localization. In: 2019 International conference on robotics and automation (ICRA), IEEE, pp 8958–8964

Ammour N, Alhichri H, Bazi Y et al. (2017) Deep learning approach for car detection in UAV imagery. Remote Sensing 9(4):312. https://doi.org/10.3390/rs9040312

Amponis G, Lagkas T, Zevgara M et al. (2022) Drones in b5G/6G networks as flying base stations. Drones 6(2):39

Antonio GP, Maria-Dolores C (2022) Multi-agent deep reinforcement learning to manage connected autonomous vehicles at tomorrow's intersections. IEEE Trans Veh Technol 71(7):7033–7043

Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: Proceedings of the 34th international conference on machine learning-volume 70, pp 214–223

Arrieta AB, Díaz-Rodríguez N, Del Ser J et al. (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Artan Y, Alkan B, Balci B et al. (2019) Deep learning based vehicle make, model and color recognition using license plate recognition camera images. In: 2019 27th signal processing and communications applications conference (SIU), pp 1–4, https://doi.org/10.1109/SIU.2019.8806465

Assion F, Schlicht P, Greßner F et al. (2019) The attack generator: A systematic approach towards constructing adversarial attacks. In: 2019 IEEE/CVF Conference on computer vision and pattern recognition workshops (CVPRW), pp 1370–1379, https://doi.org/10.1109/CVPRW.2019.00177

Asvadi A, Garrote L, Premebida C et al. (2017) Depthcn: Vehicle detection using 3D-LiDAR and ConvNet. In: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC), pp 1–6, https://doi.org/10.1109/ITSC.2017.8317880

Athalye A, Carlini N, Wagner D (2018a) Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In: International conference on machine learning, PMLR, pp 274–283

Athalye A, Engstrom L, Ilyas A et al. (2018b) Synthesizing robust adversarial examples. In: International conference on machine learning, PMLR, pp 284–293

Avola D, Cinque L, Diko A et al. (2021) Ms-faster R-CNN: multi-stream backbone for improved faster R-CNN object detection and aerial tracking from UAV images. Remote Sensing 13(9):1670

Azar AT, Koubaa A, Ali Mohamed N et al. (2021) Drone deep reinforcement learning: a review. Electronics 10(9):999

Baheti B, Gajre S, Talbar S (2018) Detection of distracted driver using convolutional neural network. In: 2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), pp 1145–11456, https://doi.org/10.1109/CVPRW.2018.00150

Bai W, Quan C, Luo Z (2017) Alleviating adversarial attacks via convolutional autoencoder. 2017 18th IEEE/ACIS international conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD) pp 53–58

Bansal M, Krizhevsky A, Ogale A (2018) Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. arXiv preprint arXiv:1812.03079

Barmpounakis E, Geroliminis N (2020) On the new era of urban traffic monitoring with massive drone data: the pneuma large-scale field experiment. Transport Res Part C Emerg Technol 111:50–71

Bechtel MG, McEllhiney E, Kim M et al. (2018) Deeppicar: A low-cost deep neural network-based autonomous car. In: 2018 IEEE 24th international conference on embedded and real-time computing systems and applications (RTCSA), IEEE, pp 11–21

Benjdira B, Khursheed T, Koubaa A et al. (2019) Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and yolov3. In: 2019 1st International conference on unmanned vehicle systems-oman (UVS), IEEE, pp 1–6

Benjumea A, Teeti I, Cuzzolin F et al. (2021) Yolo-z: Improving small object detection in yolov5 for autonomous vehicles. arXiv preprint arXiv:2112.11798

Bewley A, Rigley J, Liu Y et al. (2019) Learning to drive from simulation without real world labels. In: 2019 International conference on robotics and automation (ICRA), IEEE, pp 4818–4824

Biggio B, Fumera G, Roli F (2010) Multiple classifier systems for robust classifier design in adversarial environments. J Mach Learn Cybern 1:27–41. https://doi.org/10.1007/s13042-010-0007-7

Biggio B, Corona I, He ZM et al. (2015) One-and-a-half-class multiple classifier systems for secure learning against evasion attacks at test time. In: International workshop on multiple classifier systems, Springer, pp 168–180

Biswas A, Reon MO, Das P et al. (2022) State-of-the-art review on recent advancements on lateral control of autonomous vehicles. IEEE Access 10:114759–114786

Bithas PS, Michailidis ET, Nomikos N et al. (2019) A survey on machine-learning techniques for UAV-based communications. Sensors 19(23):5170

Bochkovskiy A, Wang CY, Liao H (2020) Yolov4: optimal speed and accuracy of object detection. arXiv:2004.10934

Bojarski M, Del Testa D, Dworakowski D et al. (2016) End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316

Bolte JA, Bar A, Lipinski D et al. (2019) Towards corner case detection for autonomous driving. In: 2019 IEEE Intelligent vehicles symposium (IV), IEEE, pp 438–445

Bonsignorio F, Hsu D, Johnson-Roberson M et al. (2020) Deep learning and machine learning in robotics [from the guest editors]. IEEE Robotics Automation Magazine 27(2):20–21. https://doi.org/10.1109/MRA.2020.2984470

Bouguettaya A, Zarzour H, Kechida A et al. (2021) Vehicle detection from UAV imagery with deep learning: a review. IEEE Trans Neural Netw Learning Syst 33(11):6047–6067

Bouguettaya A, Zarzour H, Kechida A et al. (2022) Vehicle detection from UAV imagery with deep learning: a review. IEEE Trans Neural Netw Learning Syst 33(11):6047–6067. https://doi.org/10.1109/TNNLS.2021.3080276

Bradshaw J, Matthews AGdG, Ghahramani Z (2017) Adversarial examples, uncertainty, and transfer testing robustness in Gaussian process hybrid deep networks. arXiv preprint arXiv:1707.02476

Brik B, Ksentini A, Bouaziz M (2020) Federated learning for UAVs-enabled wireless networks: use cases, challenges, and open problems. IEEE Access 8:53841–53849. https://doi.org/10.1109/ACCESS.2020.2981430

Brown TB, Mané D, Roy A et al. (2017) Adversarial patch. arXiv preprint arXiv:1712.09665

Butt FA, Chattha JN, Ahmad J et al. (2022) On the integration of enabling wireless technologies and sensor fusion for next-generation connected and autonomous vehicles. IEEE Access 10:14643–14668. https://doi.org/10.1109/ACCESS.2022.3145972

Cai K, Qu T, Gao B et al. (2023) Consensus-based distributed cooperative perception for connected and automated vehicles. IEEE Trans Intell Transp Syst 24(8):8188–8208. https://doi.org/10.1109/TITS.2023.3264608

Capellier E, Davoine F, Cherfaoui V et al. (2019a) Evidential deep learning for arbitrary LiDAR object classification in the context of autonomous driving. In: 2019 IEEE intelligent vehicles symposium (IV), pp 1304–1311, https://doi.org/10.1109/IVS.2019.8813846

Capellier E, Davoine F, Cherfaoui V et al. (2019b) Transformation-adversarial network for road detection in LIDAR rings, and model-free evidential road grid mapping. In: 11th Workshop on Planning, Perception, Navigation for Intelligent Vehicle (PPNIV-IROS 2019), pp 47–52

Carlini N, Wagner D (2017) Towards evaluating the robustness of neural networks. In: 2017 IEEE symposium on security and privacy (SP), IEEE, pp 39–57

Carlini N, Katz G, Barrett C et al. (2018) Provably minimally-distorted adversarial examples. In: International conference on machine learning, PMLR, pp 799–808

Carranza-García M, Torres-Mateo J, Lara-Benítez P et al. (2021) On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. Remote Sensing 13(1):89. https://doi.org/10.3390/rs13010089

Carrio A, Sampedro Pérez C, Rodríguez Ramos A et al. (2017) A review of deep learning methods and applications for unmanned aerial vehicles. J Sensors 2017:1–13. https://doi.org/10.1155/2017/3296874

Cawthorne D, Juhl PM (2022) Designing for calmness: Early investigations into drone noise pollution management. In: 2022 International conference on unmanned aircraft systems (ICUAS), IEEE, pp 839–848

Challita U, Ferdowsi A, Chen M et al. (2019) Machine learning for wireless connectivity and security of cellular-connected UAVs. IEEE Wirel Commun 26(1):28–35. https://doi.org/10.1109/MWC.2018.1800155

Chang S, Jia Z, Yu Y et al. (2019a) UAV path planning design based on deep learning. In: International conference in communications, signal processing, and systems, Springer, pp 1280–1288

Chang WJ, Chen LB, Su KY (2019b) Deepcrash: a deep learning-based internet of vehicles system for head-on and single-vehicle accident detection with emergency notification. IEEE Access 7:148163–148175. https://doi.org/10.1109/ACCESS.2019.2946468

Chehimi M, Saad W (2022) Quantum federated learning with quantum data. In: ICASSP 2022 - 2022 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 8617–8621, https://doi.org/10.1109/ICASSP43922.2022.9746622

Chen X, Ma H, Wan J et al. (2017) Multi-view 3D object detection network for autonomous driving. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1907–1915

Chen PY, Sharma Y, Zhang H et al. (2018) Ead: elastic-net attacks to deep neural networks via adversarial examples. In: Proceedings of the AAAI conference on artificial intelligence

Chen B, Carvalho W, Baracaldo N et al. (2021) Detecting backdoor attacks on deep neural networks by activation clustering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11631–11640

Chen D, Qi Q, Fu Q et al. (2024a) Transformer-based reinforcement learning for scalable multi-uav area coverage. IEEE Trans Intell Transport Syst 25(8):10062–10077

Chen L, Sinavski O, Hünermann J et al. (2024b) Driving with llms: Fusing object-level vector modality for explainable autonomous driving. In: 2024 IEEE International conference on robotics and automation (ICRA), IEEE, pp 14093–14100

Chib PS, Singh P (2023) Recent advancements in end-to-end autonomous driving using deep learning: a survey. IEEE Trans Intelligent Vehicles 9(1):103–118

Chi L, Mu Y (2017) Deep steering: Learning end-to-end driving model from spatial and temporal visual cues. arXiv preprint arXiv:1708.03798

Chun D, Choi J, Kim H et al. (2019) A study for selecting the best one-stage detector for autonomous driving. 2019 34th international technical conference on circuits/systems, computers and communications (ITC-CSCC) pp 1–3

Cisse M, Bojanowski P, Grave E et al. (2017) Parseval networks: Improving robustness to adversarial examples. In: International conference on machine learning, PMLR, pp 854–863

