IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 February 2025, accepted 12 March 2025, date of publication 17 March 2025, date of current version 27 March 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3551686*

RESEARCH ARTICLE

# Advancing Object Detection: A Narrative Review of Evolving Techniques and Their Navigation Applications

**SHANELLE TENNEKOON[1], (Student Member, IEEE), NUSHARA WEDASINGHA[2], ANURADHI WELHENGE[1], (Member, IEEE), NIMSIRI ABHAYASINGHE[1], (Member, IEEE), AND IAIN MURRAY AM[1], (Senior Member, IEEE)**

[1]School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, Perth, WA 6102, Australia
[2]Department of Electrical and Electronic Engineering, Center of Excellence in Informatics, Electronics Transmission, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe 10115, Sri Lanka

Corresponding author: Shanelle Tennekoon (h.tennekoon@postgrad.curtin.edu.au)

This work was supported in part by the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University.

**ABSTRACT** Object detection plays a pivotal role in advancing computer vision systems by enabling machines to perceive and interact intelligently with their environments. Despite significant advancements, comprehensive exploration of its evolution and applications in navigation remains underrepresented. This review paper examines the evolution of object detection technologies, from early methodologies to contemporary advancements, and their critical role in navigation tasks. The emphasis was on the significance of contextual learning in enhancing object detection performance by leveraging spatial and temporal information. Furthermore, the limitations of conventional approaches that rely heavily on hand-engineered features are examined. It is then demonstrated that contextual learning facilitates automated feature extraction, resulting in improved accuracy exceeding a 50% increase and adaptability in diverse applications. The review concludes by outlining future trends and opportunities for further advancements in object detection and, underscoring its transformative impact on autonomous navigation and beyond. In summary, this review contributes to a comprehensive understanding of object detection technologies by offering insights into their evolution, highlighting their applications in navigation, and providing guidance for future research in context-aware systems.

**INDEX TERMS** Autonomous navigation, computer vision, contextual features, convolutional neural networks (CNN), deep learning, evolution, navigation, object detection, review.

## I. INTRODUCTION

Accurately identifying the types of objects within a scene is crucial for enabling computer-aided devices to intelligently interact with the visual world [1], [2]. This capability, known as object detection, serves as a core function of advanced AI, enabling machines not only to detect and count objects but also to understand their nature and respond appropriately. Object detection has garnered significant attention in recent years as one of the most fundamental and challenging problems in computer vision [1], [2], [3], [4], [5], [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Berdakh Abibullaev.

The origin of object detection in computer vision traces back to 1966, when a platform was developed to automatically differentiate between the foreground and background in images and extract distinct, non-overlapping objects from real-world scenes [7]. Since then, object detection has evolved from simple detection algorithms to sophisticated deep learning networks, revolutionizing various sectors, including autonomous navigation [8], [9], [10], robotics [11], [12], [13], surveillance systems [14], [15], [16], and assistive technologies [17], [18].

Despite significant advancements, many sophisticated object detection algorithms are still limited in their ability to understand the relationships between objects and their

environment during classification. Understanding these relationships is critical to improve the accuracy of decision making, improve semantic understanding, and resolve ambiguity when objects are visually similar. To develop more advanced algorithms, it is essential to first comprehend the spatio-temporal features of objects, explore the evolution of object detectors, analyze how current AI-based models classify objects, identify their limitations, and propose ways to address these challenges.

This paper presents an in-depth review of the features analyzed during object detection tasks and identifies key features that should be examined to understand the relationship between the detected objects and their environments. By doing so, we aim to gain deeper insight into the nature of objects. We will explore the various frameworks used for object detection, the datasets employed to train and test these models, and the metrics used to evaluate their performance. In addition, we describe the advantages and limitations of each model and examine how contextual learning can enhance performance. Given the critical role of object detection in autonomous driving and robotic navigation, we discuss how current algorithms support state-of-the-art navigation systems, address their limitations, and explore how integrating contextual learning can further improve their effectiveness.

The data presented in Figure 1 form the core body of the literature analyzed in this review. From 1985 to 2024, the studies included foundational work, algorithmic advancements, and applications of object detection in navigation, as well as emerging trends and challenges. By systematically categorizing these publications into key areas of Detection Fundamentals, Algorithms, Applications, and Challenges, the graph provides a statistical understanding of the focus of the study on the evolution and emerging developments in autonomous systems. By linking past, present, and future trends, Figure 1 represents the transformative impact of object detection on navigation applications and highlights the ongoing evolution of the fields discussed in this review.

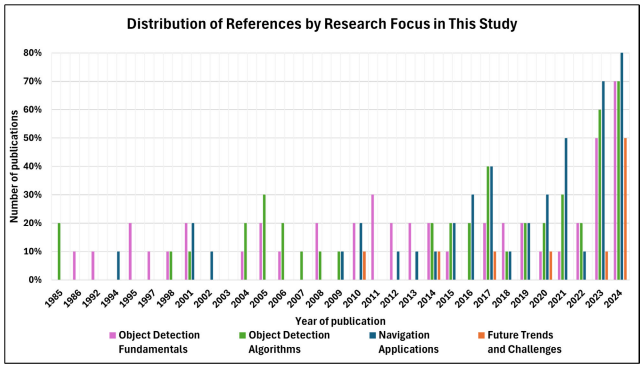


**FIGURE 1. Statistics of publications analyzed in this review paper, categorized by Detection Fundamentals, Algorithms, Applications, and Future Trends.**

The main motivation of this study is to provide a foundation for researchers and engineers seeking to enhance the performance of object detection algorithms through contextual learning. We also aim to discuss how integrating contextual learning into navigation systems can enable more precise decision making. The contributions of this study are as follows.

- Discuss the spatial and temporal features that aid in distinguishing and providing a deeper understanding of an object's nature during classification.
- Propose a comprehensive and current survey on the evolution of object detection algorithms.
- Examine the limitations of current object detectors and analyze how contextual learning can enhance performance.
- Discuss the working principles, advantages, and disadvantages of both older and more recent navigation systems.
- Elaborate how contextual learning can further improve navigation systems.

The remainder of this paper is organized as follows (see Figure 2). Section II covers the fundamentals of object detection, including a review of feature frameworks used over the past two decades. Section III discusses the evolution of object detection techniques from traditional detection methods to machine learning and deep learning techniques. Section IV provides a comprehensive discussion on the evolution of navigation applications. Finally, in Section V, we explore the current challenges and future trends in object detection, emphasizing how contextual learning can enhance both accuracy and efficiency.

## II. EXPLORING THE FUNDAMENTALS OF OBJECT DETECTION

Object detectors closely mimic the sophisticated process of the human visual system in object identification. Humans naturally begin by perceiving local features such as edges, textures, shapes, and colors, and then integrate these details with contextual cues such as depth, lighting, and the surrounding environment to fully understand an object and its relationship to the scene [19]. This intricate process allows rapid and accurate recognition, even in dynamic or complex settings. Similarly, object detection models rely on foundational features learned during training to make precise predictions. This section explores the key local, contextual, and temporal features that automated detectors should consider for the precise classification of objects.

### A. LOCAL FEATURES

Local features are small-scale elements that provide detailed information regarding an object's appearance, which is essential for accurate classification in computer vision systems. Initially, pixel-based methods such as template matching [20] were used, where a template slides over an image to find matching regions based on the pixel intensity. However, this method struggles with variations in rotation and scale [21], [22], [23].

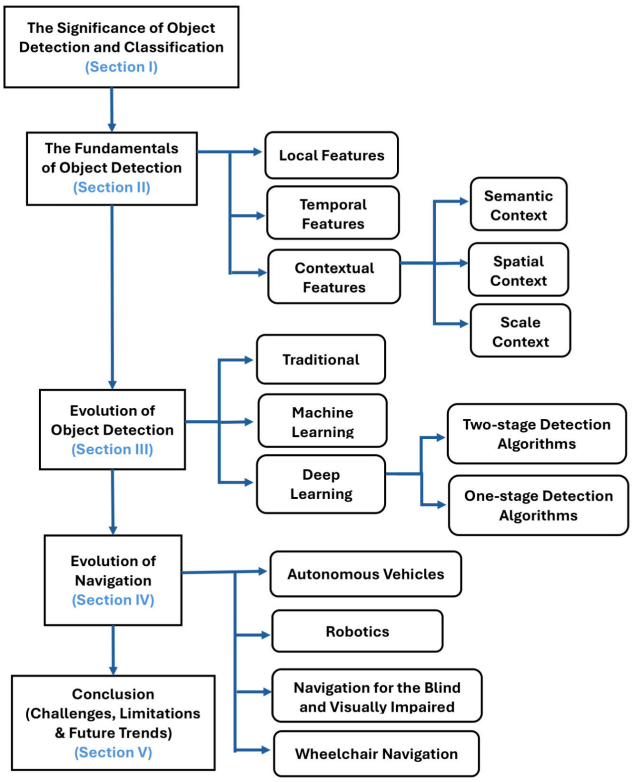


**FIGURE 2.** Structured overview of the review paper: A section-wise flowchart depicting the logical progression of topics.

Among these local features, edges play a pivotal role in defining the shape of an object, making them fundamental in object detection [24]. Early edge detection methods, such as Roberts, Prewitt, and Sobel [25], were limited by noise sensitivity and poor edge sharpness [26]. The Canny Edge Detector [24], introduced in 1986, improved this by applying a multi-stage process that included noise reduction with Gaussian filters, gradient calculation, and double thresholding to capture essential boundary information. Despite its efficiency, the Canny method has limitations, such as incomplete edge detection. To address this, the Hough Transform-based Canny (HT-Canny) [27] replaced the double threshold method to improve edge accuracy using edge-point gradient calculations and connected edges through the Hough Transform [28].

Although these methods were foundational, advancements in feature-based methods brought even more robust solutions to the forefront, outperforming pixel-based approaches. In 2001, Viola and Jones [29] introduced Haar-like features for facial recognition based on Papageorgiou's work [30] using an integral image to sum the pixel intensities of rectangular regions for pattern identification. In 2004, Lowe introduced the Scale Invariant Feature Transform (SIFT) [31], which extracts key features invariant to scaling, rotation, and lighting changes. Despite its robustness, SIFT's computation was slower. Zhao et al. addressed this by introducing a flip-invariant SIFT (F-SIFT) [32], preserving properties of SIFT while tolerating flips.

Subsequently, in 2005, Dalal and Triggs [33] introduced a Histogram of Oriented Gradients (HOG), where local gradients characterize object shape without precise edge positions, thereby proving effective in pedestrian detection. In 2006, Bay et al. [34] introduced Speeded-up Robust Features (SURF), improving efficiency with Hessian matrix-based detection and integral images and achieving faster keypoint detection than SIFT.

Color features are also important for object detection and for complementing shape-based methods [35]. Swain and Ballard's color histograms [36] revolutionized object recognition by describing color distribution across an image. Recent advancements include RGB-D detection methods that combine color and depth information [37] for accurate object detection in real-world applications.

Although analyzing individual features provides valuable insights, it often lacks the ability to capture the broader context necessary for a comprehensive understanding of objects within a scene. Convolutional Neural Networks (CNNs) address this limitation by integrating multiple features across layers, allowing for a more holistic representation that incorporates both local details and contextual relationships. Its early layers extract low-level features such as edges, whereas deeper layers capture high-level semantic information, such as object parts. Architectures such as AlexNet [38], VGGNet [39], and ResNet [40] further enhance feature extraction by progressively refining the details through deeper layers.

### B. TEMPORAL FEATURES

Human visual perception relies on observing object movements and changes over time, allowing the human brain to recognize and predict their actions. This natural ability to track motion and detect patterns is mirrored in object detection systems using temporal features. By analyzing the sequences of frames, these systems capture dynamic changes and interactions, similar to how our brains process visual information. Incorporating temporal features enables object detection models to achieve a deeper understanding of motion and behavior, improving accuracy in scenarios with moving or occluded objects, much like our visual system enhances our awareness of and response to our environment.

One of the earliest mentions of temporal features was in [41], where the study indicated that while spectral and spatial domain features were derived from a single image of a scene, temporal domain features required multiple images of the scene [42].

In 1995, Gil et al. [43] describes a motion-analysis system for vehicle tracking in real-world highway scenarios. They utilized temporal information to analyze the movement of objects between frames. By examining how objects move and change their positions, the system was able to track their trajectories and predict future locations. They integrated temporal features to enhance the tracking process, specifically addressing challenges such as occlusions and

object interactions over time. Gunsel et al. [44] presented a content-based temporal video segmentation system that combines syntactic and semantic features to manage video data. It involves scene change detection and shot classification through a two step process of tracking user-defined semantic objects and an unsupervised method to detect cuts and edits. Object tracking uses region matching, whereas an unsupervised clustering aids scene change detection without threshold selection.

One of the most recent studies by Zadaianchuk et al. [45] emphasized the importance of temporal features in object detection by capturing motion dynamics and inter frame relationships in video data. The incorporation of temporal information such as future predictions has been shown to significantly enhance detection performance. Temporal similarity loss improves object tracking by maintaining object identity, even when the size changes, thus addressing fragmentation issues in backgrounds and large objects. Effective scene decomposition integrates motion (via optical flow) and semantic information, underscoring the necessity of temporal features for real-world object discovery. Many unsupervised video object-centric models fail to effectively utilize temporal correlations, but self-supervised losses based on feature similarities address this shortcoming. The study also reveals that integrating temporal features can improve performance metrics, such as a 20-point increase in the foreground adjusted Rand index (FG-ARI) [46] and a 7-point rise in mean object tracking performance (mBO) [45]. These methods also demonstrate robustness to camera motion and generalize well to unseen datasets, thereby highlighting their effectiveness.

Although temporal features are essential in object detection to capture motion and changes over time, they have limitations. These include difficulty handling occlusions, variations in speed, and inability to robustly model complex temporal dependencies, often leading to misclassifications in dynamic scenes [47]. Furthermore, temporal data can be computationally intensive and require significant resources for accurate processing. Contextual learning can address these shortcomings by integrating spatial and semantic cues to enhance a model's understanding of object relationships across frames. This combination allows for more accurate detection in dynamic environments, improved robustness, and reduced false positives.

