frontiers | Frontiers in Robotics and AI

TYPE Review
PUBLISHED 28 October 2024
DOI 10.3389/frobt.2024.1430740

OPEN ACCESS

EDITED BY
Aleksandar Vakanski,
University of Idaho, United States

REVIEWED BY
Mark Post,
University of York, United Kingdom
Kaya Kuru,
University of Central Lancashire,
United Kingdom

*CORRESPONDENCE
Alexandros Gkillas,
gkillas@isi.gr

RECEIVED 10 May 2024
ACCEPTED 30 September 2024
PUBLISHED 28 October 2024

CITATION
Piperigkos N, Gkillas A, Arvanitis G, Nousias S, Lalos A, Fournaris A, Radoglou-Grammatikis P, Sarigiannidis P and Moustakas K (2024) Distributed intelligence in industrial and automotive cyber–physical systems: a review. *Front. Robot. AI* 11:1430740. doi: 10.3389/frobt.2024.1430740

Distributed intelligence in industrial and automotive cyber–physical systems: a review

Nikos Piperigkos[1], Alexandros Gkillas[1]*, Gerasimos Arvanitis[2], Stavros Nousias[3], Aris Lalos[1], Apostolos Fournaris[1], Panagiotis Radoglou-Grammatikis[4,5], Panagiotis Sarigiannidis[4] and Konstantinos Moustakas[2]

[1]Industrial Systems Institute, Athena Research Center, Patras, Greece, [2]Department of Electrical and Computer Engineering, University of Patras, Patras, Greece, [3]Chair of Computational Modeling and Simulation, School of Engineering and Design, Technical University of Munich, Munich, Germany, [4]Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani, Greece, [5]K3Y Ltd., Sofia, Bulgaria

Cyber–physical systems (CPSs) are evolving from individual systems to collectives of systems that collaborate to achieve highly complex goals, realizing a cyber–physical system of systems (CPSoSs) approach. They are heterogeneous systems comprising various autonomous CPSs, each with unique performance capabilities, priorities, and pursued goals. In practice, there are significant challenges in the applicability and usability of CPSoSs that need to be addressed. The decentralization of CPSoSs assigns tasks to individual CPSs within the system of systems. All CPSs should harmonically pursue system-based achievements and collaborate to make system-of-system-based decisions and implement the CPSoS functionality. The automotive domain is transitioning to the system of systems approach, aiming to provide a series of emergent functionalities like traffic management, collaborative car fleet management, or large-scale automotive adaptation to the physical environment, thus providing significant environmental benefits and achieving significant societal impact. Similarly, large infrastructure domains are evolving into global, highly integrated cyber–physical systems of systems, covering all parts of the value chain. This survey provides a comprehensive review of current best practices in connected cyber–physical systems and investigates a dual-layer architecture entailing perception and behavioral components. The presented perception layer entails object detection, cooperative scene analysis, cooperative localization and path planning, and human-centric perception. The behavioral layer focuses on human-in-the-loop (HITL)-centric decision making and control, where the output of the perception layer assists the human operator in making decisions while monitoring the operator's state. Finally, an extended overview of digital twin (DT) paradigms is provided so as to simulate, realize, and optimize large-scale CPSoS ecosystems.

KEYWORDS

cyber–physical systems, cyber–physical system of systems, cyber–physical–social system, human-in-the-loop, human-machine interfaces

1 Introduction

In the past few years, there has been a significant investment in cyber–physical systems of systems (CPSoSs) in various domains, like automotive, industrial manufacturing, railways, aerospace, smart buildings, logistics, energy, and industrial processes, all of which have a significant impact on the economy and society. The automotive domain is transitioning to the system of systems approach, aiming to provide a series of emergent functionalities like traffic management, collaborative car fleet management, or large-scale automotive adaptation to the physical environment, thus providing significant environmental benefits (e.g., air pollution reduction) and achieving significant societal impact.

Similarly, large infrastructure domains, like industrial manufacturing (Nota et al., 2020), are evolving into global, highly integrated CPSoSs that go beyond pure production and cover all parts of the value chain, including research, design, and service provision. This novel approach can enable a high level of flexibility, allowing for rapid adaptation to customer requirements, a high degree of product customization, and improved industrial sustainability. Furthermore, achieving collective behavior in CPSoS-based solutions for large-scale control processes will help citizens improve their quality of life through smart, safe, and secure cities, energy-efficient buildings, and green infrastructures (i.e., lighting, water, and waste management), as well as smart devices and services for smart home functionality, home monitoring, health services, and assisted living.

However, in practice, there are significant challenges in the applicability and usability of CPSoSs that need to be addressed to take full advantage of the CPSoS benefits and sustain/extend their growth. The fact that even a small CPSoS (e.g., a connected car) consists of several subsystems and executes thousands of lines of code highlights the complexity of the system-of-systems solution and the extremely elaborate CPSoS orchestration required, highlighting the need for an approach beyond traditional control and management centers (Engell et al., 2015). Given this complexity, having a centralized authority that handles all CPSoS processes, subsystems, and control loops seems to be challenging to capture and implement, thus pointing to the need for a different design, control, and management approach. The decentralization of CPSoS processes and overall functionality by assigning tasks to individual cyber–physical systems (CPSs) within the system of systems can be a reasonable solution. However, the collaborative mechanisms between CPSs (that constitute the CPSoS behavior) remain a point of research since appropriate tools and methodologies are needed to ensure that the expected system-of-systems functional requirements are met (the CPSoS operates as it should be) and that non-functional requirements are fulfilled (the CPSoS remains resilient, safe, and efficient). Another critical challenge in effectively developing CPSoSs is the need for integrating social and human factors into the design process of CPSs so that the cyber, physical, and human layers can be efficiently coordinated and operated (Zhou et al., 2020). Compared with traditional CPSs, cyber–physical–social systems (CPSSs) regard humans as an important factor of the system and, therefore, incorporate human-in-the-loop (HITL) mechanisms into system operations so as to increase the trustworthiness of the overall CPSoS. To be more concise, creating an intelligent CPSoS environment relies on both modern technology and the natural resources provided by its inhabitants. Specifically, both "things" (objects and devices) and "humans" are essential for making smart environments even smarter. People benefit from smart services made possible by today's technology while simultaneously contributing to the enhancement of business intelligence. In this context, CPSS, as the human-in-the-loop counterpart of CPSs, can be used to gather information from humans and provide services with user awareness, creating a more responsive and personalized intelligent environment.

To address the complex challenges in CPSoSs, researchers have proposed a two-layer architecture consisting of a perception layer and a behavioral layer. This approach serves as a foundation for advancing CPSoS research and development across multiple critical areas. The proposed architecture aims to enhance functionality, reliability, adaptability, and the overall situational awareness (Chen and Barnes, 2014) of CPSoSs in various domains, from automotive and industrial manufacturing to smart cities and healthcare. Situational awareness refers to the collective understanding of an environment shared among multiple agents or entities—whether human or machine—who work together to achieve a common goal. More specifically, this concept of collective understanding emphasizes that situational awareness is not confined to individual knowledge but is distributed across team members. By cooperating, participants can better understand complex environments, adapt to dynamic changes, and respond more efficiently to evolving situations. As such, by focusing on these two interconnected layers, researchers can tackle the intricate issues of system integration, human–machine interaction, and real-time decision-making that are essential for the next generation of CPSoSs. The perception layer focuses on enhancing situational awareness through sophisticated algorithms for object detection, cooperative scene analysis, cooperative localization, and path planning. Research in this layer aims to develop effective perception mechanisms that are foundational for achieving higher levels of autonomy and reliability in CPSoSs. These advancements will enable CPSoSs to interact more intelligently with their environment and make informed decisions based on comprehensive situational awareness. Additionally, the behavioral layer focuses on integrating human operators into CPSoSs, recognizing the crucial role of human knowledge, senses, and expertise in ensuring operational excellence. This layer introduces the HITL approach, which allows continuous interaction between humans and CPSoS control loops. It addresses applications in which humans directly control CPSoS functionality, systems passively monitor human actions and biometrics, and hybrid combinations of both. The behavioral layer explores advanced Human-Machine Interfaces (HMI), including speech recognition, gesture recognition, and extended reality technologies, to enhance situational awareness and enable seamless human–system interaction. Furthermore, it investigates the prediction of operator intentions to improve collaboration between humans and CPSoSs, particularly in industrial scenarios where safety and efficiency are paramount. By integrating human expertise and intuition alongside automated processes, this layer aims to create a true human–machine symbiosis, vital for maintaining system flexibility and responsiveness in dynamic environments and unforeseen events. Key research directions within this two-layer framework include decentralized control and management, human-in-the-loop integration, data analytics and cognitive computing,

real-time processing and decision making, and collaborative mechanisms between individual CPSs. These areas of study aim to develop more intelligent, responsive, and human-centric systems that can adapt to the complex demands of our interconnected world.

In addition to these approaches, digital twins (DTs) are an emerging technology that assists in addressing the challenges within CPSoSs (Mylonas et al., 2021). DTs create accurate virtual replicas of physical systems, allowing for continuous observation of system performance, real-time data integration, and predictive maintenance, thus improving the reliability and efficiency of CPSoSs (Tao et al., 2019). DTs facilitate better decision-making by providing a comprehensive view of the entire system and enabling the simulation of various scenarios for proactive planning and response (Wang et al., 2023b). Furthermore, the integration of multiple CPSoSs through DTs can significantly enhance task performance. By enabling seamless communication and coordination among different CPSoSs, DTs ensure that each subsystem can efficiently share data and resources, leading to improved overall system performance. For example, in smart city environments, integrating transportation systems, energy grids, and public safety networks through DTs can optimize urban operations, reduce response times in emergencies, and enhance the overall quality of life for residents (Jafari et al., 2023). By integrating DTs into the CPSoS framework, systems can achieve higher efficiency, reliability, and adaptability. This synergy between DTs and CPSoSs leads to smarter, more resilient, and more efficient systems across various domains, providing robust solutions to complex challenges and contributing to the overall improvement of system performance and human wellbeing.

The remainder of this survey is organized as follows. In Section 2, we present the related work and outline the contributions of this study. Section 3 describes the conceptual architecture of CPSs. Section 4 delves into the perception layer, summarizing relevant works and providing detailed insights. Section 5 focuses on the behavioral layer, offering a comprehensive summary of pertinent research. Section 6 discusses the role of digital twins in optimizing the CPSoS ecosystem. Section 7 identifies open research questions, while Section 8 highlights key lessons learned. In Section 9, we provide a detailed discussion of various aspects of the study. Finally, Section 10 concludes the survey.

2 Related work and contribution

Many of the recent review papers discuss how the CPSs are utilized in emerging applications. Chen (2017) conducted an extensive literature review on CPS applications. Sadiku et al. (2017) provided a brief introduction to CPSs, their applications, and challenges. Yilma et al. (2021) presented the SoA perspectives on CPSs regarding definitions, underlining principles, and application areas. Other survey papers focus more on very specific areas. Haque et al. (2014) presented a survey of CPS in healthcare applications, characterizing and classifying different components and methods that are required for the application of CPS in healthcare, while Oliveira et al. (2021) presented the use of CPSs in the chemical industry. On the other hand, architecture and CPS characteristics are also common issues that are discussed in many survey papers. Hu et al. (2012) reviewed previous works of CPS architecture and introduced the main challenges, which are real-time control, security assurance, and integration mechanisms. CPS characteristics (like generic architecture, design principles, modeling, dependability, and implementation) and their application domains are also presented by Liu and Wang (2020). Lozano and Vijayan (2020) presented the current state-of-the-art, intrinsic features (like autonomy, stability, robustness, efficiency, scalability, safe trustworthiness, reliable consistency, and accurate high precision), design methodologies, applications, and challenges for CPS. Liu et al. (2017b) discussed the development of CPS from the perspectives of the system model, information processing technology, and software design. Oliveira et al. (2021) discussed the use of artificial intelligence to confer cognition to the system. Topics such as control and optimization architectures and digital twins are presented as components of the CPSs. Al-Mhiqani et al. (2018) investigated the current threats on CPSs (e.g., the type of attack, impact, intention, and incident categories). Leitão et al. (2022) provided an analysis of the main aspects, challenges, and research opportunities to be considered for implementing collective intelligence in industrial CPSs. Estrada-Jimenez et al. (2023) explored the concept of smart manufacturing, focusing on self-organization and its potential to manage the complexity and dynamism of manufacturing environments. It presents a systematic literature review to summarize current technologies, implementation strategies, and future research directions in the field. Pulikottil et al. (2023) explored the integration of multi-agent systems in cyber–physical production systems for smart manufacturing, offering a thorough review and SWOT analysis validated by industry experts to assess their potential benefits and challenges. Hamzah et al. (2023) provided a comprehensive overview of CPSs across 14 critical domains, including transportation, healthcare, and manufacturing, highlighting their integration into modern society and their role in advancing the fourth industrial revolution. Additionally, DT-based survey works focus on realizing automotive CPS (Xie et al., 2022a), achieving high adaptability with a short development cycle, low complexity, and high scalability, which meet various design requirements during the development process, how the interconnection between different components in CPSs and DTs affect the smart manufacturing domain (Tao et al., 2019), as well as presenting the potential of DTs as a means to reinforce and secure CPSs and Industry 4.0 in general (Lampropoulos and Siakas, 2023).

In this survey, we focus on the CPS architecture and its modules that are used to increase the situational awareness of the CPSoS users. Considering the importance of human integration in CPSs, we include the HITL component and human–machine interaction to realize the CPSS paradigm. Additionally, we emphasize the critical role of DTs in optimizing CPSoS ecosystems. The main contributions of this paper can be summarized as follows:

- Comprehensive review of current best practices in connected CPSs.
- Investigation of a dual-layer architecture encompassing a perception layer and a behavioral layer, where the perception layer focuses on enhancing situational awareness and the behavioral layer integrates human operators through HITL mechanisms and advanced HMI technologies.

- Presentation of different datasets and sources of data available to the research community. Perception algorithms related to scene understanding (object detection and tracking, pose estimation), localization mapping, and path planning are thoroughly investigated. The behavioral part focuses on decision making and human-in-the-loop control.
- Discussion on the integration of DTs into CPSoSs, highlighting their applications in smart cities, intelligent transportation systems, and aerial traffic monitoring.

3 Conceptual architecture

3.1 CPSoSs are heterogeneous systems

They consist of various autonomous CPSs, each with unique performance capabilities, criticality levels, priorities, and pursued goals. CPSs, in general, are self-organized and, on several occasions, may have conflicting goals, thus competing to get access to common resources. However, from a CPSoS perspective, all CPSs must also harmonically pursue system-based achievements and collaborate to make system-of-system-based decisions and implement the CPSoS behavior. Considering that CPSoSs consists of many CPSs, finding the methodology to achieve such an equilibrium in a decentralized way is not an easy task. The above issue becomes more complex when we also consider the amount of data to be exchanged between CPSs and the processing of those data. The collection of data and the data analytics need to be refined in such a way that only the important information is extracted and forwarded to other CPSs and the overall system. Furthermore, mechanisms to handle large amounts of data in a distributed way are needed to extract cognitive patterns and detect abnormalities. Thus, local data classification, labeling, and refinement mechanisms should be implemented in each CPS to offload the complexity and communication overhead at the system-of-systems level (Atat et al., 2018).

In the above-described setup, we cannot overlook the fact that CPSoSs depend on humans since humans are part of the CPSoS functionality and services, interact with the CPSs, and contribute to the CPSoS behavior. Operators and managers play a key role in the operation of CPSoSs and make many important decisions, while in several cases, human CPS users are the key players in the CPSoS main role (thus forming cyber–physical human systems). Thus, we need to structure a close symbiosis between computer-based systems and human operators/users and constantly enhance human situational awareness as well as devise a collaborative mechanism for handling CPSoS decisions, forcing the CPSoSs to comply with human guidelines and reactions. Novel approaches to human–machine interfaces that employ eXtended Reality (XR) principles need to be devised to help humans gain fast and easy-to-grasp insights into CPSoS processes while also enabling their seamless integration into CPSoS operations.

As shown in Figure 1, it is assumed that a CPSoS consists of interconnected CPSs, each acting independently while also collectively functioning as part of the CPSoS. We also assume that each individual CPS is composed of a *perception module* and a *behavioral or decision-making module* while bearing actuation capacities represented by the *physical layer*; this configuration facilitates the coordination with the other connected CPSs toward the collective implementation of a common goal or mission. A key aspect of the proposed architecture is the integration of various sensors, including redundant, complimentary, and cooperative sensors across nodes. More specifically, incorporating redundant sensors in interconnected CPSs is necessary to improve reliability, fault tolerance, and system safety. Redundant sensors provide backup in the event of failure of primary sensors, ensuring continued operation without disruption and maintaining the integrity and reliability of CPS (Bhattacharya et al., 2023). Additionally, the use of complementary sensors in heterogeneous systems allows for cross-verification of data, better coverage of sensor limitations, and improved decision-making in dynamic environments (Alsamhi et al., 2024). For example, combining visuals with depth sensors in autonomous vehicles helps enhance object detection, environmental mapping, and path planning. The diverse data from complementary sensors can be fused to produce more accurate and comprehensive situational awareness. Finally, cooperative sensors on individual CPSs work together to improve the accuracy and robustness of measurements and operations. These sensors share information and collaborate to address limitations inherent in individual sensors. For instance, in robotics, multiple sensors like cameras, inertial measurement units (IMUs), and proximity sensors can cooperate to provide accurate localization and object detection (Zhang and Singh, 2015).

The aggregation of the individual perception modules formulates the perception layer of the CPSoS, while the sensor, behavioral, and physical layers represent the summation of sensing, decision making, and actuation capabilities of the CPSoS. The perception layer can also be envisioned as a cognitive engine that employs appropriate algorithmic approaches for effective scene understanding, a task predominately accomplished today by deep neural network architectures. To this end, data collection and annotation are crucial for the training of AI models to undertake such tasks. This review paper sheds light upon all of the aforementioned aspects of interconnected CPSs.

4 Perception layer

4.1 Cooperative scene analysis

4.1.1 Background on object detection from 2D and 3D data

Object detection has evolved considerably since the appearance of deep convolutional neural networks (Zhao et al., 2019b). Nowadays, there are two main branches of proposed techniques, namely, two-stage and single-stage detectors.

In the first one, the object detectors, using two stages, generate region proposals, which are subsequently classified into the categories that are determined by the application at hand (e.g., vehicles, cyclists and pedestrians, in the case of autonomous driving). Some important, representative, high-performance examples of this first branch are R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2016), spatial pyramid pooling net (He et al., 2015), region-based fully convolutional network (R-FCN) (Dai et al., 2016), feature

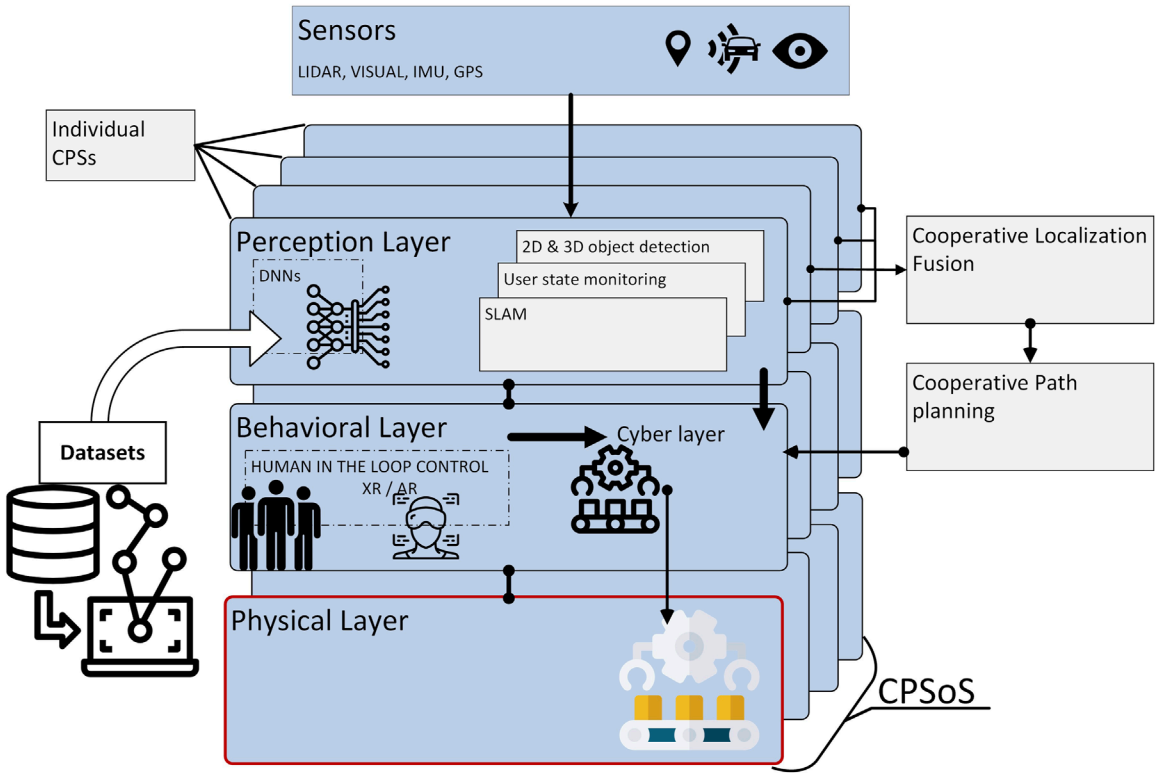

FIGURE 1
Cognitive cyber–physical system of system architecture with human-in-the-loop control.

pyramid network (FPN) (Lin et al., 2017), and mask R-CNN (He et al., 2017). In the second branch, object detection is cast to a single-stage, regression-like task with the aim to provide directly both the locations and the categories of the detected objects. Notable examples, here, are Single Shot MultiBox Detector (SSD) (Liu et al., 2016), SqueezeDet (Wu et al., 2017),YOLO (Redmon et al., 2016), YOLOv3 (Redmon and Farhadi, 2018) and EfficientDet (Tan et al., 2020). A recent review on object detection using deep learning (Zhao et al., 2019b) provides inquisitive insight into the aforementioned approaches to object detection.

Object detection in LiDAR point clouds is a three-dimensional problem where the sampled points are not uniformly distributed over the objects in the scene and do not directly correspond to a Cartesian grid. Three-dimensional object detection is dominantly performed using 3D convolutional networks due to the irregularity and lack of apparent structure in the point cloud. Several transformations take place to match the point cloud to feature maps that are forwarded into deep networks. Commendable detection outcomes have appeared in the literature as early as 2016. Li et al. (2016a) projected the 3D points onto a 2D map and employed 2D fully convolutional networks to successfully detect cars in a LiDAR point cloud, reaching an accuracy of 71.0% for car detection of moderate difficulty. A follow-up paper by Li (2017) proposed 3D fully convolutional networks, reporting an accuracy of 75.3% for car detection of moderate difficulty. However, since dense 3D fully convolutional networks demonstrate high execution times, Yan et al. (2018) investigated an improved sparse convolution method for such networks, which significantly increases the speed of both training and inference. According to KITTI benchmarks, the reported accuracy reaches 78.6% for car detection of moderate difficulty. To revisit 2D convolutions in 3D object detection, Lang et al. (2019) proposed a novel encoder called PointPillars that utilizes PointNets to learn a representation of point clouds organized in vertical columns (pillars) and subsequently employs a series of 2D convolutions. PointPillars reported an accuracy of 77.28% in the same category. Shi et al. (2019) proposed PointRCNN for 3D object detection from raw point clouds. They devised a two-stage approach where the first stage generates bottom-up 3D proposals and the second stage refines these proposals in the canonical coordinates to obtain the final detection results, reporting an accuracy of 78.70%. An extended variation of PointRCNN is the part-aware and aggregation neural network (Part-A^2 Net). The part-aware stage, for the first time, fully utilizes free-of-charge part supervisions derived from 3D ground-truth boxes to simultaneously predict high-quality 3D proposals and accurate intra-object part locations. Then, the part-aggregation stage learns to re-score the box and refines the box location by exploring the spatial relationship of the pooled intra-object part locations. The reported accuracy reaches 79.40%. Yang et al. (2020b) introduced the 3D single-stage detection (3DSSD) framework, which employed a unique fusion sampling strategy that included farthest point sampling in both feature and Euclidean spaces. PointGNN (Shi and Rajkumar, 2020) extended the application of graph neural networks to 3D object

detection. PV-RCNN (Shi et al., 2020) and its subsequent research (Shi et al., 2023) derived point-wise features from voxel abstraction networks to refine proposals generated by the 3D voxel backbone. Additionally, HVPR (Noh et al., 2021), a single-stage 3D detector, implemented an efficient memory module to enhance point-based features, achieving a better compromise between accuracy and efficiency. Qian et al. (2022) developed a lightweight region aggregation refine network (BANet) through local neighborhood graph construction, which resulted in more precise box boundary predictions.

