IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 25 February 2025, accepted 27 April 2025, date of publication 30 April 2025, date of current version 8 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3565901

SURVEY

Trending Machine Learning Methods for Vehicle, Pedestrian, and Traffic for Detection and Tracking Task in the Post-Covid Era: A Literature Review

JOÃO VITOR DE ANDRADE PORTO[1], **PÉTER TAMÁS SZEMES**[2], **HEMERSON PISTORI**[1], **AND JÓZSEF MENYHÁRT**[2]

[1]INOVISÃO, Dom Bosco Catholic University, Campo Grande 79117-900, Brazil
[2]Faculty of Engineering, Vehicles and Mechatronics Institute, Department of Vehicles Engineering, University of Debrecen, 4032 Debrecen, Hungary

Corresponding author: João Vitor de Andrade Porto (jvaporto@gmail.com)

This work was supported in part by Dom Bosco Catholic University; and in part by the Foundation for the Support and Development of Education, Science and Technology from the State of Mato Grosso do Sul (FUNDECT).

ABSTRACT This study, aimed at professionals in research and development in the fields of computer vision, artificial intelligence, and intelligent transportation, presents a systematic literature review on recent machine learning methodologies applied to the detection and tracking of vehicles, pedestrians, and traffic flow. The analysis of articles published between 2022 and 2025 (early access) in the post-COVID era explored the integration of machine learning and deep learning to address traffic challenges, allowing for the comparison of different approaches and the formulation of hypotheses based on the 46 articles that comprised the review corpus. Furthermore, the evaluation of the reported metrics revealed inconsistencies in the methodologies employed, attributed to the lack of standardization across the studies. In light of this, this work proposes alternatives for future experiments, emphasizing the emerging potential of the field through the adoption of new standardization systems and the exploration of experimental combinations.

INDEX TERMS Deep learning, detection, machine learning, tracking, urban mobility.

I. INTRODUCTION

The transformation of the global economy is undeniable, as the accelerated gains driven by the COVID-19 pandemic have placed significant pressure on all manufacturing sectors, from basic goods to luxury products, including the automotive industry. This shift has necessitated the development of competencies, transparency, and sustainability within automotive supply chains, both to enhance routine operations in response to the multi-level supply crisis and to support market policymakers in addressing these challenges [1].

Additionally, the post-COVID recovery of the European, North American, and Asian economies has been accompanied by the need to control global warming and increase electricity usage at the expense of fossil fuels. This energy and climate policy perspective has accelerated adaptations for a new era, highlighting vehicle sharing, electrification, and automation as key elements of the urban transportation revolution. The technological advancements associated with the deployment of shared autonomous vehicles further enhance mobility [2], [3].

Reference [4] confirm that one of the largest contributions to the global increase in air pollution is due to carbon emissions from vehicles and diesel engines, which serve as triggers to initiate or exacerbate respiratory diseases. Therefore, societal participation is crucial in the development of policies necessary for a high quality of life, identifying which drivers should be prioritized for change and providing data for the validation of corrective actions.

Even almost two decades after the publication of the Organisation for Economic Co-operation and Development (OECD) in 2008 [5] addressing issues related to the economic development of societies while respecting ecosystems and natural resources, the topic remains relevant. It proposes that

The associate editor coordinating the review of this manuscript and approving it for publication was Angel F. García-Fernández.

governments and societies find the right tools and policies to promote correct production and consumption practices, through the availability of data on urban logistics flows and the key tasks performed by the population, serving to guide the environmental, economic, and social directions of a city, promoting sustainability.

In this context, human activities have a defined impact on the environment and urban aesthetics in developed countries, and serve as a guide for defining planning systems, particularly regarding deterioration due to the expansion and restructuring of urban space. Associated with this growth is the increase in automobile traffic, and the imminent need for spaces used as vehicle parking, altering the geography of spaces not only in the urban core but also in suburban areas [6].

The automotive sector is a significant driver of economic growth for Hungary, playing an important role in Europe in the production of road vehicles. It is one of the key production hubs for the automotive industry, particularly for German companies, with recent investments focused on electromobility. This provides an economic gain, as within the global value chain, the automotive industry plays a prominent role in the country's average income [7].

It is undeniable that growth and development contribute to the restriction of mobility and urban access, directly impacting infrastructure with limitations on vehicle flow, whether private or public. This also includes freight vehicles and other forms of transportation. Projections based on historical data patterns of human-driven vehicle ownership and GDP per capita, as highlighted by [8], suggest that autonomous passenger vehicles in Hungary, although seen as a means to reduce carbon emissions, will lead to saturation in the automotive market and directly affect traffic flow.

Therefore, understanding transportation models is essential for the formulation of urban policies and the implementation of urban mobility modes to achieve the European Union's zero-emission target for the transportation sector by 2050. The initial focus lies on the effectiveness of urban transportation policies and the transition from internal combustion engine vehicles to battery electric vehicles; however, in the long term, road space will require appropriate pricing to accommodate increasingly sophisticated car flow and collective mobility systems [9].

Willing to contribute to the understanding of pedestrian and vehicle dynamics under traffic conditions in cities, this literature review identifies the trend of how machine learning methods are effective tools for detecting and tracking study targets, and, in a precise manner, collecting metrics that will support actions and drivers for intelligent urban growth in this post-COVID era.

II. MATERIALS AND METHODS

Regarding the methodological criteria, this study adopts a qualiquantitative approach, following the methodology proposed by [10], which is based on the selection of relevant sources from the specialized scientific literature. The selected elements were organized into instructional content, emphasizing practical applications for solving concrete problems and seeking viable solutions [11]. In the initial phase of the research, the main descriptors were defined using the following search query: ("vehicle" OR "car" OR "automobile" OR "truck" OR "bus" OR "traffic") AND ("detection" OR "tracking") AND ("computer vision"). This search was conducted in the Web of Science database and supplemented by direct access to journals to obtain complete citations when necessary. The formulation of the search query prioritized detection and tracking tasks, enabling a more in-depth analysis of these topics and the proposal of future experiments related to these issues.

The initial search yielded a total of 34,715 papers, which underwent a rigorous refinement process through the application of four successive filters, ensuring the selection of the most relevant studies for the scope of this review. Initially, the analysis was restricted to publications from the period between 2022 and 2025, including those available in early access, in order to encompass only works from the post-COVID era. This criterion led to the exclusion of a significant portion of the documents, leaving 1,692 papers. Subsequently, the removal of 27 studies written in languages other than English further reduced this number to 1,665 publications.

Subsequently, a manual scanning of the 1,665 abstracts was conducted, simultaneously evaluating two main factors. The first factor involved assessing the type of publication, retaining only articles focused on experimentation through the practical development of methodologies, which led to the exclusion of 924 studies. The second factor examined the alignment of the remaining publications with the research scope of this review, prioritizing those addressing the detection and tracking of vehicles and pedestrians in urban traffic scenarios using artificial intelligence and machine learning methodologies, resulting in the removal of an additional 274 articles.

As a result of this screening process, the abstract scanning yielded a final set of 467 publications, which were subsequently used in the in-depth analysis to construct the textual corpus. All procedures, from the initial time-based filtering to the abstract scanning stage, are graphically represented in the step diagram shown in Figure 1, in the "Filter Step" section.

For the composition of the textual corpus of the literature review, the 467 resulting articles were read and evaluated based on the following inclusion criteria: (1) analysis of titles, keywords, and abstracts to identify the central problem addressed and the proposed solution, verifying whether their combination aligns with the scope of the review; (2) assessment of the methodology and obtained results to ensure coherence and continuity between these elements in solving the presented problems; and (3) examination of the discussion of results and conclusion, considering the applicability and interpretations provided by the authors, with a focus on the investigated problem and perspectives for future research.

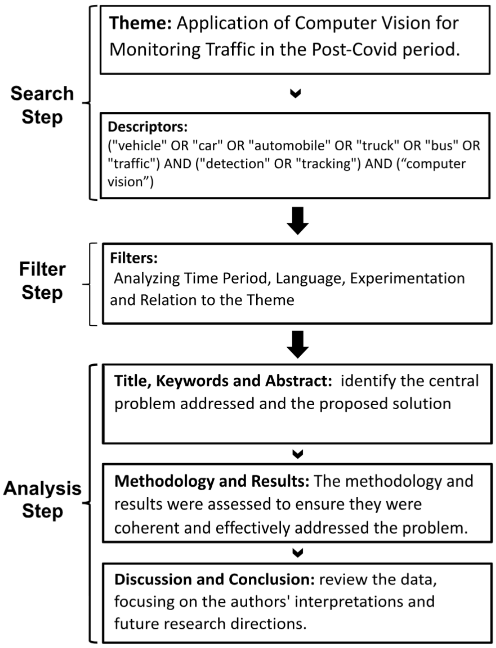


FIGURE 1. Flowchart of the steps taken for the formation of the paper's textual corpus.