### C. CONTEXTUAL FEATURES

Contextual information refers to the additional data surrounding an object or event that enhances the understanding of its role and significance within a specific environment. In computer vision and machine learning, this involves analyzing the relationships and interactions between objects, their spatial arrangements, and the overall scene context. For example, in a street image, contextual information includes the presence of road signs, pedestrians, and vehicles, which are essential for accurate scene interpretation and decision making. By incorporating these contextual cues, systems can more precisely recognize, predict, and respond to elements within the environment. Galleguillos and Belongie [48] identified three types of contextual learning that enhanced scene understanding beyond isolated object features.

1) **Semantic context:** Features that help understand the meaning and relationships of objects within a scene, focusing on their roles, functions, and interactions with other elements.
2) **Spatial context:** Attributes that describe the arrangement, position, and relationships of objects within a scene, enabling a more accurate interpretation of their locations and interactions.
3) **Size/scale context:** Characteristics involving the recognition of objects based on their relative sizes and scales within a scene, as well as the interpretation of their dimensions compared to other objects or the surrounding environment.

#### 1) SEMANTIC CONTEXT

In the semantic context domain, Farhadi et al. [49] examined how distributional semantics could mitigate the issue of out-of-vocabulary words by linking sentences and images through semantic similarity. They proposed a model that maps sentences and images to a shared semantic space, thereby improving the recognition of unknown objects and actions using known categories. This model leverages features such as geometry-based classification scores, HOG features, Felzenszwalb detection responses [50], and gist-based scene classifications [51]. A linear Support Vector Machine (SVM) predicts node features and k-nearest neighbor averaging is used for similar images and sentences. While the model effectively correlates images with sentences, it struggles with out-of-vocabulary words and complex sentence structures, indicating the need for further refinement.

Li and Hoiem [52] explored the "Learning without Forgetting" (LwF) method, which preserves the performance on previously learned tasks while adapting to new ones in machine learning models. Focusing on image classification, this study examines how task similarity influences the retention of old task performance. Using datasets such as PASCAL VOC [53] and Caltech-UCSD Birds [54], the authors demonstrated that LwF outperformed traditional fine-tuning methods, effectively maintaining the performance on old tasks while learning new ones. This study highlights the importance of task similarity and suggests extending LwF to applications such as semantic segmentation and object detection. However, it has limitations, including the need for all new task data to be available before updating old task responses, and the gradual decline in performance on old tasks as new tasks are introduced.

Liu et al. [55] presented a novel approach to pedestrian detection by framing it as a high-level semantic feature-detection task, moving away from traditional low-level feature methods. They proposed a convolution-based detector

designed to identify pedestrian central points and predict their scales, simplifying the detection process and reducing the need for complex configurations and extensive post-processing. This method utilizes multiscale feature maps generated by a CNN, combining shallow maps for precise localization with deeper maps for semantic contexts. The study concluded that this approach, known as the Center and Scale Prediction (CSP) detector, achieved state-of-the-art performance in pedestrian detection benchmarks, demonstrating strong performance across various scales and conditions, including occlusions. While effective, the study notes challenges such as the reliance on high-level abstractions, which may falter in cluttered environments, and the influence of training data quality on model performance.

Chen et al. [19] introduced an iterative contextualization procedure based on the Context-SVM model to enhance object classification and detection by mutually reinforcing these tasks. Their method effectively learns context models across various conditions and achieves superior results on benchmark datasets, such as VOC 2007 [56], VOC 2010 [57], and SUN09 [58], outperforming state-of-the-art methods. In addition, they introduced ambiguity modeling techniques and refined the max-margin model to improve contextual learning.

#### 2) SPATIAL CONTEXT

Spatial features are crucial for understanding the semantics of a scene, as they describe the interaction between an object and its environment in a given time period [59]. This makes them an essential component of the discussion of object detection. In this section, we examine how spatial features enhance object classification by enabling automated models to learn the relationship between objects and their environments.

As one of the early influential studies on incorporating spatial information, Hoiem et al. [60] in 2005 pioneered the use of geometric context in object detection, leveraging spatial relationships between objects and their surroundings. By analyzing typical object configurations and interactions, their method integrates contextual information to refine localization and classification and improve detection in occlusions and complex scenes. The study showed that geometric context can be effectively applied even in unstructured outdoor images.

Similarly, Hoiem et al. [61] and Bao et al. [62] explored the use of object interdependence, three-dimensional (3D) spatial geometry, and camera orientation as contextual information. Galleguillos et al. [63] expanded on this by investigating contextual interactions at the pixel, region, and object levels and integrating these insights using a multi-kernel learning algorithm [63], [64]. Despite the significant advantages of contextual learning for object detection, these approaches commonly require extensive manual efforts, such as labeling contextual objects or scene parts and specifying the locations and spatial relationships between target objects and their context [65].

To address this limitation, Zheng et al. [65] proposed a context modeling framework that enhances object detection using a polar geometric context descriptor and the Maximum Margin Context (MMC) model. These models excel in ambiguous contexts, especially for detecting people, and outperform existing methods on the PASCAL VOC 2005 [66] and 2007 [57] benchmarks. The study highlighted the importance of spatial context, using techniques such as Markov Random Fields (MRF) and Conditional Random Fields (CRF), with examples such as ''a boat on a river'' or ''a car on a road.'' This framework allows unsupervised detection without manual labeling. Notably, the MMC model increased average precision for people detection by 4.22% compared to the proposed 'Things and Stuff' (TAS) model, outperforming it in 8 out of 10 categories.

As the interest in contextual learning has grown, research in this area has expanded. In 2019, Liu et al. introduced a deep salient object detection network [67] that improved feature discriminability using a novel guidance strategy and Group Convolutional Module (GCM). The network integrates shallow feature maps for low-level spatial details and deep maps for high-level semantic understanding, thereby enhancing detection accuracy in complex scenes. This method outperformed previous techniques, particularly in terms of background suppression and salient object identification. However, they struggled with highly complex semantic scenes, indicating the need for future improvements in scene understanding and parsing.

Subsequently, Xu et al. [68] advanced object detection by integrating a Single Shot MultiBox Detector (SSD) with ensemble learning, context modeling, and multiscale feature representation. They introduced Non-maximum suppression (NMS) Ensembling and Feature Ensembling, achieving high accuracy for PASCAL VOC [57] (mAP = 81.1%) and MS COCO [69] (mAP = 78.1%). The framework combines spatial features from shallow and deep layers using deconvolution and concatenation to enhance detection accuracy. While deconvolution fusion outperformed upsampling, the study noted increased computational overhead, network complexity, and potential noise from ensemble methods, which affected model speed and performance.

#### 3) SCALE CONTEXT

The scale context plays a critical role in scene understanding by providing information about the relative size of objects within an image. The scale of an object can indicate its distance from the camera, importance in the scene, or relationship with other objects. The effective interpretation of scale helps models differentiate between small, distant objects and large, close objects, which is crucial for tasks such as object detection. The ability to understand and adapt to scale variations is essential for accurate scene comprehension, particularly in complex environments where objects appear at diverse distances, orientations, and sizes. Models such as Feature Pyramid Networks (FPNs) [70], [71] use multiscale feature extraction to detect objects of different

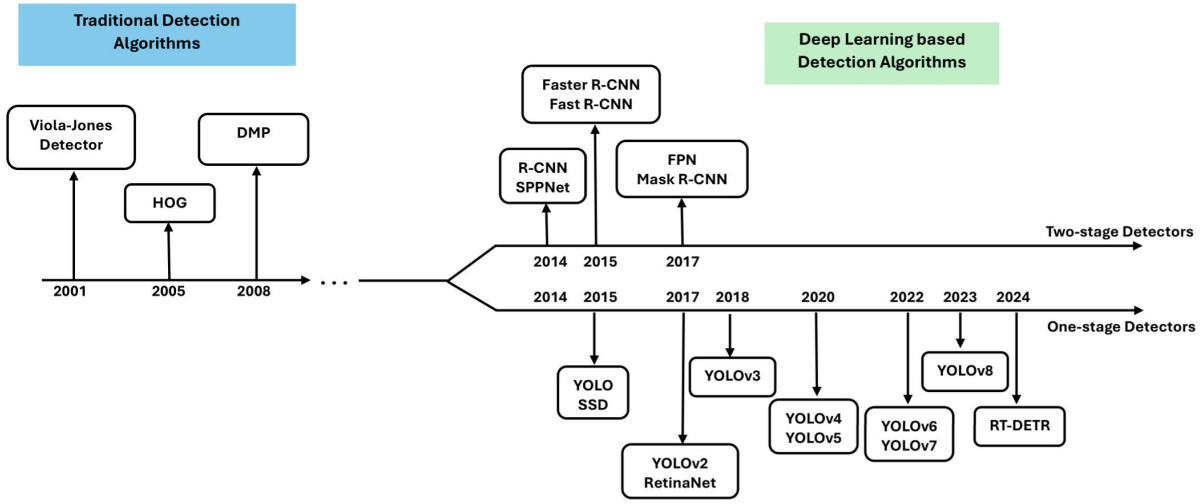


**FIGURE 3.** Evolution of object detection algorithms from 2001 to 2024: A visual timeline of traditional and deep learning based detection algorithms.

sizes. The use of hierarchical features allows models to capture both large and small objects within the same image better.

Understanding the importance of this, Divvala et al. [72] emphasized that object size is a key contextual factor in improving detection accuracy, with the strongest impact compared to other cues, such as location and presence. Excluding the object size results in a significant decline in average precision, particularly for small and occluded objects, highlighting its crucial role in enhancing object localization and classification. Subsequently, Wang et al. presented a novel approach for salient object detection by integrating edges and superpixel regions to boost performance [73]. It underscored the critical role of multiscale context, demonstrating that larger scale regions are essential for preserving object compactness, an important factor in achieving accurate detection. The authors contended that traditional methods, often dependent on hand-crafted features, fail to optimize detection performance by adequately considering object size and scale, highlighting a gap in existing approaches.

Similarly, in 2018, Kim et al. introduced a Scale Aware Network (SAN) [74] designed to improve CNN-based object detection by projecting convolutional features from multiple scales onto a scale-invariant subspace. At the time, this innovative approach enhanced detection accuracy by addressing a key limitation in traditional CNN models, which often falters in handling scale variations, leading to inconsistent feature representation. By mitigating the impact of object size on feature extraction, the SAN strengthens the robustness and effectiveness of object detection across a wide range of object scales, offering a more adaptable solution to this fundamental challenge.

More recently, Qiu et al. [75] underscored the critical role of scale in object detection, demonstrating how multiscale features enhance the detection performance across small, medium, and large objects. The hierarchical context embedding (HCE) module, which integrates segmentation features from different scales, improves the understanding of complex scenes and detection accuracy. The experimental results revealed an increase in the average precision (AP50:95) from 40.5% to 44.7%, with notable gains for small objects (1.2%), medium objects (3.4%), and large objects. The proposed method outperformed Faster R-CNN [5] by 5.9% and showed a 4.7% improvement, even without segmentation features [75]. The ability of the HCE module to explain the relationship between contexts further refines detection, particularly for obscured or challenging objects. These findings highlight the importance of scale and contextual features for improving object detection across diverse object sizes and complex scenes. Recent advancements in Vision Transformers (ViTs) [76], [77], [78] have further demonstrated the importance of global context in object detection, enabling models to capture long-range dependencies and relationships between objects. By incorporating contextual learning, these models achieve better localization accuracy and reduce the number of false positives in dynamic environments.

## III. EVOLUTION OF OBJECT DETECTION

In the past two decades, object detection has been addressed in two periods/techniques: 'Traditional object detection algorithms' and 'object detection algorithms based on machine learning and deep learning'. This section explores the evolution of detection algorithms over time, tracing their development from the traditional techniques to the advanced methods used today. Early traditional detection algorithms relied primarily on sliding windows for region detection and feature extractors [33]. An overview of this evolution is shown in Figure 3, and the algorithms are discussed in the following subsection.

### A. TRADITIONAL DETECTION ALGORITHMS

Traditional object detection techniques have laid the groundwork for modern advancements by introducing various approaches to enhance the accuracy and efficiency. These

methods employ innovative strategies such as feature selection, edge direction characterization, and part-based modeling to address different detection challenges. VJ-Det [29] is a pioneering real-time face detection algorithm proposed by Viola and Jones. It uses AdaBoost for feature selection, integral images for fast computation, and a cascade classifier to efficiently detect faces. However, the sliding window technique and its computational demands pose limitations [1]. Histogram of Oriented Gradient (HOG) detector proposed by Dalal and Triggs [33] improves object detection by leveraging edge direction distributions, making it robust against lighting and pose variations, particularly for pedestrian detection. The Deformable Part Model (DPM) by Felzenszwalb et al. [79] extended HOG by treating objects as collections of deformable parts, thereby enhancing detection accuracy. Although highly influential, the DPM has since been surpassed by deep learning methods. A detailed comparison of these techniques, highlighting their strengths and limitations, is shown in Table 1.

### B. OBJECT DETECTION ALGORITHMS BASED ON MACHINE LEARNING

Traditional object detection algorithms rely heavily on manual feature engineering, which often results in biased and brittle models. This approach, constrained by predefined patterns, limits the ability of the model to generalize and adapt effectively. In contrast, the advent of machine learning and deep learning offers a transformative solution by automating the feature-learning process. These advanced methods enable models to independently discover and optimize feature relationships from data, thereby overcoming the limitations of hand-engineered approaches. Recent advancements demonstrate that automating feature learning through these techniques not only enhances robustness and flexibility, but also improves the overall accuracy of object detection, marking a significant evolution in the field.

These models have evolved significantly over time driven by advancements in computational power and theoretical insights. Early models such as linear and logistic regression laid the groundwork for simplicity and interpretability. Algorithms such as k-Nearest Neighbors (k-NN) [83], [84], and SVM [85] were introduced in the 1990s with advanced classification capabilities. The 2000s saw the rise of ensemble methods such as Random Forests [86] and boosting techniques, which improved accuracy and robustness. In recent decades, the dominance of deep learning has enabled unprecedented performance in complex tasks. A detailed comparison of these techniques, highlighting their strengths and limitations is presented in Table 2.