Codevilla F, Müller M, López A et al. (2018) End-to-end driving via conditional imitation learning. In: 2018 IEEE international conference on robotics and automation (ICRA), IEEE, pp 4693–4700

Coelho D, Oliveira M (2022) A review of end-to-end autonomous driving in urban environments. IEEE Access 10:75296–75311
Cohen J (2023) Breakdown: How Tesla will transition from modular to End-To-End deep learning. https://www.thinkautonomous.ai/blog/tesla-end-to-end-deep-learning/, accessed: 2025-6-30
Cortese A (2025) Autonomous driving: The future is getting closer (Part I). https://archivemacropolo.org/analysis/autonomous-driving-the-future-is-getting-closer-part-i/?rp=e, Accessed: 2025-6-30
Cui Q, Gong Z, Ni W et al. (2019) Stochastic online learning for mobile edge computing: learning from changes. IEEE Commun Mag 57(3):63–69. https://doi.org/10.1109/MCOM.2019.1800644
Cui C, Du H, Jia Z et al. (2023a) Data poisoning attacks with hybrid particle swarm optimization algorithms against federated learning in connected and autonomous vehicles. IEEE Access 11:136361–136369. https://doi.org/10.1109/ACCESS.2023.3337638
Cui Y, Huang S, Zhong J et al. (2023b) Drivellm: Charting the path toward full autonomous driving with large language models. IEEE Trans Intell Veh
Cui C, Ma Y, Cao X et al. (2024) A survey on multimodal large language models for autonomous driving. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 958–979
Dalwai FS, Kumar GS, Sharief ZSA et al. (2023) ViT-CD: A vision transformer approach to vehicle collision detection. In: 2023 Global conference on information technologies and communications (GCITC), IEEE, pp 1–7
De Curtò J, De Zarza I, Calafate CT (2023) Semantic scene understanding with large language models on unmanned aerial vehicles. Drones 7(2):114
Deng Y, Zheng X, Zhang T et al. (2020) An analysis of adversarial attacks and defenses on autonomous driving models. In: 2020 IEEE international conference on pervasive computing and communications (PerCom), IEEE, pp 1–10
Deshmukh P, Satyanarayana G, Majhi S et al. (2023) Swin transformer based vehicle detection in undisciplined traffic environment. Expert Syst Appl 213:118992
Dickmann J, Appenrodt N, Bloecher H et al. (2014) RADAR contribution to highly automated driving. In: 2014 44th European microwave conference, pp 1715–1718, https://doi.org/10.1109/EuMC.2014.6986787
Ding W, Chen B, Xu M et al. (2020) Learning to collide: An adaptive safety-critical scenarios generating method. In: 2020 IEEE/RSJ International conference on intelligent robots and systems (IROS), IEEE, pp 2243–2250
Do QV, Pham QV, Hwang WJ (2021) Deep reinforcement learning for energy-efficient federated learning in UAV-enabled wireless powered networks. IEEE Commun Lett pp 1–1
Dong Y, Liao F, Pang T et al. (2018) Boosting adversarial attacks with momentum. In: 2018 IEEE/CVF conference on computer vision and pattern recognition, IEEE, pp 9185–9193
Dong J, Chen S, Zong S et al. (2021) Image transformer for explainable autonomous driving system. In: 2021 IEEE International intelligent transportation systems conference (ITSC), IEEE, pp 2732–2737
Dosovitskiy A, Ros G, Codevilla F et al. (2017) CARLA: an open urban driving simulator. In: Conference on robot learning, PMLR, pp 1–16
Du L, Chen W, Pei Z et al. (2020a) Learning-based lane-change behaviour detection for intelligent and connected vehicles. Comput Intell Neurosci 2020
Du Z, Wu C, Yoshinaga T et al. (2020b) Federated learning for vehicular internet of things: recent advances and open issues. IEEE Open J Comput Soc 1:45–61. https://doi.org/10.1109/OJCS.2020.2992630
El Khatib A, Ou C, Karray F (2019) Driver inattention detection in the context of next-generation autonomous vehicles design: A survey. IEEE Transactions on Intelligent Transportation Systems pp 1–14. https://doi.org/10.1109/TITS.2019.2940874
Engelhardt N, Pérez R, Rao Q (2019) Occupancy grids generation using deep RADAR network for autonomous driving. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp 2866–2871, https://doi.org/10.1109/ITSC.2019.8916897
Engstrom L, Ilyas A, Athalye A (2018) Evaluating and understanding the robustness of adversarial logit pairing. arXiv preprint arXiv:1807.10272
Eraqi HM, Moustafa MN, Honer J (2017) End-to-end deep learning for steering autonomous vehicles considering temporal dependencies. arXiv preprint arXiv:1710.03804
Feng D, Rosenbaum L, Dietmayer K (2018) Towards safe autonomous driving: capture uncertainty in the deep neural network for LiDAR 3D vehicle detection. In: 2018 21st international conference on intelligent transportation systems (ITSC), IEEE, pp 3266–3273
Feng S, Haykin S (2019a) Anti-jamming V2V communication in an integrated UAV-CAV network with hybrid attackers. In: International Conference on Communications, pp 1–6, https://doi.org/10.1109/ICC.2019.8761101
Feng S, Haykin S (2019b) Anti-jamming V2V communication in an integrated UAV-CAV network with hybrid attackers. In: International Conference on Communications (ICC), IEEE, pp 1–6

Feng S, Haykin S (2019c) Anti-jamming V2V communication in an integrated UAV-CAV network with hybrid attackers. In: International Conference on Communications, pp 1–6, https://doi.org/10.1109/ICC.2019.8761101
Feng S, Feng Y, Sun H et al. (2020a) Testing scenario library generation for connected and automated vehicles: an adaptive framework. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2020.3023668
Feng S, Feng Y, Sun H et al. (2020b) Testing scenario library generation for connected and automated vehicles, part II: Case studies. IEEE Trans Intell Transport Syst. https://doi.org/10.1109/TITS.2020.2988309
Feng S, Feng Y, Yan X et al. (2020c) Safety assessment of highly automated driving systems in test tracks: a new framework. Accid Anal Prev 144:105664. https://doi.org/10.1016/j.aap.2020.105664
Feng S, Feng Y, Yu C et al. (2021) Testing scenario library generation for connected and automated vehicles, part I: methodology. IEEE Trans Intell Transp Syst 22(3):1573–1582. https://doi.org/10.1109/TITS.2020.2972211
Fu CY, Liu W, Ranga A et al. (2017) Dssd: deconvolutional single shot detector. arXiv preprint arXiv:1701.06659
Fu C, Li S, Yuan X et al. (2022) $\mathrm{Ad}^2$ attack: adaptive adversarial attack on real-time UAV tracking. In: 2022 International Conference on Robotics and Automation (ICRA), pp 5893–5899, https://doi.org/10.1109/ICRA46639.2022.9812056
Furda A, Vlacic L (2011) Enabling safe autonomous driving in real-world city traffic using multiple criteria decision making. Intell Transp Syst Magazine IEEE 3:4–17. https://doi.org/10.1109/MITS.2011.940472
Gangopadhyay A, Tripathi SM, Jindal I et al. (2015) SA-CNN: dynamic scene classification using convolutional neural networks. arXiv preprint arXiv:1502.05243
Gao Y, Xu C, Wang D et al. (2019) Strip: A defence against trojan attacks on deep neural networks. In: Proceedings of the 35th annual computer security applications conference, pp 113–125
Gao J, Yi J, Murphey YL (2022) M2-Conformer: Multi-modal CNN-transformer for driving behavior detection. In: 2022 5th International symposium on autonomous systems (ISAS), IEEE, pp 1–6
Giordan D, Adams M, Aicardi I et al. (2020) The use of unmanned aerial vehicles (UAVs) for engineering geology applications. Bull Eng Geol Env 79:3437–3481. https://doi.org/10.1007/s10064-020-01766-2
Girdhar M, Hong J, Moore J (2023) Cybersecurity of autonomous vehicles: a systematic literature review of adversarial attacks and defense models. IEEE Open J Vehicular Technol 4:417–437. https://doi.org/10.1109/OJVT.2023.3265363
Girshick R, Donahue J, Darrell T et al. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
Girshick R (2015) Fast R-CNN. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448
GK SK, Muniyal B, Rajarajan M et al. (2025) Explainable federated framework for enhanced security and privacy in connected vehicles against advanced persistent threats. IEEE Open J Veh Technol
Goodfellow I, Shlens J, Szegedy C (2015) Explaining and harnessing adversarial examples. In: Proceedings of the 2015 international conference on learning representations (ICLR)
Goodfellow I, Pouget-Abadie J, Mirza M et al. (2020) Generative adversarial networks. Commun ACM 63(11):139–144
Grigorescu SM, Trasnea B, Marina L et al. (2019) Neurotrajectory: a neuroevolutionary approach to local state trajectory learning for autonomous vehicles. IEEE Robotics Automation Lett 4(4):3441–3448
Grigorescu S, Trasnea B, Cocias T et al. (2020) A survey of deep learning techniques for autonomous driving. J Field Robotics 37(3):362–386
Gu T, Dolan JM, Lee J (2016) Human-like planning of swerve maneuvers for autonomous vehicles. In: 2016 IEEE Intelligent Vehicles Symposium (IV), pp 716–721, https://doi.org/10.1109/IVS.2016.7535466
Guillen-Perez A, Sanchez-Iborra R, Cano MD et al. (2016) WiFi networks on drones. In: 2016 ITU Kaleidoscope: ICTs for a Sustainable World (ITU WT), IEEE, pp 1–8
Guillen-Perez A, Cano MD (2018) Flying ad hoc networks: a new domain for network communications. Sensors 18(10):3571
Guillen-Perez A, Cano MD (2019) Counting and locating people in outdoor environments: a comparative experimental study using wifi-based passive methods. In: ITM Web of Conferences, EDP Sciences, p 01010
Guillen-Perez A, Montoya AM, Sanchez-Aarnoutse JC et al. (2021) A comparative performance evaluation of routing protocols for flying ad-hoc networks in real conditions. Appl Sci 11(10):4363
Gulrajani I, Ahmed F, Arjovsky M et al. (2017) Improved training of Wasserstein Gans. Adv Neural Inf Process Syst 30:5767–5777
Guo C, Rana M, Cisse M et al. (2017) Countering adversarial images using input transformations. arXiv preprint arXiv:1711.00117