Based on this analysis, the textual corpus was composed of 46 articles. The progressive reduction in the number of publications is represented in Figure 2, where each refinement step is illustrated. The corpus analysis involved the detailed classification of each article based on fundamental aspects for understanding and comparing the investigated proposals. Initially, the primary object of study of each work was identified and categorized as Vehicles, Pedestrians, or Traffic, according to the focus adopted by the authors. Additionally, the tasks addressed in the studies were analyzed and classified into four main categories: Detection, Tracking, Counting, and Classification, reflecting the investigated problems.

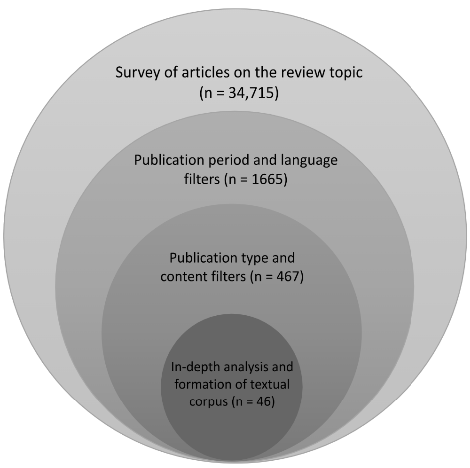


FIGURE 2. Diagram representing the number of publications analyzed at each stage of the methodology in a visual format.

Regarding the models employed, the analysis was conducted in two complementary dimensions. Initially, a precise identification of the architecture proposed in each article was made. Subsequently, these architectures were grouped into one or more of the following general categories: Region-based, Attention-based, Transformer-based, Feature-based, and YOLO (You Only Look Once) [12] family, with the aim of establishing patterns and relationships among the adopted approaches. Although the Attention-based and Transformer-based categories share similarities in their foundations, they were differentiated based on the use of the attention mechanism.

When the mechanism was applied locally to specific parts of the image as an additional module, the architecture was classified as Attention-based. On the other hand, when the mechanism was used as the main structure of the network, replacing conventional convolution operations and learning from the relationships between all input data, the Transformer-based category was assigned. Additionally, the quality metrics presented in each study were analyzed, as well as the datasets used for training, validation, and testing—key elements for assessing the robustness and effectiveness of the employed methodologies.

Thus, the main contributions of this work, related to the theme "Trending Machine Learning Methods for Vehicle, Pedestrian, and Traffic Detection and Tracking Task in the Post-Covid Era: A Literature Review", can be described as follows: (1) a search key was strategically formulated, not only considering terms like "Traffic," but also specific variations of "Vehicle." This focus allowed for an in-depth analysis of the application of machine learning and deep learning methodologies, with particular emphasis on detection and tracking techniques, which are crucial for urban mobility in the post-Covid context. From the Web of Science database, it was possible to compile the key studies that directly address this issue, establishing a solid foundation for detailed comparisons between the most relevant works; (2) the thorough analysis of the textual corpus allowed for a critical evaluation of the studies, considering not only the architectures used but also the datasets employed, which is essential to understand the effectiveness and robustness of the proposed solutions in vehicle, pedestrian, and traffic detection and tracking contexts; (3) finally, the review revealed important gaps in the existing literature, suggesting promising directions for future experiments and proposing innovative solutions that can fill these gaps, with a focus on the potential of machine learning and computer vision technologies applied to the evolution of urban mobility in the post-Covid period.

III. RESULTS

Based on the information and methods presented and detailed in the methodology of this study, Table 1 was created as a relational indicator between the analyzed works, their respective areas of focus, the tasks they aim to address, and the categories of networks employed. This organization

TABLE 1. Organizing the papers of the textual corpus according to their respective data collected during the in-depth analysis step.

REF	SUBJECT	TASKS	CATEGORIES	DATASETS
[13]	Vehicles	Detection, Tracking and Classification	Region-based	BIT-Vehicle and Apollo Scape
[14]	Traffic	Detection, Tracking and Classification	YOLO family	Particular
[15]	Vehicles	Detection and Tracking	YOLO family	UA-DETRAC
[16]	Vehicles	Detection, Tracking and Classification	YOLO family	UA-DETRAC and Track4
[17]	Vehicles	Detection, Tracking and Counting	YOLO family and Attention-based	Particular
[18]	Vehicles	Detection and Classification	Attention-based	VLD-45
[19]	Vehicles	Detection and Classification	YOLO family and Transformer-based	Particular
[20]	Traffic	Detection and Tracking	YOLO family	UA-DETRAC
[21]	Vehicles	Detection and Classification	YOLO family and Region-based	FLIR
[22]	Traffic	Classification	Transformer-based	KITTI
[23]	Traffic	Detection	Region-based	Particular
[24]	Traffic	Detection	Transformer-based and Attention-based	K-lane
[25]	Traffic	Detection and Classification	YOLO family and Attention-based	FLIR and M3FD
[26]	Vehicles	Detection and Classification	Attention-based	Particular
[27]	Pedestrians	Detection	Attention-based	KAIST and LLVIP
[28]	Traffic	Detection and Classification	YOLO family and Attention-based	Particular
[29]	Pedestrians	Detection	Attention-based	KAIST, LLVIP, FLIR and M3FD
[30]	Pedestrians	Detection	Region-based	CVC-14 night/visible, night/FIR and INRIA
[31]	Traffic	Detection and Tracking	YOLO family	KITTI, LASIESTA, PESMOD and MOCS
[32]	Traffic	Detection and Classification	Transformer-based	ViWi
[33]	Vehicles	Detection	Region-based and YOLO family	KITTI and DAIR-V2X
[34]	Vehicles	Detection	Region-based and YOLO family	Caltech Cars 1999 and Caltech Cars 2001
[35]	Vehicles	Detection	YOLO family	VisDrone2019
[36]	Vehicles	Detection and Tracking	Region-based	Particular
[37]	Vehicles	Detection	YOLO family and Attention-based	KITTI
[38]	Vehicles	Detection and Tracking	Region-based	Particular
[39]	Vehicles	Detection	Region-based	VSAI
[40]	Vehicles	Detection	Attention-based	SODA 10 M and BDD100K
[41]	Traffic	Detection and Classification	YOLO family	UA-DETRAC
[42]	Vehicles	Detection	YOLO family	Particular
[43]	Vehicles	Detection	YOLO family	BIT-Vehicle and UA-DETRAC
[44]	Vehicles	Detection and Tracking	Region-based	VAID, UAVDT and AU-AIR
[45]	Vehicles	Detection and Tracking	YOLO family	MS COCO
[46]	Vehicles	Detection and Tracking	Region-based and YOLO family	MS COCO
[47]	Traffic	Detection and Classification	YOLO family	MS COCO
[48]	Traffic	Detection and Classification	Region-based and YOLO family	Particular
[49]	Traffic	Detection	Transformer-based	CULane
[50]	Vehicles	Detection and Tracking	Region-based	Particular
[51]	Traffic	Detection	YOLO family	U.S. benchmark detection
[52]	Vehicles	Detection and Tracking	YOLO family	MS COCO and DAWN
[53]	Vehicles	Detection	YOLO family	VEDAI
[54]	Vehicles	Detection, Classification and Tracking	YOLO family	Particular
[55]	Traffic	Detection	YOLO family	KITTI and DATS_2022
[56]	Traffic	Detection and Tracking	YOLO family and Transformer-based	Particular
[57]	Vehicles	Detection, Classification and Tracking	YOLO family	Particular
[58]	Traffic	Detection, Classification and Counting	YOLO family	UA-DETRAC and KITTI

facilitates the identification of intrinsic relationships among the studies in the textual corpus, enabling the formulation of conclusions based on the prevalence and frequency of each previously described category. Furthermore, it allows for the observation of preferences in the selection of specific categories to solve similar tasks, according to the subjects of study adopted by each article, directly observed through the databases utilized during their experimentation.

The structure of the columns in Table 1 is directly linked to the final stage of the methodology, as depicted in Figure 1, in the "Analysis Step" section. During this phase of the process, the selected papers were thoroughly analyzed, with a careful assessment of their construction and results. The primary factors that contributed to the development of the entries in the table were predominantly the network architecture chosen and the type of task it aimed to address.

Thus, papers whose objectives did not align with the main focus of the review were excluded from the corpus. This decision is exemplified by the works of [59] and [60], which, although presenting relevant discussions, robust experimental methodologies, and appropriate metrics to support their conclusions, were not included in the analyzed set due to their predominant focus on regression related to speed data and accident risk.

The acronyms used to describe the datasets employed in each study and their respective purposes were detailed in a specific subsection of this review, accompanied by a dedicated table that links the datasets to their key characteristics. Following the systematization presented in Table 1, the discussion was conducted based on the columns corresponding to each study and the values obtained from the in-depth evaluation.

This procedure allowed for the identification of recurring patterns among the studies, as well as the inference of potential directions for future research, focusing on the exploration of new combinations of subjects and tasks, thereby expanding the scope for experimentation and methodological advancement in the field.

IV. DISCUSSION

This discussion was organized into subsections for greater clarity. First, an overview was provided, explaining the most commonly used datasets and their main characteristics (IV-A). This was followed by an analysis of the distribution of the studied focuses and the tasks performed to solve the problems (IV-B).