### C. OBJECT DETECTION ALGORITHMS BASED ON DEEP LEARNING

With the continuing influence of traditional object detection algorithms, a new era of deep learning algorithms was developed in 2012 [1]. After deep learning algorithms were introduced for object detection, CNNs have been widely used for this task. This method utilizes multi-structured network models that feed the features extracted from images into a classification network [96]. This has significantly improved the accuracy and efficiency of object detection. These algorithms can be further divided into two categories. The first is the 'Two-stage detection algorithms' that generate candidate boxes and feature extraction using CNN. This method can achieve high accuracy. However, it cannot perform real-time object detection and is slow in speed [96]. Overcoming this, the second type, 'One-stage object detection algorithms' were developed based on direct regression. A popular algorithm that falls into this category is the 'You Only Look Once' (YOLO) series [97]. This improves the speed of detection and enables real-time object detection. The primary difference between the two categories is that the former generates a region proposal where the objects of interest consist of a set of regions in the image before classification [98], and the latter directly obtains classification and position without generating a region proposal. A brief description of this evolution is provided below.

#### 1) TWO-STAGE DETECTION ALGORITHMS

Two-stage detection algorithms represent a pivotal advancement in object detection by systematically improving the accuracy and precision. In the first stage, the algorithms generate region proposals or candidate regions that are likely to contain objects. The second stage classifies the proposals and refines their boundaries. This approach typically results in a higher detection performance than single-stage methods, as it allows for a more detailed analysis of potential object locations. In 2014, Girshick et al., introduced R-CNN [3] which processes object detection in three stages: generating region proposals via selective search [99], extracting features using a CNN, and classifying regions with a linear SVM [1] [100]. Despite achieving 53.3% mAP on PASCAL VOC 2012 [101], R-CNN is slow and computationally expensive, requiring 14s per image on a GPU [1]. He et al. addressed these limitations using SPPNet [102], which introduced spatial pyramid pooling to improve speed and efficiency. Fast R-CNN further streamlined the process, achieving 70% mAP [1] on PASCAL VOC 2007 [57] and faster training time [4]. Faster R-CNN introduced a Region Proposal Network [5], significantly enhancing speed and accuracy. Mask R-CNN added instance segmentation [103], whereas Feature Pyramid Networks improved multiscale object detection [70]. A detailed comparison of these techniques, highlighting their strengths and limitations, is shown in Table 3.

#### 2) ONE-STAGE DETECTION ALGORITHMS

Two-stage detection algorithms generate region proposals in which the objects of interest are located within specific regions of an image. This approach improves accuracy, but is slow and complex, making it unsuitable for real-time

**TABLE 1. Overview of traditional detection algorithms: Key features, Strengths, and Limitations.**

| Algorithm | Description | Key Features | Strengths | Limitations |
|---|---|---|---|---|
| Viola-Jones Detector (VJ-Det) | The first real-time face detector developed by Viola and Jones [29] | • Feature selection with AdaBoost [80]<br>• Feature computation with integral image<br>• Cascade structure for classifiers [81] | • High frame rate: 15 fps on 384x288 pixel images on a conventional 700 MHz Intel Pentium 111 [29]<br>• Approximately 15 times faster than other methods at the time [29] | Complex calculations due to sliding window technique requiring significant power and computation [1] |
| Histogram of Oriented Gradient (HOG) Detectors | Developed by Dalal and Triggs [33], using edge direction distribution for object characterization. | • Fine orientation binning [33]<br>• Robust in varied conditions [33]<br>• High-quality local contrast normalization [33] | • Real-time detection<br>• Robust to diverse conditions (lighting, backgrounds, pose)<br>• High accuracy with SVM [82] | • Primarily used for pedestrian detection only [33]<br>• Limited focus on broader object classes |
| Deformable Part Model Detector (DPM) | Introduced by Felzenszwalb et al. [79] extending HOG with part-based representation. | • Object represented as a collection of parts<br>• Deformable part configuration<br>• Uses 'divide and conquer' approach [1] | • Highly accurate and efficient [33]<br>• Significantly higher recognition rates [50]<br>• Flexibility in capturing object variations [50] | Outperformed by modern deep learning techniques and considered less advanced |

**TABLE 2. Overview of machine learning algorithms: Key features, Strengths, and Limitations.**

| Algorithm | Functionality | Strengths | Limitations | References |
|---|---|---|---|---|
| Bayesian Networks (BN) | Graphical model that represents probability relationships among variables. | Captures probabilistic dependencies and prior knowledge, handling complex relationships. | Not suitable for high-dimensional datasets due to time and space constraints | [87] |
| Naïve Bayes | BN with strong independence assumptions among features. | • Less computational time for training<br>• Simplified BN<br>• Effective for large datasets<br>• Probabilistic output | • Assumes independence among features, which may not hold.<br>• Less effective with feature interactions | [88]–[91] |
| Logistic Regression | A statistical model where a logistic curve is fitted to the dataset to binary outcome data. | • Probabilistic interpretation<br>• Easy to update with new data<br>• Handles interaction and non-linear effects. | • Requires large sample sizes for stable results<br>• Less effective with highly complex data | [86], [92] |
| Decision Trees | Classifies data by splitting it based on feature values. | • Interpretable<br>• Handles interactions and various data types<br>• Robust to noise. | • Prone to overfitting<br>• High-dimensional data can lead to fragmentation<br>• Errors propagate through the tree | [89], [90], [93] |
| Random Forests | Ensemble method using multiple decision trees to improve classification performance. | • Scalable<br>• Robust to noise<br>• Reduces overfitting<br>• Easy to interpret | • Slower for real-time predictions as the number of trees increase.<br>• Computationally intensive | [86], [94] |
| Support Vector Machines (SVM) | Classifies data by finding the hyperplane that maximizes the margin between classes. | • High accuracy<br>• Works well with high-dimensional data<br>• Robust to overfitting. | • Requires careful parameter tuning<br>• Training speed is slow<br>• Complex for large datasets | [92], [95] |
| K-Nearest Neighbors (k-NN) | Classifies data based on the majority class of nearest neighbors. | • Simple architecture<br>• Effective for multi-modal classes | • Computationally expensive<br>• Sensitive to noise and irrelevant features<br>• Performance depends on 'k' | [86]–[88] |

applications. By contrast, one-stage detection algorithms detect all objects in a frame in a single step, making them more efficient and better suited for real-time detection [1].

One of the groundbreaking object detection algorithms, You Only Look Once (YOLO), was introduced by Joseph et al. [6] with real-time detection capabilities. It revolutionizes object detection by framing it as a regression problem, enabling a single neural network to predict bounding boxes in one pass. This approach significantly improves both accuracy and speed. YOLO demonstrated excellent performance on high-end GPUs, but struggled with lower-end systems like CPUs, as noted by Redmon et al. [6]. A comparative analysis of YOLO versions and evaluation of their performance metrics, strengths, and limitations to provide a comprehensive understanding of their evolution and effectiveness across different conditions is presented in Table 4.

The field of one-stage object detection algorithms has witnessed significant advancements beyond YOLO, with several notable approaches emerging to address various

**TABLE 3. Overview of two-stage detection algorithms: Key features, Strengths, and Limitations.**

| Algorithm | Description | Key Features | Performance | Limitations |
|---|---|---|---|---|
| **R-CNN** | Proposed by Girshick et al. [3]. Its 3 step functionality:<br>1. Selective search [99] to generate region proposals<br>2. Re-scaling to a fixed resolution [1].<br>3. CNNs and SVMs for prediction within regions [1] [100] | • Region proposals generated via selective search.<br>• Rescaling of regions.<br>• 4096-dimensional CNN features as final representation [100]<br>• Linear SVM for classification | • mAP=53.3% [3] on PASCAL VOC 2012 [101].<br>• Detection speed ∼14 seconds per image with GPU [1].<br>• 2.5 GPU-days for 5k images of the VOC 2007 trainval set with VGG16 [4] | • Expensive training and storage requirements [100]<br>• Slow detection speed [1] |
| **SPPNet** | He et al. [102] introduced spatial pyramid pooling to avoid fixed-size image requirement and improves speed. | • Spatial pyramid pooling for arbitrary image sizes [102]<br>• Does not require re-scaling [102]<br>• Single forward pass through CNN [102]<br>• Faster than R-CNN | • mAP=59.2% on PASCAL VOC 2007 [102]<br>• 20 times faster than R-CNN [102]<br>• Improved speed and reduced computational cost. | • Multistage training process.<br>• Limited to passing feature extraction only once through convolutional layers, despite applying fully connected layers to each region proposal [1] |
| **Fast R-CNN** | Introduced by Girshick et al. [4] to improve speed and accuracy of R-CNN by incorporating multi-task loss for single-stage training. | • Single-stage training.<br>• Region of Interest (RoI) pooling to extract feature maps of fixed sizes for each region proposal.<br>• Shared feature map for for classification and rounding box regression.<br>• Pooled features fed into fully connected layers.<br>• Faster training and inference. | • mAP=70% on PASCAL VOC 2007 [1]<br>• Detection time 200 times faster than R-CNN [1]<br>Training time:<br>• 9 times faster than R-CNN [4]<br>• 3 times faster than SPPNet [4] | Limited by proposal detection [1] |
| **Faster R-CNN** | Ren et al. [5] introduced a Region Proposal Network (RPN) for nearly cost-free region proposals and improved accuracy. | • Region Proposal Network (300 proposals per image).<br>• Shares convolutional features with detection network.<br>• Near-real-time detection | • mAP=73.2% on PASCAL VOC 2007 [5]<br>• mAP=70.4% on PASCAL VOC 2012 [5]<br>• 17 fps with ZF-Net [104] | Computational redundancy at subsequent detection stages [1] |
| **Mask R-CNN** | He et al. [103] added object mask prediction in parallel with bounding box detection for better perforamnce. | • Parallel object mask prediction.<br>• Two-stage process with RPN and mask prediction.<br>• High-quality segmentation masks.<br>• Overcame challenges where classification depended on mask predictions [105]–[107] | • Surpassed COCO 2016 keypoint competition results [103]<br>• Runtime of 5 fps. | Not optimized for speed [103] |
| **Feature Pyramid Networks (FPNs)** | Proposed by Lin et al. [70], this generates multiscale feature maps to handle objects at different scales. | • Top-down architecture with lateral connections [70]<br>• Builds high-level semantic feature maps at all scales [70]<br>• Addresses scale variability.<br>• Improves object detection with a wide variety of scales since CNNs naturally form feature pyramids through forward propagation [1]. | • mAP=59.1% on COCO dataset [1]<br>• Runs at 5 fps on a GPU [70]<br>• Improved detection further with improved versions, Parallel FPN (PFPNet) [108], GraphFPN [109], Improved FPN (ImFPN) [110], Bi-directional FPN (BiFPN) [111] | computationally expensive to generate feature maps at multiple scales. |

challenges. These algorithms have been developed to enhance speed, accuracy, and efficiency. For a better understanding, a concise overview of key one-stage detection algorithms, outlining their unique features and performance metrics and illustrating their evolutionary progress is presented in Table 5.

Over the past decade, the evolution of object detection algorithms has demonstrated significant improvements in accuracy, measured by the mean Average Precision (mAP) on benchmark datasets such as PASCAL VOC 2007 [53], PASCAL VOC 2012 [101], and MS COCO [69]. These improvements are illustrated in Figure 4, which compares the

**TABLE 4. Comparison of YOLO versions: Key features, Performance metrics, Strengths, and Limitations.**

| YOLO Version | Year | Key Features | Speed (fps) | mAP (Mean Average Precision) | Strengths | Limitations |
|---|---|---|---|---|---|---|
| YOLO | 2015 | • Single neural network for predicting bounding boxes<br>• Treated object detection as a regression problem | • 155 fps (PASCAL VOC 2007) [6]<br>• 45 fps (Fast YOLO) [6] | • 52.7% (YOLO) [6]<br>• 63.4% (Fast YOLO) [6] | • High accuracy<br>• Extremely fast | • Struggled with accurate object localization [6]<br>• Poor detection of smaller objects [112] |
| YOLOv2 | 2016 | • Inspired by Network-in-Network [113] and VGG [39]<br>• Added data augmentation techniques [112] | 40 fps (VOC 2007) [114] | 78.6% (VOC 2007) [114] | Enhanced accuracy | Struggled with detecting smaller objects [112] |
| YOLOv3 | 2018 | • Based on Darknet-53 framework [115]<br>• Better for small object detection | 22 fps [116] | 82.1% mAP@0.5 (COCO) [116] | Outperformed previous versions in small object detection | Struggled with medium and larger object accuracy [112] |
| YOLOv4 | 2020 | Faster and more accurate than previous versions [97] | 33 fps (MS COCO) [112] | 65.70% [112] | Significant improvements in speed and accuracy | High computation in terms of memory and processing power |
| YOLOv5 | 2020 | • Improvements in model size, flexibility, and data enhancement.<br>• Introduced Hardswish activation function [117] | 129–248 fps (COCO) [116] | 57.9% mAP@0.5 (COCO) [116] | • Flexible model control<br>• Improved activation function | Challenges in detecting smaller objects or objects that are densely packed |
| YOLOv7 | 2022 | • Follow-up work from YOLOv4 team<br>• Introduced optimized structures [1] | 161 fps [118] | 56.8% (COCO dataset, AP50 metric) [118] | Outperformed previous versions in accuracy | Higher computational demand |
| YOLOv8 | 2023 | Similar backbone to YOLOv5 | 280 fps (NVIDIA A100 and TensorRT) [119] | 53.90% [119] | • High speed detection<br>• Real-time object detection | Decreased inference speed due to separate training requirements [96] |

mAP values of the object detection algorithms across these benchmark datasets. The observed trends in accuracy reflect not only advancements in algorithmic design, but also the varying complexity of the datasets themselves.

It was observed that the mAP values varied significantly across datasets, reflecting differences in their complexity and scope. On PASCAL VOC 2007, algorithms generally achieved higher mAP values (e.g., 76.9% for Faster R-CNN) than the MS COCO dataset (e.g., 56.8% for YOLOv8). This discrepancy can be attributed to the inherent challenges of the COCO dataset, which includes 80 object classes [69], diverse object scales, and cluttered scenes. In contrast, the PASCAL VOC datasets are smaller and less complex, making them more amenable to higher accuracy. PASCAL VOC 2012, but more challenging than PASCAL VOC 2007, still yielded higher mAP values than MS COCO, as shown in Figure 4. These dataset-specific trends highlight the importance of considering dataset characteristics when evaluating object detection algorithms.