Gupta A, Anpalagan A, Guan L et al. (2018) Deep learning for object detection and scene perception in self-driving cars. IEEE Commun Mag 56(9):121–127
Gürel NM, Qi X, Rimanic L et al. (2021) Knowledge enhanced machine learning pipeline against diverse adversarial attacks. In: International Conference on Machine Learning, PMLR, pp 3976–3987
Gurghian A, Koduri T, Bailur SV et al. (2016) Deeplanes: End-to-end lane position estimation using deep neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 38–45, https://doi.org/10.1109/CVPRW.2016.12
Gu S, Rigazio L (2014) Towards deep neural network architectures robust to adversarial examples. arXiv preprint arXiv:1412.5068
Gu S, Yi P, Zhu T et al. (2019) Detecting adversarial examples in deep neural networks using normalizing filters. In: ICAART (2): 164–173
Hafeez A, Topolovec K, Awad S (2019) ECU fingerprinting through parametric signal modeling and artificial neural networks for in-vehicle security against spoofing attacks. In: 2019 15th International Computer Engineering Conference (ICENCO), pp 29–38, https://doi.org/10.1109/ICENCO48310.2019.9027298
Hafeez F, Sheikh UU, Alkhaldi N et al. (2020) Insights and strategies for an autonomous vehicle with a sensor fusion innovation: a fictional outlook. IEEE Access 8:135162–135175
Harun MH, Abdullah SS, Aras MSM et al. (2022) Sensor fusion technology for unmanned autonomous vehicles (UAV): a review of methods and applications. In: 2022 IEEE 9th international conference on underwater system technology: theory and applications (USYS), pp 1–8, https://doi.org/10.1109/USYS56283.2022.10072667
Hawke J, Shen R, Gurau C et al. (2020) Urban driving with conditional imitation learning. In: 2020 IEEE international conference on robotics and automation (ICRA), IEEE, pp 251–257
He Y, Zhai D, Huang F et al. (2021) Joint task offloading, resource allocation, and security assurance for mobile edge computing-enabled UAV-assisted VANETs. Remote Sensing 13(8):1547
Hecker S, Dai D, Van Gool L (2018) Learning driving models with a surround-view camera system and a route planner. arXiv preprint arXiv:1803.10158
Hildmann H, Kovacs E (2019) Review: using unmanned aerial vehicles (UAVs) as mobile sensing platforms (MSPs) for disaster response, civil security and public safety. Drones 3:59. https://doi.org/10.3390/drones3030059
Hinton G, Vinyals O, Dean J et al. (2015) Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 2(7)
Hu J, Chen C, Cai L et al. (2021) UAV-assisted vehicular edge computing for the 6G internet of vehicles: architecture, intelligence, and challenges. IEEE Commun Standards Magazine 5(2):12–18
Huang R, Tan X, Xu Q (2022) Quantum federated learning with decentralized data. IEEE Journal of Selected Topics in Quantum Electronics 28(4: Mach. Learn. in Photon. Commun. and Meas. Syst.):1–10. https://doi.org/10.1109/JSTQE.2022.3170150
Huizing A, Heiligers M, Dekker B et al. (2019) Deep learning for classification of mini-UAVs using micro-doppler spectrograms in cognitive RADAR. IEEE Aerosp Electron Syst Mag 34(11):46–56. https://doi.org/10.1109/MAES.2019.2933972
Hussain T, Muhammad K, Ullah A et al. (2020) Cloud-assisted multiview video summarization using CNN and bidirectional lstm. IEEE Trans Indus Inf 16(1):77–86. https://doi.org/10.1109/TII.2019.2929228
Hussain K, Moreira C, Pereira J et al. (2025) A comprehensive literature review on modular approaches to autonomous driving: deep learning for road and racing scenarios. Smart Cities 8(3):79. https://doi.org/10.3390/smartcities8030079
Iftikhar S, Asim M, Zhang Z et al. (2023) Target detection and recognition for traffic congestion in smart cities using deep learning-enabled UAVs: A review and analysis. Appl Sci 13(6):3995
Islam Z, Abdel-Aty M, Anik BTH (2023) Transformer-conformer ensemble for crash prediction using connected vehicle trajectory data. IEEE Open J Intell Transp Syst 4:979–988
Jacob S, Menon VG, KS FS et al. (2020) Intelligent vehicle collision avoidance system using 5G-enabled drone swarms. In: Proceedings of the 2nd ACM MobiCom workshop on drone assisted wireless communications for 5G and beyond, pp 91–96
Jain A, Ramaprasad R, Narang P et al. (2021) AI-enabled object detection in UAVs: challenges, design choices, and research directions. IEEE Network 35(4):129–135. https://doi.org/10.1109/MNET.011.2000643
Jaritz M, De Charette R, Toromanoff M et al. (2018) End-to-end race driving with deep reinforcement learning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp 2070–2075
Javaid S, Fahim H, He B et al. (2024a) Large language models for UAVs: Current state and pathways to the future. IEEE Open J Veh Technol
Javaid S, Khalil RA, Saeed N et al. (2024b) Leveraging large language models for integrated satellite-aerial-terrestrial networks: recent advances and future directions. arXiv preprint arXiv:2407.04581

Jeon S, Shin JW, Lee YJ et al. (2017) Empirical study of drone sound detection in real-life environment with deep neural networks. In: 2017 25th European signal processing conference (EUSIPCO), pp 1858–1862, https://doi.org/10.23919/EUSIPCO.2017.8081531
Ji P, Zhang C, Zhang Z (2024) A method based on vision transformer and multiple image information for vehicle lane-changing recognition in mixed traffic and connected environment. Transp Lett 17:1–13
Jiang W, Li H, Liu S et al. (2020) Poisoning and evasion attacks against deep learning algorithms in autonomous vehicles. IEEE Trans Veh Technol 69(4):4439–4449. https://doi.org/10.1109/TVT.2020.2977378
Jiang F, Peng Y, Dong L et al. (2024) Large language model enhanced multi-agent systems for 6G communications. IEEE Wireless Commun
Jiao L, Zhang F, Liu F et al. (2019) A survey of deep learning-based object detection. IEEE Access 7:128837–128868. https://doi.org/10.1109/ACCESS.2019.2939201
Jin G, Shen S, Zhang D et al. (2019) APE-GAN: Adversarial perturbation elimination with GAN. ICASSP 2019–2019 IEEE International Conference on Acoustics. IEEE, Speech and Signal Processing (ICASSP), pp 3842–3846
Johansson R, Williams D, Berglund A et al. (2004) Carsim: A system to visualize written road accident reports as animated 3D scenes. In: Proceedings of the 2nd workshop on text meaning and interpretation. Association for Computational Linguistics, USA, p 57–64
Kang H, Joung J, Kim J et al. (2020) Protect your sky: a survey of counter unmanned aerial vehicle systems. IEEE Access 8:168671–168710. https://doi.org/10.1109/ACCESS.2020.3023473
Kang M, Lee W, Hwang K et al. (2022) Vision transformer for detecting critical situations and extracting functional scenario for automated vehicle safety assessment. Sustainability 14(15):9680
Kannan H, Kurakin A, Goodfellow I (2018) Adversarial logit pairing. arXiv preprint arXiv:1803.06373
Karim Amer MS, Shaker M, ElHelw M (2019) Deep convolutional neural network based autonomous drone navigation. In: Thirteenth International Conference on Machine Vision, p 1160503
Karunakaran D, Worrall S, Nebot E (2020) Efficient statistical validation with edge cases to evaluate highly automated vehicles. In: 2020 IEEE 23rd International conference on intelligent transportation systems (ITSC), IEEE, pp 1–8
Kavas-Torris O, Gelbal SY, Cantas MR et al. (2021) Connected UAV and CAV coordination for improved road network safety and mobility. Tech. rep, SAE Technical Paper
Kavas-Torris O, Gelbal SY, Cantas MR et al. (2022a) V2x communication between connected and automated vehicles (cavs) and unmanned aerial vehicles (uavs). Sensors 22(22):8941
Kavas-Torris O, Gelbal SY, Cantas MR et al. (2022b) V2X communication between connected and automated vehicles (CAVs) and unmanned aerial vehicles (UAVs). Sensors 22(22):8941. https://doi.org/10.3390/s22228941
Kendall A, Hawke J, Janz D et al. (2019) Learning to drive in a day. In: 2019 international conference on robotics and automation (ICRA), IEEE, pp 8248–8254
Khabbaz M, Antoun J, Assi C (2019) Modeling and performance analysis of UAV-assisted vehicular networks. IEEE Trans Veh Technol 68(9):8384–8396
Khan AI, Al-Mulla Y (2019) Unmanned aerial vehicle in the machine learning environment. J King Saud Univ Comput Inf Sci 31(4):486–492
Khan AA, Laghari AA, Shafiq M et al. (2022) Vehicle to everything (V2X) and edge computing: a secure lifecycle for uav-assisted vehicle network and offloading with blockchain. Drones 6(12):377. https://doi.org/10.3390/drones6120377
Khezaz A, Hina MD, Ramdane-Cherif A (2022a) Perception enhancement and improving driving context recognition of an autonomous vehicle using uavs. J Sens Actuator Netw 11(4):56
Khezaz A, Hina MD, Ramdane-Cherif A (2022b) Perception enhancement and improving driving context recognition of an autonomous vehicle using UAVs. J Sens Actuator Netw 11(4):56
Khrulkov V, Oseledets I (2018) Art of singular vectors and universal adversarial perturbations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8562–8570
Kim NV, Chervonenkis M (2015) Situation control of unmanned aerial vehicles for road traffic monitoring. Math Models Methods Appl Sci 9:1
Kim W, Choi H, Jang B et al. (2017) Driver distraction detection using single convolutional neural network. In: 2017 International Conference on Information and Communication Technology Convergence (ICTC), pp 1203–1205, https://doi.org/10.1109/ICTC.2017.8190898
Kim S, Lee K, Doo S, et al. (2019) Automotive RADAR signal classification using bypass recurrent convolutional networks. In: 2019 IEEE/CIC International conference on communications in China (ICCC), pp 798–803
Kim YY, Kim H, Lee W et al. (2021) Black-box expectation-maximization algorithm for estimating latent states of high-speed vehicles. J Aerospace Inf Syst 18(4):175–192
Koay HV, Chuah JH, Chow CO et al. (2021) Yolo-rtUAV: towards real-time vehicle detection through aerial images with low-cost edge devices. Remote Sensing 13(21):4196. https://doi.org/10.3390/rs13214196