Finally, an analysis of the architectures and categories was conducted, with a focus on comparative metrics and their lack of standardization, alongside a sugestion for a possible solution to solve this problem in future works with the same theme (IV-C).

A. DATASET OVERVIEW

Analyzing the publications in the textual corpus revealed a significant diversity in the datasets used, with approximately 30.30% of the studies (14 out of 46) opting to conduct their research with a proprietary dataset. In some cases, the exact number of images used was reported, as in the work of [28], while in other cases, such as the study by [17], the number of images was left unspecified, with the focus directed towards describing the image set.

Despite this peculiarity, it was possible to identify more than 30 distinct datasets described throughout the studies, exhibiting significant variation in the number of images, capture type, presence or absence of videos, and even some containing multispectral images. To establish parallels between the datasets, they were organized into a table based on sensor type, annotation type, number of images, and study object, as shown in Table 2, adapted from the works of [61] and [62], and populated with information from the textual corpus, providing further details in the paragraphs that follow.

The analysis of the table reveals significant associative factors, with the first being related to the type of sensor used in capturing the data that make up the image dataset. The majority of the datasets analyzed were acquired through cameras, whether stationary or mobile, highlighting the importance of computer vision in image feature extraction within this field. In fact, less than 10% of the datasets (2 out of 33) do not adopt this approach, as indicated by Figure 3, where numerical categories 3 and 5 account for approximately 6% of the total, with the remaining 94% utilizing visual information.

Furthermore, when examining the type of sensor, it is observed that the most prominent datasets in the textual corpus are those incorporating additional technologies alongside cameras, such as the KITTI [69] and KAIST [72] datasets, which combine GPS (Global Positioning System) and LiDAR (Light Detection and Ranging) data with traditional camera channels. These resources provide robust support for learning the proposed architectures, offering a significant advancement in the effectiveness of the developed models.

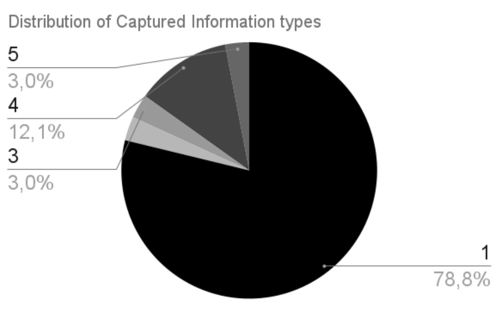


FIGURE 3. Pie chart illustrating the distribution of captured information types. The numbered categories represent: (1) Visual only, (2) Depth only, (3) Depth, Radar, and Motion, (4) Visual, Depth, and Motion, and (5) Visual and Radar, with a big tendency to visual information.

Regarding the type of annotation, a predominant trend is observed in the use of 2D annotations for detection and tracking processes, with a strong emphasis on the use of bounding boxes to identify objects of interest in images and videos, representing approximately 81.8% of the datasets

TABLE 2. Relevant information from the datasets used in the experiments within the textual corpus.

REF	NAME	SENSOR	LABEL	IMAGES	CONTENT
[63]	BIT-Vehicle	Camera	2D	9850	Vehicles
[64]	Apollo Scape	Camera, IMU/GNSS and LiDAR	2D and 3D	140000+	Traffic
[65]	UA-DETRAC	Camera	2D	140000	Vehicles
[66]	Track4	LiDAR, Radar and GNSS	3D	4000	Traffic
[67]	VLD-45	Camera	2D	45000	Vehicles
[68]	FLIR	Camera	2D	26000	Traffic
[69]	KITTI	Camera, GPS/IMU and LiDAR	2D and 3D	15000	Traffic
[70]	K-lane	LiDAR	2D and 3D	15000	Traffic
[71]	M3FD	Camera	2D	4200	Traffic
[72]	KAIST	Camera, GPS and LiDAR	2D	25086	Traffic
[73]	LLVIP	Camera	2D	30976	Pedestrians
[74]	CVC-14	Camera	2D	5000	Traffic
[75]	INRIA	Camera	2D	2573	Pedestrians
[76]	LASIESTA	Camera	2D	Not Specified	General Scenarios
[77]	PESMOD	Camera	2D	4107	General Scenarios
[78]	MOCS	Camera	2D	31000	Traffic
[79]	ViWi	Camera and mmWave Antennas	2D and Channel Data	Not Specified	Traffic
[80]	DAIR-V2X	Camera, GPS/IMU and LiDAR	3D	142508	Vehicles
[81]	Caltech Cars 1999	Camera	2D	126	Vehicles
[82]	Caltech Cars 2001	Camera	2D	256	Vehicles
[83]	VisDrone2019	Camera	2D	272117	Traffic
[39]	VSAI	Camera	2D	444	Vehicles
[84]	SODA 10 M	Camera	2D	10000000+	Vehicles
[85]	BDD100K	Camera	2D	100000	Vehicles
[86]	VAID	Camera	2D	6000	Vehicles
[87]	AU-AIR	Camera	2D	32823	Vehicles
[88]	UAVDT	Camera	2D	80000	Vehicles
[89]	MS COCO	Camera	2D	328000	General Scenarios
[90]	CULane	Camera	2D	133235	Traffic
[91]	US Benchmark Det	Camera	2D	Not Specified	Traffic
[92]	DAWN	Camera	2D	1000	Vehicles
[93]	VEDAI	Camera	2D	1246	Vehicles
[94]	DATS_2022	Camera	2D	10000+	Traffic

that adopt this format. Most of these datasets use only 2D annotations, disregarding other information for training. In contrast, the use of 3D annotations is significantly lower, with 2D annotations being replaced by three-dimensional approaches, such as point clouds, for the same purposes.

This type of annotation is present in only five datasets, of which only DAIR-V2X [80] and Track4 [66] exclusively use 3D data, as evidenced in Figure 4, which illustrates these trends in the types of data presented. During the analysis, an interesting peculiarity was identified: the ViWi [79] dataset was the only one to adopt a completely distinct type of annotation, associated with the 2D format but innovating by incorporating wireless channel information. This strategy aimed to enrich the available data for model learning, offering an innovative approach in combining data sources.

Finally, when examining the focus of each dataset, as illustrated by the 'CONTENT' column and highlighted by the distribution presented in Figure 5, it is evident that most of the datasets depict everyday scenarios characterized by heterogeneous information, such as complex traffic environments involving pedestrians and vehicles

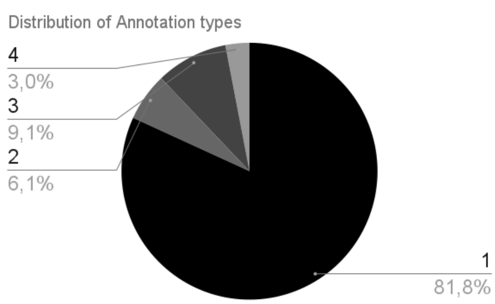


FIGURE 4. Pie chart illustrating the distribution of annotation types. The numbered categories correspond to: (1) 2D only, (2) 3D only, (3) Both 2D and 3D, and (4) 2D and other annotation types. A strong bias toward 2D annotations is observed.

simultaneously. This category, termed 'Traffic,' represents an approximately 42.4% of the total analyzed, sharing the same value as 'Vehicle' only data. Although this first type of content indirectly encompasses aspects covered by the other categories, these remain present due to differences in focus.

While 'Traffic' describes complete urban mobility situations, incorporating all annotation categories, the other types of content focus exclusively on their main theme. Thus, these categories together represent the remaining 57.6% of the data listed in Table 2, with their focus divided between 'Pedestrians,' 'Vehicles,' and 'General Scenarios.' Although relevant, these categories are not necessarily related to the traffic theme addressed in this review. This distribution highlights the diversity of approaches and objectives present in the analyzed datasets and can be observed in the annotated samples shown in Figure 6.

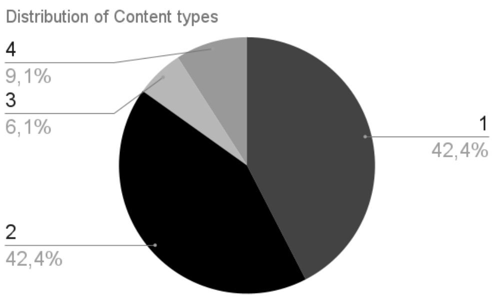
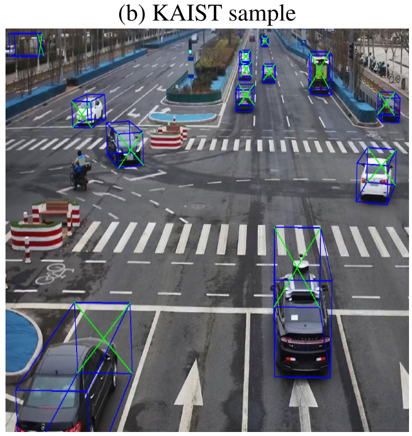


FIGURE 5. Pie chart illustrating the distribution of content categories. The numbered categories correspond to: (1) Vehicles, (2) Traffic, (3) Pedestrians, and (4) General Scenarios. This chart highlights the balance between traffic-focused datasets and those primarily centered on vehicles.