4.1.2 Cooperative object detection and fusion

Object detection from a single point of view of a single agent is definitely vulnerable to a series of sensor limitations that can significantly affect the outcome. These limitations entail occlusion, limited field-of-view, and low-point density at distant regions. The transition from isolated CPSs to CPSoSs, enabling the collaboration among various agents, offers the opportunity to tackle such problems. Chen et al. (2019b) proposed a cooperative sensing scheme where each CPS combines its sensing data with those of other connected vehicles to help enhance perception. Furthermore, to tackle the increased amount of data, the authors propose a sparse point-cloud object detection method. It is important to highlight that the agents share on-board V2V information and fuse these data locally. Feature-level fusion is examined in a follow-up work by Chen et al. (2019a). The authors propose F-Cooper framework, a method that improves the autonomous vehicle's detection precision without introducing much computational overhead. This framework aims to utilize the capacity of feature maps, especially for 3D LiDAR data generated by autonomous vehicles as the feature maps are used for object detection only by single vehicles. F-Cooper is an end-to-end 3D object detection system with feature-level fusion supporting voxel feature fusion and spatial feature fusion. Voxel feature fusion achieves almost the same detection precision improvement compared to the raw-data level fusion solution, which offers the ability to dynamically adjust the size of feature maps to be transmitted. A unique characteristic of F-Cooper is that it can be deployed and executed on in-vehicle and roadside edge systems. Hurl et al. (2020) proposed TruPercept to tackle malicious attacks against cooperative perception systems. In their trust scheme, each agent reevaluates the detections originating from its neighboring agents using data from its position and perspective. Arnold et al. (2020) proposed a central system that fuses data from multiple infrastructure sensors, facilitating the management of both sensor and processing costs through shared resources while addressing evaluations of pose sensor configurations, the number of sensors, and the sensor field-of-view. The authors deploy VoxelNet (Zhou and Tuzel, 2018) and claim to have reached an average precision score of $AP_{3D} = 98\%$ for the early fusion strategy and $AP_{3D} = 81\%$ for the late fusion strategy in a T-Junction. In a more recent study, Guo et al. (2021) proposed cooperative spatial feature fusion (CoFF) to address F-Cooper limitations. F-Cooper's Chen et al. (2019a) *maxout* function treats feature maps from different sources and conditions similarly, leading to misclassifications. CoFF calculates and assigns importance weights to the received feature maps based on the data available to the ego vehicle. The authors claim to reach a 90% improvement in average precision for far object cases with respect to F-Cooper.

4.2 Cooperative localization, cooperative path planning, and SLAM

Unmanned vehicles, either ground (UGV), aerial (UAV), or underwater (UUV), are prominent CPSoSs. Typical examples include autonomous vehicles and robots, operating for a variety of different civilian and military challenging tasks. At the same time, the prototyping of 5G and V2X (e.g., V2V and V2I) related communication protocols enables the close collaboration of vehicles to address their main individual or collective goals. Autonomous vehicles with inter-communication and network abilities are known as connected and automated vehicles (CAVs), being part of the more general concept of connected CPSoSs. The main focus of CAV's related technologies is to increase and improve the safety, security, and energy consumption of (cooperative or not) autonomous driving by the strict control of the vehicle's position and motion (Montanaro et al., 2018). At a higher level, CAV have the potential for a further enhancement of the transportation sector's overall performance.

Perception and scene analysis ability are fundamental for a vehicle's reliable operation. Computer vision-based object detection and tracking should be seen as a first (though necessary) pre-processing step, feeding more sophisticated operational modules of vehicles (Eskandarian et al., 2021). The latter is imperative to have accurate knowledge of both its own and its neighbors' (vehicles, pedestrians, or static landmarks) position in order to design efficiently the future motion actions, i.e., to determine the best possible velocity, acceleration, yaw rate, etc. These motion actions primarily focus on, e.g., keeping safe inter-vehicular distances, eco-friendly driving by reducing gas emissions, etc. The above challenges can be addressed in the context of localization, SLAM, and Path planning, which are discussed below.

4.2.1 Cooperative localization

The localization module is responsible for providing absolute position information to the vehicles. Global Navigation Satellite Systems (GNSSs), like GPS, Beidou, Glonass, etc., are usually exploited for that purpose. The GPS sensor is currently employed as the most common commercial device. It is straightforward to couple or fuse GPS information with IMU readings (Noureldin et al., 2013) to design a complete inertial navigation system (INS) providing positioning, velocity, and timing (PVT) solutions. The IMU sensor consists of gyroscopes and accelerometers for measuring the yaw rate and acceleration (in x, y, z directions) of vehicles. Additionally, odometers and wheel sensors (Skog and Handel, 2009) can also be utilized. However, even highly reliable IMU sensors suffer from accumulative or drift error, significantly reducing their consistency as the vehicle is moving. Another limitation of stand-alone GPS localization is directly related to GPS itself. Its accuracy is highly degraded in dense urban canyons or tunnels (Kuutti et al., 2018), even exceeding $10m$ errors. The main sources of GPS signal degradation are due to Noureldin et al. (2013) satellite clock error, receiver clock error, ionosphere delay, tropospheric delay, multi-path, etc. Moreover, it is vulnerable to cyber-attacks (Ren et al., 2020), like spoofing or jamming. The former causes an intentionally "wrong" position, even kilometers away from the expected GPS measurement. The latter poses a rather more severe threat since it totally blocks the GPS signal. Several alternative approaches

relying on ground base stations have been developed for enhancing localization accuracies, such as assisted GPS (AGPS) or differential GPS (DGPS). However, they are also susceptible to multi-path effects and signal blockage (Alam and Dempster, 2013). The desired localization error, as it has been reported in the literature, should be lower than $1m$ (where in-lane accuracy) (Neto et al., 2020) to meet the standards of autonomous driving. For example, if a vehicle is localized on the curb instead of the road, it may lead to a serious accident with pedestrians or other vehicles. Therefore, it is quite clear that for obtaining the desired positioning solutions, other types of advanced sensors, like LiDAR, cameras, and RADAR, must be additionally taken into account. Moreover, the emergence of V2V communications in the context of the Internet of Things (IoT) facilitates the exploitation of both onboard and off-board information in order to design a more robust localization system. This collaborating multi-modal fusion of heterogeneous measurements is known as cooperative localization (CL), a rather recent and very promising technique that can tackle the limitations and drawbacks of GPS/IMU localization. Each vehicle can now receive external information (like absolute position, relative distance, velocity, and acceleration) from nearby vehicles, infrastructure, or pedestrians, effectively assisting its localization system.

There are many existing works (Kuutti et al., 2018; Buehrer et al., 2018; Wymeersch et al., 2009; Safavi et al., 2018; Gao, 2019) that survey-related aspects, challenges, and algorithms of CL. For example, Kuutti et al. (2018) provided an overview of current trends and future applications of localization (not only CL) in autonomous vehicle environments. The discussed techniques are mainly distinguished on the basis of the utilized sensor. Ranging measurements like relative distance and angle can also be extracted through the V2V abilities of CL. Common ranging techniques include time of arrival (TOA), angle of arrival (AoA), time difference of arrival (TDOA), and received signal strength (RSS). The works of Buehrer et al. (2018), Wymeersch et al. (2009) delve into detailed mathematical modeling of CL tasks. More specifically, Buehrer et al. (2018) exploit various criteria to categorize related algorithms:

- **Measurement type:** The sensor or ranging technique being used for localization. V2V communications enable different ranging methods to be used (as mentioned above).
- **Centralized vs. distributed:** Centralized algorithms require nodes/vehicles of the network to broadcast their measurements to a fusion center (e.g., cloud or some leader-vehicle), responsible for all the computations. Although higher accuracy can be achieved, limitations like communication overhead, computational power, network size, and fusion center malfunctioning must be taken into account. On the contrary, with distributed processing architecture, the computations are assigned to each vehicle, which interacts only with close neighbors.
- **One-shot vs. tracking:** One-shot refers to methods that do not exploit any past information. On the other hand, tracking has to do with algorithms that, apart from measurements, employ kinematic models in order to approximate the actual movement of vehicles. Tracking methods exploit Bayesian estimators, as mentioned below.
- **Fusion estimator:** Multi-modal fusion is vital for increased location estimation accuracy. Fusion can be effectively performed using well-known estimators like least squares (LS), maximum likelihood (ML), minimum mean square error (MMSE), and maximum a posteriori (MAP). The one-shot ML estimator coincides with (weighted by measurement noise variance) LS when the measurements are corrupted by Gaussian noise. MMSE and MAP are common Bayesian estimators that treat the unknown vehicle's position as a random variable instead of a deterministic value as one-shot do. Kalman, extended Kalman, and unscented Kalman Filters (KF, EKF, and UKF) are prominent examples of MMSE estimators. Belief propagation and factor graph optimization are also important MAP tools.

Wymeersch et al. (2009) formulated a distributed gradient descent (GD) algorithm as an LS solution and the Bayesian factor graph approach of the sum–product algorithm over wireless networks (SPAWNs). In general, distributed and tracking/Bayesian algorithms are more attractive to perform CL. An overview of distributed localization algorithms in IoT is also given by Safavi et al. (2018). Additionally, the authors discuss the proposed distributed geometric framework of DILOC, as well as the extended versions of DLRE and DILAND, which facilitate the design of a linear localization algorithm. These methods require the vehicle to be inside the convex full formed by three neighboring anchors (nodes with known and fully accurate positions) and to compute its barycentric coordinates with respect to neighbors. However, major challenges are related to mobile scenarios due to varying topologies, as well as how feasible the presence of anchors will be in automotive applications. An interesting approach is discussed by Meyer et al. (2016), where mobile agents, in general, try to cooperatively estimate their position and track non-cooperative objects. The authors developed a distributed particle filter-based belief propagation approach with message passing although they consider the presence of anchor nodes. Furthermore, the computational and communication overhead may be a serious limitation toward real-time implementation. Soatti et al. (2018) proposed a novel distributed technique to improve the stand-alone GNSS accuracy of vehicles. Once again, non-cooperative objects or features (e.g., trees and pedestrians) are exploited in order to improve location accuracy. Features are cooperatively detected by vehicles using their onboard sensors (e.g., LiDAR), where a perfect association is assumed. These vehicle-to-feature measurements are fused with GNSS in the context of a Bayesian message-passing approach and KF. Experimental evaluation was assessed using the SUMO simulator; however, the number of detected features, as well as communication overhead, should be taken into serious account. The work of Brambilla et al. (2020) extends that of Soatti et al. (2018) by proposing a distributed data association framework for features and vehicles. Data association was based on belief propagation. Validation was performed in realistic urban traffic conditions. The main aspect of Brambilla et al. (2020) and Soatti et al. (2018) is that vehicles must reach a consensus about feature states in order to improve their location. Graph Laplacian CL has been introduced in Piperigkos et al. (2020b) and Piperigkos et al. (2020a). Centralized or distributed Laplacian localization formulates an

LS optimization problem, which fuses the heterogeneous inter-vehicular measurements along with the V2V connectivity topology through the linear Laplacian operator. EKF- and KF-based solutions have been proposed for addressing CL in tunnels (Elazab et al., 2017; Yang et al., 2020a) when the GPS signal may be blocked. A distributed robust cubature KF enhanced by Huber M-estimation is presented by Liu et al. (2017a). The method is used to tackle the challenges of data fusion in the presence of outliers. Pseudo-range measurements from satellites are also considered during the fusion process. Zhao et al. (2020) developed a distributed Bayesian CL method for localizing vehicles in the presence of non-line-of-sight range measurements and spoofed vehicles. They focused primarily on ego vehicle location estimation and abnormal vehicle detection rates. Potential applications of localization in various domains like wireless sensor networks (WSNs), intelligent transportation systems (ITS), and robotics are demonstrated in Figure 2. Table 1 summarizes the above mentioned works.

4.2.2 SLAM

Simultaneous localization and mapping (SLAM) is also a relevant task of localization. It refers to the problem of mapping an environment using measurements from sensors (e.g., a camera or LiDAR) onboard the vehicle or robot while at the same time estimating the position of that sensor relative to the map. Although when stated in this way, SLAM can appear to be quite an abstract problem robust, and efficient solutions to the SLAM problem are critical to enabling the next wave of intelligent mobile devices. SLAM, in its general, form tries to estimate over a time period the poses of the vehicle/sensor and the landmarks' position of the map, given control input measurements provided by odometry sensors onboard the vehicle and measurements with respect to landmarks. Therefore, we have mainly two subsystems: the front-end, which detects the landmarks of the map and correlates them with the poses, and the back-end, which casts an optimization problem in order to estimate the pose and the location of landmarks. SLAM techniques can be distinguished to either visual or LiDAR based odometry (VO and LO) solutions, reflecting camera or LiDAR as the main sensor to be exploited:

- To compute the local position and motion of a camera, VO algorithms must estimate the transformation that the camera undergoes between the current frame and a reference frame. The reference frame can be defined by the previous frame in the input sequence, some key frame in the recent past, or a collection of frames from the recent past. In each case, the task is to estimate the transformation that takes information in the camera's current frame into the frame of reference of the past frame(s). This task can be seen as an optimization problem where the cost is given by the residual between the information measured in the current frame and the corresponding information derived and reprojected from the reference frames. The vast majority of VO algorithms use feature-based approaches (Klein and Murray, 2007). The image is decomposed into a sparse set of interesting points, with each interest point's location described by a feature vector that remains invariant to camera transformations. The feature vectors are associated with the input frame and reference to form a set of geometric constraints from which we can derive the camera motion and scene structure. In this case, the cost function is formulated by the difference between the measured reprojection locations of these interest points between frames, referred to as the reprojection error. However, some known limitations of feature-based methods include i) extraction of interesting points and feature vectors may be expensive (using well-known algorithms like SHIFT or SURF), ii) they are prone to errors in areas where there is a low number of interesting points, etc. On the contrary, dense or direct VO approaches (Steinbrücker et al., 2011; Whelan et al., 2013) focus on minimizing the (geometric) reprojection error, aiming to directly minimize the photometric error between pixels in the optimization problem. State-of-the-art VO algorithms include Direct Sparse Odometry (DSO) (Engel et al., 2018), ORB-SLAM (Mur-Artal et al., 2015), and ORB-SLAM2 (Mur-Artal and Tardos, 2017).
- The LiDAR sensor provides dense 3D point clouds of the vehicle's surroundings. The goal of LO is to estimate the pose of the vehicle by accumulating the transformation between consecutive frames of 3D point clouds. The existing LO solutions can be divided into two groups: point-wise and feature-wise methods. Point-wise methods estimate the relative transformation directly using the raw 3D points, while feature-wise methods try to utilize more sophisticated characteristics of the point cloud such as the edge and planar feature points. The most well-known pointwise LO method is the iterative close point (ICP) (Besl and McKay, 1992). ICP operates at a point-wise level and directly matches two frames of the point cloud by finding the correspondences. One of the major drawbacks of the ICP is that when the frames include large quantities of points, ICP may suffer from a high computational load arising from the point cloud registration. Many variants of ICP have been proposed to improve its efficiency and accuracy, such as the trimmed ICP (Makihara et al., 2002) and normal ICP (Serafin and Grisetti, 2015). To avoid the high computational load resulting from using the entire set of raw points, the feature-based LO methods extract a set of representative features from the raw points. The fast point feature histogram (FPFH) was proposed by Rusu et al. (2009) to extract and describe important features. The FPFH enables the exploration of the local geometry and the transformation is optimized by matching the one-by-one FPFH-based correspondence. Another well-known feature-based LO method is LOAM (Zhang and Singh, 2014). Theoretically, LOAM integrates the properties of both the point- and feature-wise methods. On the one hand, to decrease the computational load of typical ICP, LOAM proposed to extract two types of feature points, the edge and planar, respectively. The extraction of the feature is simply based on the smoothness of a small region near a given feature point. Different from the FPFH which provides multiple categories of features based on its descriptors, LOAM involves only two feature groups. Another popular variant of LOAM is Lego-LOAM (Shan and Englot, 2018).

Cooperative SLAM approaches are, in general, more immature with respect to CL since they are usually applied in indoor experimental environments with small-scale robots. A thorough

10.3389/frobt.2024.1430740

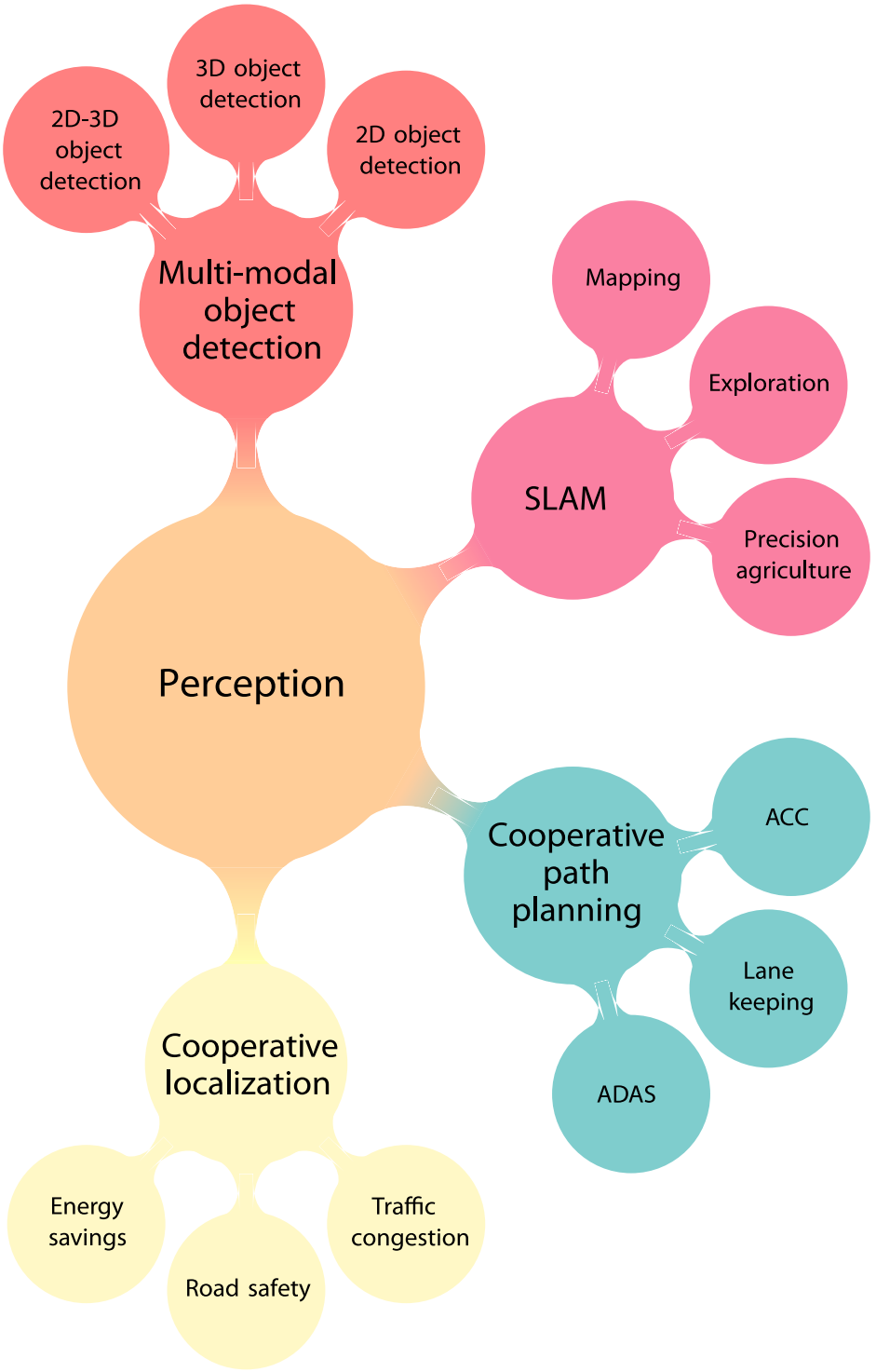

FIGURE 2
Key components of the perception layer in CPSoSs.

overview of multiple-robot SLAM methods is provided by Saeedi et al. (2016), focusing mainly on agents equipped with cameras or 2D LiDARs (Mourikis and Roumeliotis, 2006; Zhou and Roumeliotis, 2008; Estrada et al., 2005). In addition, cooperative SLAM approaches using 3D LiDAR sensors are discussed by Kurazume et al. (2017), Michael et al. (2012), and Nagatani et al. (2011). Table 2 summarizes the above mentioned SLAM-based methods.

TABLE 1 Cooperative localization methods.

Fusion algorithm(s)	Survey	Centralized solution	Distributed solution	Benefits	Limitations	Reference
LS, GD, and SPAWN	—	—	✓	Two state-of-the art algorithms	Large number of iterations and information exchange are required to reach good solutions	Wymeersch et al. (2009)
Particle filter-based belief propagation	—	—	✓	Distributed tracking of mobile nodes and non-cooperative objects	Nodes have to reach a consensus on objects' position	Meyer et al. (2016)
EKF	—	—	✓	Overall location estimation under harsh conditions and realistic network simulation	Lacks evaluation for the individual vehicle	Elazab et al. (2017)
Cubature KF and Huber M-estimation	—	—	✓	Robust location estimation in the presence of measurement outliers	Not considering the impact of dynamic VANET's topology	Liu et al. (2017a)
—	✓	—	—	Complete survey of different fusion algorithms and technologies for CL	—	Buehrer et al. (2018)
—	✓	—	—	Complete survey of different fusion algorithms and technologies for CL, including SLAM methods	—	Kuutti et al. (2018)
Geometric algorithms	✓	—	✓	Linear and distributed approach based on sophisticated selection of neighbors	Developed mainly for static scenarios	Safavi et al. (2018)
Gaussian message passing and KF	—	✓	✓	Distributed CL method relying on the cooperative detection of features	Vehicles have to reach a consensus on features' position	Soatti et al. (2018)
—	✓	—	✓	Detailed book about the current and potential status of CL methods	—	Gao (2019)
Particle filter-based belief propagation	—	—	✓	Distributed data association approach	Vehicles have to reach a consensus on features' position	Brambilla et al. (2020)
Graph Laplacian processing	—	✓	—	Fusion of three measurement modalities via linear LS	No motion model is concerned	Piperigkos et al. (2020b)

(Continued on the following page)

TABLE 1 (*Continued*) Cooperative localization methods.

Fusion algorithm(s)	Survey	Centralized solution	Distributed solution	Benefits	Limitations	Reference
Graph Laplacian processing	—	✓	✓	Fusion of three measurement modalities via linear LS	No motion model is concerned	Piperigkos et al. (2020a)
KF and ML	—	—	✓	Effective and simple implementation of cooperative awareness	Measurement model is rather abstract, not discussing in detail how it can be formulated	Yang et al. (2020a)
Bayesian approach	—	—	✓	Accurate location estimation under harsh conditions	Only ego vehicle location is assessed	Zhao et al. (2020)

TABLE 2 SLAM methods based on VO and LO solutions.