Furthermore, the progression from early region-based methods, such as R-CNN [3] and SPPNet [102] to modern one-stage and transformer-based models has led to substantial improvements in accuracy. It was observed that R-CNN achieved an mAP of 53.3% [3] on PASCAL VOC 2007, while Faster R-CNN [5], which introduces an RPN, improves this to 76.9%. Similarly, Mask R-CNN [103], which extends Faster R-CNN with mask prediction, achieved competitive mAP

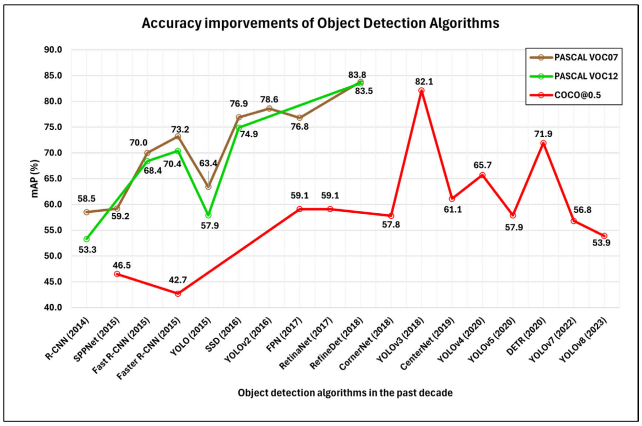


**FIGURE 4. Evolution of object detection accuracy over the past decade across PASCAL VOC [53], [101] and MS COCO [69] Datasets.**

values (e.g., 74.9% on VOC 2007) while enabling instance segmentation. These advancements have demonstrated the impact of architectural innovations on accuracy.

The introduction of one-stage detectors such as YOLO [97] and SSD [56] has marked a significant leap in both speed and accuracy. As shown in Figure 4, YOLOv5 and YOLOv8 achieved mAP values of 74.9% and 76.8% on PASCAL VOC 2007 respectively, while achieving real-time performance. These algorithms eliminate the need for region proposals, reduce the computational overhead, and enable faster inference without sacrificing accuracy. This balance

**TABLE 5. Overview of one-stage detection algorithms: Key features and Performance metrics.**

| Algorithm | Description | Key Features | Performance (mAP@.5) |
|---|---|---|---|
| **Single-Shot Multi-box Detector (SSD)** | SSD combines region proposals and detection into a single network, improving speed and efficiency. It detects objects at multiple scales and layers [56]. | • Multiscale and multiresolution detection techniques [1].<br>• Significantly faster than two-stage methods | 74.3% (300x300 input), 76.9% (512x512 input) on PASCAL VOC 2007 [56] |
| **RetinaNet** | Addressed class imbalance in one-stage detectors with 'Focal Loss,' focusing on harder examples during training [120] | Focal Loss reshapes cross-entropy loss to improve accuracy on complex scenes. | 59.1% on COCO dataset [1] |
| **CornerNet** | Detects bounding boxes as pairs of keypoints (top left and bottom right corners) without anchor boxes [121] | Introduces corner pooling for better corner localization [121]. | 57.8% on COCO dataset [1] |
| **CenterNet** | Represents objects as single center points of bounding boxes, simplifying detection and eliminating anchor boxes [122]. | Uses center points for detection; simpler and faster with end-to-end differentiability. At the post-processing stage, it uses NMS [123] to filter out redundant center points and refine the detections | 61.1% on COCO dataset [1] |
| **Detection Transformer (DETR)** | • Treats object detection as a set prediction problem with a transformer architecture [124].<br>•Avoids anchor boxes and NMS.<br>• The presence and location of objects within the image is predicted by a fixed set of leaned object queries that is sent to the transformer decoder | • Uses transformer encoder-decoder architecture; bipartite matching loss for unique object assignment<br>• DETR directly outputs the final set of predictions in parallel [124] | 71.9% on MS COCO dataset (with Deformable DETR) [1] |

between speed and accuracy makes one-stage detectors particularly suitable for real-time applications.

More recently, the advent of transformer-based models, such as DETR [124], represents a paradigm shift in object detection. DETR leverages self-attention mechanisms to capture the global context and relationships between objects, achieving competitive mAP@0.5 values on MS COCO (e.g., 61.1%). However, the accuracy of DETR still lags behind that of YOLO variants, particularly in real-time applications [125]. Recent improvements, such as RT-DETR [125], have bridged this gap by optimizing the transformer architectures for speed. These developments [126], [127], [128] underscore the potential of transformer-based models, while highlighting the challenges of balancing accuracy and computational efficiency.

These advancements in object detection accuracy, evident by rising mAP values across datasets, can be attributed to a combination of architectural innovations, improved training methodologies, and optimization techniques. These factors have collectively driven the field forward, enabling algorithms to achieve higher performance while addressing challenges such as class imbalance, computational efficiency, and generalization across diverse datasets. Specifically, the following key developments played a pivotal role in enhancing accuracy:

- **Backbone Networks:** The adoption of powerful backbone networks, such as ResNet [40], CSPNet [129], and EfficientNet [130], has significantly improved feature extraction capabilities. For example, YOLOv8's CSPDarknet [129] backbone contributes to its superior performance.
- **Training Techniques:** Advances in training methods, including data augmentation, focal loss (used in RetinaNet [120]), and self-supervised learning, have addressed challenges such as class imbalance and improved generalization across datasets.
- **Real-Time Optimization:** Modern algorithms, particularly YOLO variants [97], are optimized for both accuracy and speed. Techniques such as model pruning, quantization, and the use of mixed precision (FP16) [131] have enabled these algorithms to achieve high mAP values while maintaining real-time performance.

Although it is evident that object detection algorithms depict continuous improvement, incorporating contextual learning can further enhance these models by enabling them to understand objects within their surroundings rather than in isolation. By integrating scene context, spatial relationships, and semantic dependencies, the models can achieve better localization accuracy and reduce false detections.

## IV. EVOLUTION OF NAVIGATION APPLICATIONS

Object detection plays a critical role in navigation applications, enabling systems to recognize and track objects in real-time. Therefore, it is essential to identify obstacles and ensure safe maneuvering in complex environments. In robotics, object detection helps machines navigate through dynamic spaces by identifying and avoiding obstacles, and ensuring efficient path planning. Advancements in object detection algorithms have significantly enhanced their application in various fields. This section discusses several applications and their evolution using object detection algorithms.

Among these applications, autonomous navigation has emerged as one of the most transformative, leveraging advancements in object detection to enable machines to perceive and interact with their environment in real time. The evolution of object detection for navigation applications is demonstrated in Figure 5, which shows the exponential

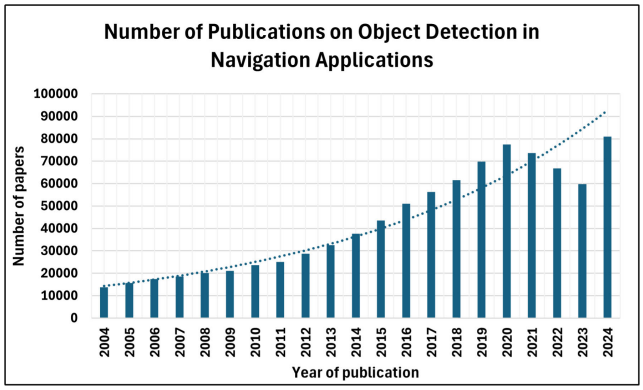


**FIGURE 5.** Increasing number of publications on object detection in navigation applications. (Data from Google scholar advanced search: allwords: "object detection in navigation applications".)

growth of research publications over the past two decades. This growth reflects the increasing importance of object detection in enabling safe and efficient navigation in autonomous vehicles, drones, and robotic systems.

### A. AUTONOMOUS VEHICLES

Real-time obstacle detection in autonomous vehicles has been challenging since its outset. Badal et al. [132] made an early attempt with a generic obstacle detection system that uses stereo images to compute a disparity map, enabling the vehicle to detect and navigate around obstacles. In 2001, Fu et al. [133] improved this approach by introducing a simplified binocular stereopsis method that eliminated the need for disparity calculations and provided more accurate distance measurements. The VJ-detector laid the foundation for pedestrian detection [29], becoming essential in autonomous vehicle technology. Object detection is critical for maintaining safety, efficiency, and compliance with the traffic rules. Kato et al. [134] developed a system using Distance Measurement Point (DMP) technology, which excelled in PASCAL challenges. The system employs Euclidean clustering on point-cloud data to detect objects and obtain distance information, which is then enhanced by sensor fusion with image classification results to improve detection accuracy.

Ess et al. [135] addressed vision-based navigation in urban environments and emphasized the importance of semantic information. Their approach focused on categorizing and tracking moving objects to predict their behavior and plan dynamic paths. By integrating geometric world mapping with object detection and multi-object tracking using a stereo rig, their system achieved high accuracy in detecting and tracking cars and pedestrians, thereby enhancing location prediction and path planning. However, real-time processing is hindered by occlusion and computational complexity, particularly in densely populated areas.

Recently, Wang et al. [136] developed MV2_S_YE, a novel object detection algorithm that enhances YOLOv4 by improving the accuracy and speed of road-object detection. By replacing YOLOv4's CSPDarknet53 [129] backbone with MobileNetV2 [137], the proposed model was able to reduce model complexity. A channel attention mechanism incorporating the Squeeze-and-excitation Networks (SENet) [138] module into Path Aggregation Network (PANet) [139] was used to optimize detection precision. The model was tested on the PASCAL VOC [57], Udacity [140], and KAIST (Korea Advanced Institute of Science and Technology) [141] datasets, where it was observed that MV2_S_YE outperformed YOLOv8s on Udacity and KAIST, achieving mAP@0.5 = 80.9%, 66.7%, and 94.8%, respectively. With a detection speed of 45 FPS on VOC2007, it demonstrated superior performance and efficiency in real-time applications.

Similarly, Wang et al. [142] proposed the YOLOv8-QSD network, which is an advanced anchor-free object detection model for self-driving vehicles that enhance both accuracy and efficiency. Building on YOLOv8, this model utilizes structural reparameterization with a diverse branch block (DBB) backbone for optimized feature extraction. To improve small-object detection, the model integrates a bidirectional feature pyramid network (BiFPN). Furthermore, a novel query-based pipeline was introduced to enhance the long-range detection. Tested on the SODA-A dataset, it outperforms YOLOv8 with a 64.5% accuracy [142] and lower computational demand (7.1 GFLOPs) [142], making it ideal for high-speed autonomous driving by balancing speed, precision, and cost-effectiveness.

### B. ROBOTICS

As autonomous and intelligent agents capable of navigating various environments, robots have become a key area of research. The two fundamental challenges in robot navigation are object detection and obstacle avoidance. Navigations can be classified into two types.

1) **Global navigation:** where prior knowledge of the environment is used [143].
2) **Local navigation:** where the robot autonomously determines its motion and orientation.

One approach to local navigation is Fuzzy Logic [144], which processes imprecise sensor data to make navigation decisions based on a set of fuzzy rules that mimic human reasoning, using degrees of truth rather than binary values. Fuzzy logic has been widely applied in robot navigation. Ren et al. [145] developed an intelligent fuzzy logic controller for dynamic environments to address the challenges in wheeled mobile robot navigation. Bao et al. [146] created a behavior-based fuzzy method for mobile robots that includes obstacle avoidance. Recently, Iwanowsky et al. [147] introduced an image search method that retrieves images based on text queries that describe object classes and their spatial relationships. The input query specifies the desired objects and their relative positions, which are used to score and rank the relevant images, with the most relevant placed at the top. This approach combines object detection techniques with a

fuzzy logic system to evaluate object positions relative to one another, thereby enabling efficient and context-aware image retrieval. Similarly, in 2024, Silva et al. [148] reviewed Fuzzy Logic systems for edge detection to address uncertainties inherent in edge detection. Pandey et al. [149] conducted an extensive review of obstacle avoidance and robot navigation in this domain.

With the advancement of autonomous mobile systems, object detection and identification in indoor environments have become crucial and challenging tasks. Hernández et al. [150] developed a mobile robot for indoor navigation that detects and classifies objects using RGB and depth images as input. The system employs SVM [82] for classification and compares two feature extraction methods: geometric shape descriptors and the Bag of Words (BoW) approach. Geometric shape descriptors focus on low-level features such as contours and spatial properties, whereas BoW clusters local features into a visual vocabulary, represented as a histogram for classification. An experiment demonstrated that geometric shape descriptors outperformed BoW in terms of both speed and accuracy, making the system suitable for real-time mobile robot navigation [150].

To enhance the detection accuracy in robotics, Coates et al. [151] developed a multi-camera view object detection method. This approach integrates multiple camera views into a single-image detection algorithm. Trained on large synthetic datasets using a distributed, parallel learning algorithm, the system processes up to 100 million image patches [151], resulting in a robust object detector. The probabilistic method achieved a notable performance improvement, with accuracy increasing from 0.86 to 0.97 in area-under-curve evaluations. However, the method's high computational complexity, owing to data processing from multiple cameras and synchronization challenges, impacted its real-time performance.

Underwater object detection (UOD) and navigation have presented significant challenges owing to issues such as low visibility, high noise, low contrast, and blurred edges in underwater images. UOD methods generally rely on two techniques: generic object detection (GOD) for locating and identifying objects, and underwater image enhancement (UIE) methods [152], [153], [154] to improve image quality. Liu et al. [155] addressed the limitations of existing datasets, such as insufficient test set annotations and incomplete labels, by introducing a novel dataset with more accurate annotations. This advancement enhanced the efficiency and accuracy of the UOD. Xu et al. [156] provided a comprehensive review of current methods used for UOD. Recently, a novel computer vision-based approach for estimating the position of an Autonomous Underwater Vehicle (AUV) was proposed by Enrico et al. [157]. This method leverages computer vision and deep learning techniques to reconstruct the surroundings of a vehicle during brief surfacing events at its current location.

### C. NAVIGATION FOR THE VISUALLY IMPAIRED (VI)

Technological advancements have led to various solutions designed to assist individuals with visual impairment in their daily lives. These solutions include smartphone applications [158], wearable devices [18], smart canes [17], indoor navigation systems utilizing computer vision [159], guide dogs [160], and smart glasses [161]. Each of these tools addresses the challenges of autonomous navigation for visually impaired users. Ongoing research and development continue to enhance and expand these assistive technologies.