Thus, it can be concluded that there is a clear trend in the adoption of complex databases that involve complete urban scenarios represented visually by camera images, accompanied by 2D annotations, for the development of future research in the area addressed in this review. Although these complexities may initially appear as an additional challenge for the model to overcome, they reflect real-world situations that any applications potentially targeting these scenarios will face. Therefore, these elements are of utmost relevance for training, becoming highly representative, as evidenced by [95] and [96], who highlighted the direct relationship between data representativeness and the accuracy of the trained model.

(a) FLIR sample

(b) KAIST sample



(c) DAIRV2X sample

FIGURE 6. Annotated samples from the FLIR [68] (a), KAIST [72] (b), and DAIRV2X [80] (c) datasets.

B. STUDY OF SUBJECT TYPES

The trend observed when analyzing the datasets used according to their content, at the end of subsection IV-A, does not fully align when examining the subjects studied by the articles in the textual corpus through the "SUBJECT" column of Table 1. By observing the synthesis presented in Figure 7, it becomes evident that despite the datasets having a stronger focus on complex traffic situations, the primary focus of most of the works studied remained on performing tasks related to vehicles themselves. This suggests that the primary use of these highly heterogeneous datasets can assist models in learning relevant information about urban environments, which is challenging to extract from simpler and more specialized datasets, as seen in works that opted for datasets focused solely on vehicles.

Another relevant insight presented by the careful analysis of the table is the direct connection between works focused on pedestrian detection and the use of image banks and data belonging to the category with the highest complexity. This is exemplified by the study of [27], where part of the experiment

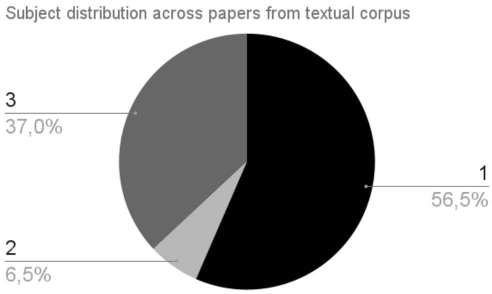


FIGURE 7. Pie chart depicting the percentage distribution of publications by subject type within the textual corpus. The numbered categories correspond to: (1) Vehicles, (2) Traffic, and (3) Pedestrians. This chart further emphasizes the relatively balanced distribution between traffic-focused studies and those primarily centered on vehicles, though with a more pronounced tendency toward the latter.

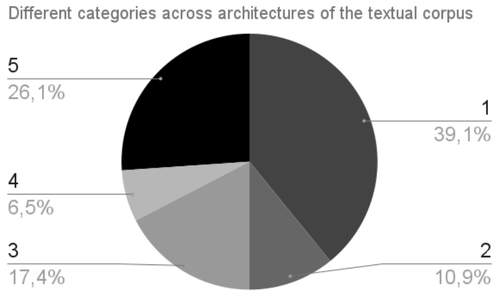


FIGURE 8. Pie chart illustrating the distribution of categories within the textual corpus. The numbered categories correspond to: (1) YOLO family, (2) Attention-based, (3) Region-based, (4) Transformer-based, and (5) Combinations of multiple categories. This chart reveals a relatively balanced distribution among categories 1 and 5, with the YOLO family exhibiting a slight predominance.

was conducted using the KAIST [72] dataset, which primarily focuses on complex traffic situations, while the second part utilized the LLVIP [73] dataset, targeted towards the "Pedestrians" content category. As this pattern was observed in all works sharing the same "SUBJECT" as [27], a similar deduction can be made to those focused on vehicles, suggesting that the use of complex and heterogeneous information provided by traffic-focused datasets can have a positive impact on the model's learning process when applied to real-world scenarios.

C. ARCHITECTURES: TASKS, CATEGORIES, CONNECTIONS, AND PERFORMANCE METRICS

Regarding the factor of architectures and their respective categories, the corpus reveals a significant tendency towards the use of models from the "YOLO family," either by using the YOLO architecture in one of its versions (39.1% of the studies) or by combining it with structures from other categories. Ten out of the eleven studies that combine categories feature the "YOLO family" as one of them, as observable in the pie chart of Figure 8. Another notable point concerns the type of task proposed by each study, where the overwhelming majority involves detection, whether related to vehicles as a whole or parts of them, as in [19], pedestrians, as in [29], or moving objects, as in [31]. Detection is an integral part of the study, regardless of its specific category.

Nevertheless, an exception exists. The experiment proposed by [22] bypasses the detection stage and directly addresses classification through an alternative approach. Rather than utilizing conventional image classifiers such as ResNet101 [97] or VGG16 [98], or employing state-of-the-art detection and classification methods like YOLOv8 [99] or Faster R-CNN [100], the study adopts an innovative strategy by training the ChatGPT-4V model [101] to classify risk in traffic images. This approach yielded promising results, achieving a correlation coefficient (r) of 0.83, indicating a strong positive correlation between the predicted outcomes and actual conditions. These findings prompt further inquiry into the potential application of tools like ChatGPT in expanding the scope of future research endeavors.

The in-depth analysis revealed that one of the main challenges faced was identifying the studies that provided an affirmative answer to the question: "Was the proposed problem solved through the architecture and experiment indicated?" This challenge became particularly evident due to the lack of standardization in the performance metrics used. While studies such as those by [14] and [15] adopted standard classification and detection metrics, others opted for alternative metrics, such as the "S4" proposed by [16] and the "MR" used by [27] and [29].

The diversity in the metrics reported across these studies adds a significant layer of complexity to quantitative comparisons, even between experiments that feature similar architectures. In some cases, such variation in metrics makes quantitative analysis unfeasible, forcing the adoption of qualitative approaches to synthesize results through the inference of additional aggregating metrics.

Although the use of standardized metrics can increase the visibility of a study due to the ease of comparison between different experiments, in some cases these metrics may not be sufficient to fully demonstrate a model's capabilities. In the study by [16], for instance, in addition to the traditionally used metric, the "F1 score," the "S4" metric was introduced, specifically aimed at anomaly detection. This case, along with other examples in the literature, raises the question of whether it would be advantageous to use both popular and domain-specific metrics simultaneously.

Such an approach not only maintains the specificity of certain model aspects but also enables a more robust qualitative comparative analysis, without losing the possibility of quantitative assessments. This strategy facilitates the integration of the article into both qualitative and quantitative analyses, promoting a more comprehensive and contextualized evaluation of the experiments conducted in comparison with others.

Finally, a growing trend can be observed among the "TASKS" that the articles in the textual corpus aim to

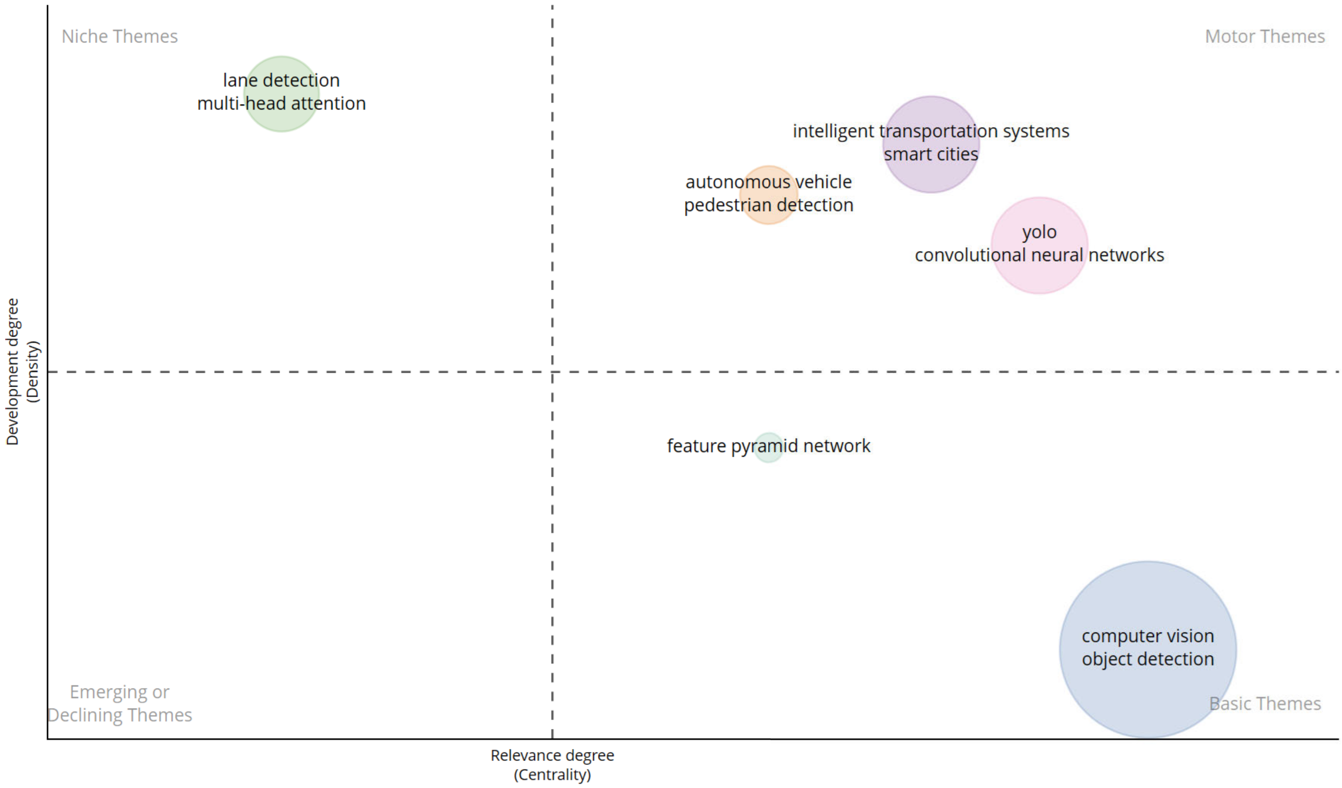


FIGURE 9. **Cartesian graph illustrating thematic connections between articles, correlating levels of development and relevance.**