Camera	LiDAR	Benefits	Limitations	Reference
	✓	Fundamental work	High computational load	ICP (Besl and McKay, 1992)
—	✓	Variant of ICP	Improves the computational complexity of ICP	TICP (Makihara et al., 2002)
✓	—	Fundamental feature-based approach	Challenging the extraction of feature points	Klein and Murray (2007)
—	✓	Exploits a set of representative features from the raw point cloud	Lacks evaluation under different weather and lighting conditions	FPFH (Rusu et al., 2009)
✓	—	Directly minimize the photometric error between pixels	Sensitive to image noise	Steinbrücker et al. (2011)
✓	—	Directly minimize the photometric error between pixels	Sensitive to image noise	Whelan et al. (2013)
—	✓	State-of-the-art LO solution	Lacks evaluation under different weather and lighting conditions	LOAM (Zhang and Singh, 2014)
✓	—	State-of-the-art VO solution	Lacks evaluation under different weather and lighting conditions	ORB-SLAM (Mur-Artal et al., 2015)
—	✓	Variant of ICP	Improves the computational complexity of ICP	NICP (Serafin and Grisetti, 2015)
✓	—	State-of-the-art VO solution	Lacks evaluation under different weather and lighting conditions	ORB-SLAM2 (Mur-Artal and Tardos, 2017)
✓	—	State-of-the-art VO solution	Lacks evaluation under different weather and lighting conditions	DSO (Engel et al., 2018)
—	✓	State-of-the-art LO solution	Lacks evaluation under different weather and lighting conditions	LeGO-LOAM (Shan and Englot, 2018)

4.2.3 Cooperative path planning

Connected advanced driver assistance systems (ADASs) help reduce road fatalities and have received considerable attention in research and industrial societies (Uhlemann, 2016). Recently, there has been a shift of focus from individual drive-assist technologies like power steering, anti-lock braking systems (ABS), electronic stability control (ESC), and adaptive cruise control (ACC) to features with a higher level of autonomy like collision avoidance, crash mitigation, autonomous drive, and platooning. More importantly, grouping vehicles into platoons (Halder et al., 2020; Wang et al., 2019a) has received considerable interest since it seems to be a promising strategy for efficient traffic management and road

transportation, offering several benefits in highway and urban driving scenarios related to road safety, highway utility, and fuel economy.

To maintain the cooperative motion of vehicles in a platoon, the vehicles exchange their information with the neighbors using V2V and V2I (Hobert et al., 2015). The advances in V2X communication technology (Hobert et al., 2015; Wang et al., 2019b) enable multiple automated vehicles to communicate with one another, exchanging sensor data, vehicle control parameters, and visually detected objects facilitating the so-called 4D cooperative awareness (e.g., identification/detection of occluded pedestrians, cyclists, or vehicles).

Several works have been proposed for tackling the problems of cooperative path planning. Many of them focus on providing spacing policy schemes using both centralized and decentralized model predictive controllers. However, very few take into account the effect of network delays, which are inevitable and can significantly deteriorate the performance of distributed controllers.

Viana et al. (2019) presented a unified approach to cooperative path-planning using nonlinear model predictive control with soft constraints at the planning layer. The framework additionally accounts for the planned trajectories of other cooperating vehicles, ensuring collision avoidance requirements. Similarly, a multi-vehicle cooperative control system is proposed by Bai et al. (2023) and Kuriki and Namerikawa (2015) with a decentralized control structure, allowing each automated vehicle to conduct path planning and motion control separately. Halder et al. (2020) presented a robust decentralized state-feedback controller in the discrete-time domain for vehicle platoons, considering identical vehicle dynamics with undirected topologies. An extensive study of their performance under random packet drop scenarios is also provided, highlighting their robustness in such conditions. Viana and Aouf (2018) extended decentralized MPC schemes to incorporate the predicted trajectories of human-driving vehicles. Such solutions are expected to enable the co-existence of vehicles supporting various levels of autonomy, ranging from L0 (manual operation) to L5 (fully autonomous operation) (Taeihagh and Lim, 2019). Furthermore, a distributed motion planning approach based on the artificial potential field is proposed by Xie et al. (2022b), where its innovation is related to developing an effective mechanism for safe autonomous overtaking when the platoon consists of autonomous and human-operated vehicles.

In additional to the cooperative path planning mechanisms, spacing policies and controllers have also received increased interest in ensuring collision avoidance by regulating the speeds of the vehicles forming a platoon. Two different types of spacing policies can be found in the literature, i.e., the constant-spacing policy (Liu et al., 2017; Shen et al., 2022) and the constant-time-headway spacing policy (e.g., focusing on maintaining a time gap between vehicles in a platoon resulting in spaces that increase with velocity) (Wang et al., 2023a). In both categories, most works use a one-direction control strategy. At this point, it should be mentioned that in a one-directional strategy, the vehicle controller processes the measurements that are received from leading vehicles. Similarly, a bidirectional platoon control scheme takes into consideration the state of vehicles in front and behind (see Ghasemi et al., 2013). In most of the cooperative platooning approaches, the vehicle platoons are formulated as double-integrator systems that deploy decentralized bidirectional control strategies similar to mass–spring–damper systems. This model is widely deployed since it is capable of characterizing the interaction of the vehicles with uncertain environments and, thus, is more efficient in stabilizing the vehicle platoon system in the presence of modeling errors and measurement noise. However, it should be noted that the effect of network delays on the performance of such systems has not been extensively studied. Time delays, including sensor detective delay, braking delay, and fuel delay, not only seem to be inevitable but are also expected to deteriorate significantly the performance of the distributed controllers. Table 3 summarizes the above mentioned Path planning-based methods.

TABLE 3 Path planning methods.

Cooperative path planning	Spacing controller mechanism	Year	Reference
—	✓	2013	Ghasemi et al. (2013)
✓	—	2015	Kuriki and Namerikawa (2015)
—	✓	2017	Liu et al. (2017)
✓	—	2018	Viana and Aouf (2018)
✓	—	2019	Viana et al. (2019)
✓	—	2019	Taeihagh and Lim (2019)
✓	—	2022	Xie et al. (2022b)
—	✓	2022	Shen et al. (2022)
✓	—	2023	Bai et al. (2023)
—	✓	2023	Wang et al. (2023a)

4.3 Human centric perception

Humans, as a part of a CPSoS, play an important role in the functionality of the system. The humans' role in such complicated systems (e.g., CPSoSs) is vital since they react and collaborate with the machines, providing them with useful feedback and affecting the way that these systems work. Humans can provide valuable input both in an active (on purpose) or a passive (without consideration) way (Figure 3). For example, an input such as a gesture or voice can be used as an order or command to control the operation of a system via an HMI. On the other hand, pose estimation or biometrics, like heart rate, could be taken into account by a decision component, resulting in a corresponding change in the system's functionality for security reasons (e.g., when a user's fatigue has been detected).

The following sections present some human-related inputs (e.g., behavior and characteristics) that can be beneficially used in CPSoSs according to the literature.

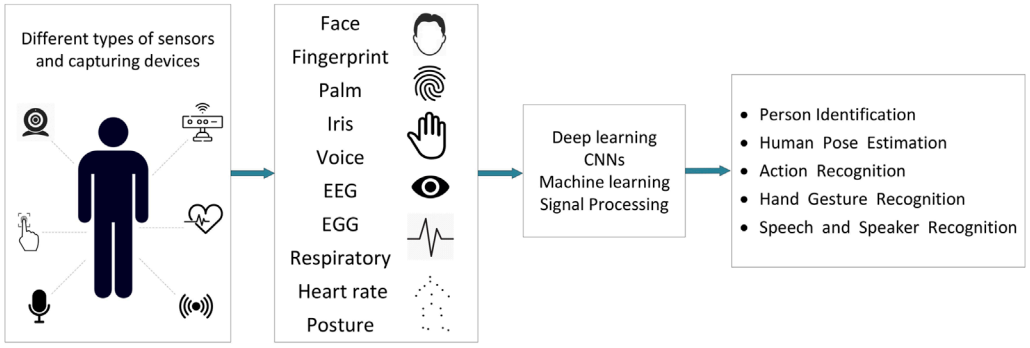

FIGURE 3
Human biometrics and applications in CPSoSs.

- **Biometrics and biometric recognition.** The most well-known and most frequently used biometrics related to humans are face, fingerprint, iris, EEG, EGG, respiratory, and heart rate. Some of them are unique for each person, so they can be used for human identification, while others can be used for monitoring the humans' state or the special cognitive situation of a specific time period. The use of biometrics covers a large variety of tasks and applications in CPSoSs.

Regarding the face of a human as a biometric, the related tasks can be face detection (Claudi et al., 2013; Isern et al., 2020; Chen et al., 2016; Galbally et al., 2019), face alignment (Kaburlasos et al., 2020a; Kaburlasos et al., 2020b), face recognition (Makovetskii et al., 2020), face tracking (Li et al., 2016b; Pereira Passarinho et al., 2015), face classification/verification (Arvanitis et al., 2016), and face landmarks extraction (Jeong et al., 2018; Eskimez et al., 2020; Lee et al., 2020). Fingerprint (Okpara and Bekaroo, 2017; Valdes-Ramirez et al., 2019; Preciozzi et al., 2020), palmprint (Wang et al., 2006), and iris/gaze (Vicente et al., 2015; Lai et al., 2015) are mainly used for user's identification tasks due to their uniqueness for each person. EEG (Laghari et al., 2018; Pandey et al., 2020; Lou et al., 2017), EGG, respiratory (Jiang et al., 2020; Saatci and Saatci, 2021; Meng et al., 2020), and heart rate (Rao et al., 2012; Prado et al., 2018; Majumder et al., 2019) are used for the user's state monitoring. In addition to the fact that they can provide valuable information, their usage in real applications is difficult to apply due to the special wearable devices that it is required for the capturing.

The choice of which specific biometric to utilize depends on the use case scenario, including factors such as the availability and feasibility of using a sensor (e.g., whether it will be placed in a stationary location or worn constantly during the operation), the special power consumption needs of each sensor, the accuracy, and the latency. One other important issue that needs to be taken under serious consideration before the use of a biometric in real systems is the privacy and security of these sensitive data since they must be protected via encoding in order to be anonymously stored or used.

- **Person identification.** Person identification is a common image retrieval problem, where the objective is the recognition of a person's identity using only a single image captured by a camera.

Generally, the person identification task is a more complicated and challenging problem in comparison to the face identification since face identification is applied in a more controlled environment (e.g., use of a smaller captured frame, the user has to remove glasses, hats, and other accessories to be identified). On the other hand, person identification has to deal with more complex issues like the different points of view, light and weather conditions, different resolutions of the camera, types of clothes, and a large variety of background contexts. Person identification has shown great usability in applications related to CPSoSs, mostly for security purposes. Its utility has been marked specifically when it is applied "in the wild" and in uncontrolled environments where other biometrics are not feasible to be used due to technical constraints. Nowadays, approaches usually use deep networks to perform reliable and accurate results.

Li et al. (2020a) proposed an additive distance constraint approach with similar label loss to learn highly discriminative features for person re-identification. Ye and Yuen (2020) proposed a deep model (PurifyNet) to address the issue of the person re-identification task with label noise, which has limited annotated samples for each identity. Li et al. (2020b) used an unsupervised re-identification deep learning approach capable of incrementally discovering discriminative information from automatically generated person tracklet data. Table 4 and Table 5 summarize relevant datasets for the face recognition, detection, and facial landmark extraction problems.

- **Human pose estimation and action recognition.**

Human pose estimation and action recognition have been proved to be particularly valuable tasks in modern video-captured applications related to CPSoSs. They can be utilized in a variety of fields such as ergonomics assessment, safe training of new operators, fatigue and drowsiness detection of the user, HMIs, and the prediction of an operator's next action for avoiding accidents by changing the operation of a machine. They are also useful for monitoring dangerous movements in insecure workspace areas.

A restriction that can negatively affect and obstruct the quality of the results of these tasks is the limited coverage area of the camera. Nevertheless, this limitation can be overcome using new types of sensors and tools like IMUs and whole-body tracking systems (e.g., SmartsuitPro and Xsens) (Hu et al., 2017).

TABLE 4 Datasets for face recognition, detection, and facial landmark extraction tasks.

Dataset	Short description	Link of the dataset	Paper name
Helen	Helen dataset consists of 2,330 images (400 × 400 pixels) with labeled facial components, which are manually annotated, containing contours near the eyes, eyebrows, nose, lips, and jawline	http://www.ifp.illinois.edu/~vuongle2/helen/	Interactive Facial Feature Localization (Le et al., 2012)
AFW	AFW (Annotated Faces in the Wild) is a face detection dataset consisting of 205 images with 468 faces. Each face image is labeled with at most 6 landmarks with visibility labels, as well as a bounding box	https://www.ics.uci.edu/~xzhu/face/	Face detection, pose estimation, and landmark localization in the wild (Zhu and Ramanan, 2012)
300W	The 300-W dataset consists of 300 indoor and 300 outdoor "in the wild" images, covering a large variety of identities, expressions, illumination conditions, poses, occlusion, and face sizes	https://ibug.doc.ic.ac.uk/resources/300-W/	300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge (Sagonas et al., 2013)
LFPW	The Labeled Face Parts in the Wild (LFPW) consists of 1,432 faces from images which are downloaded from the web (e.g., google.com, flickr.com, and yahoo.com)	https://neerajkumar.org/databases/lfpw/	Localizing parts of faces using a consensus of exemplars (Belhumeur et al., 2011)
AFLW	The Annotated Facial Landmarks in the Wild (AFLW) consists of 25,000 faces that are annotated with up to 21 landmarks per image. The images have been gathered from Flickr, covering a large variety of poses, expressions, ethnicities, ages, genders, and environmental conditions	https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/aflw/	Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization (Köstinger et al., 2011)
AFLW 2000-3D	AFLW 2000-3D dataset consists of 2,000 images that have been annotated using 68 points representing 3D facial landmarks. This dataset is usually used for the evaluation of 3D facial landmark detection models	http://www.cbsr.ia.ac.cn/users/xiangyuzhu/projects/3DDFA/main.htm	Face Alignment Across Large Poses: A 3D Solution (Zhu et al., 2016)
300-VW	300 Videos in the Wild (300-VW) is a dataset for evaluating facial landmark tracking algorithms in the wild. Each video of this dataset is almost 1 min in duration (at 25–30 fps). Each frame of all videos has been annotated in the same way as the 300-W dataset	https://ibug.doc.ic.ac.uk/resources/300-VW/	Offline Deformable Face Tracking in Arbitrary Videos (Chrysos et al., 2015)
COCO-WholeBody	This dataset is an extension of COCO dataset, covering a whole-body annotation (i.e., face, hand, and feet)	https://github.com/jin-s13/COCO-WholeBody	Whole-Body Human Pose Estimation in the Wild (Jin et al., 2020)
MALF	MALF consists of 5,250 images with 11,931 faces in total. This dataset is the first face detection dataset that supports fine-gained evaluation	http://www.cbsr.ia.ac.cn/faceevaluation/	Fine-grained Evaluation on Face Detection in the Wild (Yang et al., 2015)
FDDB	FDDB dataset consists of 2,845 images with 5,171 annotated faces	http://vis-www.cs.umass.edu/fddb/index.html	FDDB: A Benchmark for Face Detection in Unconstrained Settings (Jain and Learned-Miller, 2010)

TABLE 5 Datasets of images with the iris.

Dataset	Short description	Link of the dataset	Paper name
UBIRIS.v2	The UBIRIS.v2 dataset consists of 11,102 images of the iris that were captured from 261 subjects, with 10 images for each subject. The images were acquired using a variety of different conditions like distance, motion, and different visible wavelengths. They have also been affected by real noise	http://iris.di.ubi.pt/ubiris2.html	The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-the-Move and At-a-Distance (Proenca et al., 2010)
OpenEDS	Open Eye Dataset (OpenEDS) consists of images with eyes captured using a virtual-reality head display. This dataset was collected from 152 individual participants and is divided into four subsets	https://research.fb.com/programs/	OpenEDS: Open Eye Dataset (Garbin et al., 2019)

Islam et al. (2019) presented an approach that exploits visual cues from human pose to solve industrial scenarios for safety applications in CPSs. El-Ghaish et al. (2018) integrated three modalities (i.e., 3D skeletons, body part images, and motion history images) into a hybrid deep learning architecture for human action recognition. Nikolov et al. (2018) proposed a skeleton-based approach utilizing spatio-temporal information and CNNs for the classification of human activities. Deniz et al. (2020) presented an indoor monitoring reconfigurable CPS that uses embedded local nodes (Nvidia Jetson TX2), proposing learning architectures to address human action recognition. Table 6 and Table 7 summarize relevant datasets for the pose estimation and action recognition problem.

- **Hand gesture recognition.**

Hand gesture recognition tasks can be a very useful tool for interactions with machines or subsystems in CPSoSs (Horváth and Erdős, 2017), particularly in applications where the user is not allowed to have physical hand contact with a machine due to security reasons. This task mainly consists of three sequential steps, which are hand detection, hand tracking, and gesture recognition. Gesture recognition can occur either by a single image (i.e., static gesture recognition) or a sequence of images (i.e., dynamic gesture recognition). The first strategy looks more like a retrieval problem where the gesture of the image has to match with a known predefined gesture from a dataset of gestures. The second is a more complicated problem, but it is more useful since it can cover the requirements of a wider variety of real-life use cases (Choi and Kim, 2017). Gesture recognition is a very common task in human–computer interaction. Nonetheless, the recognition of complex patterns demands accurate sensors and sufficient computational power (Grützmacher et al., 2016). Additionally, we have to refer to the fact that visual computing plays an important role in CPSoSs, especially in these applications where the visual gesture recognition system relies on multi-sensor measurements (Posada et al., 2015; Aviles-Arriaga et al., 2006).

Horváth and Erdős (2017) presented a control interface for cyber–physical systems that interprets and executes commands in a human-robot shared workspace using a gesture recognition approach. Lou et al. (2016) tried to address the problem of personalized gesture recognition for cyber–physical environments, proposing an event-driven service-oriented framework. However, in other gesture recognition applications, a body-worn setup was proposed, which supplements the omnipresent 3 DoF motion sensors with a set of ultrasound transceivers (Putz et al., 2020). Table 8 summarizes relevant datasets for the hand and gesture recognition problems.

- **Speech and speaker recognition.**

Speech recognition is a sub-category of a more generic research area related to the domain of natural language processing (NLP). The main objective of speech recognition is to automatically translate the content of the entire speech (or the most significant part of it) into text or other recognizable forms from the computers. Assuming that the recording and processing of speech do not require a special sensor, but just a simple audio recorder, we can understand how easy to use this information is. Additionally, speech can be applied without any physical contact interaction, making it an ideal signal for HMI applications.

Speech recognition tasks can be utilized in the smart input system (Wang, 2020; Han et al., 2016), automatic transcription system (Chaloupka et al., 2012; Chaloupka et al., 2015), smart voice assistant (Subhash et al., 2020), computer-assisted speech (Tejedor-García et al., 2020), rehabilitation (Aishwarya et al., 2018; Mariya Celin et al., 2019), and language teaching.

Similar to the face recognition task that focuses on the recognition of an individual human using facial information, the speaker recognition task tries to achieve the same goal using the vocal tone information of the subject. Speaker recognition is one of the most basic components for human identification, which has various applications in many CPSoSs. Additionally, fusion schemes can be used combining both speaker recognition and face recognition for more secure integrations (Zhang and Tao, 2020).

TABLE 6 Datasets for pose estimation.

Dataset	Short description	Link of the dataset	Paper name
COCO	The Microsoft Common Objects in Context (MS COCO) consists of 328,000 images. This dataset is a general-proposed, large-scale object detection, segmentation, key-point detection, and captioning dataset containing labeled human's poses	https://cocodataset.org/	Microsoft COCO: Common Objects in Context (Lin et al., 2014)
MPII	The MPII Human Pose Dataset consist of 25,000 images, of which 15,000 images are training samples, 3,000 images are validation samples, and the remaining 7,000 images are testing samples. The single-person poses are manually annotated with up to 16 body joints. The images are taken from YouTube videos, covering 410 different human activities	http://human-pose.mpi-inf.mpg.de/	2D Human Pose Estimation: New Benchmark and State of the Art Analysis (Andriluka et al., 2014)
DensePose	DensePose-COCO is a large-scale ground-truth dataset with image-to-surface correspondences, which are manually annotated from 50,000 images of the COCO dataset and train DensePose-RCNN, to densely regress part-specific UV coordinates within every human region at multiple frames per second	http://densepose.org/	DensePose: Dense Human Pose Estimation in the Wild (Güler et al., 2018)
LSP	The Leeds Sports Pose (LSP) dataset consists of 2,000 images of sportspersons in total gathered from Flickr, 1,000 for training and 1,000 for testing. This dataset is used for human pose estimation, and each image is annotated with 14 joint locations	https://dbcollection.readthedocs.io/en/latest/datasets/leeds_sports_pose_extended.html	Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation (Johnson and Everingham, 2010)
JHMDB	JHMDB is a recognition dataset that consists of 960 video sequences belonging to 21 actions. This dataset is a subset of the larger HMDB51 dataset, which has been collected from digitized movies and YouTube videos	http://jhmdb.is.tue.mpg.de/	Towards Understanding Action Recognition (Jhuang et al., 2013)
Unite the People	Unite The People dataset is mainly used for 3D body estimation. The images come from an extended version of the LSP dataset, as well as the single person-tagged people from the MPII Human Pose Dataset. The images are labeled with different types of annotations such as segmentation labels, poses, or 3D representation	https://files.is.tuebingen.mpg.de/classner/up/	Unite the People: Closing the Loop Between 3D and 2D Human Representations (Lassner et al., 2017)

A speaker recognition system consists of three separate parts, namely, the speech acquisition module, the feature extraction and selection module, and finally the pattern matching and classification module. In CPSoSs, the implementation of an automatic speech recognition system relies on a voice user interface so that humans can interact with robots or other CPS components. Nevertheless, this type of interface cannot replace the classical GUIs, but it can intensify them by providing, in some cases, a more efficient way of interaction.

Kozhirbayev et al. (2018) developed a technique to train a neural network (NN) on the extracted mel-frequency cepstral coefficient (MFCC) features from audio samples to increase the recognition accuracy of the short utterance speaker recognition system. Wang et al. (2020) tried to improve the robustness of speaker identification using a stacked sparse denoising auto-encoder. Table 9 summarizes relevant datasets for the speech recognition problem.

TABLE 7 Datasets of action recognition.

Dataset	Short description	Link of the dataset	Paper name
UCF101	This dataset consists of 13,320 video clips (~ 27 h) from Youtube, classified into 101 categories and into 5 types (i.e., body motion, human–human interactions, human–object interactions, playing musical instruments, and sports)	https://www.crcv.ucf.edu/data/UCF101.php	UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild (Soomro et al., 2012)
Kinetics	It is a high-quality dataset of videos used for human action recognition. The dataset consists of approximately 500,000 labeled video clips of 10 s, covering 600 human action classes with at least 600 video clips for each action class	https://deepmind.com/research/open-source/kinetics	The Kinetics Human Action Video Dataset (Kay et al., 2017)
HMDB51	The HMDB51 is a dataset consisting of 6,849 video clips from 51 action categories (such as "jump," "kiss," and "laugh"). Each category containing at least 101 clips	https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/	HMDB: A large video database for human motion recognition (Kuehne et al., 2011)
ActivityNet	The ActivityNet contains 200 different types of activities and a total of 849 h of videos collected from YouTube. It is one of the largest datasets based on the number of activity categories and videos	http://activity-net.org/	ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding (Heilbron et al., 2015)
NTU RGB + D	NTU RGB + D consists of 56,880 video clips of 60 action classes collected from 40 subjects. The actions can be generally divided into 3 categories: 40 daily actions (e.g., drinking, eating, and reading), 9 health-related actions (e.g., sneezing, staggering, and falling down), and 11 mutual actions (e.g., punching, kicking, and hugging)	http://rose1.ntu.edu.sg/datasets/actionrecognition.asp	NTU RGB + D: A Large Scale Dataset for 3D Human Activity Analysis (Shahroudy et al., 2016)
KTH	The KTH dataset contains six actions: walk, jog, run, box, hand-wave, and hand clap by 25 different individuals in different environments: outdoor (s1), outdoor with scale variation (s2), outdoor with different clothes (s3), and indoor (s4)	https://www.csc.kth.se/cvap/actions/	Recognizing Human Actions: A Local SVM Approach (Schuldt et al., 2004)
Composable activity dataset	This dataset consists of 693 annotated videos of activities in 16 classes performed by 14 individuals	https://ialillo.sitios.ing.uc.cl/ActionsCVPR2014/	Discriminative Hierarchical Modeling of Spatio-Temporally Composable Human Activities (Lillo et al., 2014)
HACS	HACS dataset contains 504 K videos (shorter than 4 min) collected from YouTube, categorized in 200 action classes. It is used in human action recognition	http://hacs.csail.mit.edu/	HACS: Human Action Clips and Segments Dataset for Recognition and Temporal Localization (Zhao et al., 2019a)

5 Behavioral layer

In each CPSoS, human knowledge, senses, and expertise constitute important informative values that can be taken into account for the assurance of its operational excellence. However, a substantial concern that needs to be addressed at an early age of CPSoS evolution is determining how these abstract human features can be made accessible and understandable to the system.