***Smartphone Apps***: Among smartphone apps for VI, BlindSquare [162] is a top GPS application providing detailed site and junction information for safe navigation indoors and outdoors, and it integrates with third-party navigation apps [158]. Microsoft's Seeing AI app [163] uses AI to describe the environment, recognize people, read texts, depict scenes, and identify coins. NavCog3 [164] offers high-accuracy indoor navigation with turn-by-turn guidance and semantic spatial understanding, achieving an average error of 1.65 meters in tests with over 50 participants. The ASSIST app [165] enhances indoor navigation with customizable interfaces and multiple feedback modes, showing significant improvements in accuracy and convenience.

***Smart White Canes***: The white cane, essential for VI, has historically lacked obstacle detection capabilities, leading to reliance on GPS and smartphone apps. Recent advancements include "smart white canes," which integrate modern technology with traditional tools. Chen et al. [166] developed the 'City College of New York' (CCNY) smart cane with simultaneous localization and mapping (SLAM) [167] for indoor navigation, using haptic feedback for guidance. Nguyen et al. [168] introduced a visual SLAM system for localization in Wi-Fi or GPS-free areas by employing fast-appearance-based mapping and a Kalman filter. Hosny et al. [169] created an indoor navigation system that allowed users to choose routes and receive guidance. Subbiah et al. [170] designed a smart cane with object, light, staircase, and heat detection capabilities using IR sensors, GSM, GPS, and ultrasonic sensors, with feedback provided through a Bluetooth headset. Messaoudi et al. [171] developed a system using ultrasonic sensors and Internet of Things (IoT) wireless scanners for obstacle avoidance and wayfinding, whereas Liao et al. [172] used RFID technology for environmental information and voice command navigation. Despite these advancements, smart cane systems still face challenges such as limited detection range, insufficient feedback, short battery life, and bulkiness.

In 2024, Atitallah et al. [173] proposed an obstacle detection system for (VI) individuals utilizing a modified YOLOv5 neural network. The system recognizes and locates key indoor and outdoor objects and aids navigation to VI individuals. The model was optimized using model-width scaling, quantization, and channel pruning to enhance its performance on embedded devices. The model was tested on

the MS COCO dataset and a mAP@0.5 = 81.02% with an inference speed of 67fps was achieved, making the system effective for real-time applications in assistive technologies for BVI navigation.

### D. WHEELCHAIR NAVIGATION

Argyros et al. [174] developed a semi-autonomous wheelchair system using sonar for distance measurements and a 360-degree camera for target tracking. The system determines the object orientation via color histograms and considers non-holonomic kinematic constraints for stability. Baklouti et al. [175] created an autonomous wheelchair using Fuzzy Logic controllers and Kinect Xbox 360 sensors for real-time obstacle avoidance by integrating RGB images with depth data into a Deformable Virtual Zone (DVZ) for reactive control.

Kim Eun introduced an intelligent wheelchair (IW) navigation system [176] that improves the safety of disabled or elderly users by avoiding obstacles using a camera sensor and eight ultrasonic sensors. The system achieved 98.3% accuracy in recognizing outdoor environments and 92.0% accuracy in generating avoidable paths. Field tests with five disabled users revealed high satisfaction, reduced physical effort, and successful navigation. However, the author emphasized the use of machine learning algorithms for contextual learning to enhance obstacle recognition and path generation.

A more recent study in 2021 by Lecronsnier et al. introduced an innovative system tailored for a smart wheelchair aimed at refining navigation within indoor settings [177]. It harnessed the YOLOv3 [71] algorithm for precise object detection, coupled it with Intel RealSense technology for depth discernment, and employed a Simple Online Realtime Tracking (SORT) [178] algorithm for 3D object tracking. This integration ensured adept identification and monitoring of critical navigational aids such as doors and handles, significantly bolstering the wheelchair's semi-autonomous capabilities. Moreover, the discussion extends to semantic mapping, merging semantic insights with environmental contours, and enhancing the wheelchair's spatial intelligence.

However, current wheelchair systems face issues with context-awareness and real-time operation. Liang et al. [179] developed a global planning mechanism for mapless navigation using Deep Reinforcement Learning (DRL); however, it has limitations in smooth motion control. Furthermore, Beale et al. [180] identified challenges related to varying terrain and obstacles that affect wheelchair navigation. To address some of these issues, Kasemsuppakorn et al. [181] proposed the Absolute Restriction Method (ARM), which considers the environmental characteristics and user preferences for route selection, suggesting the need for more personalized navigation solutions.

In 2024, Gallo et al. [182] proposed a system that focuses on a lightweight, embedded navigation system using a monocular RGB camera and deep learning to detect a caregiver's feet and measure their distance from a powered wheelchair (PW). The system achieved metrological accuracy comparable to LiDAR and stereo cameras, with measurement uncertainties within 10cm [182]. By reducing the data volume and detection complexity, the proposed system simplifies calibration and deployment, making real-time object detection and distance estimation more efficient for assistive mobility applications.

Despite numerous applications with advanced detection capabilities, there is still significant room for improvement. These applications rely heavily on object detection algorithms to understand and interact with their environment. However, many of them face challenges in complex, real-world scenarios where objects may be occluded or where environments may change rapidly. To address this, contextual information is crucial for making accurate decisions. By incorporating semantic, spatial, and temporal context into detection algorithms, systems can better interpret relationships between objects, predict movement patterns, and adapt to new environments. Overall, leveraging contextual learning can enhance the robustness and adaptability of such systems in dynamic environments.

Therefore, navigation technologies have advanced significantly, playing a critical role in autonomous vehicles, robotics, BVI navigation, and wheelchair navigation. These technologies integrate computer vision, deep learning, sensor fusion, and IoT-based approaches to improve real-time decision making, obstacle detection, and path planning.

The network diagram in Figure 6 visually represents the interconnections between navigation applications and technologies, showing how methods such as deep learning and computer vision are applied across multiple domains. This highlights their versatility in addressing various navigation challenges.

In autonomous vehicles, disparity mapping improves obstacle detection but is computationally expensive [132], [133], [183], [184], [185]. Sensor fusion enhances detection precision by integrating data through DMP and Euclidean clustering but suffers from high processing time [134], [186], [187]. Deep learning-based YOLO object detection enables rapid obstacle recognition, though occlusions in dense traffic remain a limitation [136], [142], [184]. Feature mapping, including multi-object tracking and SLAM, enhances movement prediction but increases complexity, especially in urban settings [135], [188], [189].

Robotics employs similar advances in navigation. Fuzzy logic systems provide adaptability in dynamic environments, but face scalability challenges [145], [145], [146], [147], [148], [149], [190]. Multi-camera views enhance object recognition through parallel learning but require significant computational resources [151], [152], [153], [154], [155], [156]. RGB-D image classification, using SVM and BoW, improves feature representation, but is slower than deep learning methods [150], [191], [192]. Reinforcement learning and deep learning have enhanced real-time navigation, but are


**FIGURE 6.** Network diagram of navigation technologies across applications: The diagram illustrates the interconnections between navigation technologies and their applications, highlighting the shared methodologies and unique implementations across autonomous vehicles, robotics, BVI navigation, and wheelchair navigation.

further constrained by lighting conditions and computational limitations [150], [184], [193], [194], [195].

For BVI navigation, smart canes with SLAM, RFID, and ultrasonic sensors improve indoor navigation, but have a limited detection range [17], [196], [197], [198]. Wearable AI, using YOLO-based vision models and IoT, provides real-time guidance, but is hindered by battery life constraints [18], [171], [199], [200], [201].Smartphone applications integrate GPS and AI-based scene recognition for affordability and accessibility but are less effective in indoor environments [158], [202], [203], [204]. Computer vision-based navigation aids text recognition and scene comprehension, but is affected by lighting variations and computational demands [159], [205], [206], [207], [208].

In wheelchair navigation, object tracking with YOLO, SORT, and RealSense cameras improves tracking accuracy, but depends on ideal lighting [177], [209], [210], [211], [212]. Semantic mapping, which combines deep learning with environmental contours, enhances path planning, but has high computational overhead [209], [213], [214], [215], [216]. Reinforcement-based navigation allows real-time adaptive decision-making, but struggles with motion smoothness in complex paths [179], [209], [217], [218], [219], [220]. IoT-enabled smart wheelchairs leverage sensors and cloud computing for adaptive control but face challenges related to network dependency, power consumption, and data security [221], [222], [223].

Across all applications, trends emphasize improving computational efficiency, robustness in varying environments, and multi-modal sensor integration. Although deep learning and reinforcement learning have revolutionized real-time decision-making, challenges related to computational demands and energy efficiency persist. IoT and cloud computing improve navigation intelligence but raise concerns regarding security, latency, and network reliability. Sensor fusion continues to advance in autonomous vehicles and

robotics, highlighting the need for optimized processing and reduced system complexity.

The growing emphasis on innovation is reflected in the research landscape, as evidenced by a bibliometric study conducted using Google Scholar. The search queries focused on publications from 2004 to 2024, covering four key domains of navigation applications: autonomous vehicles, robotics, BVI navigation, and wheelchair navigation. The number of publications in each category over the past two decades is shown in Figure 7. This bibliometric analysis revealed that autonomous vehicles have received the highest research attention (41%), reflecting the growing interest in deep learning and sensor fusion for autonomous navigation. In contrast, assistive technologies (e.g., navigation for the VI and wheelchair navigation) have received relatively less attention. This disparity highlights the need for further research in these areas, particularly in developing lightweight, context-aware models that can operate efficiently on low-power devices. Future studies should focus on bridging this gap to ensure that advancements in object detection benefit all users, including those with disabilities.

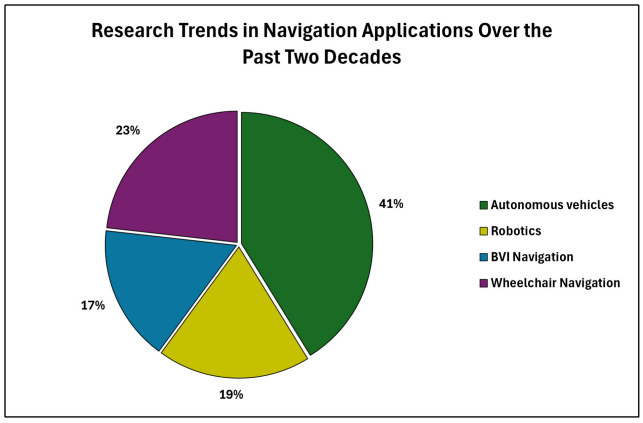


**FIGURE 7. Trends of research publications over the past two decades on Navigation applications(Data from Google scholar advanced search: allwords: "Autonomous vehicles", "Autonomous Robot Navigation", "Wheelchair Navigation", "Navigation/visually impaired").**

## V. CONCLUSION AND FUTURE TRENDS

This review underscores the importance of enhancing current automated object detection models, particularly by implementing contextual learning to provide a deeper understanding of the dynamic nature of objects during classification.

As key contributions, this study first offers an extensive literature survey on various spatial, temporal, and contextual feature types that are essential for understanding dynamic environments. We emphasize the role that these features play in improving decision making and adaptability in object detection systems. In addition, we provide a comprehensive review of current object detection algorithms and navigation systems that integrate these algorithms and explore ways in which they can be further optimized.

The literature reveals critical gaps in the integration of current computer vision-based algorithms for object detection and classification. Early image processing techniques, which relied on hand-crafted features, struggled with accuracy and efficiency in dynamic environments with challenges such as varying lighting conditions and occlusions, limiting their applicability in real-time scenarios like autonomous navigation. While modern machine learning and deep learning approaches have improved performance, they often require significant computational resources, making them less suitable for lower-end hardware. These models also rely on large, labeled datasets, which can be time-consuming to curate and less effective in dynamic or unpredictable settings. A common limitation of both traditional and modern approaches is the lack of contextual learning integration, which is crucial for understanding the dynamic nature of objects during classification. Incorporating semantic, spatial, and temporal context into these models could significantly improve navigation performance, leading to safer and more intelligent systems. Furthermore, areas such as lightweight object detectors, edge computing integration, and continual learning remain underexplored in this domain.

Emerging technologies, such as neuromorphic computing [224], [225], [226] and quantum machine learning [227], [228], [229], hold promise for further advancing object detection. Neuromorphic computing, inspired by the human brain, can enable more efficient and adaptive object detection systems, whereas quantum machine learning can revolutionize feature extraction and optimization. By exploring these technologies, researchers can unlock new possibilities for object detection, paving the way for smarter and more adaptable navigation systems. Overall, advancements in object detection continue to drive improvements in the safety and efficiency of autonomous systems and other applications. Continued innovation is expected to address existing limitations, further enhancing the real-world applicability of these models.

Furthermore, as object detection technologies continue to advance and integrate into real-world applications, ethical and privacy concerns have become increasingly critical. The deployment of AI-driven detection systems, particularly in navigation and surveillance, raises questions regarding data security, user consent, and potential biases in decision-making. Ensuring transparency in algorithmic processes, minimizing unintended biases, and safeguarding sensitive data is essential for responsible implementation. In addition, manufacturers and developers must consider regulatory compliance and ethical guidelines to foster public trust. By proactively addressing these concerns, object detection technologies can be deployed more equitably and securely, thereby enhancing their applicability in autonomous systems while maintaining ethical integrity.

In conclusion, while the field of automated object detection is progressing rapidly, challenges remain in developing solutions that are both accurate and applicable to real-world scenarios, particularly in terms of improving safety

and efficiency. Addressing these gaps will pave the way for smarter and more adaptable models that advance object detection technology.

## REFERENCES

[1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[2] M. V. Peelen, E. Berlot, and F. P. de Lange, "Predictive processing of scenes and objects," *Nature Rev. Psychol.*, vol. 3, no. 1, pp. 13–26, Nov. 2023.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[4] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Dec. 2015, pp. 91–99.

[6] 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.

[7] S. A. Papert, "The summer vision project," Massachusetts Inst. Technol. (MIT), MA, USA, Tech. Rep. AIM-100, 1966.

[8] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, "DeepDriving: Learning affordance for direct perception in autonomous driving," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2722–2730.

[9] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6526–6534.

[10] C. Che, H. Zheng, Z. Huang, W. Jiang, and B. Liu, "Intelligent robotic control system based on computer vision technology," 2024, *arXiv:2404.01116*.

[11] A. Khan, B. Rinner, and A. Cavallaro, "Cooperative robots to observe moving targets: Review," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 187–198, Jan. 2018.