accomplish. There is an increasing interest not only in detecting objects in datasets but also in determining their trajectories over time. This is seen in applications such as accident prevention, as presented by [14], or simply to improve traffic flow by reducing unnecessary vehicle idling through intelligent traffic light control, as proposed by [20]. Therefore, future experiments focused on advancing the ''Tracking'' domain become increasingly attractive.

These trends become apparent when analyzing the graph presented in Figure 9, which illustrates thematic groupings based on two primary variables. The first is the degree of relevance, also referred to as centrality, which indicates the impact of publications within each group during the analyzed period.

This variable, represented on the X-axis, is directly related to the impact factor of the publications relative to others in the corpus. The second variable considered is the degree of development, or density, represented on the Y-axis, which reflects the number of citations received by each grouping over time. In addition to these variables, the size of the circle associated with each thematic group represents the number of studies within the textual corpus belonging to that specific group, offering an additional layer of analysis.

The analysis of Figure 9 is primarily based on the location of the clusters, considering that those positioned closer to the upper right corner exhibit greater relevance for the advancement of the research field. From this perspective, the clusters in the textual corpus related to the YOLO neural network (pink), intelligent transportation systems (purple), and autonomous vehicles (orange) stand out as the most influential, displaying distinct combinations of density and centrality while remaining within this quadrant.

When comparing the purple and pink clusters, a similar overall impact is observed, differing only in their distribution: while the YOLO-related cluster exhibits lower centrality, its density index is higher than that of intelligent transportation systems, which exhibit the inverse relationship. Despite this difference, both are equally significant for the research field and are considered key drivers of investigation. Closely following, the orange cluster, representing studies focused on autonomous vehicles and accident prevention, has been gaining increasing impact and volume of publications and citations since the end of the global COVID-19 pandemic period.

Among the analyzed themes, the green cluster, associated with lane detection, stands out. This group exhibits the highest density among all themes but the lowest relevance, indicating that although it has a considerable number of citations, its impact on the field remains limited. In contrast, the largest cluster in terms of the number of articles, represented by dark blue and primarily composed of studies on fundamental concepts of computer vision and object detection, demonstrates the most significant impact among all analyzed groups. However, similar to the green cluster, it exhibits the lowest density, meaning the smallest volume of citations among the themes.

Still within the quadrant of fundamental topics, the small aqua-green cluster is identified, mainly focused on studies related to scale variation, with the central term feature pyramid network. Despite having a limited number of publications, this group has shown growth similar to the previously discussed orange cluster, albeit with lower intensity. This trend suggests its potential to become a key topic for exploration, gradually aligning with the other clusters in the upper-right quadrant in the near future.

Another promising area lies in the combinatorial analysis of thematic clusters, considering the incidence of the terms that characterize them. A trend is observed in recent studies on autonomous vehicles, where hybrid approaches are being adopted to solve complex problems, combining segmentation and other classic Computer Vision techniques to reduce noise, optimize data processing, and increase response speed. This strategy directly contributes to risk mitigation and reduces the likelihood of accidents during operation.

A particularly relevant aspect is the high frequency of the term "Segmentation" among the analyzed works, even with the filtering focused on detection and tracking. This suggests that this approach has a significant impact on the field as a whole and should be considered alongside the other techniques discussed in this review for future research. Studies such as those by [19] and [23] corroborate this perspective, emphasizing the importance and benefits of integrating intermediate techniques to enhance the architectures investigated within this application context.

V. FINAL CONSIDERATIONS

Based on the discussion and the points established throughout this study, it can be concluded that for future work focused on the detection and tracking of vehicles in urban scenarios, the use of architectures from the "YOLO family" can be highly effective, especially when trained on datasets with high complexity and heterogeneous annotations. This applies even in cases where the object of study does not require the full range of information provided by the data. The flexibility of these architectures allows them to adapt to different conditions, achieving satisfactory results in challenging urban environments, where the variability of objects and context can impact performance.

The recognition of the "YOLO family" category can be attributed to four key factors that define the architecture of the family as a whole: 1. The architecture is inherently optimized to handle problems that require high-speed information for decision-making, as highlighted by [102] in their literature review, discussing its application in autonomous vehicles; 2. High customization and optimization capacity, maintaining high metrics even in scenarios with limited data or computational resources [103]; 3. Ability to process noise, as the architecture is capable of acquiring and processing information even under low lighting conditions, as demonstrated in the experiment by [104], where the network was adapted to handle such images without any enhancement modules; 4. Adaptability to the scale of objects of interest, as shown in cases described in the textual corpus, where the architecture was used in both fixed cameras and uncrewed aerial vehicles (UAVs).

In both cases, the architecture maintained its performance, even in the presence of significant variability in annotation sizes, a feature that is further highlighted in the adaptation proposed by [105], which similarly employs the model focusing on small-scale objects.

Another potential avenue for research in the field lies in investigating the impact of segmentation on the approach to urban mobility problems. This analysis could be conducted through a literature review focused on the keyword "segmentation," evaluating the role of this methodology in the proposed solutions.

Adopting a qualitative analytical approach would highlight unexplored potentials and identify existing gaps, ultimately leading to the proposition of directions for future experiments utilizing this approach. This proposal seeks, similarly to this article, to provide a structured foundation for exploring new research directions in the field.

Regarding the standardization problem mentioned earlier, there is still no definitive answer on which approach is more effective: using only common metrics, such as "Precision", "Recall", "F1 Score", and "mAP" (mean Average Precision), or adopting problem-specific metrics that may better highlight the model's capacity, as in the case of the proposed "S4" metric. In this context, the present review article proposes a potential solution to facilitate result comparisons in future studies and, eventually, establish an optimized standard for presenting metrics. Specifically, it suggests the use of the aforementioned popular metrics in robust statistical comparisons, utilizing tools such as mean and standard deviation tables, analysis of variance, and boxplots, in order to provide sufficient data for concrete numerical conclusions, as discussed by [106] and [107].

Simultaneously, it advocates the integration of customized metrics, tailored to the specific problem at hand, enabling the consideration of potential improvements based on factors that traditional statistical metrics may not capture. This would enrich the interpretation and analysis, offering a more in-depth understanding of the results, in a manner similar to what has been demonstrated in the studies of [22], [29], and [38].

Finally, a highly relevant possibility emerges, in light of the prevailing trend to adopt widely recognized algorithms and models that have been tested and validated in public datasets, such as MS COCO [89]. This study proposes a reflection for future work in the field. In addition to formalizing the metrics and the appropriate methodologies for their proper reporting in subsequent studies, this review suggests the incorporation of explainability tools. This approach aims not only to enhance the foundation of the conclusions to be drawn but also to increase the confidence in asserting that a specific model outperforms other N models in a given experiment.

In line with the growth of the "Explainable Artificial Intelligence" (XAI) field, this conclusion advocates for the use of these tools as a means to increase the level of certainty of the presented solutions, fostering progress in the overall scenario, as evidenced by the studies of [108] and [109], which provide an in-depth analysis of the advances in this area within the context of machine learning. In light of this, the work raises a final question: why not always strive for a deeper understanding and promote greater clarity in the proposed solutions?

ACKNOWLEDGMENT

Some of the authors have been awarded with Scholarships from the Brazilian National Council of Technological and Scientific Development, CNPq, and the Coordination for the Improvement of Higher Education Personnel, CAPES. They would especially like to express their gratitude to the Department of Vehicle Engineering, Faculty of Engineering, Vehicles and Mechatronics Institute, University of Debrecen, for their support during the research and its developments.

They acknowledge the use of large language models such as ChatGPT and DeepL alongside the Oxford Online Dictionary for refining the translation of this manuscript into English. These tools were utilized to enhance linguistic clarity and ensure precise terminology, to ensure both fluency and technical accuracy in the translated text.

REFERENCES

[1] B. Adhi Santharm and U. Ramanathan, "Supply chain transparency for sustainability—An intervention-based research approach," *Int. J. Operations Prod. Manage.*, vol. 42, no. 7, pp. 995–1021, Jun. 2022.