A way to integrate the human as a separate component into a CPSoS is by introducing an anthropocentric mechanism, which is known in the literature as the HITL approach (Gaham et al., 2015; Hadorn et al., 2016). This mechanism allows a direct way for humans to continuously interact with the CPSoSs'

TABLE 8 Datasets for hand and gesture recognition.

Dataset	Short description	Link of the dataset	Paper name
HandNet	The HandNet dataset contains the depth images of 10 participants' hands non-rigidly deforming in front of a RealSense RGB-D camera. The annotations were generated using a magnetic annotation technique. 6D pose is available for the center of the hand and the five fingertips (i.e., position and orientation of each)	http://www.cs.technion.ac.il/~twerd/HandNet/	Rule of thumb: Deep derotation for improved fingertip detection (Wetzler et al., 2015)
EgoGesture	The EgoGesture dataset consists of 2,081 RGB-D videos, 24,161 gesture samples, and 2,953,224 frames from 50 distinct subjects	http://www.nlpr.ia.ac.cn/iva/yfzhang/datasets/egogesture.html	EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition (Zhang et al., 2018)
NVGesture	The NVGesture dataset consists of 1,532 dynamic gestures categorized into 25 classes. The dataset is separated into 1,050 samples for training and 482 for testing. The application in which it can be used is for touchless driver controlling	https://ieeexplore.ieee.org/document/7780825	Online Detection and Classification of Dynamic Hand Gestures With Recurrent 3D Convolutional Neural Network (Molchanov et al., 2016)
IPN Hand	The IPN Hand is a dataset consisting of videos with sufficient size, variation, and real-world elements capable to be used by deep neural networks for training and evaluation. The application on which this dataset focuses is dynamic hand gesture recognition	https://github.com/GibranBenitez/IPN-hand	Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks (Köpüklü et al., 2019)
MLGEST-URE	MlGesture consists of more than 1,300 hand gesture videos from 24 participants and features 9 different hand gesture symbols. The dataset has been recorded in a car with five different sensor types at two different viewpoints, and it can be used for hand gesture recognition tasks	https://iiw.kuleuven.be/onderzoek/eavise/mlgesture/home	Low-latency hand gesture recognition with a low resolution thermal imager (Vandersteegen et al., 2020)

control loops in both directions of the system (i.e., taking and giving inputs).

Although common CPSoSs are human-centered systems (where human constitutes an essential part of the system), unfortunately, in many real cases, these systems still consider humans as external and unpredictable elements without taking their importance into deeper consideration. The central vision of the researchers and engineers is to create a human–machine symbiosis, integrating humans as holistic beings within CPSoSs. In this way, CPSoSs have to support a tight bond with the human element through HITL controls, taking into account human features like intentions, psychological and cognitive states, emotions, and actions, all of which can be deduced through sensor data and signal processing approaches.

HITL systems integrate human feedback into the control loop of a system, allowing humans to interact with and influence automated processes in real-time. In self-driving vehicles, haptic teleoperation enables remote operators to control the vehicle with the sensation of touch. For instance, when the vehicle encounters an abnormal situation, a human operator can take over using HMI haptic feedback to feel the road conditions and obstacles, ensuring safe navigation and improving response times and overall safety (Kuru, 2021). Additionally, HITL telemanipulation of unmanned aerial vehicles (UAVs) allows operators to control drones remotely while receiving haptic feedback about the drone's interactions with its environment. For instance, an operator uses a haptic interface to control the UAV. The haptic feedback provides sensations of wind resistance, surface textures, and physical interactions with obstacles (Zhang et al., 2021). Integrating haptic feedback into HITL operations enhances human perception by providing a multisensory experience. This integration improves the realism of virtual environments and the accuracy and efficiency of tasks, requiring fine motor skills and precise control. Haptic feedback bridges the gap between the virtual and physical worlds, allowing users to interact with digital systems more naturally and intuitively. Thus, the concept of human–machine symbiosis refers to the synergistic relationship between humans and machines, where both entities work together to achieve common goals. This cooperation ensures that the unique strengths of humans (such as intuition, creativity, and decision-making) complement the capabilities of

TABLE 9 Datasets for speech recognition.

Dataset	Short description	Link of the dataset	Paper name
LibriSpeech	This dataset consist of approximately 1,000 h of audiobooks	http://www.openslr.org/12	LibriSpeech: An ASR corpus based on public domain audio books (Panayotov et al., 2015)
Speech Commands	Speech Commands consists of 65,000 of 30 short words ~, one second long. It is a collection of spoken words by thousands of different people, designed for the training and evaluation of keyword spotting systems	https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html	Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition (Warden, 2018)
MuST-C	MuST-C currently represents the largest publicly available multilingual corpus for speech translation from English into several languages. It covers eight languages. It consists of hundred hours of audio recordings from English TED Talks	https://ict.fbk.eu/must-c/	MuST-C: A multilingual corpus for end-to-end speech translation (Cattoni et al., 2021)
Common Voice	Common Voice is a dataset of 9,283 recorded hours that consists of audio files and corresponding text files including demographic metadata like age, sex, and accent	https://commonvoice.mozilla.org/en/datasets	Common Voice: A Massively-Multilingual Speech Corpus (Ardila et al., 2019)
Libri-Light	Libri-Light is a collection of over 60 K hours of spoken English suitable for training speech recognition systems under limited or no supervision	https://github.com/facebookresearch/libri-light	Libri-Light: A Benchmark for ASR with Limited or No Supervision (Kahn et al., 2020)
THCHS-30	THCHS-30 is a free Chinese speech database that can be used for speech recognition systems	http://166.111.134.19:7777/data/thchs30/README.html	THCHS-30: A Free Chinese Speech Corpus (Wang and Zhang, 2015)
VOICES	This dataset consists of speech recorded by far-field microphones in noisy room conditions for using in speech and signal processing approaches	https://registry.opendata.aws/lab41-sri-voices/	Voices Obscured in Complex Environmental Settings (VOICES) corpus (Richey et al., 2018)
LibriCSS	LibriCSS is a real recorded dataset that simulates conversations that are captured by far-field microphones	https://github.com/chenzhuo1011/libri_css	Continuous speech separation: dataset and analysis (Chen et al., 2020)
SPEECH-COCO	SPEECH-COCO contains 616,767 audios generated using text-to-speech (TTS) synthesis. The audio files are paired with images	https://zenodo.org/record/4282267	PEECH-COCO: 600k Visually Grounded Spoken Captions Aligned to MSCOCO Data Set (Havard et al., 2017)

machines (such as speed, accuracy, and endurance). By creating environments where humans and machines work together, the overall system performance can be enhanced. This human–machine symbiosis ensures that both parties contribute to the task, leading to increased productivity and innovation. The key components of behavioral layer are summarized in Figure 4.

System designers, who design and develop new generations of CPSoSs, have to understand and realize which features differentiate them from the traditional CPSs. One of these features is the HITL mechanism that allows CPSoSs to take advantage of some unique human characteristics, making them superior to the machines. The technological assessments are not mature yet to integrate these human-oriented characteristics into machines and robots. As a result, the HITL approach is essential to serve the initial goals of a CPSoS. These characteristics, as have been proposed by the literature (Sowe et al., 2016), are presented below:

- **Cognition.** Humans have a different way of observing a situation than computers do. First, they understand a problem and then make final decisions, even with missing data. Human cognition is the combined result of knowledge, experience, inspiration, and intuition, areas where no current machine can overcome or even approach in some way.
- **Predictability.** Humans are not preordained to perform the same task in the same way every time that they try. This would be a problem in some cases, especially when they have to follow

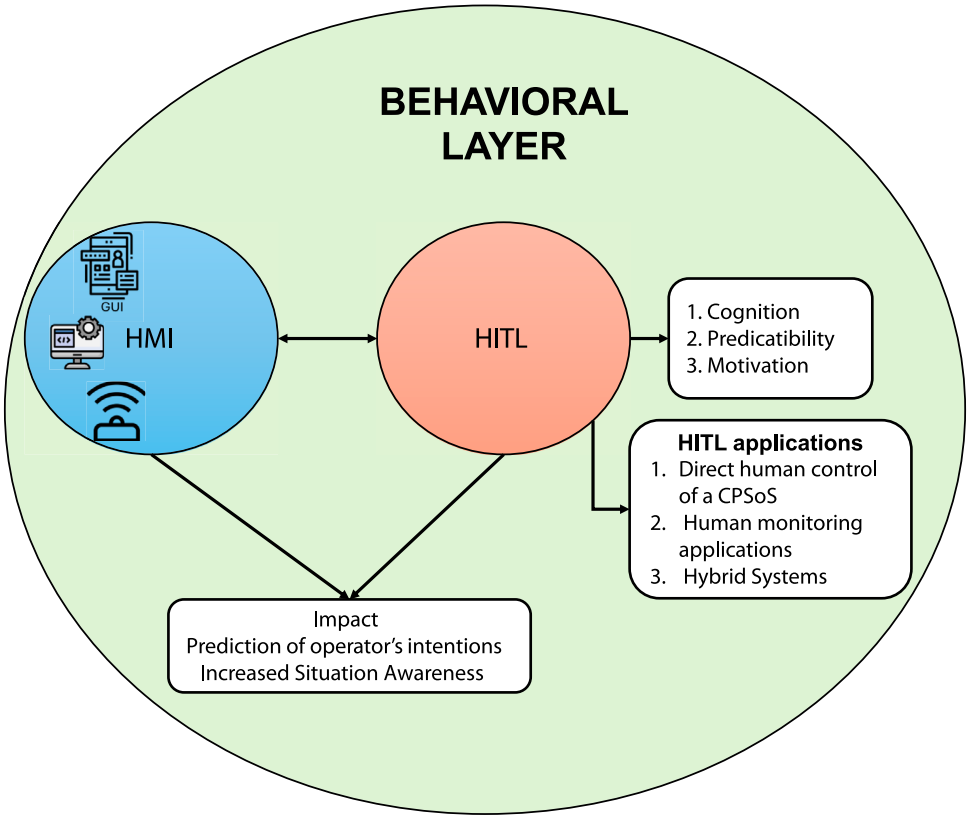

FIGURE 4
Key components of the behavioral layer in CPSoSs.

precise instructions. This feature might make them less reliable than a simple computer. However, this unpredictable behavior could be beneficial in a critical situation that has not been distinctly defined in the script of the instructions. The ability of humans to easily adapt to unknown situations makes them a perfect component to provide out-of-the-box solutions in hazardous circumstances.

- **Motivation.** Humans, by their nature, usually require incentives and become more productive when they are assured. Motivation can guide a human to perform more effort on a task than is required. On the other hand, computers and machines follow a particular pipeline of work, and they cannot change the way they perform a task to enhance their productivity.

The HITL applications can be separated into three main categories with respect to the type of input that humans provide:

1. Applications in which the human plays a leading role and directly controls the functionality of the CPSoS as an operator (Figure 5A).
2. Applications where the system passively (Figure 5B) monitors humans (e.g., biometrics, pose) and then makes decisions for appropriate actions.
3. Hybrid combination of the two types mentioned above (Figure 5C).

5.1 Direct human control of a CPSoS

The applications in this category can be separated into two different sub-categories related to CPS autonomy. In the first sub-category, operators manage a process close to an autonomous task. The system has complete control of its action, and the user is responsible for adjusting some parameters that may affect the functionality of the system when required for external reasons. An example that can describe a scenario like this is when an operator sets new values to specific parameters on a machine in the industry for changing the operation of the assembly line (e.g., for a new product).

In the second sub-category, the operator plays a more active role in the process by directly controlling several tasks and setting explicit commands for the operation of the machines or robots. An example of this scenario is when an expert operator has to remotely take complete control of a robotic arm for repair purposes.

5.2 Human monitoring applications

The applications of this category are represented by systems that passively monitor humans' actions, behavior, and biometrics. The acquired data can be used to make appropriate decisions. Based on the type of the system reaction, the applications can be separated into two types, namely, open-loop and closed-loop systems.

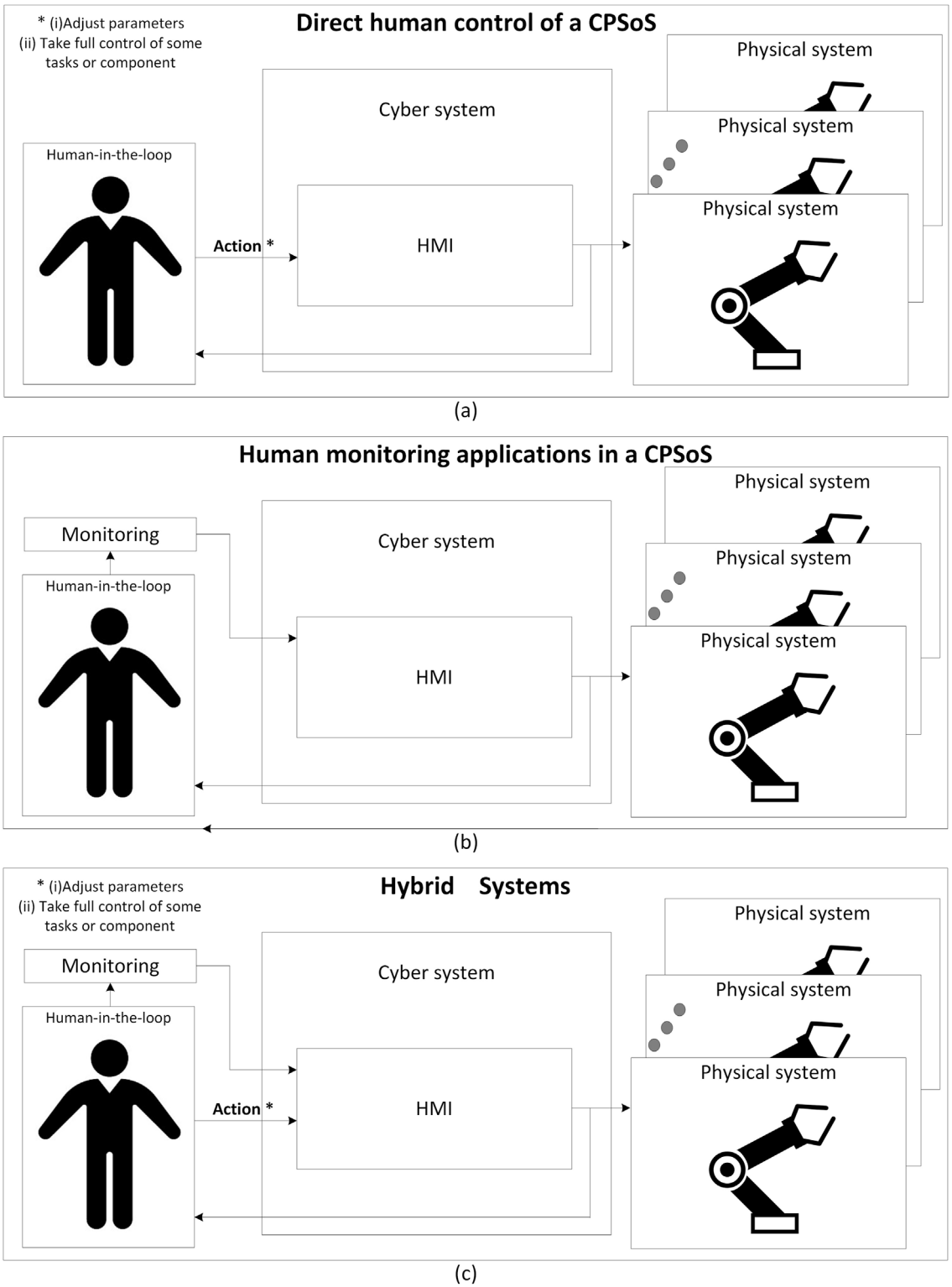

FIGURE 5
HITL applications with respect to the type of input that humans provide. **(A)**applications in which human plays a leading role of the CPSoS, **(B)**Applications where the system passively monitors humans and then makes decisions for appropriate actions. and **(C)**Hybrid combination of the two types mentioned above.

Open-loop systems continuously monitor humans and visualize (e.g., smart glasses) or send a report with relevant results, which may be helpful or interesting for the operator. The system does not take any further action in this case. The presented results can cover (i) the first level of information, (ii) the second level of information, and (iii) KPIs. First-level information includes measurements that usually are directly received by the sensors (e.g., heart rate, respiratory, and blinking of eyes). After the appropriate process of the first level of information, the second level corresponds to a higher contextual meaning (e.g., drowsiness, awareness, and anxiety level).

Closed-loop systems use the received information from the sensors and the processing results to take action. For example, in an automotive use case, if critical drowsiness of a driver is detected, the car could take complete control of the vehicle or appropriately inform the driver of his/her condition.

5.3 Hybrid systems

In manufacturing, for example, a system monitors the operators' actions while collaborating with a robot, and it can provide appropriate guided instructions. However, the level of detail of the guided assistant can be modified by the personalized preferences of the user (which may be related to the level of his/her experience, for instance). Hybrid systems take human-centric sensing information as feedback to perform an open or closed-loop action, but additionally, they also take into account the direct human inputs and preferences.

Humans play an essential role in CPSoSs. Their contribution can be summarized into three categories: (i) for data acquisition, (ii) for inference related to their state, and (iii) for the actuation of an action to complete a task of their own or to collaborate with other components of the system (Nunes et al., 2015).

1. Data acquisition:
   - Human as an informer. Humans provide the system with information through wearable sensors or other devices/sensors that monitor them.
   - Human as a communicator to transfer condensed knowledge. Humans have the special ability to easily understand complicated information, draw conclusions, and transfer filtered, useful, and deductive information to the system.
2. State inference:
   - Human as an insider component. Training algorithms and machine learning approaches can be used to recognize the human's state (e.g., cognitive, physical, emotional, and physiological), which may affect the excellent functionality of a CPSoS or put the user's safety at risk. When a critical user state is identified, the system can change the typical operation to protect the user or notify them with an appropriate message or warning.
   - Human as a feedback component. Based on the state of the users, the system may provide suggestions or recommendations to them. The acceptance of these suggestions by the users can further be utilized by the system as useful feedback, providing more personalized solutions in future similar situations.
3. Actuation:
   - Human as actuators. The actions of a human, as a part of a CPSoS, are (i) to set the values of some parameters, (ii) to execute specific tasks, or (iii) to take total control of the system, if required.

HMI is referred to as the medium that is utilized for direct communication between humans and machines, facilitating their physical interaction (Ajoudani et al., 2018). In the literature, when there are multi-human users or systems of machines instead of a separate individual machine, HMIs have also been mentioned as cooperative or collaborative HMIs (CHMIs). Nevertheless, for simplicity, they will be referred to as HMIs since the way they function and their primary features remain the same regardless of the number of humans or machines.

Typically, the classical HMI system comprises some standard hardware components, like a screen and keyboard, along with software featuring specialized functionalities, effectively performing as a graphical user interface (GUI). All sensors and wearable devices connected with humans or other components of the CPSoS are also part of an HMI. The HMI has pervasive usage in CPSoSs, allowing each part of the CPSoS to directly interact with a human and vice versa, creating a synergy loop between CPSs and humans. In the future, HMIs will also have social cohesion between humans and machines. Gorecky et al. (2014) suggested that the primary representatives of HMI tools in CPSoSs, which are mainly used for the communication between humans and machines, are automatic speech recognition, gesture recognition, and extended reality, which can be represented by augmented or virtual reality. In such implementations, a touch screen can allow human operators to pass messages to machines. Singh and Mahmoud (2017) presented a framework that is capable of visually acquiring information from HMIs in order to detect and prevent HITL errors in the control rooms of nuclear power plants. The intelligent and adaptable CPSoSs expect the automation systems to be decentralized and support "Plug-and-Produce" features. In this way, the HMIs have to dynamically update and adapt display screens and support elements to facilitate the work of the operators, like IO fields and buttons (Shakil and Zoitl, 2020). Wang et al. (2019a) proposed a graphical HMI mechanism for intelligent and connected in-vehicle systems in order to offer a better experience to automotive users. However, Pedersen et al. (2017) presented a way of connecting an HMI with a software model of an embedded control system and thermodynamic models in a hybrid co-simulation.

Naujoks et al. (2017) presented an HMI for cooperative automated driving. The cooperative perception extends the capabilities of the automated vehicles by performing tactics and strategic maneuvers independently of any driver's intervention (e.g., avoiding obstacles). The goal of the papers presented by Kharchenko et al. (2014), Orekhov et al. (2016a), and Orekhov et al. (2016b) is to increase the drivers' awareness through the development and implementation of cooperative HMIs for ITSs based on cloud computing, providing measurements of vehicle and driver's state in real-time. Johannsen (1997) dealt with HMIs for cooperative supervision and control by different human users (e.g., in control rooms or group meetings). The application in various domains (i.e., cement industry and chemical and power plants) has shown that several persons from different classes (e.g., operators, maintenance personnel, and engineers) need to cooperate.

By integrating haptic technologies into HMIs, human perception and interaction can be significantly enhanced, leading to more effective and efficient HITL operations. Haptic teleoperation allows operators to control remote or virtual systems with the sensation of touch, providing real-time tactile feedback from the remote environment. This feedback increases the operator's situational awareness and precision by mimicking the sense of touch, thereby enhancing the user's understanding and control (Cheng et al., 2022). In automotive applications, for instance, haptic feedback in steering wheels or pedals can alert drivers to hazards or guide them through automated driving tasks, thus improving response times and safety (Quintal and Lima, 2021). Additionally, techniques like haptic physical coupling can create a physical connection between the user and the system, enabling more intuitive and direct manipulation of virtual objects or remote devices. This can simulate the weight, texture, and resistance of objects in virtual environments, which is particularly beneficial in training simulations and remote manipulations, where direct human intervention is risky or impossible. For instance, it can be used in simulations for hazardous environments, such as nuclear plants or space missions. Operators can train with realistic tactile feedback, preparing them for real-world scenarios without the associated risks (Laffan et al., 2020). Incorporating haptic feedback into HITL operations can also improve human perception by providing a multisensory experience. This integration not only enhances the realism of virtual environments but also improves the accuracy and efficiency of tasks that require fine motor skills and precise control (Sagaya Aurelia, 2019).

Cooperative HMIs can be used as standalone GUI in aircraft guidance applications, allowing data and other types of graphical information (e.g., routes, route attributes, airspaces, and flight plan tracks). This real-time information enables collaborative decision-making between all associates, such as the crew (the pilot and the copilot), who have to cooperate continuously or interact with air traffic controllers. HMIs can facilitate the users in operational services, such as air traffic flow and capacity management, flight planning, and airspace management. Kraft et al. (2020) and Kraft et al. (2019) investigated the type of information that should be provided to drivers via HMIs in merging or turning left situations to support cooperative driving, facilitating each other's goal achievement. Cooperation between road users utilizing V2X communication has the potential to make road traffic safer and more efficient (Fank et al., 2021). The exchange of information enables the cooperative orchestration of crucial traffic conditions, like truck overtaking maneuvers on freeways.

In addition to cooperative decision-making facilitated by HMIs, the integration of predictive capabilities and situational awareness is paramount in enhancing the functionality and safety of CPSoSs. **Prediction of operator's intentions** and **situation awareness** are critical aspects of integrating human capabilities within CPSoSs. By understanding and anticipating human actions and maintaining a high level of awareness, these systems can significantly enhance operational safety, efficiency, and resilience. These concepts further solidify the human–machine symbiosis, ensuring that CPSoSs can effectively adapt to dynamic and unpredictable environments.

Prediction of operator's intentions is a task that can improve the effectiveness of collaboration between CPSoSs and humans. An accurate prediction can be essential, especially in industrial scenarios where the resilience and safety of all CPSoS components mainly depend on the mutual understanding between humans and CPSoSs. So, it seems necessary to design and develop reliable, robust, and accurate human behavior modeling techniques capable of predicting human actions or behavior.

On the one hand, operators are mainly responsible for their safety when they are in the same working environment with a cobot, performing collaborative tasks. However, on the other hand, CPSoSs must have intelligent components that can identify, understand, and even predict operators' intentions with the primary goal of protecting them from severe injury. A continuous video-capturing component can be used by the prediction system to detect, track, and recognize human gestures or postures, and an artificial intelligence component can be used to predict human intentions. The system can anticipate when unexpected human operations have been detected, or specific human activity patterns have been predicted (Zanchettin et al., 2019). Meanwhile, the cobot can perform other tasks (Garcia et al., 2019). In the literature, a lot of different approaches have been presented to solve the problem of the prediction operator's intentions, such as a framework for the prediction of human intentions from RGBD data (Casalino et al., 2018). A sparse Bayesian learning-based human intention predictor is used to predict the future human desired position (Li et al., 2019). A temporal CNN with a convolution operator is applied for human trajectory prediction (Zhao and Oh, 2021). A system that detects human intentions through a recursive Bayesian classifier is used, exploiting head and hand tracking data (Casalino et al., 2018). Another human intention inference system that uses an expectation-maximization algorithm with online model learning is employed (Ravichandar and Dani, 2017).