[12] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[13] M. C. Fairhurst, *Computer Vision for Robotic Systems: An Introduction*. Upper Saddle River, NJ, USA: Prentice-Hall, 1988.

[14] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transp. Res., C, Emerg. Technol.*, vol. 6, no. 4, pp. 271–288, Aug. 1998.

[15] J. Xie, Y. Zheng, R. Du, W. Xiong, Y. Cao, Z. Ma, D. Cao, and J. Guo, "Deep learning-based computer vision for surveillance in ITS: Evaluation of state-of-the-art methods," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3027–3042, Apr. 2021.

[16] G. T. S. Ho, Y. P. Tsang, C. H. Wu, W. H. Wong, and K. L. Choy, "A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities," *Sensors*, vol. 19, no. 8, p. 1796, Apr. 2019.

[17] M. Helmy Abd Wahab, A. A. Talib, H. A. Kadir, A. Johari, A. Noraziah, R. M. Sidek, and A. A. Mutalib, "Smart cane: Assistive cane for visually-impaired people," 2011, *arXiv:1110.5156*.

[18] H.-C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré, and D. Rus, "Enabling independent navigation for visually impaired people through a wearable vision-based feedback system," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 6533–6540.

[19] Q. Chen, Z. Song, J. Dong, Z. Huang, Y. Hua, and S. Yan, "Contextualizing object detection and classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 13–27, Jan. 2015.

[20] F. Jurie and M. Dhome, "A simple and efficient template matching algorithm," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Jul. 2001, pp. 544–549.

[21] G. S. Cox, "Template matching and measures of match in image processing," Dept. Elect. Eng., Univ. Cape Town, Cape Town, South Africa, Tech. Rep. 109, 1995.

[22] N. Perveen, D. Kumar, and I. Bhardwaj, "An overview on template matching methodologies and its applications," *Int. J. Res. Comput. Commun. Technol.*, vol. 2, no. 10, pp. 988–995, 2013.

[23] T. Mahalakshmi, R. Muthaiah, and P. Swaminathan, "Review article: An overview of template matching technique in image processing," *Res. J. Appl. Sci., Eng. Technol.*, vol. 4, no. 24, pp. 5469–5473, Dec. 2012.

[24] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[25] G. N. Chaple, R. D. Daruwala, and M. S. Gofane, "Comparisions of Robert, Prewitt, Sobel operator based edge detection methods for real time uses on FPGA," in *Proc. Int. Conf. Technol. Sustain. Develop. (ICTSD)*, Feb. 2015, pp. 1–4.

[26] M. Kumar and R. Saxena, "Algorithm and technique on various edge detection: A survey," *Signal Image Process. Int. J.*, vol. 4, no. 3, pp. 65–75, Jun. 2013.

[27] R. Song, Z. Zhang, and H. Liu, "Edge connection based Canny edge detection algorithm," *Pattern Recognit. Image Anal.*, vol. 27, no. 4, pp. 740–747, Oct. 2017.

[28] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, Jan. 1981.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2001, pp. 511–518.

[30] C. P. Papageorgiou, "Object and pattern detection in video sequences," Ph.D. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 1997.

[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[32] W.-L. Zhao and C.-W. Ngo, "Flip-invariant SIFT for copy and object detection," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 980–991, Mar. 2013.

[33] 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.

[34] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[35] J. V. D. Weijer and C. Schmid, "Coloring local feature extraction," in *Proc. 9th Eur. Conf. Comput. Vis.-ECCV*, Graz, Austria. Cham, Switzerland: Springer, Jan. 2006, pp. 334–348.

[36] M. J. Swain and D. H. Ballard, "Indexing via color histograms," in *Active Perception and Robot Vision*, A. K. Sood and H. Wechsler, Eds., Berlin, Germany: Springer, 1992, pp. 261–273.

[37] J.-Y. Zhu, J. Wu, Y. Xu, E. Chang, and Z. Tu, "Unsupervised object class discovery via saliency-guided multiple class learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 862–875, Apr. 2015.

[38] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*.

[39] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 169–175.

[40] B. Koonce and B. Koonce, "Resnet 50," in *Convolutional Neural Networks With Swift for Tensorflow: Image Recognition and Dataset Categorization*. Berkeley, CA, USA: Apress, 2021, pp. 63–72.

[41] M. M. Trivedi, C. Chen, and D. H. Cress, "Object detection by step-wise analysis of spectral, spatial, and topographic features," *Comput. Vis., Graph., Image Process.*, vol. 51, no. 3, pp. 235–255, Sep. 1990.

[42] D. H. Ballard and C. M. Brown, *Computer Vision*. Upper Saddle River, NJ, USA: Prentice-Hall, 1982.

[43] S. Gil, R. Milanese, and T. Pun, "Feature selection for object tracking in traffic scenes," *Proc. SPIE*, vol. 2344, pp. 253–266, Jan. 1995.

[44] B. Günsel, A. M. Ferman, and A. M. Tekalp, "Temporal video segmentation using unsupervised clustering and semantic object tracking," *J. Electron. Imag.*, vol. 7, no. 3, pp. 592–604, Jul. 1998.

[45] A. Zadaianchuk, M. Seitzer, and G. Martius, "Object-centric learning for real-world videos by predicting temporal feature similarities," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, Jan. 2023, pp. 1–7.

[46] K. Greff, R. L. Kaufman, R. Kabra, N. Watters, C. Burgess, D. Zoran, L. Matthey, M. Botvinick, and A. Lerchner, "Multi-object representation learning with iterative variational inference," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2019, pp. 2424–2433.

[47] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.

[48] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 712–722, Jun. 2010.

[49] A. Farhadi, M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth, "Every picture tells a story: Generating sentences from images," in *Proc. 11th Eur. Conf. Comput. Vis.-ECCV*, Crete, Greece. Cham, Switzerland: Springer, Jan. 2010, pp. 15–29.
[50] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[51] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Progr. Brain Res.*, vol. 155, pp. 23–36, Oct. 2006.
[52] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.
[53] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2015.
[54] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-UCSD birds-200-2011 dataset," California Inst. Technol., CA, USA, Tech. Rep. CNS-TR-2010-001, 2011.
[55] W. Liu, S. Liao, W. Ren, W. Hu, and Y. Yu, "High-level semantic feature detection: A new perspective for pedestrian detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5182–5191.
[56] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf. Comput. Vis.-ECCV*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Jan. 2016, pp. 21–37.
[57] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009.
[58] M. J. Choi, J. J. Lim, A. Torralba, and A. S. Willsky, "Exploiting hierarchical context on a large database of object categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 129–136.
[59] I. R. Olson and M. M. Chun, "Perceptual constraints on implicit learning of spatial context," *Vis. Cognition*, vol. 9, no. 3, pp. 273–302, Apr. 2002.
[60] D. Hoiem, A. A. Efros, and M. Hebert, "Geometric context from a single image," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 654–661.
[61] D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," *Int. J. Comput. Vis.*, vol. 80, no. 1, pp. 3–15, Oct. 2008.
[62] S. Y. Bao, M. Sun, and S. Savarese, "Toward coherent object detection and scene layout understanding," *Image Vis. Comput.*, vol. 29, no. 9, pp. 569–579, Aug. 2011.
[63] C. Galleguillos, B. McFee, S. Belongie, and G. Lanckriet, "Multi-class object localization by combining local contextual interactions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 113–120.
[64] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman, "Multiple kernels for object detection," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 606–613.
[65] W.-S. Zheng, S. Gong, and T. Xiang, "Quantifying and transferring contextual information in object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 762–777, Apr. 2012.
[66] M. Everingham et al., "The 2005 Pascal visual object classes challenge," in *Proc. Mach. Learn. Challenges. Evaluating Predictive Uncertainty, Vis. Object Classification, Recognising Tectual Entailment*, Southampton, U.K. Cham, Switzerland: Springer, Jan. 2006, pp. 117–176.
[67] Y. Liu, J. Han, Q. Zhang, and C. Shan, "Deep salient object detection with contextual information guidance," *IEEE Trans. Image Process.*, vol. 29, pp. 360–374, 2020.
[68] J. Xu, W. Wang, H. Wang, and J. Guo, "Multi-model ensemble with rich spatial information for object detection," *Pattern Recognit.*, vol. 99, Mar. 2020, Art. no. 107098.
[69] 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.
[70] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[71] M. Ju, H. Luo, Z. Wang, B. Hui, and Z. Chang, "The application of improved YOLO V3 in multi-scale target detection," *Appl. Sci.*, vol. 9, no. 18, p. 3775, Sep. 2019.
[72] S. K. Divvala, D. Hoiem, J. H. Hays, A. A. Efros, and M. Hebert, "An empirical study of context in object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1271–1278.
[73] X. Wang, H. Ma, X. Chen, and S. You, "Edge preserving and multi-scale contextual neural network for salient object detection," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 121–134, Jan. 2018.
[74] Y. Kim, B.-N. Kang, and D. Kim, "SAN: Learning relationship between convolutional features for multi-scale object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 316–331.
[75] H. Qiu, H. Li, Q. Wu, F. Meng, L. Xu, K. N. Ngan, and H. Shi, "Hierarchical context features embedding for object detection," *IEEE Trans. Multimedia*, vol. 22, no. 12, pp. 3039–3050, Dec. 2020.
[76] R. Azad, A. Kazerouni, M. Heidari, E. K. Aghdam, A. Molaei, Y. Jia, A. Jose, R. Roy, and D. Merhof, "Advances in medical image analysis with vision transformers: A comprehensive review," *Med. Image Anal.*, vol. 91, Jan. 2024, Art. no. 103000.
[77] A. Hatamizadeh, H. Yin, J. Kautz, and P. Molchanov, "Global context vision transformers," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2022, pp. 12633–12646.
[78] A. Hatamizadeh, J. Song, G. Liu, J. Kautz, and A. Vahdat, "DiffiT: Diffusion vision transformers for image generation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Jan. 2023, pp. 37–55.
[79] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[80] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002200009791504X
[81] Y.-Q. Wang, "An analysis of the viola-jones face detection algorithm," *Image Process. Line*, vol. 4, pp. 128–148, Jun. 2014, doi: 10.5201/ipol.2014.104.
[82] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient HOG human detection," *Signal Process.*, vol. 91, no. 4, pp. 773–781, Apr. 2011.
[83] E. Fix, *Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties*, vol. 1. Wright-Patterson AFB, OH, USA: USAF school of Aviation Medicine, 1985.
[84] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 4, pp. 580–585, Jul. 1985.
[85] M. A. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Schölkopf, "Support vector machines," *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.
[86] A. C. Lorena, L. F. O. Jacintho, M. F. Siqueira, R. D. Giovanni, L. G. Lohmann, A. C. P. L. F. de Carvalho, and M. Yamamoto, "Comparing machine learning classifiers in potential distribution modelling," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5268–5275, May 2011.
[87] F. Pernkopf, "Bayesian network classifiers versus selective K-NN classifier," *Pattern Recognit.*, vol. 38, no. 1, pp. 1–10, Jan. 2005.
[88] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed, "Investigating the performance of Naive- Bayes classifiers and K- nearest neighbor classifiers," in *Proc. Int. Conf. Converg. Inf. Technol. (ICCIT)*, Nov. 2007, pp. 1541–1546.
[89] X. Daniela, C. Hinde, and R. Stone, "Naive Bayes vs. decision trees vs. neural networks in the classification of training Web pages," *Int. J. Comput. Sci. Issues*, vol. 4, no. 1, pp. 16–23, 2009.
[90] N. B. Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs decision trees in intrusion detection systems," in *Proc. ACM Symp. Appl. Comput.*, Mar. 2004, pp. 420–424.
[91] L. I. Kuncheva, "On the optimality of naíve Bayes with dependent binary features," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 830–837, 2006.
[92] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proc. 23rd Int. Conf. Mach. Learn.-ICML*, 2006, pp. 161–168.
[93] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers—A survey," *IEEE Trans. Syst., Man Cybern., C, Appl. Rev.*, vol. 35, no. 4, pp. 476–487, Nov. 2005.
[94] M. Robnik-Šikonja, "Improving random forests," in *Proc. Eur. Conf. Mach. Learn.* Cham, Switzerland: Springer, Jan. 2004, pp. 359–370.
[95] M. Aly, "Survey on multiclass classification methods," *Neural Netw.*, vol. 19, nos. 1–9, p. 2, 2005.