[2] A. C. Goldthau and R. Youngs, "The EU energy crisis and a new geopolitics of climate transition," *JCMS, J. Common Market Stud.*, vol. 61, no. S1, pp. 115–124, Sep. 2023.

[3] Z. Shams Esfandabadi, M. Diana, and M. C. Zanetti, "Carsharing services in sustainable urban transport: An inclusive science map of the field," *J. Cleaner Prod.*, vol. 357, Jul. 2022, Art. no. 131981.

[4] L. Weimann and E. Weimann, "Traffic and air pollution: How can public participation and speed limit legislation benefit health outcomes in Germany," *Med. Res. Arch.*, vol. 11, no. 12, 2024.

[5] O. for Economic Co-operation and D. (OECD), *OECD Insights—Sustainable Development: Linking Economy, Society, Environment*. Paris, France: OECD, 2008. [Online]. Available: https://www.oecd.org/en/data/insights

[6] S. Hornyák, Z. Karancsi, A. Korom, and F. Győri, "What does a tourist see, or, an environmental-aesthetic evaluation of a street view in szeged (Hungary)," *Turizam*, vol. 27, no. 2, pp. 113–135, 2023.

[7] T. Szigetvári and G. Túry, "Can they get out of the middle-income technology trap? State strategies in Hungary and Türkiye in promoting automotive investments," in *Proc. 23rd Int. Joint Conf., Central Eastern Europe Changing Business Environ.*. Vydavatelstvo Ekonóm and the University of Economics in Bratislava, 2023, pp. 263–277.

[8] A. Alatawneh and A. Torok, "Potential autonomous vehicle ownership growth in Hungary using the Gompertz model," *Prod. Eng. Arch.*, vol. 29, no. 2, pp. 155–161, Jun. 2023.

[9] R. Hart, E. Kyriakopoulou, and T. Lu, "Urban transport policies and net zero emissions in the European union," *Annu. Rev. Resource Econ.*, vol. 16, no. 1, pp. 187–206, Oct. 2024.

[10] J. V. D. A. Porto, A. C. Dorsa, V. A. D. M. Weber, K. R. D. A. Porto, and H. Pistori, "Usage of few-shot learning and meta-learning in agriculture: A literature review," *Smart Agricult. Technol.*, vol. 5, Oct. 2023, Art. no. 100307.

[11] M. de Andrade Marconi and E. M. Lakatos, *Metodologia do trabalho científico: Procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos*. São Paulo, Brazil: Atlas, 2007.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[13] Y. Ren, "Intelligent vehicle violation detection system under human–computer interaction and computer vision," *Int. J. Comput. Intell. Syst.*, vol. 17, no. 1, p. 40, Feb. 2024.

[14] A. Karim, M. A. Raza, Y. Z. Alharthi, G. Abbas, S. Othmen, M. S. Hossain, A. Nahar, and P. Mercorelli, "Visual detection of traffic incident through automatic monitoring of vehicle activities," *World Electr. Vehicle J.*, vol. 15, no. 9, p. 382, Aug. 2024.

[15] L. You, Y. Chen, C. Xiao, C. Sun, and R. Li, "Multi-object vehicle detection and tracking algorithm based on improved YOLOv8 and ByteTrack," *Electronics*, vol. 13, no. 15, p. 3033, Aug. 2024.

[16] A. R. H. Reyna, A. J. F. Farfán, G. P. R. Filho, S. Sampaio, R. De Grande, L. H. V. Nakamura, and R. I. Meneguette, "MEDAVET: Traffic vehicle anomaly detection mechanism based on spatial and temporal structures in vehicle traffic," *J. Internet Services Appl.*, vol. 15, no. 1, pp. 25–38, Apr. 2024.

[17] J. A. Guzmán-Torres, F. J. Domínguez-Mota, G. Tinoco-Guerrero, M. C. García-Chiquito, and J. G. Tinoco-Ruíz, "Efficacy evaluation of you only learn one representation (YOLOR) algorithm in detecting, tracking, and counting vehicular traffic in real-world scenarios, the case of morelia México: An artificial intelligence approach," *AI*, vol. 5, no. 3, pp. 1594–1613, Sep. 2024.

[18] S. Yang, Y. Liu, Z. Liu, C. Xu, and X. Du, "Enhanced vehicle logo detection method based on self-attention mechanism for electric vehicle application," *World Electric Vehicle J.*, vol. 15, no. 10, p. 467, Oct. 2024.

[19] T. Mustafa and M. Karabatak, "Real time car model and plate detection system by using deep learning architectures," *IEEE Access*, vol. 12, pp. 107616–107630, 2024.

[20] W. Karoon, P. Chuasuai, P. Thipprasert, N. Khongchu, P. Kunakornjittirak, and T. Siriborvornratanakul, "Adaptive traffic light control using vision-based deep learning for vehicle density estimation," in *Proc. 6th Asia–Pacific Inf. Technol. Conf.*, Jan. 2024, pp. 37–42.

[21] C. Lv, U. Mittal, V. Madaan, and P. Agrawal, "Vehicle detection and classification using an ensemble of EfficientDet and YOLOv8," *PeerJ Comput. Sci.*, vol. 10, p. e2233, Aug. 2024.

[22] T. Driessen, D. Dodou, P. Bazilinskyy, and J. de Winter, "Putting ChatGPT vision (GPT-4V) to the test: Risk perception in traffic images," *Roy. Soc. Open Sci.*, vol. 11, no. 5, May 2024, Art. no. 231676.

[23] K. Lyu, "A nonlinear convolutional neural network algorithm for autonomous vehicle lane line detection," *Scalable Comput., Pract. Exp.*, vol. 25, no. 5, pp. 4237–4245, Aug. 2024.

[24] O. Patil, B. B. Nair, R. Soni, A. Thayyilravi, and C. R. Manoj, "BoostedDim attention: A novel data-driven approach to improving LiDAR-based lane detection," *Ain Shams Eng. J.*, vol. 15, no. 9, Sep. 2024, Art. no. 102887.

[25] C. Sun, Y. Chen, X. Qiu, R. Li, and L. You, "MRD-YOLO: A multispectral object detection algorithm for complex road scenes," *Sensors*, vol. 24, no. 10, p. 3222, May 2024.

[26] M. A. Tolba and H. A. Kamal, "SDC-Net++: End-to-end crash detection and action control for self-driving car deep-IoT-based system," *Sensors*, vol. 24, no. 12, p. 3805, Jun. 2024.

[27] S. Lee, T. Kim, J. Shin, N. Kim, and Y. Choi, "INSANet: INtra-INter spectral attention network for effective feature fusion of multispectral pedestrian detection," *Sensors*, vol. 24, no. 4, p. 1168, Feb. 2024.

[28] P. Liu, Y. Zhang, G. Guo, and J. Ding, "Enhanced detection and recognition of road objects in infrared imaging using multi-scale self-attention," *Sensors*, vol. 24, no. 16, p. 5404, Aug. 2024.

[29] X. Zhang, X. Zhang, J. Wang, J. Ying, Z. Sheng, H. Yu, C. Li, and H.-L. Shen, "TFDet: Target-aware fusion for RGB-T pedestrian detection," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 23, 2024, doi: 10.1109/TNNLS.2024.3443455.

[30] L. Kok Yee, T. Lit Ken, H. Sim Choo, Y. Asako, L. Kee Quen, H.-S. Kang, Y. S. Gan, Z.-L. Chuan, W. Y. Tey, and N. A. Che Sidik, "Effect of positive-negative image ratio on the performance of pedestrian detection model," *Malaysian J. Fundam. Appl. Sci.*, vol. 20, no. 2, pp. 266–287, Apr. 2024.

[31] M. Safaldin, N. Zaghden, and M. Mejdoub, "An improved YOLOv8 to detect moving objects," *IEEE Access*, vol. 12, pp. 59782–59806, 2024.

[32] G. Gharsallah and G. Kaddoum, "ViT LoS V2X: Vision transformers for environment-aware LoS blockage prediction for 6G vehicular networks," *IEEE Access*, vol. 12, pp. 133569–133583, 2024.

[33] J. Borau Bernad, Á. Ramajo-Ballester, and J. M. Armingol Moreno, "Three-dimensional vehicle detection and pose estimation in monocular images for smart infrastructures," *Mathematics*, vol. 12, no. 13, p. 2027, Jun. 2024.

[34] S. Reddy, N. Pillay, and N. Singh, "Comparative evaluation of convolutional neural network object detection algorithms for vehicle detection," *J. Imag.*, vol. 10, no. 7, p. 162, Jul. 2024.

[35] B. Khalili and A. W. Smyth, "SOD-YOLOv8—Enhancing YOLOv8 for small object detection in aerial imagery and traffic scenes," *Sensors*, vol. 24, no. 19, p. 6209, Sep. 2024.

[36] M. Gomathy Nayagam, K. Ramar, K. Venkatesh, and S. P. Raja, "Moving object detection and tracking algorithm using hybrid decomposition parallel processing," *Intell. Autom. Soft Comput.*, vol. 33, no. 3, pp. 1485–1499, 2022.