Koch W, Mancuso R, West R et al. (2019) Reinforcement learning for UAV attitude control. ACM Trans Cyber-Phys Syst 3(2):1–21. https://doi.org/10.1145/3301273
Kocić J, Jovičić N, Drndarević V (2019) An end-to-end deep neural network for autonomous driving designed for embedded automotive platforms. Sensors 19(9):2064. https://doi.org/10.3390/s19092064
Koren M, Alsaif S, Lee R et al. (2018) Adaptive stress testing for autonomous vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV), IEEE, pp 1–7
Kukreja R, Rinchen S, Vaidya B et al. (2020) Evaluating traffic signs detection using faster R-CNN for autonomous driving. In: 2020 IEEE 25th International workshop on computer aided modeling and design of communication links and networks (CAMAD), pp 1–6, https://doi.org/10.1109/CAMAD50429.2020.9209289
Kumari A, Tanwar S, Tyagi S et al. (2017) Multimedia big data computing and internet of things applications: a taxonomy and process model. Multimed Tools Appl 76(8):10795–10827
Kurakin A, Goodfellow I, Bengio S (2016) Adversarial machine learning at scale. arXiv preprint arXiv:1611.01236
Kurakin A, Xu K, Wang W et al. (2017) Adversarial attacks and defences competition. In: International Conference on Learning Representations (ICLR) Workshop Track, https://arxiv.org/abs/1804.00097
Kurakin A, Goodfellow I, Bengio S et al. (2018) Adversarial attacks and defences competition. In: The NIPS'17 Competition: Building Intelligent Systems. Springer, p 195–231
Kusenbach M, Luettel T, Wuensche HJ (2020) Fast object classification for autonomous driving using shape and motion information applying the dempster-shafer theory. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE, pp 1–6
Kuutti S, Bowden R, Jin Y et al. (2020) A survey of deep learning applications to autonomous vehicle control. IEEE Trans Intell Transp Syst 22(2):712–733
Lai-Dang QV (2024) A survey of vision transformers in autonomous driving: Current trends and future directions. arXiv preprint arXiv:2403.07542
Le THN, Zheng Y, Zhu C et al. (2016) Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection. In: 2016 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 46–53, https://doi.org/10.1109/CVPRW.2016.13
Lecuyer M, Atlidakis V, Geambasu R et al. (2019) Certified robustness to adversarial examples with differential privacy. In: 2019 IEEE symposium on security and privacy (SP), IEEE, pp 656–672
Lee J, Wang J, Crandall D et al. (2017) Real-time, cloud-based object detection for unmanned aerial vehicles. In: 2017 First IEEE International conference on robotic computing (IRC), pp 36–43, https://doi.org/10.1109/IRC.2017.77
Lee T, Mckeever S, Courtney J (2021) Flying free: a research overview of deep learning in drone navigation autonomy. Drones 5(2):52
LG-Autonomous SS (2023) Autonomous and robotics real-time sensor simulation, LiDAR, camera simulation for ROS1, ROS2, Autoware. Baidu Apollo, Perception, Planning, Localization, SIL and HIL Simulation, Open Source and Free
Li S, Liu T, Zhang C et al. (2017) Learning unmanned aerial vehicle control for autonomous target following. arXiv preprint arXiv:1709.08233
Li Q, Mou L, Xu Q et al. (2018) $R^3$-net: a deep network for multi-oriented vehicle detection in aerial images and videos. arXiv preprint arXiv:1808.05560
Li M, Hu T (2021a) Deep learning enabled localization for UAV autolanding. Chin J Aeronaut 34(5):585–600. https://doi.org/10.1016/j.cja.2020.11.011
Li Y, Chen Y, Yuan S et al. (2021b) Vehicle detection from road image sequences for intelligent traffic scheduling. Comput Electr Eng 95:107406
Li G, Qiu Y, Yang Y et al. (2022) Lane change strategies for autonomous vehicles: a deep reinforcement learning approach based on transformer. IEEE Trans Intell Veh 8(3):2197–2211
Li Y, Fan Q, Huang H et al. (2023) A modified YOLOv8 detection network for UAV aerial image recognition. Drones 7(5):304
Liang X, Wang T, Yang L et al. (2018) Cirl: Controllable imitative reinforcement learning for vision-based self-driving. In: Proceedings of the European conference on computer vision (ECCV), pp 584–599
Liao F, Liang M, Dong Y et al. (2018) Defense against adversarial attacks using high-level representation guided denoiser. In: 2018 IEEE/CVF conference on computer vision and pattern recognition, pp 1778–1787, https://doi.org/10.1109/CVPR.2018.00191
Liao Z, Ma Y, Huang J et al. (2023) Energy-aware 3D-deployment of UAV for IOV with highway interchange. IEEE Trans Commun 71(3):1536–1548. https://doi.org/10.1109/TCOMM.2022.3232512
Lim H, Bae B, Han LD et al. (2021a) A data-fusion method using bayesian approach to enhance raw data accuracy of position and distance measurements for connected vehicles. In: 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), pp 1018–1023

Lim WYB, Garg S, Xiong Z et al. (2021b) UAV-assisted communication efficient federated learning in the era of the artificial intelligence of things. IEEE Network pp 1–8. https://doi.org/10.1109/MNET.002.2000334
Lim WYB, Huang J, Xiong Z et al. (2021c) Towards federated learning in UAV-enabled internet of vehicles: a multi-dimensional contract-matching approach. IEEE Trans Intell Transp Syst 22(8):5140–5154. https://doi.org/10.1109/TITS.2021.3056341
Lin CM, Tai CF, Chung CC (2014) Intelligent control system design for UAV using a recurrent wavelet neural network. Neural Comput Appl 24:487–496. https://doi.org/10.1007/s00521-012-1242-5
Lin TY, Cui Y, Belongie S et al. (2015) Learning deep representations for ground-to-aerial geolocalization. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 5007–5015, https://doi.org/10.1109/CVPR.2015.7299135
Lin TY, Goyal P, Girshick R et al. (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988
Lin N, Fu L, Zhao L et al. (2020) A novel multimodal collaborative drone-assisted VANET networking model. IEEE Trans Wireless Commun 19(7):4919–4933
Liu S, Liu L, Tang J et al. (2019a) Edge computing for autonomous driving: opportunities and challenges. Proc IEEE 107(8):1697–1716. https://doi.org/10.1109/JPROC.2019.2915983
Liu S, Wang S, Shi W et al. (2019b) Vehicle tracking by detection in UAV aerial video. Sci China Inf Sci 62(2):24101
Liu X, Yang H, Liu Z et al. (2021) Dpatch: an adversarial patch attack on object detector. IEEE Trans Dependable Secure Comput 18(4):2374–2385
Liu Y, Zhou J, Tian D et al. (2022) Joint communication and computation resource scheduling of a UAV-assisted mobile edge computing system for platooning vehicles. IEEE Trans Intell Transp Syst 23(7):8435–8450. https://doi.org/10.1109/TITS.2021.3082539
Liu W, Anguelov D, Erhan D et al. (2016a) SSD: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, pp 21–37
Liu Y, Chen X, Liu C et al. (2016b) Delving into transferable adversarial examples and black-box attacks
Liu X, Cheng M, Zhang H, et al. (2018a) Towards robust neural networks via random self-ensemble. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 369–385
Liu Y, Ma S, Aafer Y et al. (2018b) Trojaning attack on neural networks. In: Proceedings 2018 Network and distributed system security symposium, internet society
Lombacher J, Hahn M, Dickmann J et al. (2016) Potential of RADAR for static object classification using deep learning methods. In: 2016 IEEE MTT-S International conference on microwaves for intelligent mobility (ICMIM), pp 1–4, https://doi.org/10.1109/ICMIM.2016.7533931
Luo W, Tang Q, Fu C et al. (2018) Deep-sarsa based multi-UAV path planning and obstacle avoidance in a dynamic environment. In: International Conference on Swarm Intelligence, Springer, pp 102–111
Lu W, Zhou Y, Wan G et al. (2019) L3-Net: Towards learning based LiDAR localization for autonomous driving. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 6382–6391, https://doi.org/10.1109/CVPR.2019.00655
Ma Z, Xiong J, Gong H et al. (2024) Mission planning of uavs and cavs based on graph neural network transformer model. IEEE Internet Things J 11(24):40532–40546. https://doi.org/10.1109/JIOT.2024.3451248
Maanpää J, Taher J, Manninen P et al. (2021) Multimodal end-to-end learning for autonomous steering in adverse road and weather conditions. In: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, pp 699–706
Mahaur B, Mishra K (2023) Small-object detection based on yolov5 in autonomous driving systems. Pattern Recognition Lett 168:115–122
Mahmood A, Zen H, Hilles SM (2018) Big Data and privacy issues for connected vehicles in intelligent transportation systems, pp 1–7. https://doi.org/10.1007/978-3-319-63962-8_234-1
Major B, Fontijne D, Ansari A et al. (2019) Vehicle detection with automotive RADAR using deep learning on range-azimuth-doppler tensors. In: 2019 IEEE/CVF International conference on computer vision workshop (ICCVW), pp 924–932,https://doi.org/10.1109/ICCVW.2019.00121
Malawade AV, Mortlock T, Al Faruque MA (2022) Hydrafusion: Context-aware selective sensor fusion for robust and efficient autonomous vehicle perception. In: 2022 ACM/IEEE 13th International conference on cyber-physical systems (ICCPS), pp 68–79, https://doi.org/10.1109/ICCPS54341.2022.00013
Manzoor MA, Morgan Y, Bais A (2019) Real-time vehicle make and model recognition system. Mach Learn Knowledge Extraction 1:611–629. https://doi.org/10.3390/make1020036
Maqueda AI, Loquercio A, Gallego G et al. (2018) Event-based vision meets deep learning on steering prediction for self-driving cars. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 5419–5427, https://doi.org/10.1109/CVPR.2018.00568