[96] P. Liu, Q. Wang, H. Zhang, J. Mi, and Y. Liu, "A lightweight object detection algorithm for remote sensing images based on attention mechanism and YOLOv5s," *Remote Sens.*, vol. 15, no. 9, p. 2429, May 2023.
[97] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[98] L. Du, R. Zhang, and X. Wang, "Overview of two-stage object detection algorithms," *J. Phys., Conf. Ser.*, vol. 1544, no. 1, May 2020, Art. no. 012033.
[99] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.
[100] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
[101] S. Vicente, J. Carreira, L. Agapito, and J. Batista, "Reconstructing Pascal VOC," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 41–48.
[102] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[103] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[104] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Eur. Conf. Comput. Vis.-ECCV*, Zurich, Switzerland. Cham, Switzerland: Springer, Jan. 2014, pp. 818–833.
[105] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3150–3158.
[106] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4438–4446.
[107] P. O. Pinheiro, R. Collobert, and P. Dollár, "Learning to segment object candidates," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Jan. 2015, pp. 1–17.
[108] S. Kim, H.-K. Kook, J.-Y. Sun, M.-C. Kang, and S.-J. Ko, "Parallel feature pyramid network for object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 239–256.
[109] G. Zhao, W. Ge, and Y. Yu, "GraphFPN: Graph feature pyramid network for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2743–2752.
[110] L. Zhu, F. Lee, J. Cai, H. Yu, and Q. Chen, "An improved feature pyramid network for object detection," *Neurocomputing*, vol. 483, pp. 127–139, Apr. 2022.
[111] Q. Ma, S. Jin, G. Bian, and Y. Cui, "Multi-scale marine object detection in side-scan sonar images based on BES-YOLO," *Sensors*, vol. 24, no. 14, p. 4428, Jul. 2024.
[112] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.
[113] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[114] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[115] Z. Li, L. Zhao, X. Han, M. Pan, and F.-J. Hwang, "Lightweight ship detection methods based on YOLOv3 and DenseNet," *Math. Problems Eng.*, vol. 2020, pp. 1–10, Sep. 2020.
[116] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, "Statistical analysis of design aspects of various YOLO-based deep learning models for object detection," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, p. 126, Aug. 2023.
[117] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of YOLO algorithm developments," *Proc. Comput. Sci.*, vol. 199, pp. 1066–1073, Aug. 2022.
[118] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
[119] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023.
[120] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[121] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 765–781.
[122] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.
[123] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, Oct. 2006, pp. 850–855.
[124] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Jan. 2020, pp. 213–229.
[125] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 16965–16974.
[126] F. Li, H. Zhang, S. Liu, J. Guo, L. M. Ni, and L. Zhang, "DN-DETR: Accelerate DETR training by introducing query DeNoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13609–13617.
[127] Q. Chen, X. Chen, J. Wang, S. Zhang, K. Yao, H. Feng, J. Han, E. Ding, G. Zeng, and J. Wang, "Group DETR: Fast DETR training with group-wise one-to-many assignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 6610–6619.
[128] Y. Pu, W. Liang, Y. Hao, Y. Yuan, Y. Yang, C. Zhang, H. Hu, and G. Huang, "Rank-DETR for high quality object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, Jan. 2023, pp. 1–8.
[129] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1571–1580.
[130] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2019, pp. 6105–6114.
[131] A. Haidar, S. Tomov, J. Dongarra, and N. J. Higham, "Harnessing GPU tensor cores for fast FP16 arithmetic to speed up mixed-precision iterative refinement solvers," in *Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal.*, Nov. 2018, pp. 603–613.
[132] S. Badal, S. Ravela, B. Draper, and A. Hanson, "A practical obstacle detection and avoidance system," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Aug. 1994, pp. 97–104.
[133] L.-C. Fu and C.-Y. Liu, "Computer vision based object detection and recognition for vehicle driving," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, vol. 3, Nov. 2001, pp. 2634–2641.
[134] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada, "An open approach to autonomous vehicles," *IEEE Micro*, vol. 35, no. 6, pp. 60–68, Nov. 2015.
[135] A. Ess, K. Schindler, B. Leibe, and L. Van Gool, "Object detection and tracking for autonomous navigation in dynamic environments," *Int. J. Robot. Res.*, vol. 29, no. 14, pp. 1707–1725, Dec. 2010.
[136] P. Wang, X. Wang, Y. Liu, and J. Song, "Research on road object detection model based on YOLOv4 of autonomous vehicle," *IEEE Access*, vol. 12, pp. 8198–8206, 2024.
[137] W. Zhou, Y. Lv, J. Lei, and L. Yu, "Embedded control gate fusion and attention residual learning for RGB–thermal urban scene parsing," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 4794–4803, May 2023.
[138] X. Jin, Y. Xie, X.-S. Wei, B.-R. Zhao, Z.-M. Chen, and X. Tan, "Delving deep into spatial pooling for squeeze-and-excitation networks," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108159.
[139] L. Tang, Y. Wang, and L.-P. Chau, "Weakly-supervised part-attention and mentored networks for vehicle re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8887–8898, Dec. 2022.
[140] X. Wang and J. Song, "ICIoU: Improved loss based on complete intersection over union for bounding box regression," *IEEE Access*, vol. 9, pp. 105686–105695, 2021.
[141] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "KAIST multi-spectral day/night data set for autonomous and assisted driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 934–948, Mar. 2018.
[142] H. Wang, C. Liu, Y. Cai, L. Chen, and Y. Li, "YOLOv8-QSD: An improved small object detection algorithm for autonomous vehicles based on YOLOv8," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–16, 2024.
[143] J. Ni, L. Wu, X. Fan, and S. X. Yang, "Bioinspired intelligent algorithm and its applications for mobile robot control: A survey," *Comput. Intell. Neurosci.*, vol. 2016, no. 1, pp. 1–16, 2016.

[144] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, Jan. 1975.
[145] L. Ren, W. Wang, and Z. Du, "A new fuzzy intelligent obstacle avoidance control strategy for wheeled mobile robot," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, Aug. 2012, pp. 1732–1737.
[146] Q.-Y. Bao, S.-M. Li, W.-Y. Shang, and M.-J. An, "A fuzzy behavior-based architecture for mobile robot navigation in unknown environments," in *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, vol. 2, Nov. 2009, pp. 257–261.
[147] M. Iwanowski, A. Haidukievich, M. Leszczyński, and B. Wnorowski, "Fuzzy approach to object-detection-based image retrieval," in *Proc. Int. Conf. Comput. Vis. Graph. (ICCVG)*, in Lecture Notes in Networks and Systems, vol. 598, L. J. Chmielewski and A. Orlowski, Eds., Cham, Switzerland: Springer, Jan. 2023, pp. 121–135, doi: 10.1007/978-3-031-22025-8_9.
[148] M. A. S. Silva, G. P. Dimuro, E. N. Borges, G. Lucca, and C. Marco-Detchart, "Analysis of fuzzy techniques in edge detection," in *Proc. Int. Conf. Enterprise Inf. Syst. (ICEIS)*, in Lecture Notes in Bus. Information Processing, vol. 518, J. Filipe, M. Smialek, A. Brodsky, and S. Hammoudi, Eds., Cham, Switzerland: Springer, Jan. 2024, pp. 161–180, doi: 10.1007/978-3-031-64748-2_8.
[149] A. Pandey, "Mobile robot navigation and obstacle avoidance techniques: A review," *Int. Robot. Autom. J.*, vol. 2, no. 3, p. 22, May 2017.
[150] A. Hernández, C. Gómez, J. Crespo, and R. Barber, "Object detection applied to indoor environments for mobile robot navigation," *Sensors*, vol. 16, no. 8, p. 1180, Jul. 2016.
[151] A. Coates and A. Y. Ng, "Multi-camera object detection for robotics," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 412–419.
[152] L. Chen, L. Tong, F. Zhou, Z. Jiang, Z. Li, J. Lv, J. Dong, and H. Zhou, "A benchmark dataset for both underwater image enhancement and underwater object detection," 2020, *arXiv:2006.15789*.
[153] L. Jiang, Y. Wang, Q. Jia, S. Xu, Y. Liu, X. Fan, H. Li, R. Liu, X. Xue, and R. Wang, "Underwater species detection using channel sharpening attention," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 4259–4267.
[154] C. Liu, Z. Wang, S. Wang, T. Tang, Y. Tao, C. Yang, H. Li, X. Liu, and X. Fan, "A new dataset, Poisson GAN and AquaNet for underwater object grabbing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2831–2844, May 2022.
[155] C. Liu, H. Li, S. Wang, M. Zhu, D. Wang, X. Fan, and Z. Wang, "A dataset and benchmark of underwater object detection for robot picking," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2021, pp. 1–6.
[156] S. Xu, M. Zhang, W. Song, H. Mei, Q. He, and A. Liotta, "A systematic review and analysis of deep learning-based underwater object detection," *Neurocomputing*, vol. 527, pp. 204–232, Mar. 2023.
[157] R. Enrico, P. Ricioppo, M. Mancini, and S. Primatesta, "Computer vision-based autonomous navigation for UAV vineyard row following," in *Proc. IEEE Int. Workshop Metrology Agricult. Forestry (MetroAgriFor)*, Dec. 2024.
[158] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Smartphone navigation support for blind and visually impaired people—A comprehensive analysis of potentials and opportunities," in *Proc. Universal Access Hum.-Comput. Interaction. Appl. Pract.*, M. Antona and C. Stephanidis, Eds., Cham, Switzerland: Springer, Jan. 2020, pp. 568–583.
[159] D. Plikynas, A. Žvironas, M. Gudauskis, A. Budrionis, P. Daniušis, and I. Sliesoraityte, "Research advances of indoor navigation for blind people: A brief review of technological instrumentation," *IEEE Instrum. Meas. Mag.*, vol. 23, no. 4, pp. 22–32, Jun. 2020.
[160] W. Jeamwatthanachai, M. Wald, and G. Wills, "Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings," *Brit. J. Vis. Impairment*, vol. 37, no. 2, pp. 140–153, May 2019.
[161] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 258–266, Aug. 2017.
[162] M. Gebresselassie and T. W. Sanchez, "'Smart' tools for socially sustainable transport: A review of mobility apps," *Urban Sci.*, vol. 2, no. 2, p. 45, May 2018.
[163] C. Granquist, S. Y. Sun, S. R. Montezuma, T. M. Tran, R. Gage, and G. E. Legge, "Evaluation and comparison of artificial intelligence vision aids: Orcam MyEye 1 and seeing AI," *J. Vis. Impairment Blindness*, vol. 115, no. 4, pp. 277–285, Jul. 2021.
[164] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa, "NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment," in *Proc. 19th Int. ACM SIGACCESS Conf. Comput. Accessibility*, Oct. 2017, pp. 270–279.
[165] V. Nair, G. Olmschenk, W. H. Seiple, and Z. Zhu, "ASSIST: Evaluating the usability and performance of an indoor navigation assistant for blind and visually impaired people," *Assistive Technol.*, vol. 34, no. 3, pp. 289–299, May 2022.
[166] Q. Chen, M. Khan, C. Tsangouri, C. Yang, B. Li, J. Xiao, and Z. Zhu, "CCNY smart cane," in *Proc. IEEE 7th Annu. Int. Conf. CYBER Technol. Autom., Control, Intell. Syst. (CYBER)*, Jul. 2017, pp. 1246–1251.
[167] H. Taheri and Z. C. Xia, "SLAM; definition and evolution," *Eng. Appl. Artif. Intell.*, vol. 97, Jan. 2021, Art. no. 104032.
[168] Q.-H. Nguyen, H. Vu, T.-H. Tran, D. V. Hamme, P. Veelaert, W. Philips, and Q.-H. Nguyen, "A visual SLAM system on mobile robot supporting localization services to visually impaired people," in *Proc. Comput. Vis.-ECCV Workshops*, Zurich, Switzerland. Cham, Switzerland: Springer, Jan. 2015, pp. 716–729.
[169] M. Hosny, R. Alsarrani, and A. Benabid, "Indoor wheelchair navigation for the visually impaired," in *Proc. Int. Conf. HCI Int.-Posters' Extended Abstr.*, Los Angeles, CA, USA. Cham, Switzerland: Springer, Jan. 2015, pp. 411–417.
[170] S. Subbiah, S. Ramya, G. P. Krishna, and S. Nayagam, "Smart cane for visually impaired based on IoT," in *Proc. 3rd Int. Conf. Comput. Commun. Technol. (ICCCT)*, Feb. 2019, pp. 50–53.
[171] M. Messaoudi, B.-A. Menelas, and H. Mcheick, "Autonomous smart white cane navigation system for indoor usage," *Technologies*, vol. 8, no. 3, p. 37, Jun. 2020.
[172] C. Liao, P. Choe, T. Wu, Y. Tong, C. Dai, and Y. Liu, "RFID-based road guiding cane system for the visually impaired," in *Proc. 5th Int. Conf. Cross-Cultural Design Methods, Pract., Case Stud.*, Las Vegas, NV, USA. Cham, Switzerland: Springer, Jan. 2013, pp. 86–93.
[173] A. Ben Atitallah, Y. Said, M. A. Ben Atitallah, M. Albekairi, K. Kaaniche, and S. Boubaker, "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation," *Ain Shams Eng. J.*, vol. 15, no. 2, Feb. 2024, Art. no. 102387.
[174] A. Argyros, P. Georgiadis, P. Trahanias, and D. Tsakiris, "Semi-autonomous navigation of a robotic wheelchair," *J. Intell. Robotic Syst.*, vol. 34, pp. 315–329, Aug. 2002.
[175] E. Baklouti, N. B. Amor, and M. Jallouli, "Autonomous wheelchair navigation with real time obstacle detection using 3D sensor," *Automatika*, vol. 57, no. 3, pp. 761–773, Jan. 2016.
[176] E. Kim, "Wheelchair navigation system for disabled and elderly people," *Sensors*, vol. 16, no. 11, p. 1806, Oct. 2016.
[177] L. Lecrosnier, R. Khemmar, N. Ragot, B. Decoux, R. Rossi, N. Kefi, and J.-Y. Ertaud, "Deep learning-based object detection, localisation and tracking for smart wheelchair healthcare mobility," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 91, Dec. 2020.
[178] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3464–3468.
[179] J. Liang, Z. Wang, Y. Cao, J. Chiun, M. Zhang, and G. A. Sartoretti, "Context-aware deep reinforcement learning for autonomous robotic navigation in unknown area," in *Proc. Conf. Robot Learn.*, 2023, pp. 1425–1436.
[180] L. Beale, K. Field, D. Briggs, P. Picton, and H. Matthews, "Mapping for wheelchair users: Route navigation in urban spaces," *Cartographic J.*, vol. 43, no. 1, pp. 68–81, Mar. 2006.
[181] P. Kasemsuppakorn, H. A. Karimi, D. Ding, and M. A. Ojeda, "Understanding route choices for wheelchair navigation," *Disab. Rehabil., Assistive Technol.*, vol. 10, no. 3, pp. 198–210, May 2015.
[182] V. Gallo, I. Shallari, M. Carratù, V. Laino, and C. Liguori, "Design and characterization of a powered wheelchair autonomous guidance system," *Sensors*, vol. 24, no. 5, p. 1581, Feb. 2024.
[183] W. Chuah, R. Tennakoon, R. Hoseinnezhad, D. Suter, and A. Bab-Hadiashar, "Semantic guided long range stereo depth estimation for safer autonomous vehicle applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18916–18926, Oct. 2022.