[37] W.-J. Yang, W.-J. Liow, S.-F. Chen, J.-F. Yang, P.-C. Chung, and S. Mao, "Improved vehicle detection systems with double-layer LSTM modules," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, p. 7, Dec. 2022.

[38] Z. Dong, Y. Lu, G. Tong, Y. Shu, S. Wang, and W. Shi, "WatchDog: Real-time vehicle tracking on geo-distributed edge nodes," *ACM Trans. Internet Things*, vol. 4, no. 1, pp. 1–23, Feb. 2023.

[39] J. Wang, X. Teng, Z. Li, Q. Yu, Y. Bian, and J. Wei, "VSAI: A multi-view dataset for vehicle detection in complex scenarios using aerial images," *Drones*, vol. 6, no. 7, p. 161, Jun. 2022.

[40] H. Xu, S. Lai, X. Li, and Y. Yang, "Cross-domain car detection model with integrated convolutional block attention mechanism," *Image Vis. Comput.*, vol. 140, Dec. 2023, Art. no. 104834.

[41] Y. Song, S. Hong, C. Hu, P. He, L. Tao, Z. Tie, and C. Ding, "MEB-YOLO: An efficient vehicle detection method in complex traffic road scenes," *Comput., Mater. Continua*, vol. 75, no. 3, pp. 5761–5784, 2023.

[42] S. Li and H.-S. Yoon, "Vehicle localization in 3D world coordinates using single camera at traffic intersection," *Sensors*, vol. 23, no. 7, p. 3661, Mar. 2023.

[43] J. Zhao, S. Hao, C. Dai, H. Zhang, L. Zhao, Z. Ji, and I. Ganchev, "Improved vision-based vehicle detection and classification by optimized YOLOv4," *IEEE Access*, vol. 10, pp. 8590–8603, 2022.

[44] A. Mehmood Qureshi, N. Abdullah Almujally, S. S. Alotaibi, M. Hamad Alatiyyah, and J. Park, "Intelligent traffic surveillance through multi-label semantic segmentation and filter-based tracking," *Comput., Mater. Continua*, vol. 76, no. 3, pp. 3707–3725, 2023.

[45] T. Jin, X. Ye, Z. Li, and Z. Huo, "Identification and tracking of vehicles between multiple cameras on bridges using a YOLOv4 and OSNet-based method," *Sensors*, vol. 23, no. 12, p. 5510, Jun. 2023.

[46] K. A. Redmill, E. Yurtsever, R. G. Mishalani, B. Coifman, and M. R. McCord, "Automated traffic surveillance using existing cameras on transit buses," *Sensors*, vol. 23, no. 11, p. 5086, May 2023.

[47] S. Chuprov, P. Belyaev, R. Gataullin, L. Reznik, E. Neverov, and I. Viksnin, "Robust autonomous vehicle computer-vision-based localization in challenging environmental conditions," *Appl. Sci.*, vol. 13, no. 9, p. 5735, May 2023.

[48] M. I. Basheer Ahmed, R. Zaghdoud, M. S. Ahmed, R. Sendi, S. Alsharif, J. Alabdulkarim, B. A. Albin Saad, R. Alsabt, A. Rahman, and G. Krishnasamy, "A real-time computer vision based approach to detection and classification of traffic incidents," *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 22, Jan. 2023.

[49] B. Liu, L. Feng, Q. Zhao, G. Li, and Y. Chen, "Improving the accuracy of lane detection by enhancing the long-range dependence," *Electronics*, vol. 12, no. 11, p. 2518, Jun. 2023.

[50] S. S. Ram, "Fusion of inverse synthetic aperture radar and camera images for automotive target tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 2, pp. 431–444, Feb. 2022.

[51] D. Oh, K. Kang, S. Seo, J. Xiao, K. Jang, K. Kim, H. Park, and J. Won, "Low-cost object detection models for traffic control devices through domain adaption of geographical regions," *Remote Sens.*, vol. 15, no. 10, p. 2584, May 2023.

[52] E. O. Appiah and S. Mensah, "Object detection in adverse weather condition for autonomous vehicles," *Multimedia Tools Appl.*, vol. 83, no. 9, pp. 28235–28261, Aug. 2023.

[53] M. A. Momin, M. H. Junos, A. S. Mohd Khairuddin, and M. S. Abu Talip, "Lightweight CNN model: Automated vehicle detection in aerial images," *Signal, Image Video Process.*, vol. 17, no. 4, pp. 1209–1217, Jun. 2023.

[54] V. V. Rathod, D. P. Rana, R. G. Mehta, and V. Nath, "A computer vision approach to vehicle detection, classification, and tracking from UAV data for Indian traffic analysis," *IETE J. Res.*, vol. 2024, pp. 1–14, Jul. 2024.

[55] S. Mishra and D. K. Yadav, "Vehicle detection in high density traffic surveillance data using YOLO.v5," *Recent Adv. Electr. Electron. Eng., Formerly Recent Patents Electr. Electron. Eng.*, vol. 17, no. 2, pp. 216–227, 2024.

[56] L. Ye and L. Pei, "Real-time vehicle pedestrian detection and tracking algorithm based on computer vision," in *Proc. 3rd Int. Conf. Frontiers Artif. Intell. Mach. Learn.*, Apr. 2024, pp. 17–22.

[57] S. Pudaruth, I. M. Boodhun, and C. W. Onn, "Reducing traffic congestion using real-time traffic monitoring with YOLOv8," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 10, p. 1068, 2024.

[58] P. Premaratne, R. Blacklidge, and M. Lee, "Vehicle detection, classification and counting on highways-accuracy enhancements," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, 2022, pp. 394–408.

[59] J. de Winter, J. Hoogmoed, J. Stapel, D. Dodou, and P. Bazilinskyy, "Predicting perceived risk of traffic scenes using computer vision," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 93, pp. 235–247, Feb. 2023.

[60] R. Itu and R. Danescu, "Fully convolutional neural network for vehicle speed and emergency-brake prediction," *Sensors*, vol. 24, no. 1, p. 212, Dec. 2023.

[61] S. Y. Alaba, A. C. Gurbuz, and J. E. Ball, "Emerging trends in autonomous vehicle perception: Multimodal fusion for 3D object detection," *World Electr. Vehicle J.*, vol. 15, no. 1, p. 20, Jan. 2024.

[62] L. Liang, H. Ma, L. Zhao, X. Xie, C. Hua, M. Zhang, and Y. Zhang, "Vehicle detection algorithms for autonomous driving: A review," *Sensors*, vol. 24, no. 10, p. 3088, May 2024.

[63] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015.

[64] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The ApolloScape open dataset for autonomous driving and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 10, pp. 2702–2719, Oct. 2020.

[65] L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M.-H. Yang, and S. Lyu, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Comput. Vis. Image Understand.*, vol. 193, Apr. 2020, Art. no. 102907.

[66] M. Naphade, M.-C. Chang, A. Sharma, D. C. Anastasiu, V. Jagarlamudi, P. Chakraborty, T. Huang, S. Wang, M.-Y. Liu, R. Chellappa, J.-N. Hwang, and S. Lyu, "The 2018 NVIDIA AI city challenge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 53–537.

[67] S. Yang, C. Bo, J. Zhang, P. Gao, Y. Li, and S. Serikawa, "VLD-45: A big dataset for vehicle logo recognition and detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 25567–25573, Dec. 2022.

[68] I. Flir Systems. (2018). *Flir's Starter Dataset for Machine Learning Adas*. Accessed: Nov. 28, 2024. [Online]. Available: https://www.flir.com/adasdataset

[69] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "The KITTI vision benchmark suite," *Int. J. Comput. Vis.*, vol. 2, no. 5, pp. 1–13, 2015.

[70] D.-H. Paek, S.-H. Kong, and K. T. Wijaya, "K-lane: LiDAR lane dataset and benchmark for urban roads and highways," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 4449–4458.

[71] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, and Z. Luo, "Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5792–5801.

[72] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1037–1045.

[73] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "LLVIP: A visible-infrared paired dataset for low-light vision," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 3489–3497.

[74] A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, J. Xu, and A. López, "Pedestrian detection at day/night time with visible and FIR cameras: A comparison," *Sensors*, vol. 16, no. 6, p. 820, Jun. 2016.

[75] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.

[76] C. Cuevas, E. M. Yáñez, and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA," *Comput. Vis. Image Understand.*, vol. 152, pp. 103–117, Nov. 2016.

[77] I. Delibasoglu, "PESMOD: Small moving object detection benchmark dataset for moving cameras," in *Proc. 7th Int. Conf. Frontiers Signal Process. (ICFSP)*, Sep. 2022, pp. 23–29.

[78] A. Xuehui, Z. Li, L. Zuguang, W. Chengzhi, L. Pengfei, and L. Zhiwei, "Dataset and benchmark for detecting moving objects in construction sites," *Autom. Construct.*, vol. 122, Feb. 2021, Art. no. 103482.