Awareness: Situational awareness in HMIs is used to describe the level of awareness that operators/drivers/users have of the situation in order to perform tasks successfully (Endsley, 1995). Based on the definition provided by Vidulich et al. (1994), situational awareness needs to include four specific requirements:

1. to easily receive information from the environment.
2. to integrate this information with relevant internal knowledge, creating a mental model of the current situation.
3. to use this model to direct further perceptual exploration in a continual perceptual cycle.
4. to anticipate future events.

Taking these four requirements into account, situational awareness is defined as the continuous extraction of environmental information, integrating this information with previous knowledge to form a coherent mental picture and using that picture in directing perception and anticipating future events. The system will be able to monitor and understand the user's state (e.g., fatigue and cognitive level) to produce personalized alarms, warnings, information, and suggestions to the users. A situational awareness application could also provide

- Information streams regarding the task underway, improving focus.

- Personalized reminders regarding other parallel or scheduled tasks significantly improve response time.
- Notifications and visual aids regarding imminent dangers or accident-related factors.
- Environmental values and real-time measurements of sensors.
- KPIs visualizing the effectiveness of the CPSoS functionality.

Situation awareness is essential in cases where a user must intervene in operations and co-operations with highly automated systems in order to correct failed autonomous decisions in CPSoSs (Horváth, 2020). It is also an effective method to keep the mechanical parts of a system and its operators secure and safe; it can be classified into two groups, human and computer awareness (Yang et al., 2018). Moreover, situational awareness for security reasons is critical since it can be used to inform the user about a cyber-attack that takes place in real time (Joo et al., 2018).

6 Role of digital twins in optimizing CPSoS ecosystems

Building upon the previously described perception and behavioral layers of the CPSoS general architecture, DTs emerge as a crucial technology that realizes and optimizes this framework. The perception layer's role in enhancing situational awareness through object detection, cooperative scene analysis, and effective path planning and the behavioral layer's focus on integrating human operators and supporting HITL interactions are both significantly augmented by the capabilities of DTs. DTs provide a dynamic and real-time simulation of physical systems, creating accurate virtual replicas that enable continuous monitoring and data integration across both layers. More specifically, DTs in the perception layer offer a comprehensive view of the environment by synchronizing data from various sensors and subsystems, ensuring more precise and reliable situational awareness. This integration allows for improved responsiveness to environmental changes and anomalies, enhancing the autonomy and reliability of CPSoSs. In the behavioral layer, DTs facilitate seamless human-machine interfaces by delivering real-time feedback and predictive insights. This supports the HITL approach by allowing human operators to interact with the system using real-time simulations and up-to-date information. Advanced HMIs such as gesture recognition and eXtended Reality technologies are further empowered by DTs, making interactions more intuitive and efficient. Moreover, the predictive capabilities of DTs help anticipate operator intentions, improving the collaborative efforts between humans and the system, especially in critical scenarios where safety and efficiency are paramount.

As such, by integrating DT frameworks into the proposed two-layered architecture, the CPSoS ecosystem is not only optimized but also becomes more resilient and adaptive to the complex demands of various domains, including automotive, industrial manufacturing, and smart cities. This synergy between DTs and the CPSoS architecture leads to smarter, more efficient systems capable of addressing modern challenges with greater efficiency. In the following sections, we will present how DTs can be employed to realize indicative large-scale CPSoSs like smart cities, transportation systems, and aerial traffic monitoring.

6.1 Digital twins in prominent examples of CPSoSs

DTs will play a pivotal role in optimizing large CPSoSs by creating virtual replicas of physical systems, allowing for real-time monitoring, simulation, and predictive maintenance. This capability is particularly important for large CPSoSs, where the integration of numerous interconnected subsystems demands precise coordination and management. DTs enhance the functionality and efficiency of these complex systems by providing a unified platform for data integration, analysis, and visualization. By enabling continuous feedback loops between the physical and digital realms, DTs improve decision-making processes, enhance system reliability, and optimize operational performance across diverse domains, including but not limited to smart cities, intelligent transportation systems, and aerial traffic monitoring (Mylonas et al., 2021; Kušić et al., 2023; Wang et al., 2021).

6.1.1 DTs and smart cities

The management and development of smart cities can be revolutionized by DTs as they provide detailed digital replicas of urban environments. These digital models integrate various data sources to deliver real-time insights and simulations, enhancing urban planning, infrastructure maintenance, and environmental monitoring. More specifically, DTs enable city planners to test different scenarios and make data-driven decisions, optimizing the layout and functionality of urban spaces. By modeling traffic flows, optimizing traffic light timings, and reducing congestion, DTs improve urban mobility and air quality. For example, DTs can analyze data from traffic cameras, sensors, and GPS to provide real-time traffic management solutions, facilitating efficient and sustainable urban traffic control (Schwarz and Wang, 2022). Furthermore, DTs allow for the detailed simulation of urban infrastructure, helping planners optimize the layout and design of utilities such as water, electricity, and waste management systems. By providing a virtual model of the city's infrastructure, DTs enable predictive maintenance and efficient resource allocation, reducing operational costs and improving service delivery (Broo and Schooling, 2023). Additionally, DTs play a crucial role in enhancing public safety and emergency response (Aluvalu et al., 2023). By integrating data from surveillance systems, emergency services, and environmental sensors, DTs provide real-time situational awareness, enabling faster and more coordinated responses to emergencies. For instance, DTs can simulate natural disaster scenarios and help plan effective response strategies, thereby improving the efficiency and effectiveness of emergency management. Another important aspect is that DTs support smart building management by monitoring energy consumption, predicting maintenance needs, and improving overall building efficiency. By providing a virtual model of buildings, DTs help optimize heating, ventilation, and air conditioning (HVAC) systems, lighting, and other building services, contributing to energy savings and improved occupant comfort (Eneyew et al., 2022). Finally, DTs may be utilized for environmental monitoring by integrating data from air quality sensors, weather stations, and other environmental monitoring tools (Purcell et al., 2023). They provide real-time insights into environmental conditions, helping cities monitor pollution levels, manage natural resources, and implement sustainability initiatives. For instance, DTs can help track and

manage water quality in urban water systems, ensuring safe and clean water for residents. Overall, DTs enhance the functionality and efficiency of smart cities by providing a unified platform for data integration, analysis, and visualization. This technology enables cities to become more resilient, sustainable, and responsive to the needs of their residents.

6.1.2 DTs and intelligent transportation systems

The complexity of modern transportation systems necessitates sustainable technological innovations. DT technology, as an innovative architecture, is well-suited to examine the lifecycle of various systems in a digital format. DTs support numerous aspects of transportation infrastructure, including transport system monitoring, energy management, traffic forecasting, EV energy consumption forecasting, subway regenerative braking energy forecasting, parking lot management, driver behavior analysis, pedestrian behavior investigation, health system control, and cyber–physical attack detection (Jafari et al., 2023). By enhancing traffic forecasting accuracy through real-time data collection and high-quality models, DTs improve traffic planning and management. This enhancement results in time and cost savings, reduced energy consumption, improved driver wellbeing, and overall better performance. By organizing data and algorithms, DTs can support sustainable urban traffic formation and efficient control (Jiang et al., 2022). For example, they optimize traffic light timings and provide accurate traffic information, facilitating optimal traffic management and extensive EV traffic planning (Kušić et al., 2023; Chomiak-Orsa et al., 2023). Furthermore, DTs are instrumental in predicting and optimizing the energy consumption and production patterns of electrical transportation systems. They enhance the management and optimization of energy consumption, thus improving the operation and performance of these systems (Ketzler et al., 2020; Bhatti et al., 2021). DTs are crucial for the development and operation of autonomous vehicles (Almeaibed et al., 2021). They simulate vehicle behavior under various conditions, allowing for extensive testing and optimization without the risks associated with real-world testing. DTs help in refining algorithms for navigation, obstacle detection, and collision avoidance, making autonomous vehicles safer and more reliable. By providing a comprehensive digital environment, DTs enable the testing of autonomous vehicles in a wide range of scenarios, including adverse weather conditions, complex urban environments, and interactions with other vehicles and pedestrians. This capability is essential for improving the robustness and safety of autonomous driving systems (Bhatti et al., 2021). Another notable application of DT technology is in analyzing and investigating real-time driver and pedestrian behavior, enhancing security and environmental sustainability. By assembling real-time data from drivers and vehicles, DTs transfer crucial information to the physical world, addressing security concerns and promoting sustainability (Yan et al., 2022). Furthermore, DTs play a vital role in detecting cyber and physical attacks in transportation systems, increasingly targeted by hackers due to the integration of wireless and IoT technologies. This capability ensures a secure and reliable environment for all transportation agents (Liu et al., 2020; Damjanovic-Behrendt, 2018). As transportation systems evolve with advanced technologies, they become more vulnerable to cyber and physical attacks. DTs enable the dynamic analysis of transportation systems, providing real-time detection of such attacks. This functionality helps construct a secure and reliable environment for transportation systems, ensuring the safety of all users, including pedestrians (Almeaibed et al., 2021).

6.1.3 DTs and aerial traffic monitoring

DTs are increasingly recognized for their transformative potential in managing aerial traffic, particularly for UAVs and drones. DTs provide a real-time digital replica of physical assets, enabling precise navigation, real-time monitoring, and effective management of aerial operations. This technology is crucial for ensuring the safety, efficiency, and reliability of aerial traffic, especially in urban environments where the density of aerial traffic is high (Wang et al., 2021). In more detail, DTs can enable real-time monitoring of UAVs and drones by providing a comprehensive digital model that mirrors the physical asset. This model integrates data from various sensors, including GPS, cameras, and environmental sensors, to offer a holistic view of the status and environment of UAVs. These real-time data allow for precise navigation, collision avoidance, and efficient route planning, ensuring safe operations even in complex and dynamic environments (Glaessgen and Stargel, 2012; Soliman et al., 2023). DTs also play a critical role in airspace management by providing a comprehensive and integrated view of all aerial activities within a given area. They can simulate different flight scenarios, optimize airspace usage, and manage the traffic flow to prevent collisions (He et al., 2019). DTs support the coordination of multiple UAVs, ensuring that flight paths are optimized and comply with airspace regulations. This capability is particularly important in urban areas where multiple UAVs may be operating simultaneously (Tang et al., 2023; Lv et al., 2021). In emergency response scenarios, DTs are invaluable for enhancing situational awareness and coordination among various entities. For instance, during a natural disaster, DTs can be used to deploy UAVs for search and rescue operations, assess damage, and deliver essential supplies. The real-time data provided by DTs help emergency responders make informed decisions quickly, thereby improving the efficiency and effectiveness of the response (Piperigkos et al., 2023; Wen et al., 2024; Ariyachandra and Wedawatta, 2023).

6.2 Digital twins-based integration of smaller CPSoSs into larger CPSoSs

In the previous examples, DTs have been exploited not only to model and optimize the performance of individual CPSoSs like autonomous ground and aerial vehicles but also to integrate these smaller CPSoSs into larger CPSoS ecosystems, like smart city environments. This integration features great potential for creating a synergistic ecosystem where various systems work together seamlessly to enhance overall functionality, efficiency, and resilience. For example, the integration of autonomous ground vehicles into the infrastructure of smart cities exemplifies the convergence of smaller CPSoSs into larger CPSoSs. Autonomous vehicles operate as part of a larger network, interacting with smart traffic management systems, connected infrastructure, and other smart devices to optimize urban mobility. Autonomous vehicles communicate with

traffic lights, road sensors, and central management systems to navigate efficiently, reduce traffic congestion, and enhance road safety. This level of integration allows for real-time data sharing and coordinated decision-making, significantly improving the performance of urban transportation systems (Piperigkos et al., 2023; Huang et al., 2024; Hu et al, 2024). Furthermore, a key advantage of integrating smaller CPSoSs (e.g., autonomous vehicles) into larger CPSoSs (e.g., smart cities) is the seamless data integration and management it facilitates. Data from various sources, such as vehicles, smart buildings, environmental sensors, and public transportation systems, can be collected, analyzed, and utilized in a unified platform. This integrated data ecosystem enables better decision-making and proactive management of urban systems. For example, data from autonomous vehicles can be combined with environmental data to monitor and manage urban air quality more effectively (Kopelias et al., 2020; Bayat et al., 2017). In general, the urban ecosystem can be enhanced by enabling real-time data sharing and collaborative decision-making across various systems. This interconnected network of CPSoSs facilitates efficient resource management, improves public services, and increases resilience against disruptions. Integrating transportation systems, energy grids, and public safety networks through DTs allows for optimized urban operations, reduced response times in emergencies, and an overall enhancement in the quality of life for residents (Lv et al., 2022). In the context of urban transportation, multiple CPSoSs can collaboratively work to streamline traffic flow, reduce congestion, and lower emissions by sharing real-time data and predictive analytics (Liu et al., 2023). This synergy allows for dynamic adjustment of traffic signals, real-time rerouting of vehicles, and efficient public transport scheduling. Energy grids can interact with transportation systems to manage the charging of electric vehicles, ensuring that the energy supply meets demand without overloading the grid (Bhatti et al., 2021). This holistic integration supports disaster management, improves emergency response times, and enhances overall urban resilience.

7 Challenges and open research issues

The advancement of CPSoSs involves overcoming technical and integration challenges in areas such as cooperative object detection and fusion, cooperative localization and path planning, cooperative SLAM, and HITL integration. These areas are crucial for enhancing the functionality, reliability, and efficiency of CPSoSs. Cooperative object detection and fusion integrate data from various sensors, improving situational awareness in the perception layer. Cooperative localization and path planning optimize navigation and traffic management, aligning with the perception layer's goals. Cooperative SLAM enables accurate environmental mapping. HITL integration enhances decision-making, linking to the behavioral layer's focus on human interaction and control. Addressing these challenges is essential for improving performance, safety, and adaptability in diverse and dynamic environments, driving significant advancements in the development and deployment of CPSoSs across various sectors.

7.1 Cooperative object detection and fusion

Integrating data from heterogeneous sensors, such as cameras, LiDAR, and radar, remains complex due to differences in data formats, resolutions, and sampling rates, necessitating the development of robust fusion algorithms. Furthermore, developing methods for multi-modal object representation that reconcile discrepancies in perception across sensors is crucial for cohesive and accurate environmental understanding (Arnold et al., 2020; Guo et al., 2021). Managing uncertainties in sensor measurements and fusion processes is vital to enhance the reliability of object detection. Establishing standardized benchmarks and metrics for evaluation is necessary to effectively assess and compare the performance of cooperative object detection systems. Optimizing computational efficiency and energy consumption of algorithms while maintaining high accuracy poses additional challenges. Lastly, understanding human interactions with autonomous systems equipped with cooperative object detection capabilities is essential for ensuring safe integration into mixed-traffic environments.

7.2 Cooperative localization and path planning

Cooperative localization for connected CPSs is advancing, but several key research challenges remain. Algorithms need to maintain accurate positioning even when communication is disrupted or networks are patchy. Integrating data from GPS, LiDAR, and cameras is crucial for precise localization. Algorithms must also work well in GPS-unavailable areas. Efficiently processing large data on limited computing power is essential for inter-vehicle communication. Urban areas present signal issues that need advanced processing techniques. Finally, creating realistic tests and benchmarks is vital for ensuring system reliability.

Cooperative path planning involves multiple agents working together to navigate from their respective starting points to designated endpoints. Despite significant advancements, several open issues remain in this domain. For application to real conditions, scalability and efficient coordination require algorithms capable of managing large fleets of CPSs (Halder et al., 2020; Wang et al., 2019b) without compromising performance opting for real-time decision-making for a swift response to dynamic environments (Viana and Aouf, 2018), such as sudden obstacles (Viana et al., 2019). Robust algorithms are needed to maintain path planning when inter-vehicle communication is unreliable. Effectively integrating heterogeneous sensor data from LiDAR, cameras, and GPS can improve decision-making accuracy, but it also requires conflict resolution mechanisms. Furthermore, adapting path planning algorithms to diverse conditions and regulatory environments requires the definition and introduction of constraints. Real-world testing and validation are essential to validate algorithms under diverse conditions, ensuring they meet the stringent requirements of safe and efficient behavior of CPSoSs.

7.3 Cooperative SLAM

As mentioned by Saeedi et al. (2016), cooperative SLAM has to face quite some challenges in order to fully exploit the potential of collaboration.

7.3.1 Data distribution

Further investigation is needed to determine which processing architecture, either centralized or distributed, is more efficient for the cooperative fusion of different SLAM solutions. The optimal choice is closely related to the three other major challenges.

7.3.2 Relative poses of robots

The map provided by each robot in its own reference coordinates is called the local map. Each robot aims to integrate all individual local maps to generate a global map of the environment. However, this difficult task requires *a priori* unknown transformation matrices, which relate these maps to one another. The problem of the relative pose of the robot is coupled with the multiple-robot data association problem. Knowledge of one makes the other a simple problem.

7.3.3 Updating maps and poses

Once the relative transformation is determined, the fusion of local maps is necessary. The resulting map should integrate all the information contained within local maps. As a result of updating the maps, the poses of the robots should also be updated. This requires considering the current full trajectory of the robots and new information received from other maps.

7.3.4 Communication requirements

The availability of a medium for data sharing among robots is an important requirement in multiple-robot SLAM. Information between robots can be exchanged via communication channels. The quality of the communication channels is dependent on the (harsh or not) environment. Additionally, the amount of data that needs to be exchanged may have a significant impact on the efficiency of communication. For instance, a local map of thousands of 3D points that needs to be transmitted to a group of robots is not a trivial task. Therefore, rich communication resources are also needed in order to realize cooperative SLAM.

7.4 Human in the loop in CPSoSs

In CPSoSs, integrating human feedback into the control loop of a system allows humans to interact with and influence automated processes in real time. However, a series of challenges appear with respect to this component that CPSoSs have to address.

7.4.1 Processing in real time

The complexity of CPSoSs, consisting of a variety of different components, leads to the instantaneous production of a huge amount of data. The processing of these data and real-time decision-making are challenging tasks, considering that human safety is the most important issue. The processing of data in batches could be a solution to this challenge. However, this approach is not reliable in critical situations within CPSoSs, where vital and accurate decisions have to be made quickly to protect human life and security. In other words, the real-time data processing framework requires the system to handle large amounts of data with very low latency while maintaining relatively high performance.

7.4.2 Online streaming of data

CPSoSs are systems of systems that are interconnected, collaborating, and transferring in real-time helpful information and data. Online streaming also requires real-time data processing. In the case of online streaming, the challenge originates from data transfer in an ordered sequence of instances that can usually be accessed once or a few times due to limited computing and storage capabilities. The tremendous growth of data demands switching from traditional data processing solutions to systems that can process a continuous stream of real-time data.

7.4.3 High-dimensional data

High-dimensional data are becoming a prevalent issue in many real-world applications of CPSoSs. The processing of high-dimensional data acquired by different sensors and devices presents a fundamental challenge, leading to more sophisticated methods being developed. High-frequency data refer to data that usually appear as time series, with their values updating very rapidly (i.e., new observations take place every milliseconds-second). The appropriate management of high-frequency data is essential for contemporary CPSoSs. Processing these data introduces new challenges to decision-making tasks, especially when a human takes part in the CPSoS as a HITL component.

7.4.4 Unsupervised learning in data of CPSoSs

Unsupervised learning is a type of learning that tries to discover hidden patterns in untagged data autonomously. This can be beneficially used in real-time applications where the observed data possess a large variety of classes compared to those in a restricted dataset. However, applying this in CPSoSs is a very challenging task as they require accurate and precise results, and usually, there are no "ground truth" data for the evaluation of the method's accuracy (Ma et al., 2018).

8 Lessons learned

In the connected CPSoS, each node perceives the environment and generates data streams shared among all connected nodes, facilitating collaborative perception, localization, and path planning. This interconnectedness allows each node to access a wealth of information about the common environment that would otherwise be unavailable. In practice, the sensing range of each individual CPS is extended according to the sensing capabilities of the other interconnected CPSs (Piperigkos et al., 2021), thus leading to complementary data fusion. The integration of cooperative perception, localization, SLAM, path planning, and HITL forms the foundation of distributed intelligence in CPSoSs. The integration of various sensors across nodes significantly enhances cooperative situational awareness, and the benefits of that type of collaboration have been quantified both theoretically and algorithmically. For example, Fisher information

matrix-based analysis (Buehrer et al., 2018) provides indicative insights about the fundamental nature of collaboration and the scaling with network size, anchor placement, neighbor selection, etc., demonstrating how the integration of sensors facilitates swarm navigation in Mars exploration missions (Zhang et al., 2020). Therefore, this interconnected data sharing results in a more comprehensive understanding of the environment, leading to improved decision-making capabilities. For instance, in autonomous driving, shared data from multiple vehicles can provide a clearer picture of road conditions and traffic patterns, enhancing safety and efficiency (Arvanitis et al., 2023). Cooperative localization enables precise positioning even in challenging environments where traditional GPS signals may be unreliable (Piperigkos et al., 2023).

By sharing location data among nodes, CPSoSs can achieve more accurate and reliable navigation. This is crucial for applications such as autonomous vehicles and drones, which rely on precise localization for safe operation. Cooperative SLAM extends the capabilities of individual systems by allowing multiple nodes to collaboratively map their environment; as stated by Saeedi et al. (2016), continuously updating and sharing maps among nodes enhances the overall system's adaptability to dynamic environments. However, this fact also emphasizes the need for more accurate and up-to-date maps, which are essential for navigation and obstacle avoidance. In the same context, incorporating human feedback into CPSoS operations could potentially enhance system performance in dynamic situations where algorithms might lack situational awareness or adaptability (Alsamhi et al., 2024). This is particularly important in scenarios where automated systems face uncertain or complex situations. HITL systems ensure that human operators can intervene when necessary, providing a safety net for critical operations.

It is expected that the collective intelligence of connected CPSoS nodes will lead to more robust and resilient systems. By distributing computational tasks and decision-making processes across multiple nodes, CPSoSs will be able to handle larger and more complex tasks with greater efficiency. In this distributed type of approach, system scalability can also be enhanced, making it easier to expand CPSoSs to accommodate more nodes and diverse sensors. In the same context, DTs will play a crucial role in optimizing CPSoS ecosystems by creating virtual replicas of physical systems (i.e., smart cities, transportation, etc.), enabling real-time monitoring, predictive maintenance, and scenario simulation. In practice, DTs are expected to enhance the interaction between smaller and larger CPSoSs, facilitating efficient data sharing and collaborative decision-making across various systems, with a direct impact on the quality of life (Lv et al., 2022). Despite significant advancements, several challenges remain, including managing high-dimensional data, ensuring real-time processing, and developing robust algorithms for data fusion and localization. Future research should focus on addressing these challenges to further enhance the capabilities of CPSoSs. Overall, the lessons learned from the development and implementation of CPSoSs highlight the importance of collaboration, data sharing, and human integration in creating intelligent, adaptive, and efficient systems. These insights provide a roadmap for future research and development, aiming to optimize CPSoSs for various applications, from smart cities to autonomous transportation.

9 Discussion

In this survey, we examine CPSoSs and their components that improve situational awareness for users, an aspect not thoroughly discussed in previous review papers. By focusing on human integration into CPSs, we include the HITL element and HMI in the CPSoS concept. We also emphasize the crucial role of DTs in optimizing CPSoS ecosystems. The key contributions of this paper consist of an extensive review of current leading practices in connected CPSs and an analysis of a dual-layer architecture with a perception layer for situational awareness and a behavioral layer for incorporating human operators through HITL mechanisms and sophisticated HMI technologies. Furthermore, we provide various datasets and data sources accessible to the research community, concentrating on perception algorithms for scene understanding, localization, mapping, and path planning, along with decision-making and HITL control. We also discuss the incorporation of DTs into CPSoSs, showcasing their applications in smart cities, intelligent transportation systems, and aerial traffic monitoring.