Marti E, de Miguel MA, Garcia F et al. (2019) A review of sensor technologies for perception in automated driving. IEEE Intell Transp Syst Mag 11(4):94–108. https://doi.org/10.1109/MITS.2019.2907630
Maturana D, Scherer SA (2015) 3D convolutional neural networks for landing zone detection from LiDAR. 2015 IEEE International conference on robotics and automation (ICRA) pp 3471–3478
Mendis GJ, Randeny T, Wei J et al. (2016) Deep learning based doppler RADAR for micro UAS detection and classification. In: MILCOM 2016-2016 IEEE Military Communications Conference, pp 924–929, https://doi.org/10.1109/MILCOM.2016.7795448
Menouar H, Guvenc I, Akkaya K et al. (2017a) UAV-enabled intelligent transportation systems for the smart city: applications and challenges. IEEE Commun Mag 55(3):22–28. https://doi.org/10.1109/MCOM.2017.1600238CM
Menouar H, Guvenc I, Akkaya K et al. (2017b) UAV-enabled intelligent transportation systems for the smart city: applications and challenges. IEEE Commun Mag 55(3):22–28
Milioto A, Vizzo I, Behley J et al. (2019) Rangenet ++: Fast and accurate LiDAR semantic segmentation. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 4213–4220, https://doi.org/10.1109/IROS40897.2019.8967762
Mirza M, Osindero S (2014) Conditional generative adversarial nets. International conference on neural information processing systems pp 2672–2680
Mishra D, Natalizio E (2020) A survey on cellular-connected UAVs: design challenges, enabling 5G/B5G innovations, and experimental advancements. Comput Netw 182:107451. https://doi.org/10.1016/j.comnet.2020.107451
Mittal P, Singh R, Sharma A (2020) Deep learning-based object detection in low-altitude UAV datasets: a survey. Image Vis Comput 104:104046. https://doi.org/10.1016/j.imavis.2020.104046
Mohseni S, Pitale M, Singh V et al. (2019) Practical solutions for machine learning safety in autonomous vehicles. arXiv preprint arXiv:1912.09630
Montañez OJ, Suarez MJ, Fernandez EA (2023) Application of data sensor fusion using extended Kalman filter algorithm for identification and tracking of moving targets from LiDAR-RADAR data. Remote Sensing 15(13):3396
Montavon G, Samek W, Müller KR (2018) Methods for interpreting and understanding deep neural networks. Digital Signal Process 73:1–15
Moosavi-Dezfooli SM, Fawzi A, Frossard P (2016) Deepfool: A simple and accurate method to fool deep neural networks. IEEE Conference on Computer Vision and Pattern Recognition pp 2574–2582
Moosavi-Dezfooli SM, Fawzi A, Fawzi O et al. (2017) Universal adversarial perturbations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1765–1773
Moosavi-Dezfooli SM, Shrivastava A, Tuzel O (2018) Divide, denoise, and defend against adversarial attacks. arXiv preprint arXiv:1802.06806
Mopuri KR, Ojha U, Garg U et al. (2018) Nag: Network for adversary generation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 742–751
Morito T, Sugiyama O, Kojima R et al. (2016) Partially shared deep neural network in sound source separation and identification using a UAV-embedded microphone array. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 1299–1304, https://doi.org/10.1109/IROS.2016.7759215
Motlagh NH, Taleb T, Arouk O (2016) Low-altitude unmanned aerial vehicles-based internet of things services: comprehensive survey and future perspectives. IEEE Internet Things J 3(6):899–922
Moukahal LJ, Elsayed MA, Zulkernine M (2020) Vehicle software engineering (vse): Research and practice. IEEE Internet of Things Journal pp 1
Mozaffari M, Saad W, Bennis M et al. (2019) A tutorial on UAVs for wireless networks: applications, challenges, and open problems. IEEE Commun Surv Tutorials 21(3):2334–2360. https://doi.org/10.1109/COMST.2019.2902862
Muhammad K, Lloret J, Baik SW (2019) Intelligent and energy-efficient data prioritization in green smart cities: current challenges and future directions. IEEE Commun Mag 57(2):60–65. https://doi.org/10.1109/MCOM.2018.1800371
Muhammad K, Ullah A, Lloret J et al. (2021) Deep learning for safe autonomous driving: current challenges and future directions. IEEE Trans Intell Transp Syst 22(7):4316–4336. https://doi.org/10.1109/TITS.2020.3032227
Müller M, Dosovitskiy A, Ghanem B et al. (2018) Driving policy transfer via modularity and abstraction. arXiv preprint arXiv:1804.09364
Nagabandi A, Finn C, Levine S (2018) Deep online learning via meta-learning: Continual adaptation for model-based rl
Nagpal R, Krishna C, Jayababu VR et al. (2019) Real-time traffic sign recognition using deep network for embedded platforms. Electron Imaging 2019:33. https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-033

Narottama B, Shin SY (2022a) Quantum neural networks for resource allocation in wireless communications. IEEE Trans Wireless Commun 21(2):1103–1116. https://doi.org/10.1109/TWC.2021.3102139
Narottama B, Duong TQ (2022b) Quantum neural networks for optimal resource allocation in cell-free MIMO systems. In: GLOBECOM 2022-2022 IEEE Global Communications Conference, pp 2444–2449, https://doi.org/10.1109/GLOBECOM48099.2022.10001726
Nazemi A, Azimifar Z, Shafiee MJ et al. (2020) Real-time vehicle make and model recognition using unsupervised feature learning. IEEE Trans Intell Transp Syst 21(7):3080–3090. https://doi.org/10.1109/TITS.2019.2924830
Nazib RA, Moh S (2020) Routing protocols for unmanned aerial vehicle-aided vehicular ad hoc networks: A survey. IEEE Access 8:77535–77560
Ng JS, Bryan Lim WY, Dai HN et al. (2020) Communication-efficient federated learning in UAV-enabled IOV: a joint auction-coalition approach. In: GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp 1–6, https://doi.org/10.1109/GLOBECOM42002.2020.9322584
Ng JS, Lim WYB, Dai HN et al. (2021) Joint auction-coalition formation framework for communication-efficient federated learning in UAV-enabled internet of vehicles. IEEE Trans Intell Transp Syst 22(4):2326–2344. https://doi.org/10.1109/TITS.2020.3041345
Nguyen A, Yosinski J, Clune J (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 427–436
Ni J, Chen Y, Chen Y et al. (2020) A survey on theories and applications for self-driving cars based on deep learning methods. Appl Sci 10:2749. https://doi.org/10.3390/app10082749
Nie Y, Zhao J, Gao F et al. (2021) Semi-distributed resource management in UAV-aided MEC systems: A multi-agent federated reinforcement learning approach. IEEE Transactions on Vehicular Technology pp 1. https://doi.org/10.1109/TVT.2021.3118446
Nie W, Guo B, Huang Y et al. (2022) Diffusion models for adversarial purification. arXiv preprint arXiv:2205.07460
Niranjan D, VinayKarthik B et al. (2021) Deep learning based object detection model for autonomous driving research using CARLA simulator. In: 2021 2nd international conference on smart electronics and communication (ICOSEC), IEEE, pp 1251–1258
Noor-A-Rahim M, Liu Z, Lee H et al. (2022) 6G for vehicle-to-everything (V2X) communications: enabling technologies, challenges, and opportunities. Proc IEEE 110(6):712–734. https://doi.org/10.1109/JPROC.2022.3173031
Norkobil Saydirasulovich S, Abdusalomov A, Jamil MK et al. (2023) A YOLOv6-based improved fire detection approach for smart city environments. Sensors 23(6):3161
Odena A, Olah C, Shlens J (2017) Conditional image synthesis with auxiliary classifier gans. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp 2642–2651
O'Kelly M, Sinha A, Namkoong H et al. (2018) Scalable end-to-end autonomous vehicle testing via rare-event simulation. Adv Neural Inf Process Syst 31
Ortega J, Lengyel H, Szalay Z (2020) Overtaking maneuver scenario building for autonomous vehicles with Prescan software. Transp Eng 2:100029
Osco LP, Junior JM, Ramos APM et al. (2021) A review on deep learning in UAV remote sensing. Int J Appl Earth Obs Geoinf 102:102456
Oubbati OS, Chaib N, Lakas A et al. (2019) UAV-assisted supporting services connectivity in urban VANETs. IEEE Trans Veh Technol 68(4):3944–3951
Oubbati OS, Atiquzzaman M, Baz A et al. (2021) Dispatch of UAVs for urban vehicular networks: a deep reinforcement learning approach. IEEE Trans Veh Technol 70(12):13174–13189. https://doi.org/10.1109/TVT.2021.3119070
Ounoughi C, Yahia SB (2023) Data fusion for ITS: a systematic literature review. Inf Fusion 89:267–291
Outay F, Mengash HA, Adnan M (2020) Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges. Transp Res Part A Policy Practice 141:116–129
Ouyang T, Marco VS, Isobe Y et al. (2021) Corner case data description and detection. In: 2021 IEEE/ACM 1st Workshop on AI Engineering—Software Engineering for AI (WAIN), https://doi.org/10.1109/WAIN52551.2021.00009
Padhy RP, Ahmad S, Verma S et al. (2021) Localization of unmanned aerial vehicles in corridor environments using deep learning. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp 9423–9428, https://doi.org/10.1109/ICPR48806.2021.9412096
Pan Y, Cheng CA, Saigol K et al. (2018) Agile autonomous driving using end-to-end deep imitation learning. In: Robotics: science and systems
Pang T, Xu K, Du C et al. (2019) Improving adversarial robustness via promoting ensemble diversity. In: International Conference on Machine Learning, PMLR, pp 4970–4979