[184] Y. Bai, B. Zhang, N. Xu, J. Zhou, J. Shi, and Z. Diao, "Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107584.
[185] U. Ulusoy, O. Eren, and A. Demirhan, "Development of an obstacle avoiding autonomous vehicle by using stereo depth estimation and artificial intelligence based semantic segmentation," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 106808.
[186] H. Guan, B. Wang, J. Gong, and H. Chen, "Coordinated motion planning for heterogeneous autonomous vehicles based on driving behavior primitives," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11934–11949, Nov. 2023.
[187] 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.
[188] J. Jiao, R. Geng, Y. Li, R. Xin, B. Yang, J. Wu, L. Wang, M. Liu, R. Fan, and D. Kanoulas, "Real-time metric-semantic mapping for autonomous navigation in outdoor environments," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 5729–5740, 2025.
[189] J. Wilson, Y. Fu, J. Friesen, P. Ewen, A. Capodieci, P. Jayakumar, K. Barton, and M. Ghaffari, "ConvBKI: Real-time probabilistic semantic mapping network with quantifiable uncertainty," *IEEE Trans. Robot.*, vol. 40, pp. 4648–4667, 2024.
[190] S. Zhu, R. Qin, G. Wang, J. Liu, and H. Wang, "SemGauss-SLAM: Dense semantic Gaussian splatting SLAM," 2024, *arXiv:2403.07494*.
[191] B. Şenbaşlar and G. S. Sukhatme, "DREAM: Decentralized real-time asynchronous probabilistic trajectory planning for collision-free multirobot navigation in cluttered environments," *IEEE Trans. Robot.*, vol. 41, pp. 573–592, 2025.
[192] Y. A. Singgalen, "Sentiment classification of robot hotel content using NBC and SVM algorithm," *J. Comput. Syst. Informat.*, vol. 5, no. 2, pp. 442–453, Feb. 2024.
[193] A. Alotaibi, H. Alatawi, A. Binnouh, L. Duwayriat, T. Alhmiedat, and O. M. Alia, "Deep learning-based vision systems for robot semantic navigation: An experimental study," *Technologies*, vol. 12, no. 9, p. 157, Sep. 2024.
[194] M. Y. Arafat, M. M. Alam, and S. Moh, "Vision-based navigation techniques for unmanned aerial vehicles: Review and challenges," *Drones*, vol. 7, no. 2, p. 89, Jan. 2023.
[195] S. Levine and D. Shah, "Learning robotic navigation from experience: Principles, methods and recent results," *Phil. Trans. Roy., B, Biol. Sci.*, vol. 378, no. 1869, Jan. 2023, Art. no. 20210447.
[196] C. Mai, D. Xie, L. Zeng, Z. Li, Z. Li, Z. Qiao, Y. Qu, G. Liu, and L. Li, "Laser sensing and vision sensing smart blind cane: A review," *Sensors*, vol. 23, no. 2, p. 869, Jan. 2023.
[197] C.-E. Panazan and E.-H. Dulf, "Intelligent cane for assisting the visually impaired," *Technologies*, vol. 12, no. 6, p. 75, May 2024.
[198] A. Mocanu, V. Sita, C. Avram, and A. Aştilean, "Enhanced cane for blind people mobility assistance," in *Proc. IEEE Int. Conf. Autom., Quality Test., Robot. (AQTR)*, May 2024, pp. 1–6.
[199] P. Xu, G. A. Kennedy, F.-Y. Zhao, W.-J. Zhang, and R. Van Schyndel, "Wearable obstacle avoidance electronic travel aids for blind and visually impaired individuals: A systematic review," *IEEE Access*, vol. 11, pp. 66587–66613, 2023.
[200] P. Xu, A. Song, and K. Wang, "Intelligent head-mounted obstacle avoidance wearable for the blind and visually impaired," *Sensors*, vol. 23, no. 23, p. 9598, Dec. 2023.
[201] J. Yuvanesh, S. Sherine, and I. Kala, "Implantable and wearable devices for IoT applications—A prototype of integrated multi-feature smart shoes and glass for the safe navigation of blind people," in *Proc. Big Data Anal. Intell. IoT Cyber-Phys. Syst.* Cham, Switzerland: Springer, Nov. 2023, pp. 151–165.
[202] B. Chaudary, S. Pohjolainen, S. Aziz, L. Arhippainen, and P. Pulli, "Teleguidance-based remote navigation assistance for visually impaired and blind people—Usability and user experience," *Virtual Reality*, vol. 27, no. 1, pp. 141–158, Mar. 2023.
[203] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "DeepNAVI: A deep learning based smartphone navigation assistant for people with visual impairments," *Expert Syst. Appl.*, vol. 212, Feb. 2023, Art. no. 118720.
[204] B. Pydala, T. P. Kumar, and K. K. Baseer, "Smart_Eye: A navigation and obstacle detection for visually impaired people through smart app," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 2, pp. 992–1011, Jun. 2023.
[205] M. M. Valipoor and A. de Antonio, "Recent trends in computer vision-driven scene understanding for VI/blind users: A systematic mapping," *Universal Access Inf. Soc.*, vol. 22, no. 3, pp. 983–1005, Aug. 2023.
[206] G. Jain, Y. Teng, D. H. Cho, Y. Xing, M. Aziz, and B. A. Smith, "'I want to figure things out': Supporting exploration in navigation for people with visual impairments," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. 1, pp. 1–28, Apr. 2023.
[207] Y. Said, M. Atri, M. A. Albahar, A. Ben Atitallah, and Y. A. Alsariera, "Obstacle detection system for navigation assistance of visually impaired people based on deep learning techniques," *Sensors*, vol. 23, no. 11, p. 5262, Jun. 2023.
[208] I. Patel, M. Kulkarni, and N. Mehendale, "Review of sensor-driven assistive device technologies for enhancing navigation for the visually impaired," *Multimedia Tools Appl.*, vol. 83, no. 17, pp. 52171–52195, Nov. 2023.
[209] S. K. Sahoo and B. B. Choudhury, "Autonomous navigation and obstacle avoidance in smart robotic wheelchairs," *J. Decis. Anal. Intell. Comput.*, vol. 4, no. 1, pp. 47–66, Feb. 2024.
[210] M. Kutbi, H. Li, Y. Chang, B. Sun, X. Li, C. Cai, N. Agadakos, G. Hua, and P. Mordohai, "Egocentric computer vision for hands-free robotic wheelchair navigation," *J. Intell. Robotic Syst.*, vol. 107, no. 1, p. 10, Jan. 2023.
[211] C. V. Giménez, S. Krug, F. Z. Qureshi, and M. O'Nils, "Evaluation of 2D-/3D-feet-detection methods for semi-autonomous powered wheelchair navigation," *J. Imag.*, vol. 7, no. 12, p. 255, Nov. 2021.
[212] V. Ramaraj, A. Paralikar, E. J. Lee, S. M. Anwar, and R. Monfaredi, "Development of a modular real-time shared-control system for a smart wheelchair," *J. Signal Process. Syst.*, vol. 96, no. 3, pp. 203–214, Mar. 2024.
[213] C. Prandi, B. R. Barricelli, S. Mirri, and D. Fogli, "Accessible wayfinding and navigation: A systematic mapping study," *Universal Access Inf. Soc.*, vol. 22, no. 1, pp. 185–212, Mar. 2023.
[214] D. Correia, A. Pereira, and E. Pedrosa, "Semantic navigation applied to narrow passages for intelligent wheelchairs," in *Proc. IEEE Int. Conf. Auto. Robot Syst. Competitions (ICARSC)*, Apr. 2023, pp. 151–156.
[215] E. Mohamed, K. Sirlantzis, and G. Howells, "Indoor/outdoor semantic segmentation using deep learning for visually impaired wheelchair users," *IEEE Access*, vol. 9, pp. 147914–147932, 2021.
[216] C. Messiou, D. Fusaro, G. Beraldo, and L. Tonin, "Real-time free space semantic segmentation for detection of traversable space for an intelligent wheelchair," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2022, pp. 1–6.
[217] C.-L. Lu, Z.-Y. Liu, J.-T. Huang, C.-I. Huang, B.-H. Wang, Y. Chen, N.-H. Wu, H.-C. Wang, L. Giarré, and P.-Y. Kuo, "Assistive navigation using deep reinforcement learning guiding robot with UWB/voice beacons and semantic feedbacks for blind and visually impaired people," *Frontiers Robot. AI*, vol. 8, Jun. 2021, Art. no. 654132.
[218] N. Rodrigues, A. Sousa, L. P. Reis, and A. Coelho, "Intelligent wheelchairs rolling in pairs using reinforcement learning," in *Proc. Iberian Robot. Conf.* Cham, Switzerland: Springer, Nov. 2022, pp. 274–285.
[219] F. Pacini, P. Dini, and L. Fanucci, "Design of an assisted driving system for obstacle avoidance based on reinforcement learning applied to electrified wheelchairs," *Electronics*, vol. 13, no. 8, p. 1507, Apr. 2024.
[220] P. De Almeida Afonso and P. R. Ferreira, "Autonomous navigation of wheelchairs in indoor environments using deep reinforcement learning and computer vision," in *Proc. Latin Amer. Robot. Symp. (LARS), Brazilian Symp. Robot. (SBR), Workshop Robot. Educ. (WRE)*, Oct. 2023, pp. 260–265.
[221] M. A. K. Al Shabibi and S. M. Kesavan, "IoT based smart wheelchair for disabled people," in *Proc. Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN)*, Jul. 2021, pp. 1–6.
[222] J. Cui, L. Cui, Z. Huang, X. Li, and F. Han, "IoT wheelchair control system based on multi-mode sensing and human-machine interaction," *Micromachines*, vol. 13, no. 7, p. 1108, Jul. 2022.
[223] L. Hou, J. Latif, P. Mehryar, S. Withers, A. Plastropoulos, L. Shen, and Z. Ali, "An autonomous wheelchair with health monitoring system based on Internet of thing," *Sci. Rep.*, vol. 14, no. 1, p. 5878, Mar. 2024.
[224] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photon.*, vol. 15, no. 2, pp. 102–114, Feb. 2021.
[225] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, "Opportunities for neuromorphic computing algorithms and applications," *Nature Comput. Sci.*, vol. 2, no. 1, pp. 10–19, Jan. 2022.

[226] J. Yuan, C. Wu, S. Wang, F. Wu, C. K. Tan, and D. Guo, "Enhancing plasticity in optoelectronic artificial synapses: A pathway to efficient neuromorphic computing," *Appl. Phys. Lett.*, vol. 124, no. 2, Jan. 2024, Art. no. 021101.
[227] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, "Challenges and opportunities in quantum machine learning," *Nature Comput. Sci.*, vol. 2, no. 9, pp. 567–576, Sep. 2022.
[228] D. Peral-García, J. Cruz-Benito, and F. J. García-Peñalvo, "Systematic literature review: Quantum machine learning and its applications," *Comput. Sci. Rev.*, vol. 51, Feb. 2024, Art. no. 100619.
[229] S. Jerbi, L. J. Fiderer, H. P. Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, "Quantum machine learning beyond kernel methods," *Nature Commun.*, vol. 14, no. 1, pp. 1–8, Jan. 2023.

**SHANELLE TENNEKOON** (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from Sri Lanka Institute of Information Technology (SLIIT Uni), Sri Lanka, in 2023. She is currently pursuing the Ph.D. degree with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Australia.

In 2023, she was an Assistant Lecturer with the Electrical and Electronic Engineering Department, SLIIT Uni. Her research interests include assistive technology, machine learning, artificial intelligence, and neural networks. She is a Professional Graduate Member of Engineers Australia. She has authored a paper titled "Object Recognition and Assistance System for Visually Impaired People," which won the Best Paper Award at the Second SLIIT International Conference on Engineering and Technology, Sri Lanka, in 2023. She rceived the Best Performance Award for her Bachelor of Science in Engineering Honours Degree in Electrical and Electronic Engineering.

**NUSHARA WEDASINGHA** received the B.Sc. degree (Hons.) in electrical and electronics engineering, in 2020, and the Ph.D. degree in software engineering, in 2024.

He is currently a Lecturer and a Researcher with the Department of Electrical and Electronics Engineering, Faculty of Engineering, SLIIT, Sri Lanka. He is also collaborating with the University of Oxford and the University of Applied Sciences and Arts of Southern Switzerland to develop a tool for quantifying the severity of Parkinson's disease. Previously, he was part of the Culturally Sensitive Autism Assessment Tool (CSAAT) Group, where he developed AI models to identify anomalous repetitive movements in children. These models have been instrumental in helping medical professionals monitor and validate the progress of their intervention. His research focuses on developing AI-based tools to identify and analyze behavioral disorders and diseases, which can also be used as supportive tools for early patient screening and intervention validation.

**ANURADHI WELHENGE** (Member, IEEE) received the B.Sc. degree in electronics engineering from Asian Institute of Technology, Thailand, the master's degree in biomedical engineering from the University of New South Wales, Australia, and the Doctor of Engineering degree in telecommunications from Asian Institute of Technology. She is a Lecturer with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Australia. She was an Intern with the Data Science Research Laboratory, NEC Central Research Laboratories, Japan. Her research interests include the Internet of Things, cyber physical systems, body sensor networks, fog computing, cloud computing, and deep learning.

**NIMSIRI ABHAYASINGHE** (Member, IEEE) received the B.Sc. (Eng.) (Hons.) and M.Sc. degrees in electronic and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2003 and 2007, respectively, and the Ph.D. degree in computer engineering from Curtin University, Australia.

From 1997 to 1998, he was a Junior Research Assistant with the Department of Physics, University of Peradeniya, Sri Lanka. From 2005 to 2010, he was a Lecturer with the University of Moratuwa. From 2010 to 2020, he was a Senior Lecturer with the Department of Electrical and Computer Engineering, Sri Lanka Institute of Information Technology, Sri Lanka, and became an Assistant Professor, in 2020. He is currently a Senior Lecturer with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University. He has published more than 15 peer reviewed papers on human gait analysis and inertial navigation. His research interests include rehabilitation engineering, indoor positioning and navigation, human gait analysis, and inertial navigation.

**IAIN MURRAY AM** (Senior Member, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees in computer systems engineering from Curtin University, in 1998 and 2008, respectively. His Ph.D. thesis titled "Instructional e-Learning Technologies for the Vision Impaired."

He has worked in the field of assistive technology for nearly 35 years both as a Practitioner and a Researcher. Currently, he is with the School of Electrical Engineering, Computing and Mathematical Sciences. He is also a John Curtin Distinguished Professor. He has founded the "Cisco Academy for the Vision Impaired," in 2002, to deliver ICT training to vision impaired people globally. He has supervised 19 research students' completions and published in excess of 140 peer reviewed articles in the fields of the IoT, engineering education, and assistive technology. His research interests include learning environments for people with vision impairment, embedded sensors in health applications, the Internet of Things, and assistive technology. He is a member of the Order of Australia, a fellow of Australian Computer Society, and a Curtin Academy Fellow.