In more detail, this survey offers a comprehensive exploration of the architectural and operational intricacies of CPSoSs, highlighting the dual-layer architecture comprising the perception and behavioral layers. This approach distinguishes our study by providing a detailed examination of how these layers enhance situational awareness and integrate human elements within CPSoSs. The perception layer is meticulously designed to focus on advanced perception algorithms essential for object detection, scene analysis, cooperative localization, and path planning. These capabilities are critical for achieving higher autonomy and reliability in CPSoSs. Unlike other studies that may address these components in isolation, our research integrates these elements into a cohesive framework, demonstrating how they collectively contribute to the overall functionality and efficiency of CPSoSs. The behavioral layer emphasizes the integration of human operators through HITL mechanisms and advanced HMI technologies. This layer underscores the importance of human cognition, adaptability, and motivation, which are crucial for operational excellence in CPSoSs. Our study uniquely addresses the challenges and benefits of incorporating human feedback and interaction into automated systems, promoting a human–machine symbiosis that enhances decision-making and system flexibility. Additionally, we delve into the role of DTs in optimizing CPSoS ecosystems. Our research highlights the potential of DTs in smart cities, intelligent transportation systems, and aerial traffic monitoring, showcasing how they facilitate real-time data integration, predictive maintenance, and improved decision-making. The discussion extends to the integration of smaller CPSoS into larger systems, emphasizing that seamless communication and coordination among subsystems enhance overall system performance and urban management.

Our research provides valuable insights into how evolving CPSoSs can transform collaborative and collective decision-making and significantly impact various industries and disciplines. The integration of HITL mechanisms and HMIs enhances real-time

TABLE 10 Summary of mature, current, and future challenges in CPSoS areas of interest.

Area of interest	SOTA (Mature)	Current challenges	Future challenges
Cooperative object detection and fusion	Data fusion from multiple sensors improves situational awareness and enhances decision-making capabilities	Fusion of heterogeneous data to address discrepancies in data formats and resolutions	Robust algorithms to ensure real-time, multi-modal perception in dynamic environments
Cooperative localization and path planning	Cooperative localization improves precision in GPS-denied environment	Ensuring accurate positioning and path planning under network delay	Scalable algorithms for large CPS fleets with minimal latency and error in real-time localization and path palling
Cooperative SLAM	Cooperative SLAM enables real-time environment mapping	Managing data fusion for distributed SLAM and ensuring up-to-date maps in dynamic environments	Real-time, large-scale SLAM algorithms that are robust to communication latency and noise
HITL systems	HITL ensures human intervention in critical decision-making scenarios, improving safety	Managing the balance between human input and automation in fast-evolving situations	Seamless AI-human interaction for adaptive, safe, and autonomous decision-making in diverse environments
DTs	DTs enable real-time monitoring and predictive maintenance	Integrating smaller and larger CPSoS with DTs to enhance data sharing and decision-making	Full integration of DTs for autonomous, real-time system optimization across different sectors
Security and trust	Basic security mechanisms embedded in traditional CPSoS systems	Ensuring real-time breach detection and secure communication in interconnected environments	Developing security-by-design frameworks for heterogeneous CPSoSs
HART	Collaborative interactions between humans and semi-autonomous systems enhance productivity	Optimizing hybrid intelligence by combining human intuition with machine precision	Creating resilient environments for seamless human–machine cooperation, leveraging cognitive capabilities

interaction and feedback between humans and machines, leading to more informed and effective decision-making processes. For example, in industrial production, the ability to integrate human expertise with automated processes can create more flexible and adaptive manufacturing systems. Operators can provide real-time input and adjustments, ensuring that production lines can quickly respond to changes in demand or unforeseen issues, thus improving efficiency and reducing downtime. In the context of smart cities, CPSoSs can enhance urban mobility, energy management, and public safety. For instance, the integration of autonomous ground vehicles within smart city infrastructure allows for optimized traffic flow and reduced congestion through real-time data sharing and coordinated decision-making. This not only improves the efficiency of transportation systems but also contributes to environmental sustainability by reducing emissions. The research community can greatly benefit from this manuscript by leveraging the frameworks and methodologies presented to establish robust CPSoSs. Our detailed discussion on perception and behavioral layers, as well as the integration of DTs, offers a comprehensive guide for developing and optimizing CPSoSs. By addressing both technical and human-centric aspects, our study provides a holistic approach to understanding and implementing CPSoSs. Researchers can build on these insights to explore new opportunities, address current challenges, and develop more resilient and adaptable systems that meet the evolving demands of various industries. Overall, this survey emphasizes the transformative potential of CPSoSs in enhancing collaborative and collective decision-making, operational efficiency, and quality of life across different sectors. The insights gained from this study serve as a valuable resource for the research community, guiding future innovations and advancements in the field of CPSoSs.

9.1 Future directions

As CPSoSs continue to evolve, several key areas require further attention and development to ensure their effective implementation and optimization. One critical area is security and trust. Since CPSoSs are tightly integrated with human elements and involve diverse, interconnected systems, ensuring robust security measures is paramount. Traditional security measures often fall short due to the complexity, autonomy, and heterogeneity of CPSoSs, as well as the presence of legacy components. Therefore, a new security and trust mechanism tailored for CPSoSs is essential. This mechanism must be embedded from the design phase and continuously updated throughout the system's lifecycle to adapt to emerging cybersecurity threats. Implementing a security-by-design principle ensures that security components can accommodate varying security needs and performance capabilities across different CPSs. Additionally, specialized security monitoring tools should be integrated to detect, respond to, and mitigate security breaches in real time, ensuring resilience even under unforeseen conditions. Another promising direction is the development of systems based on the human-agent–robot teamwork (HART) framework (Demir et al., 2020). This framework emphasizes

collaborative interactions between humans and fully or semi-autonomous machines, aiming to harness hybrid intelligence—the combined strengths of human intuition and machine precision. Such synergy can significantly enhance productivity, innovation, and overall system performance. Future research should focus on creating environments that facilitate seamless co-working between humans and machines, optimizing both human cognitive abilities and machine efficiency. This cooperative model not only enhances operational effectiveness but also promotes adaptability in response to dynamic and complex environments, leading to more resilient and innovative CPSoS (Verhagen et al., 2022). By addressing these future directions, the field of CPSoSs can advance toward more secure, efficient, and human-centric systems, ensuring their relevance and effectiveness in various domains.

Finally, Table 10 summarizes the main lessons learned about the currently matured areas that we presented previously, as well as the corresponding current and future challenges derived from the previous
discussion.

10 Conclusion

This survey provides a comprehensive review of current best practices in connected CPSoSs. We present a detailed CPSoS architecture that facilitates collective intelligence through sensor fusion, scene analysis, cooperative localization and mapping, and user state monitoring. By examining all aspects of these areas, we offer insights into HITL-oriented datasets, including face analysis and landmark extraction, pose estimation, hand and gesture recognition, action recognition, and speech analysis. Through this survey, readers can gain a deep understanding of the current status, advancements, and challenges in adopting autonomous CPSs and CPSoSs within a continuously evolving technological landscape. We highlight the importance of integrating HITL mechanisms and HMIs to achieve a CPSS paradigm. Additionally, we emphasize the critical role of DTs in optimizing CPSoS ecosystems, demonstrating their applications in smart cities, intelligent transportation systems, and aerial traffic monitoring. By addressing the dual-layer architecture encompassing the perception and behavioral layers, this survey underscores the necessity of enhancing situational awareness and integrating human expertise into automated processes. The insights provided herein are intended to guide future research and development in creating more resilient, efficient, and intelligent CPSoS, ultimately contributing to improved urban ecosystems and industrial environments.

Author contributions

NP: writing–original draft and writing–review and editing. AG: writing–original draft and writing–review and editing. GA: writing–original draft and writing–review and editing. SN: writing–original draft and writing–review and editing. AL: writing–original draft and writing–review and editing. AF: writing–review and editing. PR-G: writing–review and editing. PS: writing–review and editing. KM: writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This paper was supported by the EU's HORIZON AutoTRUST (No. 101148123).

Conflict of interest

Author PR-G was employed by K3Y Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aishwarya, J., Kundapur, P. P., Kumar, S., and Hareesha, K. S. (2018). "Kannada speech recognition system for aphasic people," in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19-22 September 2018, 1753–1756. doi:10.1109/ICACCI.2018.85546572013

Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., and Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Aut. Robots* 42, 957–975. doi:10.1007/s10514-017-9677-2

Alam, N., and Dempster, A. G. (2013). Cooperative positioning for vehicular networks: facts and future. *IEEE Trans. Intelligent Transp. Syst.* 14, 1708–1717. doi:10.1109/tits.2013.2266339

Almeaibed, S., Al-Rubaye, S., Tsourdos, A., and Avdelidis, N. P. (2021). Digital twin analysis to promote safety and security in autonomous vehicles. *IEEE Commun. Stand. Mag.* 5, 40–46. doi:10.1109/mcomstd.011.2100004

Al-Mhiqani, M. N., Ahmad, R., Yassin, W., Hassan, A., Abidin, Z. Z., Ali, N. S., et al. (2018). Cyber-security incidents: a review cases in cyber-physical systems. *Int. J. Adv. Comput. Sci. Appl.* 9. doi:10.14569/IJACSA.2018.090169

Alsamhi, S. H., Kumar, S., Hawbani, A., Shvetsov, A. V., Zhao, L., and Guizani, M. (2024). Synergy of human-centered ai and cyber-physical-social systems for enhanced cognitive situation awareness: applications, challenges and opportunities. *Cogn. Comput.* 16, 2735–2755. doi:10.1007/s12559-024-10271-7

Aluvalu, R., Mudrakola, S., V, U. M., Kaladevi, A., Sandhya, M., and Bhat, C. R. (2023). The novel emergency hospital services for patients using digital twins. *Microprocess. Microsystems* 98, 104794. doi:10.1016/j.micpro.2023.104794

Andriluka, M., Pishchulin, L., Gehler, P., and Schiele, B. (2014). "2d human pose estimation: new benchmark and state of the art analysis," in Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (USA:

IEEE Computer Society), Columbus, OH, USA, 23-28 June 2014, 3686–3693. doi:10.1109/CVPR.2014.47114

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., et al. (2019). *Common voice: a massively-multilingual speech corpus*.

Ariyachandra, M. R. M. F., and Wedawatta, G. (2023). Digital twin smart cities for disaster risk management: a review of evolving concepts. *Sustainability* 15, 11910. doi:10.3390/su151511910

Arnold, E., Dianati, M., de Temple, R., and Fallah, S. (2020). Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors. *IEEE Trans. Intelligent Transp. Syst.* 23, 1852–1864. doi:10.1109/tits.2020.3028424

Arvanitis, G., Moustakas, K., and Fakotakis, N. (2016). "Online biometric identification with face analysis in web applications," in *Speech and computer*. Editors A. Ronzhin, R. Potapova, and G. Németh (Cham: Springer International Publishing), 515–522.

Arvanitis, G., Stagakis, N., Zacharaki, E. I., and Moustakas, K. (2023). Cooperative saliency-based pothole detection and ar rendering for increased situational awareness. *IEEE Trans. Intelligent Transp. Syst.* 25, 3588–3604. doi:10.1109/tits.2023.3327494

Atat, R., Liu, L., Wu, J., Li, G., Ye, C., and Yang, Y. (2018). Big data meet cyber-physical systems: a panoramic survey. *IEEE Access* 6, 73603–73636. doi:10.1109/access.2018.2878681

Aviles-Arriaga, H., Sucar, L., and Mendoza, C. (2006). "Visual recognition of similar gestures," in 18th International Conference on Pattern Recognition (ICPR'06), 1100–1103. doi:10.1109/ICPR.2006.11801

Bai, W., Xu, B., Liu, H., Qin, Y., and Xiang, C. (2023). Robust longitudinal distributed model predictive control of connected and automated vehicles with coupled safety constraints. *IEEE Trans. Veh. Technol.* 72, 2960–2973. doi:10.1109/TVT.2022.3217896

Bayat, B., Crasta, N., Crespi, A., Pascoal, A. M., and Ijspeert, A. (2017). Environmental monitoring using autonomous vehicles: a survey of recent searching techniques. *Curr. Opin. Biotechnol.* 45, 76–84. doi:10.1016/j.copbio.2017.01.009

Belhumeur, P. N., Jacobs, D. W., Kriegman, D. J., and Kumar, N. (2011). "Localizing parts of faces using a consensus of exemplars," in *Cvpr 2011*, 545–552. doi:10.1109/CVPR.2011.5995602

Besl, P., and McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Trans. Pattern Analysis Mach. Intell.* 14, 239–256. doi:10.1109/34.121791

Bhattacharya, M., Penica, M., O'Connell, E., Southern, M., and Hayes, M. (2023). Human-in-loop: a review of smart manufacturing deployments. *Systems* 11, 35. doi:10.3390/systems11010035

Bhatti, G., Mohan, H., and Raja Singh, R. (2021). Towards the future of smart electric vehicles: digital twin technology. *Renew. Sustain. Energy Rev.* 141, 110801. doi:10.1016/j.rser.2021.110801

Brambilla, M., Nicoli, M., Soatti, G., and Deflorio, F. (2020). Augmenting vehicle localization by cooperative sensing of the driving environment: insight on data association in urban traffic scenarios. *IEEE Trans. Intelligent Transp. Syst.* 21, 1646–1663. doi:10.1109/tits.2019.2941435

Broo, D. G., and Schooling, J. (2023). Digital twins in infrastructure: definitions, current practices, challenges and strategies. *Int. J. Constr. Manag.* 23, 1254–1263. doi:10.1080/15623599.2021.1966980

Buehrer, R. M., Wymeersch, H., and Vaghefi, R. M. (2018). Collaborative sensor network localization: algorithms and practical issues. *Proc. IEEE* 106, 1089–1114. doi:10.1109/jproc.2018.2829439

Casalino, A., Messeri, C., Pozzi, M., Zanchettin, A. M., Rocco, P., and Prattichizzo, D. (2018). Operator awareness in human–robot collaboration through wearable vibrotactile feedback. *IEEE Robotics Automation Lett.* 3, 4289–4296. doi:10.1109/LRA.2018.2865034

Cattoni, R., Di Gangi, M. A., Bentivogli, L., Negri, M., and Turchi, M. (2021). Must-c: a multilingual corpus for end-to-end speech translation. *Comput. Speech Lang.* 66, 101155. doi:10.1016/j.csl.2020.101155

Chaloupka, J., Červa, P., Silovský, J., Žďánský, J., and Nouza, J. (2012). "Modification of the speech feature extraction module for the improvement of the system for automatic lectures transcription," in Proceedings ELMAR-2012, Zadar, Croatia, 12-14 September 2012, 223–226.

Chaloupka, J., Nouza, J., Malek, J., and Silovsky, J. (2015). "Phone speech detection and recognition in the task of historical radio broadcast transcription," in 2015 38th International Conference on Telecommunications and Signal Processing (TSP), Prague, Czech Republic, 09-11 July 2015, 1–4. doi:10.1109/TSP.2015.7296399

Chen, H. (2017). Applications of cyber-physical system: a literature review. *J. Industrial Integration Manag.* 02, 1750012. doi:10.1142/S2424862217500129

Chen, J. Y. C., and Barnes, M. J. (2014). Human–agent teaming for multirobot control: a review of human factors issues. *IEEE Trans. Human-Machine Syst.* 44, 13–29. doi:10.1109/THMS.2013.2293535

Chen, L. W., Ho, Y. F., Tsai, M. F., Chen, H. M., and Huang, C. F. (2016). Cyber-physical signage interacting with gesture-based human–machine interfaces through mobile cloud computing. *IEEE Access* 4, 3951–3960. doi:10.1109/ACCESS.2016.2594799

Chen, Q., Ma, X., Tang, S., Guo, J., Yang, Q., and Fu, S. (2019a). "F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds," in *Proceedings of the 4th ACM/IEEE symposium on edge computing*, 88–100.

Chen, Q., Tang, S., Yang, Q., and Fu, S. (2019b). "Cooper: cooperative perception for connected autonomous vehicles based on 3d point clouds," in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 07-10 July 2019 (IEEE), 514–524.

Chen, Z., Yoshioka, T., Lu, L., Zhou, T., Meng, Z., Luo, Y., et al. (2020). Continuous speech separation: dataset and analysis

Cheng, J., Abi-Farraj, F., Farshidian, F., and Hutter, M. (2022). "Haptic teleoperation of high-dimensional robotic systems using a feedback mpc framework," in 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23-27 October 2022, 6197–6204. doi:10.1109/IROS47612.2022.9981290

Choi, H.-R., and Kim, T. (2017). Combined dynamic time warping with multiple sensors for 3d gesture recognition. *Sensors* 17, 1893. doi:10.3390/s17081893

Chomiak-Orsa, I., Hauke, K., Perechuda, K., and Pondel, M. (2023). The use of digital twin in the sustainable development of the city on the example of managing parking resources. *Procedia Comput. Sci.* 225, 2183–2193. doi:10.1016/j.procs.2023.10.209

Chrysos, G. G., Antonakos, E., Zafeiriou, S., and Snape, P. (2015). "Offline deformable face tracking in arbitrary videos," in 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 07-13 December 2015, 954–962. doi:10.1109/ICCVW.2015.12633

Claudi, A., Sernani, P., Dolcini, G., Palazzo, L., and Dragoni, A. F. (2013). "A hierarchical hybrid model for intelligent cyber-physical systems," in 2013 Proceedings of the 11th Workshop on Intelligent Solutions in Embedded Systems (WISES), Pilsen, Czech Republic, 10-11 September 2013, 1–6.

Dai, J., Li, Y., He, K., and Sun, J. (2016). R-fcn: object detection via region-based fully convolutional networks. *arXiv Prepr. arXiv:1605.06409*. doi:10.48550/arXiv.1605.06409

Damjanovic-Behrendt, V. (2018). "A digital twin-based privacy enhancement mechanism for the automotive industry," in 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25-27 September 2018 (IEEE), 272–279.

[Dataset] Garbin, S. J., Shen, Y., Schuetz, I., Cavin, R., Hughes, G., and Talathi, S. S. (2019). Openeds: open eye dataset

[Dataset] Soomro, K., Zamir, A. R., and Shah, M. (2012). Ucf101: a dataset of 101 human actions classes from videos in the wild

Demir, M., McNeese, N. J., and Cooke, N. J. (2020). Understanding human-robot teams in light of all-human teams: aspects of team interaction and shared cognition. *Int. J. Human-Computer Stud.* 140, 102436. doi:10.1016/j.ijhcs.2020.102436

Deniz, D., Barranco, F., Isern, J., and Ros, E. (2020). Reconfigurable cyber-physical system for lifestyle video-monitoring via deep learning. *2020 25th IEEE Int. Conf. Emerg. Technol. Fact. Automation (ETFA)* 1, 1705–1712. doi:10.1109/ETFA46521.2020.9211910

Elazab, M., Noureldin, A., and Hassanein, H. S. (2017). Integrated cooperative localization for vehicular networks with partial GPS access in urban canyons. *Veh. Commun.* 9, 242–253. doi:10.1016/j.vehcom.2016.11.011

El-Ghaish, H., Hussien, M. E., Shoukry, A., and Onai, R. (2018). Human action recognition based on integrating body pose, part shape, and motion. *IEEE Access* 6, 49040–49055. doi:10.1109/ACCESS.2018.2868319

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Hum. Factors* 37, 32–64. doi:10.1518/001872095779049543

Eneyew, D. D., Capretz, M. A. M., and Bitsuamlak, G. T. (2022). Toward smart-building digital twins: bim and iot data integration. *IEEE Access* 10, 130487–130506. doi:10.1109/ACCESS.2022.3229370

Engel, J., Koltun, V., and Cremers, D. (2018). Direct sparse odometry. *IEEE Trans. Pattern Analysis Mach. Intell.* 40, 611–625. doi:10.1109/tpami.2017.2658577

Engell, S., Paulen, R., Reniers, M. A., Sonntag, C., and Thompson, H. (2015). "Core research and innovation areas in cyber-physical systems of systems," in *International workshop on design, modeling, and evaluation of cyber physical systems* (Springer), 40–55.

Eskandarian, A., Wu, C., and Sun, C. (2021). Research advances and challenges of autonomous and connected ground vehicles. *IEEE Trans. Intelligent Transp. Syst.* 22, 683–711. doi:10.1109/tits.2019.2958352

Eskimez, S. E., Maddox, R. K., Xu, C., and Duan, Z. (2020). Noise-resilient training method for face landmark generation from speech. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 28, 27–38. doi:10.1109/TASLP.2019.2947741

Estrada, C., Neira, J., and Tardos, J. (2005). Hierarchical slam: real-time accurate mapping of large environments. *IEEE Trans. Robotics* 21, 588–596. doi:10.1109/TRO.2005.844673

Estrada-Jimenez, L. A., Pulikottil, T., Nikghadam-Hojjati, S., and Barata, J. (2023). Self-organization in smart manufacturing— background, systematic review, challenges and outlook. *IEEE Access* 11, 10107–10136. doi:10.1109/ACCESS.2023.3240433

Fank, J., Knies, C., and Diermeyer, F. (2021). Analysis of a human-machine interface for cooperative truck overtaking maneuvers on freeways: increase success rate and assess driving behavior during system failures. *Multimodal Technol. Interact.* 5, 69. doi:10.3390/mti5110069

Gaham, M., Bouzouia, B., and Achour, N. (2015). "Human-in-the-loop cyber-physical production systems control (hilcp 2 sc): a multi-objective interactive framework proposal," in *Service orientation in holonic and multi-agent manufacturing* (Springer), 315–325.

Galbally, J., Ferrara, P., Haraksim, R., Psyllos, A., and Beslay, L. (2019). Study on face identification technology for its implementation in the schengen information system. *Jt. Res. Cent. Ispra, Italy, Rep. JRC-34751*. doi:10.2760/661464

Gao, C. (2019). *Cooperative localization and navigation: theory, research, and practice*. Boca Raton: Taylor and Francis, CRC Press.

Garcia, M. A. R., Rojas, R., Gualtieri, L., Rauch, E., and Matt, D. (2019). A human-in-the-loop cyber-physical system for collaborative assembly in smart manufacturing. *Procedia CIRP* 81, 600–605. doi:10.1016/j.procir.2019.03.162

Ghasemi, A., Kazemi, R., and Azadi, S. (2013). Stable decentralized control of a platoon of vehicles with heterogeneous information feedback. *IEEE Trans. Veh. Technol.* 62, 4299–4308. doi:10.1109/TVT.2013.2253500

Girshick, R. (2015). "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, Santiago, Chile, 07-13 December 2015, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 23-28 June 2014, 580–587.

Glaessgen, E., and Stargel, D. (2012). "The digital twin paradigm for future nasa and us air force vehicles," in 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA, 1818. doi:10.2514/6.2012-18180

Gorecky, D., Schmitt, M., Loskyll, M., and Zühlke, D. (2014). "Human-machine-interaction in the industry 4.0 era," in 2014 12th IEEE International Conference on Industrial Informatics (INDIN), Porto Alegre, Brazil, 27-30 July 2014, 289–294. doi:10.1109/INDIN.2014.6945523

Grützmacher, F., Beichler, B., Haubelt, C., and Theelen, B. (2016). "Dataflow-based modeling and performance analysis for online gesture recognition," in 2016 2nd International Workshop on Modelling, Analysis, and Control of Complex CPS (CPS Data), Vienna, Austria, 11-11 April 2016, 1–8. doi:10.1109/CPSData.2016.74964230

Güler, R. A., Neverova, N., and Kokkinos, I. (2018). "Densepose: dense human pose estimation in the wild," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7297–7306. doi:10.1109/CVPR.2018.00762

Guo, J., Carrillo, D., Tang, S., Chen, Q., Yang, Q., Fu, S., et al. (2021). Coff: cooperative spatial feature fusion for 3d object detection on autonomous vehicles. *IEEE Internet Things J.* 8, 11078–11087. doi:10.1109/jiot.2021.3053184

Hadorn, B., Courant, M., and Hirsbrunner, B. (2016). *Towards human-centered cyber-physical systems: a modeling approach*. Fribourg, Switzerland: Département d'informatique Université de Fribourg. Tech. rep.