Panov A, Yakovlev K, Suvorov R (2018) Grid path planning with deep reinforcement learning: preliminary results. Procedia Comput Sci 123:347–353. https://doi.org/10.1016/j.procs.2018.01.054
Papernot N, McDaniel P (2017) Extending defensive distillation. arXiv e-prints pp arXiv–1705
Papernot N, McDaniel P (2018) Deep k-nearest neighbors: towards confident, interpretable and robust deep learning. arXiv preprint arXiv:1803.04765
Papernot N, McDaniel P, Goodfellow I (2016a) Limitations of the learning-while-testing paradigm for adversarial robustness. In: International Conference on Machine Learning, PMLR, pp 432–441
Papernot N, McDaniel P, Jha S et al. (2016b) Distillation as a defense to adversarial perturbations against deep neural networks. 2016 IEEE Symposium on Security and Privacy (SP) pp 582–597
Papernot N, McDaniel P, Wu X et al. (2016c) Distillation as a defense to adversarial perturbations against deep neural networks. In: IEEE Symposium on Security and Privacy (S &P), pp 582–597, https://doi.org/10.1109/SP.2016.41
Papernot N, McDaniel P, Wu X et al. (2016d) Distillation as a defense to adversarial perturbations against deep neural networks. In: 2016 IEEE symposium on security and privacy (SP), IEEE, pp 582–597
Park D, Lee S, Park S et al. (2021) RADAR-spectrogram-based UAV classification using convolutional neural networks. Sensors 21(1):210. https://doi.org/10.3390/s21010210
Parkhi O, Vedaldi A, Zisserman A (2015) Deep face recognition. In: BMVC 2015-Proceedings of the British Machine Vision Conference 2015, British Machine Vision Association
Patel K, Rambach K, Visentin T et al. (2019) Deep learning-based object classification on automotive RADAR Spectra. 2019 IEEE RADAR Conference (RADARConf) pp 1–6
Paxton C, Raman V, Hager GD et al. (2017) Combining neural networks and tree search for task and motion planning in challenging environments. In: 2017 IEEE/RSJ International conference on intelligent robots and systems (IROS), IEEE, pp 6059–6066
Pham QV, Zeng M, Ruby R et al. (2021) UAV communications for sustainable federated learning. IEEE Trans Veh Technol 70(4):3944–3948. https://doi.org/10.1109/TVT.2021.3065084
Pokhrel SR, Choi J (2020) A decentralized federated learning approach for connected autonomous vehicles. In: 2020 IEEE wireless communications and networking conference workshops (WCNCW), pp 1–6, https://doi.org/10.1109/WCNCW48565.2020.9124733
Polvara R, Patacchiola M, Sharma S et al. (2018) Toward end-to-end control for UAV autonomous landing via deep reinforcement learning. In: 2018 International Conference on Unmanned Aircraft Systems (ICUAS), pp 115–123, https://doi.org/10.1109/ICUAS.2018.8453449
Pomerleau DA (1988) Alvinn: an autonomous land vehicle in a neural network. Adv Neural Information Processing Syst 1:1988
Poudel S, Moh S (2019) Medium access control protocols for unmanned aerial vehicle-aided wireless sensor networks: a survey. IEEE Access 7:65728–65744
Punjani A, Abbeel P (2015) Deep learning helicopter dynamics models. In: 2015 IEEE international conference on robotics and automation (ICRA), pp 3223–3230, https://doi.org/10.1109/ICRA.2015.7139643
Qayyum A, Usama M, Qadir J et al. (2020) Securing connected autonomous vehicles: challenges posed by adversarial machine learning and the way forward. IEEE Commun Surveys Tutorials 22(2):998–1026. https://doi.org/10.1109/COMST.2020.2975048
Qing Z, Zhu M, Wu Z (2018) Adaptive neural network control for a quadrotor landing on a moving vehicle. In: 2018 Chinese Control And Decision Conference (CCDC), pp 28–33, https://doi.org/10.1109/CCDC.2018.8407041
Radovic M, Adarkwa O, Wang Q (2017) Object recognition in aerial images using convolutional neural networks. J Imaging 3(2):21. https://doi.org/10.3390/jimaging3020021
Raghunathan A, Steinhardt J, Liang P (2018) Certified defenses against adversarial examples. arXiv preprint arXiv:1801.09344
Rahim MA, Hassan HR (2021) A deep learning based traffic crash severity prediction framework. Transp Res Part C Emerg Technol 126:103209
Ramana K, Srivastava G, Kumar MR et al. (2023) A vision transformer approach for traffic congestion prediction in urban areas. IEEE Trans Intell Transp Syst 24(4):3922–3934
Ramos S, Gehrig S, Pinggera P et al. (2017) Detecting unexpected obstacles for self-driving cars: Fusing deep learning and geometric modeling. In: 2017 IEEE Intelligent Vehicles Symposium (IV), pp 1025–1032
Rao Q, Frtunikj J (2018) Deep learning for self-driving cars: Chances and challenges. In: 2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS), pp 35–38
Rausch V, Hansen A, Solowjow E et al. (2017) Learning a deep neural net policy for end-to-end control of autonomous vehicles. In: 2017 American Control Conference (ACC), pp 4914–4919, https://doi.org/10.23919/ACC.2017.7963716
Raza A, Bukhari SHR, Aadil F et al. (2021) An UAV-assisted VANET architecture for intelligent transportation system in smart cities. Int J Distrib Sens Netw 17(7):15501477211031750

Redmon J, Divvala S, Girshick R et al. (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767
Rehder E, Quehl J, Stiller C (2017) Driving like a human: imitation learning for path planning using convolutional neural networks. In: International Conference on Robotics and Automation Workshops, pp 1–5
Ren S, He K, Girshick R et al. (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 28:1137–1149
Ren K, Zheng T, Qin Z et al. (2020) Adversarial attacks and defenses in deep learning. Engineering 6(3):346–360. https://doi.org/10.1016/j.eng.2019.12.012
Ristea NC, Anghel A, Ionescu RT (2020) Fully convolutional neural networks for automotive RADAR interference mitigation. In: 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), IEEE, pp 1–5
Rizzoli G, Barbato F, Caligiuri M et al. (2023) Syndrone-multi-modal uav dataset for urban scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 2210–2220
Rong-xiao G, Ji-wei T, Bu-hong W et al. (2020) Cyber-physical attack threats analysis for UAVs from CPS perspective. In: 2020 International conference on computer engineering and application (ICCEA), pp 259–263, https://doi.org/10.1109/ICCEA50009.2020.00063
Rosero LA, Gomes IP, da Silva JAR et al. (2020) A software architecture for autonomous vehicles: Team LRM-B entry in the first CARLA autonomous driving challenge. arXiv preprint arXiv:2010.12598
Rosero L, Silva J, Wolf D et al. (2022) CNN-Planner: A neural path planner based on sensor fusion in the bird's eye view representation space for mapless autonomous driving. In: 2022 Latin American Robotics Symposium (LARS), 2022 Brazilian Symposium on Robotics (SBR), and 2022 Workshop on Robotics in Education (WRE), IEEE, pp 181–186
Rozsa A, Rudd EM, Boult TE (2016) Adversarial diversity and hard positive generation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 25–32
Rozsa A, Töröcsik T, Liao Z et al. (2018) Beyond pascal: A benchmark for 3d object detection in the wild. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, pp 395–404
Sadeghi K, Banerjee A, Gupta SKS (2020) A system-driven taxonomy of attacks and defenses in adversarial machine learning. IEEE Trans Emerg Topics Comput Intell 4(4):450–467. https://doi.org/10.1109/TETCI.2020.2968933
Sallab A, Abdou M, Perot E et al. (2017) Deep reinforcement learning framework for autonomous driving. Electron Imaging 19:70–76. https://doi.org/10.2352/issn.2470-1173.2017.19.avm-023
Samangouei P, Kabkab M, Chellappa R (2018) Defense-GA: Protecting classifiers against adversarial attacks using generative models. arXiv preprint arXiv:1805.06605
Samaras S, Magoulianitis V, Dimou A et al. (2019) UAV classification with deep learning using surveillance radar data, pp 744–753
Samek W, Wiegand T, Müller KR (2017) Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296
Saputro N, Akkaya K, Algin R et al. (2018) Drone-assisted multi-purpose roadside units for intelligent transportation systems. In: 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), IEEE, pp 1–5
Sarker IH, Kayes ASM, Badsha S et al. (2020) Cybersecurity data science: an overview from machine learning perspective. J Big Data 7(1):41. https://doi.org/10.1186/s40537-020-00318-5
Satar B, Dirik AE (2018) Deep learning based vehicle make-model classification. Lecture Notes in Computer Science p 544–553. https://doi.org/10.1007/978-3-030-01424-7_53,
Satti SK, Suganya Devi K, Dhar P et al. (2021) A machine learning approach for detecting and tracking road boundary lanes. ICT Express 7(1):99–103. https://doi.org/10.1016/j.icte.2020.07.007
Sauer A, Savinov N, Geiger A (2018) Conditional affordance learning for driving in urban environments. In: Conference on robot learning, PMLR, pp 237–252
Savazzi S, Nicoli M, Bennis M et al. (2021) Opportunities of federated learning in connected, cooperative, and automated industrial systems. IEEE Commun Mag 59(2):16–21. https://doi.org/10.1109/MCOM.001.2000200
Scheiner N, Appenrodt N, Dickmann J, et al. (2019) RADAR-based road user classification and novelty detection with recurrent neural network ensembles. In: 2019 IEEE Intelligent Vehicles Symposium (IV), IEEE, pp 722–729
Schoettle B (2017) Sensor fusion: a comparison of sensing capabilities of human drivers and highly automated vehicles. Report No. SWT-2017-12, University of Michigan, 2017
Scott K, Dai R, Kumar M (2016) Occlusion-aware coverage for efficient visual sensing in unmanned aerial vehicle networks. In: 2016 IEEE Global Communications Conference (GLOBECOM), pp 1–6, https://doi.org/10.1109/GLOCOM.2016.7842033

Seliem H, Shahidi R, Ahmed MH et al. (2018) Drone-based highway-VANET and DAS service. IEEE Access 6:20125–20137

Shah U, Khawad R, Krishna KM (2016) Deepfly: Towards complete autonomous navigation of mavs with monocular camera. In: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing. Association for Computing Machinery, New York, NY, USA, ICVGIP '16, https://doi.org/10.1145/3009977.3010047,

Sha H, Mu Y, Jiang Y et al. (2023) Languagempc: large language models as decision makers for autonomous driving. arXiv preprint arXiv:2310.03026

Shao H, Wang L, Chen R et al. (2023a) Safety-enhanced autonomous driving using interpretable sensor fusion transformer. In: Conference on Robot Learning, PMLR, pp 726–737

Shao H, Wang L, Chen R et al. (2023b) ReasonNet: end-to-end driving with temporal and global reasoning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 13723–13733

Sharif M, Bhagavatula S, Bauer L et al. (2016) Accessorize to a crime: real and stealthy attacks on state-of-the-art face recognition. In: 2016 IEEE Symposium on Security and Privacy (SP), IEEE, pp 152–167

Sharma P, Austin D, Liu H (2019) Attacks on machine learning: Adversarial examples in connected and autonomous vehicles. In: 2019 IEEE International Symposium on Technologies for Homeland Security (HST), pp 1–7, https://doi.org/10.1109/HST47167.2019.9032989

Sharma Y, Jaiswal S, Tiwari A (2020) Detection of adversarial examples using machine learning algorithms: a survey. IEEE Access 8:25525–25542