[79] M. Alrabeiah, A. Hredzak, Z. Liu, and A. Alkhateeb, "ViWi: A deep learning dataset framework for vision-aided wireless communications," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.

[80] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan, and Z. Nie, "DAIR-V2X: A large-scale dataset for vehicle-infrastructure cooperative 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 21361–21370.

[81] M. Weber and P. Perona, "Caltech cars 1999 dataset," California Inst. Technol. (Caltech), Pasadena, CA, USA, Tech. Rep., 2002, doi: 10.22002/D1.20084.

[82] B. Philip, P. Updike, and P. Perona, "Caltech cars 2001 dataset," California Inst. Technol. (Caltech), Pasadena, CA, USA, Tech. Rep., 2001, doi: 10.22002/D1.20085.

[83] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7380–7399, Nov. 2022.

[84] J. Han, X. Liang, H. Xu, K. Chen, L. Hong, J. Mao, C. Ye, W. Zhang, Z. Li, X. Liang, and C. Xu, "SODA10M: A large-scale 2D self/semi-supervised object detection dataset for autonomous driving," 2021, *arXiv:2106.11118*.

[85] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.

[86] H.-Y. Lin, K.-C. Tu, and C.-Y. Li, "VAID: An aerial image dataset for vehicle detection and classification," *IEEE Access*, vol. 8, pp. 212209–212219, 2020.

[87] I. Bozcan and E. Kayacan, "AU-AIR: A multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 8504–8510.

[88] H. Yu, G. Li, W. Zhang, Q. Huang, D. Du, Q. Tian, and N. Sebe, "The unmanned aerial vehicle benchmark: Object detection, tracking and baseline," *Int. J. Comput. Vis.*, vol. 128, no. 5, pp. 1141–1159, May 2020.

[89] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. 13th Eur. Conf. Comput. Vis.-ECCV*, Zurich, Switzerland. Cham, Switzerland: Springer, Jan. 2014, pp. 740–755.

[90] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, vol. 32, no. 1.

[91] S. Seo, D. Chen, K. Kim, K. Kang, D. Koo, M. Chae, and H. K. Park, "Temporary traffic control device detection for road construction projects using deep learning application," in *Proc. Construct. Res. Congr.*, Mar. 2022, pp. 392–401.

[92] M. A. Kenk and M. Hassaballah, "DAWN: Vehicle detection in adverse weather nature dataset," 2020, *arXiv:2008.05402*.

[93] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery (Vedai): A benchmark," *J. Visual Commun. Image Represent.*, 2015. [Online]. Available: https://doi.org/10.1016/j.jvcir.2015.11.002

[94] B. A. Paranjape and A. A. Naik, "DATS_2022: A versatile Indian dataset for object detection in unstructured traffic conditions," *Data Brief*, vol. 43, Aug. 2022, Art. no. 108470.

[95] G. Wen, P. Cao, L. Liu, J. Yang, X. Zhang, F. Wang, and O. R. Zaiane, "Graph self-supervised learning with application to brain networks analysis," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 8, pp. 4154–4165, Aug. 2023.

[96] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.

[97] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[98] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[99] G. Jocher, J. Qiu, and A. Chaurasia. (2023). *Ultralytics YOLO*. [Online]. Available: https://github.com/ultralytics/ultralytics

[100] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[101] OpenAI. (2024). *ChatGPT-4v: Vision and Language Model*. Accessed: Nov. 29, 2024. [Online]. Available: https://openai.com/research/chatgpt-4v

[102] M. L. Ali and Z. Zhang, "The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection," *Computers*, vol. 13, no. 12, p. 336, Dec. 2024.

[103] R. Liu, Y. Zhu, Y. Wang, and Z. Xing, "OB-YOLO: A UAV image detection model for reducing computational resource consumption," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2024, pp. 1–8.

[104] Z. Han, Z. Yue, and L. Liu, "3L-YOLO: A lightweight low-light object detection algorithm," *Appl. Sci.*, vol. 15, no. 1, p. 90, Dec. 2024.

[105] M. Hu, Z. Li, J. Yu, X. Wan, H. Tan, and Z. Lin, "Efficient-lightweight YOLO: Improving small object detection in YOLO for aerial images," *Sensors*, vol. 23, no. 14, p. 6423, Jul. 2023.

[106] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024.

[107] R. Van De Schoot, J. De Bruin, R. Schram, P. Zahedi, J. De Boer, F. Weijdema, B. Kramer, M. Huijts, M. Hoogerwerf, G. Ferdinands, A. Harkema, J. Willemsen, Y. Ma, Q. Fang, S. Hindriks, L. Tummers, and D. L. Oberski, "An open source machine learning framework for efficient and transparent systematic reviews," *Nature Mach. Intell.*, vol. 3, no. 2, pp. 125–133, Feb. 2021.

[108] W. Samek and K.-R. Müller, "Towards explainable artificial intelligence," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham, Switzerland: Springer, 2019, pp. 5–22.

[109] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

JOÃO VITOR DE ANDRADE PORTO was born in Campo Grande, Brazil, in 1999. He received the B.S. degree in computer engineering from Dom Bosco Catholic University, Campo Grande, in 2021, where he is currently pursuing the Ph.D. degree in computer engineering applied to local development. He is undertaking a Sandwich Ph.D. Program at the Faculty of Engineering, Vehicles and Mechatronics Institute, Department of Vehicles Engineering, University of Debrecen, Debrecen, Hungary. He bypassed the M.S. degree stage due to the merit and impact of his publications during the B.S. studies, in 2022. From 2019 to 2021, he was a PIBIC/UCDB/CNPq scholarship holder, conducting research using convolutional neural networks applied to agriculture and livestock. His work led to receiving the first place award in the "Outstanding Scientific Initiation Award" in the Exact and Earth Sciences category for the 2020/2021 cycle. He participated in the UCDB Fórmula Universitária Extension Project in the area of telemetry. He has been a member of the INOVISAO Research Group in computer vision and artificial intelligence, since 2019, with several published works in the field.

PÉTER TAMÁS SZEMES received the B.M.E. and M.Sc. degrees in mechanical engineering and the Ph.D. degree in robotics from the Institute of Industrial Sciences, The University of Tokyo.

He has several years of industrial engineering experience as the Technical Director and later the Chief Engineer of the Center for Invention and Research. He is currently an Associate Professor. His main research interests include robotics (smart space), behavioral recognition, mobile robot control in smart space, and the development of main tools for telemanipulation applications at micro and nano scale.

HEMERSON PISTORI received the B.S. degree in computer engineering, in 2004. He has since led the first Computer Vision (CV) and Artificial Intelligence (AI) Laboratory, INOVISÃO (inovisao.weebly.com), Center-West Region, Brazil. The laboratory has generated dozens of patents and published hundreds of scientific articles related to the development of AI algorithms for agricultural, livestock, and forestry problems. He has also co-authored articles with scientists from the USA, U.K., Italy, Canada, Australia, and Germany. In 2017, his laboratory was the first to apply deep learning (AI) in a drone experiment for identifying invasive plants in soybean fields. This work has been cited by researchers from over 40 countries and is one of the three most-cited works in this field globally. He was also the Vice-Rector for Research and Graduate Studies with Católica University, where he contributed to the creation of the first five doctoral programs at the institution, as well as the Department of International Relations, the Innovation Agency, the Undergraduate Technological Initiation Program, and the Junior Scientific Initiation Program to engage high school students in research activities. He has represented Brazil in the International Association for Pattern Recognition (IAPR), been a member of the Management Committee of the Special Committee on Computer Graphics and Image Processing of the Brazilian Computer Society, and served on the advisory board for the Research Support Foundation (FAP) in the state of Mato Grosso do Sul. He has been involved in review committees of high-impact journals, such as *Pattern Recognition Letters*, *Computers and Electronics in Agriculture*, IEEE Geoscience and Remote Sensing Letters, and *Scientific Reports* (Nature).

JÓZSEF MENYHÁRT was born in Debrecen, Hungary, in 1988. He received the B.Sc. degree in mechanical engineering, in 2011, the M.Sc. degree in construction engineering from the University of Debrecen, in 2013, the Ph.D. degree in military engineering sciences from Óbuda University, in 2018, and the M.Sc. degree in mechanical engineering from the University of Debrecen, in 2019. From 2008 to 2015, he was an Industrial Engineer in electrical and automotive industry. Since 2015, he has been an Assistant Lecturer with the University of Debrecen, where he has been a Senior Lecturer, since 2018, and an Associate Professor with the Department of Vehicles Engineering, since 2020. From 2023 to 2024, he was the Deputy Head of the Department of Vehicles Engineering. Since January 2024, he has been the full-time Head of the Department of Vehicles Engineering. He is the author/co-author of 54 scientific journal articles and six books. He publishes in Hungarian and English language. His main research interests include optimization of autonomous electric vehicles energy management system and operation with artificial intelligence, and maintenance and production systems optimization with AI in vehicle and robot industry. He is a member of the IEEE Hungary Section. He takes part as a reviewer of scientific journals.