Halder, K., Montanaro, U., Dixit, S., Dianati, M., Mouzakitis, A., and Fallah, S. (2020). Distributed h controller design and robustness analysis for vehicle platooning under random packet drop. *IEEE Trans. Intelligent Transp. Syst.* 23, 4373–4386. doi:10.1109/tits.2020.3044221

Hamzah, M., Islam, M. M., Hassan, S., Akhtar, M. N., Ferdous, M. J., Jasser, M. B., et al. (2023). Distributed control of cyber physical system on various domains: a critical review. *Systems* 11, 208. doi:10.3390/systems11040208

Han, Y., Hyun, J., Jeong, T., Yoo, J.-H., and Hong, J. W.-K. (2016). "A smart home control system based on context and human speech," in 2016 18th International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea (South), 31 January 2016 - 03 February 2016, 165–169. doi:10.1109/ICACT.2016.7423314

Haque, S. A., Aziz, S. M., and Rahman, M. (2014). Review of cyber-physical system in healthcare. *Int. J. Distributed Sens. Netw.* 10, 217415. doi:10.1155/2014/217415

Havard, W., Besacier, L., and Rosec, O. (2017). *Speech-coco: 600k visually grounded spoken captions aligned to mscoco data set. arXiv preprint arXiv:1707.08435*.

He, D., Liu, H., Chan, S., and Guizani, M. (2019). How to govern the non-cooperative amateur drones? *IEEE Netw.* 33, 184–189. doi:10.1109/mnet.2019.1800156

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22-29 October 2017, 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. pattern analysis Mach. Intell.* 37, 1904–1916. doi:10.1109/tpami.2015.2389824

Heilbron, F. C., Escorcia, V., Ghanem, B., and Niebles, J. C. (2015). "Activitynet: a large-scale video benchmark for human activity understanding," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 07-12 June 2015, 961–970. doi:10.1109/CVPR.2015.7298698

Hobert, L., Festag, A., Llatser, I., Altomare, L., Visintainer, F., and Kovacs, A. (2015). Enhancements of v2x communication in support of cooperative autonomous driving. *IEEE Commun. Mag.* 53, 64–70. doi:10.1109/mcom.2015.7355568

Horváth, G., and Erdős, G. (2017). Gesture control of cyber physical systems. *Procedia CIRP* 63, 184–188. doi:10.1016/j.procir.2017.03.312

Horváth, L. (2020). "Situation-awareness in model of cyber physical system," in 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia, 23-25 January 2020, 17–22. doi:10.1109/SAMI48414.2020.910872815

Hu, F., Hao, Q., Sun, Q., Cao, X., Ma, R., Zhang, T., et al. (2017). Cyberphysical system with virtual reality for intelligent motion recognition and training. *IEEE Trans. Syst. Man, Cybern. Syst.* 47, 1–17. doi:10.1109/TSMC.2016.2560127

Hu, L., Xie, N., Kuang, Z., and Zhao, K. (2012). "Review of cyber-physical system architecture," in 2012 IEEE 15th International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops, Shenzhen, China, 11-11 April 2012, 25–30. doi:10.1109/ISORCW.2012.15

Hu, Y., Wu, M., Kang, J., and Yu, R. (2024). D-tracking: digital twin enabled trajectory tracking system of autonomous vehicles. *IEEE Trans. Veh. Technol.*, 1–13. doi:10.1109/TVT.2024.3414410

Huang, Z., Yu, R., Ye, M., Zheng, S., Kang, J., and Zeng, W. (2024). Digital twin edge services with proximity-aware longitudinal lane changing model for connected vehicles. *IEEE Trans. Veh. Technol.*, 1–15. doi:10.1109/TVT.2024.3412119

Hurl, B., Cohen, R., Czarnecki, K., and Waslander, S. (2020). "Trupercept: trust modelling for autonomous vehicle cooperative perception from synthetic data," in *2020 IEEE intelligent vehicles symposium (IV)* (IEEE), 341–347.

Isern, J., Barranco, F., Deniz, D., Lesonen, J., Hannuksela, J., and Carrillo, R. R. (2020). Reconfigurable cyber-physical system for critical infrastructure protection in smart cities via smart video-surveillance. *Pattern Recognit. Lett.* 140, 303–309. doi:10.1016/j.patrec.2020.11.004

Islam, S. O. B., Lughmani, W. A., Qureshi, W. S., Khalid, A., Mariscal, M. A., and Garcia-Herrero, S. (2019). Exploiting visual cues for safe and flexible cyber-physical production systems. *Adv. Mech. Eng.* 11, 168781401989722. doi:10.1177/1687814019897228

Jafari, M., Kavousi-Fard, A., Chen, T., and Karimi, M. (2023). A review on digital twin technology in smart grid, transportation system and smart city: challenges and future. *IEEE Access* 11, 17471–17484. doi:10.1109/ACCESS.2023.3241588

Jain, V., and Learned-Miller, E. (2010). *FDDB: a benchmark for face detection in unconstrained settings*. Amherst: University of Massachusetts. Tech. Rep. UM-CS-2010-009.

Jeong, M., Ko, B. C., Kwak, S., and Nam, J.-Y. (2018). Driver facial landmark detection in real driving situations. *IEEE Trans. Circuits Syst. Video Technol.* 28, 2753–2767. doi:10.1109/TCSVT.2017.2769096

Jhuang, H., Gall, J., Zuffi, S., Schmid, C., and Black, M. J. (2013). "Towards understanding action recognition," in 2013 IEEE International Conference on Computer Vision, 3192–3199. doi:10.1109/ICCV.2013.396

Jiang, F., Ma, L., Broyd, T., Chen, W., and Luo, H. (2022). Digital twin enabled sustainable urban road planning. *Sustain. Cities Soc.* 78, 103645. doi:10.1016/j.scs.2021.103645

Jiang, Z., Hu, M., Gao, Z., Fan, L., Dai, R., Pan, Y., et al. (2020). Detection of respiratory infections using rgb-infrared sensors on portable device. *IEEE Sensors J.* 20, 13674–13681. doi:10.1109/JSEN.2020.3004568

Jin, S., Xu, L., Xu, J., Wang, C., Liu, W., Qian, C., et al. (2020). "Whole-body human pose estimation in the wild," in *Computer vision – eccv 2020*. Editors A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm (Cham: Springer International Publishing), 196–214.

Johannsen, G. (1997). Conceptual design of multi-human machine interfaces. *Control Eng. Pract.* 5, 349–361. doi:10.1016/S0967-0661(97)00012-9

Johnson, S., and Everingham, M. (2010). "Clustered pose and nonlinear appearance models for human pose estimation," in Proceedings of the British Machine Vision Conference (Aberystwyth, UK: BMVA Press), 12.1–12.11. doi:10.5244/C.24.12

Joo, M., Seo, J., Oh, J., Park, M., and Lee, K. (2018). "Situational awareness framework for cyber crime prevention model in cyber physical system," in 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN), 837–842. doi:10.1109/ICUFN.2018.84365912

Kaburlasos, V. G., Lytridis, C., Bazinas, C., Chatzistamatis, S., Sotiropoulou, K., Najoua, A., et al. (2020a). "Head pose estimation using lattice computing techniques," in 2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 1–5. doi:10.23919/SoftCOM50211.2020.923831522

Kaburlasos, V. G., Lytridis, C., Bazinas, C., Papakostas, G. A., Naji, A., Zaggaf, M. H., et al. (2020b). "Structured human-head pose representation for estimation using fuzzy lattice reasoning (flr)," in 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)., 1–5. doi:10.1109/ICDS50568.2020.926876022

Kahn, J., Rivière, M., Zheng, W., Kharitonov, E., Xu, Q., Mazaré, P.-E., et al. (2020). "Libri-light: a benchmark for asr with limited or no supervision," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE), 7669–7673.

Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., et al. (2017). The kinetics human action video dataset

Ketzler, B., Naserentin, V., Latino, F., Zangelidis, C., Thuvander, L., and Logg, A. (2020). Digital twins for cities: a state of the art review. *Built Environ.* 46, 547–573. doi:10.2148/benv.46.4.547

Kharchenko, V., Orehov, A., Brezhnev, E., Orehova, A., and Manulik, V. (2014). "The cooperative human-machine interfaces for cloud-based advanced driver assistance systems: dynamic analysis and assurance of vehicle safety," in *Proceedings of IEEE east-west design test symposium (EWDTS 2014)*, 1–5. doi:10.1109/EWDTS.2014.7027096

Klein, G., and Murray, D. (2007). "Parallel tracking and mapping for small ar workspaces," in *2007 6th IEEE and ACM international symposium on mixed and augmented reality* (IEEE), 225–234.

Kopelias, P., Demiridi, E., Vogiatzis, K., Skabardonis, A., and Zafiropoulou, V. (2020). Connected and autonomous vehicles – environmental impacts – a review. *Sci. Total Environ.* 712, 135237. doi:10.1016/j.scitotenv.2019.135237

Köpüklü, O., Gunduz, A., Kose, N., and Rigoll, G. (2019). "Real-time hand gesture detection and classification using convolutional neural networks," in 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), 1–8. doi:10.1109/FG.2019.8756576

Köstinger, M., Wohlhart, P., Roth, P. M., and Bischof, H. (2011). "Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization," in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2144–2151. doi:10.1109/ICCVW.2011.6130513

Kozhirbayev, Z., Erol, B. A., Sharipbay, A., and Jamshidi, M. (2018). "Speaker recognition for robotic control via an iot device," in *2018 world automation congress (WAC)*, 1–5. doi:10.23919/WAC.2018.8430295

Kraft, A.-K., Maag, C., and Baumann, M. (2019). How to support cooperative driving by hmi design? *Transp. Res. Interdiscip. Perspect.* 3, 100064. doi:10.1016/j.trip.2019.100064

Kraft, A.-K., Maag, C., and Baumann, M. (2020). Comparing dynamic and static illustration of an hmi for cooperative driving. *Accid. Analysis and Prev.* 144, 105682. doi:10.1016/j.aap.2020.105682

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., and Serre, T. (2011). "Hmdb: a large video database for human motion recognition," in 2011 International Conference on Computer Vision, 2556–2563. doi:10.1109/ICCV.2011.6126543

Kurazume, R., Oshima, S., Nagakura, S., Jeong, Y., and Iwashita, Y. (2017). Automatic large-scale three dimensional modeling using cooperative multiple robots. *Comput. Vis. Image Underst.* 157, 25–42. doi:10.1016/j.cviu.2016.05.008

Kuriki, Y., and Namerikawa, T. (2015). "Formation control with collision avoidance for a multi-uav system using decentralized mpc and consensus-based control," in 2015 European Control Conference (ECC), 3079–3084. doi:10.1109/ECC.2015.7331006

Kuru, K. (2021). Conceptualisation of human-on-the-loop haptic teleoperation with fully autonomous self-driving vehicles in the urban environment. *IEEE Open J. Intelligent Transp. Syst.* 2, 448–469. doi:10.1109/OJITS.2021.3132725

Kušić, K., Schumann, R., and Ivanjko, E. (2023). A digital twin in transportation: real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Adv. Eng. Inf.* 55, 101858. doi:10.1016/j.aei.2022.101858

Kuutti, S., Fallah, S., Katsaros, K., Dianati, M., Mccullough, F., and Mouzakitis, A. (2018). A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications. *IEEE Internet Things J.* 5, 829–846. doi:10.1109/jiot.2018.2812300

Laffan, C. F., Coleshill, J. E., Stanfield, B., Stanfield, M., and Ferworn, A. (2020). "Using the araig haptic suit to assist in navigating firefighters out of hazardous environments," in 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (IEEE), 0439–0444.

Laghari, A., Memon, Z. A., Ullah, S., and Hussain, I. (2018). Cyber physical system for stroke detection. *IEEE Access* 6, 37444–37453. doi:10.1109/ACCESS.2018.2851540

Lai, C.-C., Shih, S.-W., and Hung, Y.-P. (2015). Hybrid method for 3-d gaze tracking using glint and contour features. *IEEE Trans. Circuits Syst. Video Technol.* 25, 24–37. doi:10.1109/TCSVT.2014.2329362

Lampropoulos, G., and Siakas, K. (2023). Enhancing and securing cyber-physical systems and industry 4.0 through digital twins: a critical review. *J. Softw. Evol. process* 35, e2494. doi:10.1002/smr.2494

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., and Beijbom, O. (2019). "Pointpillars: fast encoders for object detection from point clouds," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12697–12705.

Lassner, C., Romero, J., Kiefel, M., Bogo, F., Black, M. J., and Gehler, P. V. (2017). "Unite the people: closing the loop between 3d and 2d human representations," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4704–4713. doi:10.1109/CVPR.2017.500

Le, V., Brandt, J., Lin, Z., Bourdev, L., and Huang, T. S. (2012). "Interactive facial feature localization," in *Computer vision – ECCV 2012*. Editors A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Berlin, Heidelberg: Springer Berlin Heidelberg), 679–692.

Lee, H. J., Kim, S. T., Lee, H., and Ro, Y. M. (2020). Lightweight and effective facial landmark detection using adversarial learning with face geometric map generative network. *IEEE Trans. Circuits Syst. Video Technol.* 30, 771–780. doi:10.1109/TCSVT.2019.2897243

Leitão, P., Queiroz, J., and Sakurada, L. (2022). Collective intelligence in self-organized industrial cyber-physical systems. *Electronics* 11, 3213. doi:10.3390/electronics11193213

Li, B. (2017). "3d fully convolutional network for vehicle detection in point cloud," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 1513–1518.

Li, B., Zhang, T., and Xia, T. (2016a). Vehicle detection from 3d lidar using fully convolutional network. *arXiv Prepr. arXiv:1608.07916*. doi:10.48550/arXiv.1608.07916

Li, G., Huang, L., Tang, L., Han, C., Chen, Y., Xie, H., et al. (2020a). Person re-identification using additive distance constraint with similar labels loss. *IEEE Access* 8, 168111–168120. doi:10.1109/ACCESS.2020.3023948

Li, K., Chen, R., Nuchkrua, T., and Boonto, S. (2019). "Dual loop compliant control based on human prediction for physical human-robot interaction," in 2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 459–464. doi:10.23919/SICE.2019.8859792540

Li, M., Zhu, X., and Gong, S. (2020b). Unsupervised tracklet person re-identification. *IEEE Trans. Pattern Analysis Mach. Intell.* 42, 1770–1782. doi:10.1109/TPAMI.2019.2903058

Li, S., Ngan, K. N., Paramesran, R., and Sheng, L. (2016b). Real-time head pose tracking with online face template reconstruction. *IEEE Trans. Pattern Analysis Mach. Intell.* 38, 1922–1928. doi:10.1109/TPAMI.2015.2500221

Lillo, I., Soto, A., and Niebles, J. C. (2014). "Discriminative hierarchical modeling of spatio-temporally composable human activities," in *2014 IEEE conference on computer vision and pattern recognition*, 812–819. doi:10.1109/CVPR.2014.109

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2117–2125.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: common objects in context," in *Computer vision – ECCV 2014*. Editors D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer International Publishing), 740–755.

Liu, J., gen Cai, B., and Wang, J. (2017a). Cooperative localization of connected vehicles: integrating GNSS with DSRC using a robust cubature kalman filter. *IEEE Trans. Intelligent Transp. Syst.* 18, 2111–2125. doi:10.1109/tits.2016.2633999

Liu, J., Li, C., Bai, J., Luo, Y., Lv, H., and Lv, Z. (2023). Security in iot-enabled digital twins of maritime transportation systems. *IEEE Trans. Intelligent Transp. Syst.* 24, 1–9. doi:10.1109/TITS.2021.3122566

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: single shot multibox detector," in European conference on computer vision (Springer), 21–37.

Liu, Y., Peng, Y., Wang, B., Yao, S., and Liu, Z. (2017b). Review on cyber-physical systems. *IEEE/CAA J. Automatica Sinica* 4, 27–40. doi:10.1109/JAS.2017.7510349

Liu, Y., Wang, Z., Han, K., Shou, Z., Tiwari, P., and Hansen, J. H. (2020). "Sensor fusion of camera and cloud digital twin information for intelligent vehicles," in *2020 IEEE intelligent vehicles symposium (IV) (IEEE)*, 182–187.

Liu, Y., Xu, B., and Ding, Y. (2017). Convergence analysis of cooperative braking control for interconnected vehicle systems. *IEEE Trans. Intelligent Transp. Syst.* 18, 1894–1906. doi:10.1109/TITS.2016.2615302

Liu, Z., and Wang, J. (2020). Human-cyber-physical systems: concepts, challenges, and research opportunities. *Front. Inf. Technol. Electron. Eng.* 21, 1535–1553. doi:10.1631/fitee.2000537

Lou, S., Feng, Y., Tian, G., Lv, Z., Li, Z., and Tan, J. (2017). A cyber-physical system for product conceptual design based on an intelligent psycho-physiological approach. *IEEE Access* 5, 5378–5387. doi:10.1109/ACCESS.2017.2686986

Lou, Y., Wu, W., Vatavu, R.-D., and Tsai, W.-T. (2016). Personalized gesture interactions for cyber-physical smart-home environments. *Sci. China Inf. Sci.* 60, 072104. doi:10.1007/s11432-015-1014-7

Lozano, C. V., and Vijayan, K. K. (2020). Literature review on cyber physical systems design. *Procedia Manuf.* 45, 295–300. doi:10.1016/j.promfg.2020.04.020

Lv, Z., Chen, D., Feng, H., Lou, R., and Wang, H. (2021). Beyond 5g for digital twins of uavs. *Comput. Netw.* 197, 108366. doi:10.1016/j.comnet.2021.108366

Lv, Z., Li, Y., Feng, H., and Lv, H. (2022). Deep learning for security in digital twins of cooperative intelligent transportation systems. *IEEE Trans. Intelligent Transp. Syst.* 23, 16666–16675. doi:10.1109/TITS.2021.3113779

Ma, M., Lin, W., Pan, D., Lin, Y., Wang, P., Zhou, Y., et al. (2018). Data and decision intelligence for human-in-the-loop cyber-physical systems: reference model, recent progresses and challenges. *J. Signal Process. Syst.* 90, 1167–1178. doi:10.1007/s11265-017-1304-0

Majumder, A. J., Elsaadany, M., Izaguirre, J. A., and Ucci, D. R. (2019). A real-time cardiac monitoring using a multisensory smart iot system. *2019 IEEE 43rd Annu. Comput. Softw. Appl. Conf. (COMPSAC)* 2, 281–287. doi:10.1109/COMPSAC.2019.10220

Makihara, Y., Takizawa, M., Shirai, Y., Miura, J., and Shimada, N. (2002). Object recognition supported by user interaction for service robots. *Object Recognit. supported by user Interact. Serv. robots (IEEE)* 3, 561–564. doi:10.1109/icpr.2002.1048001

Makovetskii, A., Kober, V., Voronin, A., and Zhernov, D. (2020). "Facial recognition and 3d non-rigid registration," in 2020 International Conference on Information Technology and Nanotechnology (ITNT), 1–4. doi:10.1109/ITNT49337.2020.925322410752

Mariya Celin, T. A., Anushiya Rachel, G., Nagarajan, T., and Vijayalakshmi, P. (2019). A weighted speaker-specific confusion transducer-based augmentative and alternative speech communication aid for dysarthric speakers. *IEEE Trans. Neural Syst. Rehabilitation Eng.* 27, 187–197. doi:10.1109/TNSRE.2018.2887089

Meng, F., Shi, Y., Wang, N., Cai, M., and Luo, Z. (2020). Detection of respiratory sounds based on wavelet coefficients and machine learning. *IEEE Access* 8, 155710–155720. doi:10.1109/ACCESS.2020.3016748

Meyer, F., Hlinka, O., Wymeersch, H., Riegler, E., and Hlawatsch, F. (2016). Distributed localization and tracking of mobile networks including noncooperative objects. *IEEE Trans. Signal Inf. Process. over Netw.* 2, 57–71. doi:10.1109/tsipn.2015.2511920

Michael, N., Shen, S., Mohta, K., Mulgaonkar, Y., Kumar, V., Nagatani, K., et al. (2012). Collaborative mapping of an earthquake-damaged building via ground and aerial robots. *J. Field Robotics* 29, 832–841. doi:10.1002/rob.21436

Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., and Kautz, J. (2016). "Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4207–4215. doi:10.1109/CVPR.2016.456

Montanaro, U., Dixit, S., Fallah, S., Dianati, M., Stevens, A., Oxtoby, D., et al. (2018). Towards connected autonomous driving: review of use-cases. *Veh. Syst. Dyn.* 57, 779–814. doi:10.1080/00423114.2018.1492142

Mourikis, A. I., and Roumeliotis, S. I. (2006). Predicting the performance of cooperative simultaneous localization and mapping (c-slam). *Int. J. Robotics Res.* 25, 1273–1286. doi:10.1177/0278364906072515

Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Trans. Robotics* 31, 1147–1163. doi:10.1109/tro.2015.2463671

Mur-Artal, R., and Tardos, J. D. (2017). ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-d cameras. *IEEE Trans. Robotics* 33, 1255–1262. doi:10.1109/tro.2017.2705103

Mylonas, G., Kalogeras, A., Kalogeras, G., Anagnostopoulos, C., Alexakos, C., and Muñoz, L. (2021). Digital twins from smart manufacturing to smart cities: a survey. *IEEE Access* 9, 143222–143249. doi:10.1109/ACCESS.2021.3120843

Nagatani, K., Okada, Y., Tokunaga, N., Kiribayashi, S., Yoshida, K., Ohno, K., et al. (2011). Multirobot exploration for search and rescue missions: a report on map building in robocuprescue 2009. *J. Field Robotics* 28, 373–387. doi:10.1002/rob.20389

Naujoks, F., Forster, Y., Wiedemann, K., and Neukum, A. (2017). "A human-machine interface for cooperative highly automated driving," in *Advances in human aspects of transportation*. Editors N. A. Stanton, S. Landry, G. Di Bucchianico, and A. Vallicelli (Cham: Springer International Publishing), 585–595.

Neto, J. B. P., Gomes, L. C., Ortiz, F. M., Almeida, T. T., Campista, M. E. M., Costa, L. H. M., et al. (2020). An accurate cooperative positioning system for vehicular safety applications. *Comput. and Electr. Eng.* 83, 106591. doi:10.1016/j.compeleceng.2020.106591

Nikolov, P., Boumbarov, O., Manolova, A., Tonchev, K., and Poulkov, V. (2018). "Skeleton-based human activity recognition by spatio-temporal representation and convolutional neural networks with application to cyber physical systems with human in the loop," in 2018 41st International Conference on Telecommunications and Signal Processing (TSP), 1–5. doi:10.1109/TSP.2018.844117117

Noh, J., Lee, S., and Ham, B. (2021). "Hvpr: hybrid voxel-point representation for single-stage 3d object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 14605–14614.

Nota, G., Matonti, G., Bisogno, M., and Nastasia, S. (2020). The contribution of cyber-physical production systems to activity-based costing in manufacturing. an interventionist research approach. *Int. J. Eng. Bus. Manag.* 12, 184797902096230. doi:10.1177/1847979020962301

Noureldin, A., Karamat, T. B., and Georgy, J. (2013). *Fundamentals of inertial navigation, satellite-based positioning and their integration*. Springer Berlin Heidelberg. doi:10.1007/978-3-642-30466-8

Nunes, D. S., Zhang, P., and Sá Silva, J. (2015). A survey on human-in-the-loop applications towards an internet of all. *IEEE Commun. Surv. Tutorials* 17, 944–965. doi:10.1109/COMST.2015.2398816

Okpara, O. S., and Bekaroo, G. (2017). "Cam-wallet: fingerprint-based authentication in m-wallets using embedded cameras," in 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I CPS Europe), 1–5. doi:10.1109/EEEIC.2017.797765433

Oliveira, L. M. C., Dias, R., Rebello, C. M., Martins, M. A. F., Rodrigues, A. E., Ribeiro, A. M., et al. (2021). Artificial intelligence and cyber-physical systems: a review and perspectives for the future in the chemical industry. *AI* 2, 429–443. doi:10.3390/ai2030027

Orekhov, A., Orekhova, A., and Kharchenko, V. (2016a). Cooperative human-machine interfaces for safety of intelligent transport systems: requirements development and assessment. *WSEAS Trans. Comput. Res.* 4, 183–193.

Orekhov, A., Orekhova, A., and Kharchenko, V. (2016b). Ecological design of cooperative human-machine interfaces for safety of intelligent transport systems. *MATEC Web Conf.* 76, 02049. doi:10.1051/matecconf/20167602049

Panayotov, V., Chen, G., Povey, D., and Khudanpur, S. (2015). "Librispeech: an asr corpus based on public domain audio books," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5206–5210. doi:10.1109/ICASSP.2015.7178964

Pandey, A., Sequeria, R., Kumar, P., and Kumar, S. (2020). A multistage deep residual network for biomedical cyber-physical systems. *IEEE Syst. J.* 14, 1953–1962. doi:10.1109/JSYST.2019.2923670

Pedersen, N., Bojsen, T., and Madsen, J. (2017). "Co-simulation of cyber physical systems with hmi for human in the loop investigations," in *Proceedings of the symposium on theory of modeling and simulation*, 1–12.

Pereira Passarinho, C. J., Ottoni Teatini Salles, E., and Sarcinelli Filho, M. (2015). Face tracking in unconstrained color videos with the recovery of the location of lost faces. *IEEE Lat. Am. Trans.* 13, 307–314. doi:10.1109/TLA.2015.7040663

Piperigkos, N., Anagnostopoulos, C., Lalos, A. S., and Berberidis, K. (2023). Extending online 4d situational awareness in connected and automated vehicles. *IEEE Trans. Intelligent Veh.*, 1–19doi. doi:10.1109/TIV.2023.3335605

Piperigkos, N., Lalos, A. S., and Berberidis, K. (2020a). "Graph based cooperative localization for connected and semi-autonomous vehicles," in *2020 IEEE 25th international workshop on computer aided modeling and design of communication links and networks (CAMAD)* (IEEE), 1–6.

Piperigkos, N., Lalos, A. S., and Berberidis, K. (2021). Graph laplacian diffusion localization of connected and automated vehicles. *IEEE Trans. Intelligent Transp. Syst.* 23, 12176–12190. doi:10.1109/tits.2021.3110650

Piperigkos, N., Lalos, A. S., Berberidis, K., and Anagnostopoulos, C. (2020b). "Cooperative multi-modal localization in connected and autonomous vehicles," in *2020 IEEE 3rd connected and automated vehicles symposium (CAVS)* (IEEE), 1–5.

Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., de Amicis, R., et al. (2015). Visual computing as a key enabling technology for industrie 4.0 and industrial internet. *IEEE Comput. Graph. Appl.* 35, 26–40. doi:10.1109/MCG.2015.45

Prado, B., Dantas, D., Bispo, K., Fontes, T., Santana, G., and Silva, R. (2018). "A virtual prototype semihosting approach for early simulation of cyber-physical systems," in *2018 IEEE symposium on computers and communications (ISCC)*, 00208–00213. doi:10.1109/ISCC.2018.8538621

Preciozzi, J., Garella, G., Camacho, V., Franzoni, F., Di Martino, L., Carbajal, G., et al. (2020). Fingerprint biometrics from newborn to adult: a study from a national identity database system. *IEEE Trans. Biometrics, Behav. Identity Sci.* 2, 68–79. doi:10.1109/TBIOM.2019.2962188

Proenca, H., Filipe, S., Santos, R., Oliveira, J., and Alexandre, L. A. (2010). The ubiris.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Trans. Pattern Analysis Mach. Intell.* 32, 1529–1535. doi:10.1109/TPAMI.2009.66

Pulikottil, T., Estrada-Jimenez, L. A., Ur Rehman, H., Mo, F., Nikghadam-Hojjati, S., and Barata, J. (2023). Agent-based manufacturing—review and expert evaluation. *Int. J. Adv. Manuf. Technol.* 127, 2151–2180. doi:10.1007/s00170-023-11517-8

Purcell, W., Neubauer, T., and Mallinger, K. (2023). Digital twins in agriculture: challenges and opportunities for environmental sustainability. *Curr. Opin. Environ. Sustain.* 61, 101252. doi:10.1016/j.cosust.2022.101252

Putz, V., Mayer, J., Fenzl, H., Schmidt, R., Pichler-Scheder, M., and Kastl, C. (2020). Cyber–physical mobile arm gesture recognition using ultrasound and motion data. *2020 IEEE Conf. Industrial Cyberphysical Syst. (ICPS)* 1, 203–208. doi:10.1109/ICPS48405.2020.9274795

Qian, R., Lai, X., and Li, X. (2022). Badet: boundary-aware 3d object detection from point clouds. *Pattern Recognit.* 125, 108524. doi:10.1016/j.patcog.2022.108524

Quintal, F., and Lima, M. (2021). Hapwheel: in-car infotainment system feedback using haptic and hovering techniques. *IEEE Trans. Haptics* 15, 121–130. doi:10.1109/toh.2021.3095763

Rao, I. H., Amir, N. A., Dagale, H., and Kuri, J. (2012). "e-surakshak: a cyber-physical healthcare system with service oriented architecture," in *2012 international symposium on electronic system design (ISED)*, 177–182. doi:10.1109/ISED.2012.66

Ravichandar, H. C., and Dani, A. P. (2017). Human intention inference using expectation-maximization algorithm with online model learning. *IEEE Trans. Automation Sci. Eng.* 14, 855–868. doi:10.1109/TASE.2016.2624279

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788.

Redmon, J., and Farhadi, A. (2018). *Yolov3: an incremental improvement*. *arXiv preprint arXiv:1804.02767*.

Ren, K., Wang, Q., Wang, C., Qin, Z., and Lin, X. (2020). The security of autonomous driving: threats, defenses, and future directions. *Proc. IEEE* 108, 357–372. doi:10.1109/jproc.2019.2948775

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. pattern analysis Mach. Intell.* 39, 1137–1149. doi:10.1109/tpami.2016.2577031

Richey, C., Barrios, M. A., Armstrong, Z., Bartels, C., Franco, H., Graciarena, M., et al. (2018). Voices obscured in complex environmental settings (voices) corpus.

Rusu, R. B., Blodow, N., and Beetz, M. (2009). "Fast point feature histograms (fpfh) for 3d registration," in 2009 IEEE international conference on robotics and automation (IEEE), 3212–3217.

Saatci, E., and Saatci, E. (2021). Determination of respiratory parameters by means of hurst exponents of the respiratory sounds and stochastic processing methods. *IEEE Trans. Biomed. Eng.* 68, 3582–3592. doi:10.1109/TBME.2021.3079160

Sadiku, M. N. O., Wang, Y., Cui, S., and Musa, S. M. (2017). Cyber-physical systems: a literature review. *Eur. Sci. J. ESJ* 13, 52. doi:10.19044/esj.2017.v13n36p52

Saeedi, S., Trentini, M., Seto, M., and Li, H. (2016). Multiple-robot simultaneous localization and mapping: a review. *J. Field Robotics* 33, 3–46. doi:10.1002/rob.21620

Safavi, S., Khan, U. A., Kar, S., and Moura, J. M. F. (2018). Distributed localization: a linear theory. *Proc. IEEE* 106, 1204–1223. doi:10.1109/jproc.2018.2823638

Sagaya Aurelia, P. (2019). "Haptics: prominence and challenges," in *Human behaviour analysis using intelligent systems* (Springer), 21–43.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., and Pantic, M. (2013). "300 faces in-the-wild challenge: the first facial landmark localization challenge," in 2013 IEEE International Conference on Computer Vision Workshops, 397–403. doi:10.1109/ICCVW.2013.59

Schuldt, C., Laptev, I., and Caputo, B. (2004). Recognizing human actions: a local svm approach. *Proc. 17th Int. Conf. Pattern Recognit. 2004. ICPR 2004* 3, 32–36. doi:10.1109/ICPR.2004.1334462

Schwarz, C., and Wang, Z. (2022). The role of digital twins in connected and automated vehicles. *IEEE Intell. Transp. Syst. Mag.* 14, 41–51. doi:10.1109/MITS.2021.3129524

Serafin, J., and Grisetti, G. (2015). "Nicp: dense normal based point cloud registration," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 742–749.

Shahroudy, A., Liu, J., Ng, T.-T., and Wang, G. (2016). *Ntu rgb+d: a large scale dataset for 3d human activity analysis*.

Shakil, M., and Zoitl, A. (2020). Towards a modular architecture for industrial hmis. *2020 25th IEEE Int. Conf. Emerg. Technol. Fact. Automation (ETFA)* 1, 1267–1270. doi:10.1109/ETFA46521.2020.9212011

Shan, T., and Englot, B. (2018). "Lego-loam: lightweight and ground-optimized lidar odometry and mapping on variable terrain," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 4758–4765.

Shen, Z., Liu, Y., Li, Z., and Nabin, M. H. (2022). Cooperative spacing sampled control of vehicle platoon considering undirected topology and analog fading networks. *IEEE Trans. Intelligent Transp. Syst.* 23, 18478–18491. doi:10.1109/TITS.2022.3150565

Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., et al. (2020). "Pv-rcnn: point-voxel feature set abstraction for 3d object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10529–10538.

Shi, S., Jiang, L., Deng, J., Wang, Z., Guo, C., Shi, J., et al. (2023). Pv-rcnn++: point-voxel feature set abstraction with local vector representation for 3d object detection. *Int. J. Comput. Vis.* 131, 531–551. doi:10.1007/s11263-022-01710-9

Shi, S., Wang, X., and Li, H. (2019). "Pointrcnn: 3d object proposal generation and detection from point cloud," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 770–779.

Shi, W., and Rajkumar, R. (2020). "Point-gnn: graph neural network for 3d object detection in a point cloud," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 1711–1719.

Singh, H. V., and Mahmoud, Q. H. (2017). "Eye-on-hmi: a framework for monitoring human machine interfaces in control rooms," in 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 1–5. doi:10.1109/CCECE.2017.7946695

Skog, I., and Handel, P. (2009). In-car positioning and navigation technologies—a survey. *IEEE Trans. Intelligent Transp. Syst.* 10, 4–21. doi:10.1109/tits.2008.2011712

Soatti, G., Nicoli, M., Garcia, N., Denis, B., Raulefs, R., and Wymeersch, H. (2018). Implicit cooperative positioning in vehicular networks. *IEEE Trans. Intelligent Transp. Syst.* 19, 3964–3980. doi:10.1109/tits.2018.2794405

Soliman, A., Al-Ali, A., Mohamed, A., Gedawy, H., Izham, D., Bahri, M., et al. (2023). Ai-based uav navigation framework with digital twin technology for mobile target visitation. *Eng. Appl. Artif. Intell.* 123, 106318. doi:10.1016/j.engappai.2023.106318

Sowe, S. K., Simmon, E., Zettsu, K., de Vaulx, F., and Bojanova, I. (2016). Cyber-physical-human systems: putting people in the loop. *IT Prof.* 18, 10–13. doi:10.1109/MITP.2016.14

Steinbrücker, F., Sturm, J., and Cremers, D. (2011). "Real-time visual odometry from dense rgb-d images," in 2011 IEEE international conference on computer vision workshops (ICCV Workshops) (IEEE), 719–722.

Subhash, S., Srivatsa, P. N., Siddesh, S., Ullas, A., and Santhosh, B. (2020). "Artificial intelligence-based voice assistant," in 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 593–596. doi:10.1109/WorldS450073.2020.9210344

Taeihagh, A., and Lim, H. S. M. (2019). Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transp. Rev.* 39, 103–128. doi:10.1080/01441647.2018.1494640

Tan, M., Pang, R., and Le, Q. V. (2020). "Efficientdet: scalable and efficient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 10781–10790.

Tang, X., Li, X., Yu, R., Wu, Y., Ye, J., Tang, F., et al. (2023). Digital-twin-assisted task assignment in multi-uav systems: a deep reinforcement learning approach. *IEEE Internet Things J.* 10, 15362–15375. doi:10.1109/jiot.2023.3263574

Tao, F., Qi, Q., Wang, L., and Nee, A. (2019). Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: correlation and comparison. *Engineering* 5, 653–661. doi:10.1016/j.eng.2019.01.014

Tejedor-García, C., Escudero-Mancebo, D., Cámara-Arenas, E., González-Ferreras, C., and Cardeñoso-Payo, V. (2020). Assessing pronunciation improvement in students of English using a controlled computer-assisted pronunciation tool. *IEEE Trans. Learn. Technol.* 13, 269–282. doi:10.1109/TLT.2020.2980261

Uhlemann, E. (2016). Connected-vehicles applications are emerging [connected vehicles]. *IEEE Veh. Technol. Mag.* 11, 25–96. doi:10.1109/MVT.2015.2508322

Valdes-Ramirez, D., Medina-Pérez, M. A., Monroy, R., Loyola-González, O., Rodríguez, J., Morales, A., et al. (2019). A review of fingerprint feature representations and their applications for latent fingerprint identification: trends and evaluation. *IEEE Access* 7, 48484–48499. doi:10.1109/ACCESS.2019.2909497

Vandersteegen, M., Reusen, W., Beeck, K. V., and Goedemé, T. (2020). "Low-latency hand gesture recognition with a low resolution thermal imager," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 440–449. doi:10.1109/CVPRW50498.2020.00057

Verhagen, R. S., Neerincx, M. A., and Tielman, M. L. (2022). The influence of interdependence and a transparent or explainable communication style on human-robot teamwork. *Front. Robotics AI* 9, 993997. doi:10.3389/frobt.2022.993997

Viana, I. B., and Aouf, N. (2018). Distributed cooperative path-planning for autonomous vehicles integrating human driver trajectories. , 655–661. doi:10.1109/IS.2018.8710544

Viana, I. B., Kanchwala, H., and Aouf, N. (2019). "Cooperative trajectory planning for autonomous driving using nonlinear model predictive control," in 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE), Graz, Austria, 04-08 November 2019, 1–6. doi:10.1109/ICCVE45908.2019.8965227

Vicente, F., Huang, Z., Xiong, X., De la Torre, F., Zhang, W., and Levi, D. (2015). Driver gaze tracking and eyes off the road detection system. *IEEE Trans. Intelligent Transp. Syst.* 16, 2014–2027. doi:10.1109/TITS.2015.2396031

Vidulich, M., Dominguez, C., Vogel, E., and McMillan, G. (1994). *Situation awareness: papers and annotated bibliography*. Tech. rep. Armstrong lab wright-patterson afb oh crew systems directorate.

Wang, D., and Zhang, X. (2015). Thchs-30: a free Chinese speech corpus

Wang, H., Hao, W., So, J., Chen, Z., and Hu, J. (2023a). A faster cooperative lane change controller enabled by formulating in spatial domain. *IEEE Trans. Intelligent Veh.* 8, 4685–4695. doi:10.1109/TIV.2023.3317957

Wang, P. (2020). "Research and design of smart home speech recognition system based on deep learning," in 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), Chongqing, China, 10-12 July 2020, 218–221. doi:10.1109/CVIDL51233.2020.00-98

Wang, P., Di, B., Zhang, H., Bian, K., and Song, L. (2019a). Platoon cooperation in cellular V2X networks for 5G and beyond. *IEEE Trans. Wirel. Commun.* 18, 3919–3932. doi:10.1109/TWC.2019.2919602

Wang, W., Li, X., Xie, L., Lv, H., and Lv, Z. (2021). Unmanned aircraft system airspace structure and safety measures based on spatial digital twins. *IEEE Trans. Intelligent Transp. Syst.* 23, 2809–2818. doi:10.1109/tits.2021.3108995

Wang, X., Gong, H., Zhang, H., Li, B., and Zhuang, Z. (2006). Palmprint identification using boosting local binary pattern. *18th Int. Conf. Pattern Recognit. (ICPR'06)* 3, 503–506. doi:10.1109/ICPR.2006.912

Wang, Y., Su, Z., Guo, S., Dai, M., Luan, T. H., and Liu, Y. (2023b). A survey on digital twins: architecture, enabling technologies, security and privacy, and future prospects. *IEEE Internet Things J.* 10, 14965–14987. doi:10.1109/jiot.2023.3263909

Wang, Z., Duan, S., Zeng, C., Yu, X., Yang, Y., and Wu, H. (2020). "Robust speaker identification of iot based on stacked sparse denoising auto-encoders," in 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), Rhodes, Greece, 02-06 November 2020, 252–257. doi:10.1109/ithings-greencom-cpscom-smartdata-cybermatics50389.2020.00056313

Wang, Z., Han, J.-J., and Miao, T. (2019b). "A framebuffer oriented graphical human-machine interaction mechanism for intelligent in-vehicle systems," in 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14-17 July 2019, 202–207. doi:10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00054

Warden, P. (2018). Speech commands: a dataset for limited-vocabulary speech recognition

Wen, H., Liu, S., Zheng, X., Cai, G., Zhou, B., Ding, W., et al. (2024). The digital twins for mine site rescue environment: application framework and key technologies. *Process Saf. Environ. Prot.* 186, 176–188. doi:10.1016/j.psep.2024.04.007

Wetzler, A., Slossberg, R., and Kimmel, R. (2015). "Rule of thumb: deep derotation for improved fingertip detection," in Proceedings of the British Machine Vision Conference (BMVC) (Swansea, UK: BMVA Press), 33.1–33.12. doi:10.5244/C.29.33

Whelan, T., Johannsson, H., Kaess, M., Leonard, J. J., and McDonald, J. (2013). "Robust real-time visual odometry for dense rgb-d mapping," in 2013 IEEE International Conference on Robotics and Automation (IEEE), Karlsruhe, Germany, 06-10 May 2013, 5724–5731.

Wu, B., Iandola, F., Jin, P. H., and Keutzer, K. (2017). "Squeezedet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 129–137.

Wymeersch, H., Lien, J., and Win, M. Z. (2009). Cooperative localization in wireless networks. *Proc. IEEE* 97, 427–450. doi:10.1109/jproc.2008.2008853

Xie, G., Yang, K., Xu, C., Li, R., and Hu, S. (2022a). Digital twinning based adaptive development environment for automotive cyber-physical systems. *IEEE Trans. Industrial Inf.* 18, 1387–1396. doi:10.1109/TII.2021.3064364

Xie, S., Hu, J., Bhowmick, P., Ding, Z., and Arvin, F. (2022b). Distributed motion planning for safe autonomous vehicle overtaking via artificial potential field. *IEEE Trans. Intelligent Transp. Syst.* 23, 21531–21547. doi:10.1109/TITS.2022.3189741

Yan, M., Gan, W., Zhou, Y., Wen, J., and Yao, W. (2022). Projection method for blockchain-enabled non-iterative decentralized management in integrated natural gas-electric systems and its application in digital twin modelling. *Appl. Energy* 311, 118645. doi:10.1016/j.apenergy.2022.118645

Yan, Y., Mao, Y., and Li, B. (2018). Second: sparsely embedded convolutional detection. *Sensors* 18, 3337. doi:10.3390/s18103337

Yang, B., Yan, J., Lei, Z., and Li, S. Z. (2015). Fine-grained evaluation on face detection in the wild. *2015 11th IEEE Int. Conf. Work. Automatic Face Gesture Recognit. (FG) (IEEE)* 1, 1–7. doi:10.1109/FG.2015.7163158

Yang, P., Duan, D., Chen, C., Cheng, X., and Yang, L. (2020a). Multi-sensor multi-vehicle (MSMV) localization and mobility tracking for autonomous driving. *IEEE Trans. Veh. Technol.* 69, 14355–14364. doi:10.1109/tvt.2020.3031900

Yang, Z., Li, T.-H., and Jiang, W.-Q. (2018). "Situation awareness for cyber-physical system: a case study of advanced metering infrastructure," in 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), Seattle, WA, USA, 11-13 June 2018, 1–6. doi:10.1109/ICPHM.2018.84488868

Yang, Z., Sun, Y., Liu, S., and Jia, J. (2020b). "3dssd: point-based 3d single stage object detector," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, 13-19 June 2020, 11040–11048.

Ye, M., and Yuen, P. C. (2020). Purifynet: a robust person re-identification model with noisy labels. *IEEE Trans. Inf. Forensics Secur.* 15, 2655–2666. doi:10.1109/TIFS.2020.2970590

Yilma, B. A., Panetto, H., and Naudet, Y. (2021). Systemic formalisation of cyber-physical-social system (cpss): a systematic literature review. *Comput. Industry* 129, 103458. doi:10.1016/j.compind.2021.103458

Zanchettin, A. M., Casalino, A., Piroddi, L., and Rocco, P. (2019). Prediction of human activity patterns for human–robot collaborative assembly tasks. *IEEE Trans. Industrial Inf.* 15, 3934–3942. doi:10.1109/TII.2018.2882741

Zhang, J., Shu, Y., and Yu, H. (2021). "Human-machine interaction for autonomous vehicles: a review," in *Social computing and social media: experience design and social network analysis*. Editor G. Meiselwitz (Cham: Springer International Publishing), 190–201.

Zhang, J., and Singh, S. (2014). "Loam: lidar odometry and mapping in real-time," in *Robotics: science and systems (berkeley, CA)*, 2, 1–9.

Zhang, J., and Singh, S. (2015). "Visual-lidar odometry and mapping: low-drift, robust, and fast," in 2015 IEEE international conference on robotics and automation (ICRA) (IEEE), 2174–2181.

Zhang, J., and Tao, D. (2020). Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things

Zhang, S., Pöhlmann, R., Wiedemann, T., Dammann, A., Wymeersch, H., and Hoeher, P. A. (2020). Self-aware swarm navigation in autonomous exploration missions. *Proc. IEEE* 108, 1168–1195. doi:10.1109/JPROC.2020.2985950

Zhang, Y., Cao, C., Cheng, J., and Lu, H. (2018). Egogesture: a new dataset and benchmark for egocentric hand gesture recognition. *IEEE Trans. Multimedia* 20, 1038–1050. doi:10.1109/TMM.2018.2808769

Zhao, D., and Oh, J. (2021). Noticing motion patterns: a temporal cnn with a novel convolution operator for human trajectory prediction. *IEEE Robotics Automation Lett.* 6, 628–634. doi:10.1109/LRA.2020.3047771

Zhao, H., Torralba, A., Torresani, L., and Yan, Z. (2019a). "Hacs: human action clips and segments dataset for recognition and temporal localization," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 8668–8678.

Zhao, J., Zhang, Y., Ni, S., and Li, Q. (2020). Bayesian cooperative localization with NLOS and malicious vehicle detection in GNSS-challenged environments. *IEEE Access* 8, 85686–85697. doi:10.1109/access.2020.2992338

Zhao, Z.-Q., Zheng, P., Xu, S.-t., and Wu, X. (2019b). Object detection with deep learning: a review. *IEEE Trans. neural Netw. Learn. Syst.* 30, 3212–3232. doi:10.1109/tnnls.2018.2876865

Zhou, X. S., and Roumeliotis, S. I. (2008). Robot-to-robot relative pose estimation from range measurements. *IEEE Trans. Robotics* 24, 1379–1393. doi:10.1109/TRO.2008.2006251

Zhou, Y., and Tuzel, O. (2018). "Voxelnet: end-to-end learning for point cloud based 3d object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4490–4499.

Zhou, Y., Yu, F. R., Chen, J., and Kuo, Y. (2020). Cyber-physical-social systems: a state-of-the-art survey, challenges and opportunities. *IEEE Commun. Surv. and Tutorials* 22, 389–425. doi:10.1109/COMST.2019.2959013

Zhu, X., Lei, Z., Liu, X., Shi, H., and Li, S. Z. (2016). "Face alignment across large poses: a 3d solution," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27-30 June 2016, 146–155.

Zhu, X., and Ramanan, D. (2012). "Face detection, pose estimation, and landmark localization in the wild," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16-21 June 2012, 2879–2886. doi:10.1109/CVPR.2012.6248014

Glossary

ADAS	Advanced driver assistance system
AR	Augmented reality
CAV	Connected and automated vehicle
CL	Cooperative localization
CNN	Convolutional neural network
CPS	Cyber–physical system
CPSoS	Cyber–physical system of systems
CPSS	Cyber–physical–social system
DT	Digital Twin
FPFH	Fast point feature histogram
GD	Gradient descent
GNSS	Global navigation satellite system
GPS	Global Positioning System
HITL	Human in the loop
HMI	Human–machine interface
ICP	Iterative closest point
IMU	Inertial measurement unit
IoT	Internet of Things
ITS	Intelligent transportation system
KF	Kalman filter
KPI	Key performance indicator
LiDAR	Light detection and ranging
LO	LiDAR odometry
LOAM	LiDAR odometry and mapping
LS	Least squares
RSS	Received signal strength
SLAM	Simultaneous localization and mapping
SPAWN	Sum–product algorithm over wireless networks
UAV	Unmanned aerial vehicle
UGV	Unmanned ground vehicle
UUV	Unmanned underwater vehicle
V2I	Vehicle-to-infrastructure
V2V	Vehicle-to-vehicle
V2X	Vehicle-to-everything
VANET	Vehicular *ad hoc* network
VO	Visual odometry
XR	eXtended Reality
WSN	Wireless sensor network