Shen Y, Yan WQ (2018) Blind spot monitoring using deep learning. In: 2018 International conference on image and vision computing New Zealand (IVCNZ), pp 1–5, https://doi.org/10.1109/IVCNZ.2018.8634716

Shi W, Zhou H, Li J et al. (2018) Drone assisted vehicular networks: architecture, challenges and opportunities. IEEE Netw 32(3):130–137

Shilin P, Kirichek R, Paramonov A et al. (2016) Connectivity of VANET segments using UAVs. In: Internet of Things, Smart Spaces, and Next Generation Networks and Systems: 16th International Conference, NEW2AN 2016, and 9th Conference, ruSMART 2016, St. Petersburg, Russia, September 26-28, 2016, Proceedings 16, Springer, pp 492–500

Shin J, Piran MJ, Song HK et al. (2022) UAV-assisted and deep learning-driven object detection and tracking for autonomous driving. In: Proceedings of the 5th International ACM Mobicom Workshop on Drone Assisted Wireless Communications for 5G and Beyond. Association for Computing Machinery, New York, NY, USA, DroneCom '22, p 7–12, https://doi.org/10.1145/3555661.3560856,

Shiri H, Park J, Bennis M (2020) Communication-efficient massive UAV online path control: federated learning meets mean-field game theory. IEEE Trans Commun 68(11):6840–6857. https://doi.org/10.1109/TCOMM.2020.3017281

Shrestha R, Bajracharya R, Kim S (2021) 6G enabled unmanned aerial vehicle traffic management: a perspective. IEEE Access 9:91119–91136

Sirisha U, Praveen SP, Srinivasu PN et al. (2023) Statistical analysis of design aspects of various yolo-based deep learning models for object detection. Int J Comput Intelligence Syst 16(1):126

Sligar AP (2020) Machine learning-based RADAR perception for autonomous vehicles using full physics simulation. IEEE Access 8:51470–51476. https://doi.org/10.1109/ACCESS.2020.2977922

Soll M, Hinz T, Magg S et al. (2019) Evaluating defensive distillation for defending text processing neural networks against adversarial examples. In: International Conference on Artificial Neural Networks, Springer, pp 685–696

Song Y, Wang T, Wu Y et al. (2021) Non-orthogonal multiple access assisted federated learning for UAV swarms: an approach of latency minimization. In: 2021 International wireless communications and mobile computing (IWCMC), pp 1123–1128, https://doi.org/10.1109/IWCMC51323.2021.9498792

Song R, Xu R, Festag A et al. (2023) Fedbevt: federated learning bird's eye view perception transformer in road traffic systems. IEEE Trans Intell Veh 9(1):958–969

Song Z, Zhang Z, Zhang K et al. (2024) Adversarial purification and fine-tuning for robust udc image restoration. arXiv preprint arXiv:2402.13629

Srisakaokul S, Zhang Y, Zhong Z et al. (2018) Muldef: Multi-model-based defense against adversarial examples for neural networks. arXiv preprint arXiv:1809.00065

Su Y, Liwang M, Chen Z et al. (2023) Toward optimal deployment of UAV relays in UAV-assisted IoV networks. IEEE Trans Veh Technol 72:1–14. https://doi.org/10.1109/TVT.2023.3272648

Sun H, Feng S, Yan X et al. (2021) Corner case generation and analysis for safety assessment of autonomous vehicles. Transp Res Rec 2675(11):587–600

Sun S, Nwodo K, Sugrim S et al. (2024) Vitguard: Attention-aware detection against adversarial examples for vision transformer. arXiv preprint arXiv:2409.13828

Sun L, Peng C, Zhan W et al. (2018) A fast integrated planning and control framework for autonomous driving via imitation learning. In: Dynamic Systems and Control Conference, American Society of Mechanical Engineers, p V003T37A012
Tabernik D, Skočaj D (2020) Deep learning for large-scale traffic-sign detection and recognition. IEEE Trans Intell Transp Syst 21(4):1427–1440. https://doi.org/10.1109/TITS.2019.2913588
Tahir NUA, Long Z, Zhang Z et al. (2024) PVswin-YOLOv8s: UAV-based pedestrian and vehicle detection for traffic management in smart cities using improved YOLOv8. Drones 8(3):84
Tampuu A, Matiisen T, Semikin M et al. (2020) A survey of end-to-end driving: architectures and training methods. IEEE Trans Neural Netw Learning Syst 33(4):1364–1384
Tekeli M, Yaman C, Acarman T et al. (2018) A fusion of a monocular camera and vehicle-to-vehicle communication for vehicle tracking: An experimental study. In: 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE Press, p 854–859, https://doi.org/10.1109/IVS.2018.8500449,
Telikani A, Sarkar A, Du B et al. (2024) Machine learning for UAV-aided ITS: a review with comparative study. IEEE Trans Intelligent Transportation Syst 25:15388–15406
Telikani A, Sarkar A, Du B et al. (2025) Unmanned aerial vehicle-aided intelligent transportation systems: vision, challenges, and opportunities. IEEE Commun Surv Tutorials. https://doi.org/10.1109/COMST.2025.3530913
Theile M, Bayerlein H, Nai R et al. (2020) UAV path planning using global and local map information with deep reinforcement learning. arXiv preprint arXiv:2010.06917
Tian Y, Pei K, Jana S et al. (2018) Deeptest: automated testing of deep-neural-network-driven autonomous cars. In: Proceedings of the 40th international conference on software engineering, pp 303–314
Tian J, Wang B, Guo R et al. (2022a) Adversarial attacks and defenses for deep-learning-based unmanned aerial vehicles. IEEE Internet Things J 9(22):22399–22409. https://doi.org/10.1109/JIOT.2021.3111024
Tian J, Wang B, Guo R et al. (2022b) Adversarial attacks and defenses for deep-learning-based unmanned aerial vehicles. IEEE Internet Things J 9(22):22399–22409. https://doi.org/10.1109/JIOT.2021.3111024
Tian Y, Li X, Zhang H et al. (2023) Vistagpt: generative parallel transformers for vehicles with intelligent systems for transport automation. IEEE Trans Intelligent Vehicles 8(9):4198–4207
Tong K, Solmaz S (2024) ConnectGPT: connect large language models with connected and automated vehicles. In: 2024 IEEE Intelligent Vehicles Symposium (IV), pp 581–588, https://doi.org/10.1109/IV55156.2024.10588835
Tramèr F, Papernot N, Goodfellow I et al. (2017) The space of transferable adversarial examples. arXiv preprint arXiv:1704.03453
Tran TM, Bui DC, Nguyen TV et al. (2024) Transformer-based spatio-temporal unsupervised traffic anomaly detection in aerial videos. IEEE Trans Circuits Syst Video Tech 34(9):8292–8309
Trenta F, Conoci S, Rundo F et al. (2019) Advanced motion-tracking system with multi-layers deep learning framework for innovative car-driver drowsiness monitoring. In: 2019 14th IEEE International conference on automatic face gesture recognition (FG 2019), pp 1–5
Valle F, Cooney M, Mikhaylov K et al. (2021) The integration of UAVs to the C-ITS stack. In: 2021 IEEE 29th International Conference on Network Protocols (ICNP), IEEE, pp 1–6
Velas M, Spanel M, Hradis M et al. (2018) CNN for very fast ground segmentation in velodyne LiDAR data. In: 2018 IEEE International conference on autonomous robot systems and competitions (ICARSC), IEEE, pp 97–103
Virgilio GVR, Sossa H, Zamora E (2020) Vision-based blind spot warning system by deep neural networks. In: Mexican Conference on Pattern Recognition, Springer, pp 185–194
Wang X, Zhang W, Wu X et al. (2017) Real-time vehicle type classification with deep convolutional neural networks. J Real-Time Image Proc 16:5–14
Wang X, Cheng P, Liu X et al. (2018) Fast and accurate, convolutional neural network based approach for object detection from UAV. In: IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, IEEE, pp 3171–3175
Wang H, Yu Y, Cai Y et al. (2019a) A comparative study of state-of-the-art deep learning algorithms for vehicle detection. IEEE Intell Transp Syst Mag 11(2):82–95
Wang B, Yao Y, Shan S et al. (2019b) Neural cleanse: identifying and mitigating backdoor attacks in neural networks. 2019 IEEE Symposium on Security and Privacy (SP) pp 707–723
Wang C, Yulu D, Zhou W et al. (2020a) A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition. J Adv Transp 2020:1–11. https://doi.org/10.1155/2020/9194028
Wang D, Li C, Wen S et al. (2020b) Defending against adversarial attack towards deep neural networks via collaborative multi-task training. IEEE Trans Dependable Secure Comput 19(2):953–965
Wang M, Luo X, Wang X et al. (2020c) Research on vehicle detection based on faster R-CNN for UAV images. In: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, pp 1177–1180, https://doi.org/10.1109/IGARSS39084.2020.9323323

Wang Y, Su Z, Zhang N et al. (2021) Learning in the air: secure federated learning for UAV-assisted crowdsensing. IEEE Trans Netw Sci Eng 8(2):1055–1069. https://doi.org/10.1109/TNSE.2020.3014385
Wang CY, Bochkovskiy A, Liao HYM (2023a) Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7464–7475
Wang S, Li C, Ng DWK et al. (2023b) Federated deep learning meets autonomous vehicle perception: design and verification. IEEE Netw 37(3):16–25. https://doi.org/10.1109/MNET.104.2100403
Wang Y, Jiao R, Lang C et al. (2023c) Empowering autonomous driving with large language models: a safety perspective. arXiv preprint arXiv:2312.00812
Wang CY, Yeh IH, Mark Liao HY (2024) Yolov9: learning what you want to learn using programmable gradient information. In: European conference on computer vision, Springer, pp 1–21
Wei Z, Wang C, Hao P et al. (2019) Vision-based lane-changing behavior detection using deep residual neural network. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp 3108–3113, https://doi.org/10.1109/ITSC.2019.8917158
Wong E, Kolter Z (2018) Provable defenses against adversarial examples via the convex outer adversarial polytope. In: International conference on machine learning, PMLR, pp 5286–5295
Wu B, Wan A, Yue X et al. (2018) Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3D LiDAR point cloud. In: 2018 IEEE international conference on robotics and automation (ICRA), IEEE, pp 1887–1893
Wu Y, Chen R, Liu S et al. (2019) A defense framework for deep neural networks against adversarial attacks. In: 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), IEEE, pp 302–307
Wu X, Li W, Hong D et al. (2022) Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey. IEEE Geosci Remote Sensing Magazine 10(1):91–124. https://doi.org/10.1109/MGRS.2021.3115137
Wu K, Li YB, Lou J et al. (2024) Rapid plug-in defenders. In: The Thirty-eighth annual conference on neural information processing systems, https://openreview.net/forum?id=UMPedMhKWm
Xiang C, Feng C, Xie X et al. (2023a) Multi-sensor fusion and cooperative perception for autonomous driving: a review. IEEE Intell Transp Syst Mag 15(5):36–58. https://doi.org/10.1109/MITS.2023.3283864
Xiang H, Xu R, Ma J (2023b) Hm-vit: Hetero-modal vehicle-to-vehicle cooperative perception with vision transformer. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 284–295
Xiao C, Li B, Zhu JY et al. (2018) Generating adversarial examples with adversarial networks. ACM Conference on Computer and Communications Security pp 1192–1203
Xiao Y, Codevilla F, Gurram A et al. (2020) Multimodal end-to-end autonomous driving. IEEE Trans Intell Transp Syst 23(1):537–547
Xie C, Wang J, Zhang Z et al. (2017) Adversarial examples for semantic segmentation and object detection. In: Proceedings of the IEEE international conference on computer vision, pp 1369–1378
Xie C, Wu Y, Maaten Lvd, et al. (2019) Feature denoising for improving adversarial robustness. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 501–509
Xing Y, Lv C, Wang H et al. (2019) Driver activity recognition for intelligent vehicles: a deep learning approach. IEEE Trans Veh Technol 68(6):5379–5390. https://doi.org/10.1109/TVT.2019.2908425
Xu H, Gao Y, Yu F, et al. (2017) End-to-end learning of driving models from large-scale video datasets. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2174–2182
Xu C, Liao X, Tan J et al. (2020) Recent research progress of unmanned aerial vehicle regulation policies and technologies in urban low altitude. IEEE Access 8:74175–74194
Xu W, Zhang C, Wang Q et al. (2022a) FEA-swin: foreground enhancement attention swin transformer network for accurate UAV-based dense object detection. Sensors 22(18):6993
Xu X, Feng Z, Cao C et al. (2022b) Stn-track: multiobject tracking of unmanned aerial vehicles by swin transformer neck and new data association method. IEEE J Selected Topics Appl Earth Observations Remote Sensing 15:8734–8743
Xu W, Yu Z, Wang Y et al. (2024a) RS-Agent: Automating remote sensing tasks through intelligent agents. arXiv preprint arXiv:2406.07089
Xu Z, Zhang Y, Xie E et al. (2024b) Drivegpt 4: interpretable end-to-end autonomous driving via large language model. IEEE Robotics Automation Lett 9(10):8186–8193. https://doi.org/10.1109/LRA.2024.3440097
Yan Z, Guo Y, Zhang C (2018) Deep defense: training DNNs with improved adversarial robustness. Adv Neural Inf Process Syst **31**
Yang S, Wang W, Liu C et al. (2017) Feature analysis and selection for training an end-to-end autonomous vehicle controller using deep learning approach. In: 2017 IEEE intelligent vehicles symposium (IV), IEEE, pp 1033–1038
Yang B, Luo W, Urtasun R (2018) PIXOR: Real-time 3D object detection from point clouds. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp 7652–7660

Ye T, Qin W, Zhao Z et al. (2023) Real-time object detection network in UAV-vision based on CNN and transformer. IEEE Trans Instrum Meas 72:1–13
Yeong DJ, Velasco-Hernandez G, Barry J et al. (2021) Sensor and sensor fusion technology in autonomous vehicles: a review. Sensors 21(6):2140. https://doi.org/10.3390/s21062140
Yildirim M, Dagda B, Fallah S (2024) Highwayllm: Decision-making and navigation in highway driving with RL-Informed language model. arXiv preprint arXiv:2405.13547
Yu L, Shao X, Wei Y et al. (2018) Intelligent land-vehicle model transfer trajectory planning method based on deep reinforcement learning. Sensors 18(9):2905. https://doi.org/10.3390/s18092905
Yu SY, Malawade AV, Muthirayan D et al. (2021) Scene-graph augmented data-driven risk assessment of autonomous vehicle decisions. IEEE transactions on intelligent transportation systems pp 1–11. https://doi.org/10.1109/TITS.2021.3074854
Yuan X, He P, Zhu Q et al. (2019) Adversarial examples: attacks and defenses for deep learning. IEEE Trans Neural Netw Learning Syst 30(9):2805–2824
Yurtsever E, Lambert J, Carballo A et al. (2020) A survey of autonomous driving: Common practices and emerging technologies. IEEE Access 8:58443–58469
Zanjie H, Hiroki N, Nei K et al. (2014) Resource allocation for data gathering in UAV-aided wireless sensor networks. In: 2014 4th IEEE International conference on network infrastructure and digital content, IEEE, pp 11–16
Zeggada A, Melgani F, Bazi Y (2017) A deep learning approach to UAV image multilabeling. IEEE Geoscience and Remote Sensing Letters PP:1–5. https://doi.org/10.1109/LGRS.2017.2671922
Zeng T, Semiari O, Chen M et al. (2022) Federated learning on the road autonomous controller design for connected and autonomous vehicles. IEEE Trans Wireless Commun 21(12):10407–10423
Zhang S, Wen L, Bian X et al. (2018) Single-shot refinement neural network for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4203–4212
Zhang C, Benz P, Imtiaz T et al. (2020a) Understanding adversarial examples from the mutual influence of images and perturbations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 14521–14530
Zhang H, Hanzo L (2020b) Federated learning assisted multi-UAV networks. IEEE Trans Veh Technol 69(11):14104–14109. https://doi.org/10.1109/TVT.2020.3028011
Zhang Y, Zhang Y, Qi J et al. (2021) Adversarial patch attack on multi-scale object detection for UAV. Signal Processing Image Commun 98:117056
Zhang J, Li S, Li L (2023) Coordinating cav swarms at intersections with a deep learning model. IEEE Trans Intell Transp Syst 24(6):6280–6291
Zhan Y, Xiong Z, Yuan Y (2024) Skyeyegpt: Unifying remote sensing vision-language tasks via instruction tuning with large language model. arXiv preprint arXiv:2401.09712
Zhao D, Huang X, Peng H et al. (2017a) Accelerated evaluation of automated vehicles in car-following maneuvers. IEEE Trans Intell Transp Syst 19(3):733–744
Zhao D, Lam H, Peng H et al. (2017b) Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. IEEE Trans Intell Transp Syst 18(3):595–607. https://doi.org/10.1109/TITS.2016.2582208
Zhao Q, Sheng T, Wang Y et al. (2019a) M2det: a single-shot object detector based on multi-level feature pyramid network. In: Proceedings of the AAAI conference on artificial intelligence, pp 9259–9266
Zhao Y, Bai L, Lyu Y et al. (2019b) Camera-based blind spot detection with a general purpose lightweight neural network. Electronics 8(2):233. https://doi.org/10.3390/electronics8020233
Zhao ZQ, Zheng P, St X et al. (2019c) Object detection with deep learning: a review. IEEE Trans Neural Netw Learning Syst 30(11):3212–3232
Zhao Q, Liu B, Lyu S et al. (2023) Tph-yolov5++: boosting object detection on drone-captured scenarios with cross-layer asymmetric transformer. Remote Sensing 15(6):1687
Zheng S, Song Y, Leung T et al. (2016) Improving the robustness of deep neural networks via stability training. In: Proceedings of the conference on computer vision and pattern recognition, pp 4480–4488
Zheng Z, Hong P (2018) Robust detection of adversarial attacks by modeling the intrinsic properties of deep neural networks. Adv Neural Inf Process Syst 31
Zheng T, Chen PY, Liu X et al. (2020) Distributionally adversarial attack. In: Proceedings of the 37th International Conference on Machine Learning, pp 10833–10843
Zhou Y, Cheng N, Lu N et al. (2015) Multi-UAV-aided networks: aerial-ground cooperative vehicular networking architecture. IEEE Veh Technol Mag 10(4):36–44
Zhou H, Ma A, Niu Y et al. (2022) Small-object detection for UAV-based images using a distance metric method. Drones 6(10):308
Zhu S, Gui L, Cheng N et al. (2019) Joint design of access point selection and path planning for UAV-assisted cellular networks. IEEE Internet Things J 7(1):220–233

Zhu M, Gong Y, Tian C et al. (2024) A systematic survey of transformer-based 3D object detection for autonomous driving: Methods, challenges and trends. Drones 8(8):412

Zong S, Li Y, Chen S et al. (2023) Facilitating UAV application for CAV deployment. Tech. rep., Center for Connected and Automated Transportation. Purdue University

Zou Y, Lin L, Zhang L (2022) A task offloading strategy for compute-intensive scenarios in UAV-assisted IOV. In: 2022 IEEE 5th International conference on electronic information and communication technology (ICEICT), pp 427–431, https://doi.org/10.1109/ICEICT55736.2022.9909200

Zyner A, Worrall S, Nebot E (2018) A recurrent neural network solution for predicting driver intention at unsignalized intersections. IEEE Robotics Automation Lett 3(3):1759–1764

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Muhammad Umer Zia[1] · Wei Xiang[2] · Tao Huang[1] · Jameel Ahmad[3] · Jawwad Nasar Chattha[4] · Ijaz Haider Naqvi[5] · Faran Awais Butt[6]**

✉ Wei Xiang
w.xiang@latrobe.edu.au; wei.xiang@jcu.edu.au

Muhammad Umer Zia
muhammadumer.zia@my.jcu.edu.au

Tao Huang
tao.huang1@jcu.edu.au

Jameel Ahmad
jameel.ahmad@umt.edu.pk

Jawwad Nasar Chattha
jawwad.chattha@umt.edu.pk; j.chattha@norwichuni.ac.uk

Ijaz Haider Naqvi
ijaznaqvi@lums.edu.pk

Faran Awais Butt
faranawais.butt@kfupm.edu.sa

[1] College of Science and Engineering, James Cook University, Smithfield, Cairns QLD 4878, Australia

[2] School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia

[3] Department of Computer Science, School of Systems and Technology, University of Management and Technology Lahore, Lahore 54782, Pakistan

[4] Computer Arts and Technology, Norwich University of the Arts, Norwich NR24SN, UK

[5] School of Science and Engineering, Lahore University of Management Sciences, Lahore 54792, Pakistan

[6] Center for Communication Systems and Sensing, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia