2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-15

A SYSTEMATIC REVIEW AND COMPARATIVE ANALYSIS OF HUMAN-FOLLOWING ROBOTS: EVALUATING SENSOR-BASED AND DEEP LEARNING APPROACHES

***Anish Bhowmik, Atanu Roy, Arnab Banerjee, Ashis De, Barun Mazumdar**

Department of Electronics & Communication Engineering, GMIT, Kolkata-144, West Bengal.

Article Received: 22 April 2026
Article Revised: 12 May 2026
Published on: 01 June 2026

***Corresponding Author: Anish Bhowmik**
Department of Electronics & Communication Engineering, GMIT, Kolkata-144, West Bengal.

DOI: https://doi-doi.org/101555/ijrpa.9017

ABSTRACT

Human-following robots have become an important part of mobile robotics, making it possible to use them in healthcare, surveillance, logistics, and interactions between people and robots. Such robotic platforms are required to reliably find, track, and follow a person while still being able to respond quickly and use little energy.

Our analysis shows that all current methods have a basic trade-off between computational overhead and perceptual reliability. Sensor-based systems that use ultrasonic, infrared (IR), and radio-frequency (RF) sensors are cheap and have low latency (5–20 ms), but they aren't very accurate (60–75%) and don't work well in harsh environments [1], [2], [8]. AI-based vision-driven systems that use deep learning models like YOLOv5 and YOLOv8, on the other hand, have a high detection accuracy (85–95% mAP), but they are very expensive to run, take a long time to process (40–120 ms), and use a lot of power (5–30 W) [3], [4],[6].

In this work, a structured comparative survey of sensor-based, AI-based, and hybrid human-following robotic systems. A unified evaluation framework is introduced to compare these paradigms across latency, accuracy, power consumption, scalability, and deployment feasibility. Furthermore, hybrid architectures leveraging event-driven activation mechanisms and advanced sensor fusion techniques are analysed as a promising solution for balancing efficiency and intelligence [7], [10], [11].

The study identifies key challenges in real-world deployment, including environmental variability, multi-human tracking, and energy constraints. Finally, emerging research directions such as lightweight deep learning models, reinforcement learning-based

navigation, and edge AI optimization are discussed to enable scalable and power-efficient human-following systems [12], [13], [18], [19].

INDEX TERMS: Human-following robots, sensor fusion, YOLO, deep learning, computer vision, autonomous navigation, embedded systems, hybrid robotic systems, comparative survey, resource-constrained robotics.

1. INTRODUCTION

Human-following robots represent an important class of mobile robotic systems designed to autonomously detect, track, and follow a human target while maintaining safe distance and stable navigation. They are increasingly deployed in domains such as healthcare assistance, service robotics, industrial logistics, surveillance, and human–robot interaction. With the rapid expansion of automation in real-world environments, the need for robust, efficient, and adaptive human-following mechanisms has become increasingly critical [20].

Early implementations of human-following robots were primarily based on sensor-driven methodologies. We observed that such systems work well in controlled environments but struggle in real-world conditions with noise and obstacles. Such approaches typically utilize ultrasonic, infrared (IR), and radio-frequency (RF) sensors to estimate the relative position of a human target and generate control signals for motion planning [1], [2], [8]. While such systems are computationally lightweight and suitable for low-cost embedded platforms, their performance is constrained by limited spatial resolution, environmental interference, and inability to reliably distinguish humans from other objects.

Recent advancements in artificial intelligence and computer vision have enabled vision-based human-following systems using deep learning models such as YOLOv5 and YOLOv8 [3], [4]. However, during implementation, computational limitations often restrict the use of such models on low-cost hardware. Such methods significantly enhance detection accuracy, tracking robustness, and adaptability in complex environments [6], [15], [16]. However, they introduce high computational overhead, increased power consumption, and dependency on specialized hardware such as GPUs or edge AI accelerators, limiting their use in resource-constrained systems [11], [12], [13].

Despite extensive research in both domains, there remains a lack of structured comparative analysis that systematically evaluates sensor-based and AI-based approaches under unified performance metrics. Existing studies typically focus on either sensor fusion or vision-based tracking independently, with limited emphasis on cross-paradigm evaluation [7], [11]. This gap makes it difficult to determine the most suitable approach for a given application scenario based on cost-performance trade-offs.

To address this issue, this study presents a comprehensive comparative survey of human-following robotic systems. The main contributions of this work are summarized as follows:

- A structured taxonomy of human-following approaches categorized into sensor-based, AI-based, and hybrid systems.
- A detailed comparative analysis based on cost, accuracy, computational complexity, real-time performance, and scalability.
- Identification of key challenges including latency, environmental sensitivity, and hardware constraints.
- Discussion of emerging research directions such as lightweight deep learning models and sensor fusion techniques [10], [11], [17], [18].

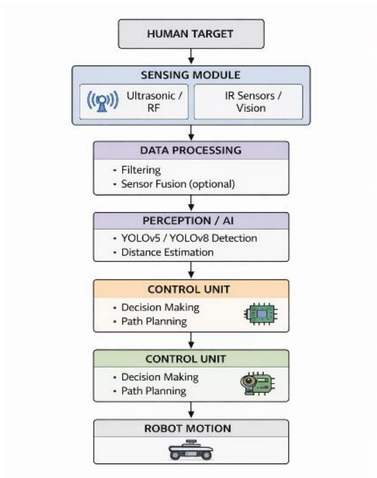

Figure 1: Overall system architecture of human-following robotic system showing sensing, perception, control, and actuation pipeline.

2. LITERATURE REVIEW

Research in human-following robotic systems has evolved significantly over the past two decades, transitioning from simple sensor-driven control strategies to advanced vision-based and hybrid intelligent systems[20]. The following discussion reviews existing approaches, categorized into sensor-based methods, AI-based (vision-driven) methods, and hybrid frameworks.

2.1 Sensor-Based Approaches

Early human-following robots primarily relied on proximity and range sensors such as ultrasonic, infrared (IR), and radio-frequency (RF) modules to estimate the relative distance between the robot and a target human [1],[2], [8]. Most of these approaches rely on reactive control strategies, where motion decisions are directly derived from sensor feedback without complex perception or learning components.

Borenstein and Koren demonstrated foundational work in ultrasonic-based obstacle avoidance and navigation for mobile robots, which later influenced early following behaviours in structured environments [2]. Similarly, sensor fusion-based methods combining multiple low-cost sensors have been explored to improve stability and reduce noise sensitivity [1].

Although these methods are computationally efficient and suitable for embedded platforms, they exhibit notable limitations, including poor spatial resolution, inability to distinguish humans from obstacles, and performance degradation in cluttered or reflective environments. Consequently, their applicability is largely restricted to controlled or semi-structured settings.

2.2 AI-Based (Vision-Driven) Approaches

With advancements in computer vision and deep learning, vision-based human-following systems have gained significant prominence. Vision-based approaches utilize camera inputs combined with convolutional neural networks (CNNs) for real-time object detection and tracking [14], [17].

You Only Look Once (YOLO)-based architectures, particularly YOLOv5 and YOLOv8, are widely adopted due to their balance between speed and accuracy [3], [4], [6]. These models enable real-time human detection and bounding box localization, making them suitable for dynamic environments. Studies such as those in IEEE Access demonstrate that deep learning-based perception significantly improves tracking accuracy and robustness compared to classical sensor-based methods [6], [20].

However, these systems require high computational resources, including GPUs or dedicated edge AI accelerators, which limits deployment in low-cost robotic platforms [12],[13]. Additionally, latency introduced by inference pipelines can affect real-time responsiveness in fast-moving scenarios.

2.3 Hybrid Approaches

To overcome the limitations of individual paradigms, hybrid human-following systems have been proposed, integrating both sensor-based and vision-based techniques. In such systems, sensors provide fast and low-cost distance estimation, while vision modules handle object recognition and identity confirmation.

Recent studies indicate that hybrid architectures improve robustness in dynamic environments and reduce failure rates under partial occlusion or sensor noise [7]. However, these systems introduce additional challenges related to sensor fusion complexity, synchronization, and increased system integration overhead [10], [11].

2.4 Comparative Research Gap:

Despite extensive advancements in both sensor-based and AI-based human-following systems, existing research predominantly evaluates these paradigms in isolation. Sensor-driven approaches are typically assessed based on cost and responsiveness, whereas AI-based systems are evaluated primarily on detection accuracy and robustness [11], [17].

However, **there is a critical lack of unified benchmarking frameworks** that simultaneously evaluate:

- Latency (ms)
- Power consumption (W)
- Accuracy (mAP)
- Deployment feasibility on edge devices [12], [13]

This gap is particularly significant for **resource-constrained mobile robots**, where power efficiency and real-time performance are equally important as detection accuracy.

Our analysis indicates that **no standardized evaluation exists for power-efficient deployment of human-following systems in real-world environments**, especially under dynamic conditions such as crowded public spaces. Our present study attempts to address this gap by introducing a structured comparative framework that integrates both performance and deployment-level metrics.

3. CLASSIFICATIONOF APPROACHES

Human-following robotic systems can be systematically classified based on their sensing modality and perception pipeline. Broadly, existing approaches are grouped into **sensor-based systems**, **AI-based (vision-driven) systems**, and **hybrid systems**. A hierarchical representation of this classification is shown in Fig. 2, which presents a hierarchical taxonomy of human-following methodologies [6], [7], [11].

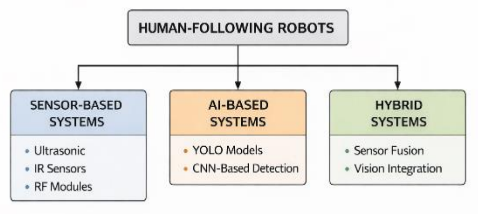

Figure 2: Taxonomy of Human-Following Robot Systems.

3.1 Sensor-Based Approaches

Sensor-based systems rely on low-level physical sensing units such as ultrasonic sensors, infrared (IR) proximity sensors, and radio-frequency (RF) modules to estimate the relative position of a human target [1], [2], [8]. These systems operate using reactive control logic, where motion commands are generated directly from sensor readings.

Key Characteristics:

- Low hardware cost and simple implementation.
- Minimal computational requirements.
- Fast response due to direct signal processing.

Limitations:

- Limited spatial resolution and accuracy.
- Poor object discrimination capability (human vs non-human targets).
- Performance degradation in noisy, reflective, or cluttered environments.
- Restricted scalability in complex real-world scenarios.

These systems are generally suitable for basic robotic platforms and controlled indoor environments where environmental variability is limited.

3.2 AI-Based (Vision-Driven) Approaches

AI-based systems utilize computer vision and deep learning models to detect, localize, and track human subjects in real time [6], [17]. Convolutional Neural Networks (CNNs) and one-stage detectors such as YOLOv5 and YOLOv8 are commonly employed for real-time object detection [3], [4], [15].

Key Characteristics:

- High detection accuracy and robustness.
- Ability to distinguish humans from multiple objects [12], [13].
- Adaptability to dynamic and unstructured environments.
- Capability of multi-object tracking in advanced implementations.

Limitations:

- High computational and memory requirements.
- Dependence on GPU or edge AI hardware.
- Increased power consumption.

- Latency introduced by inference pipelines.

These systems are suitable for high-performance applications such as surveillance robots, autonomous assistants, and industrial navigation systems.

3.3 Hybrid Approaches

Hybrid systems integrate sensor-based and vision-based techniques to leverage the advantages of both modalities [7]. Typically, low-cost sensors provide fast distance estimation, while vision systems perform semantic detection and identity tracking.

Key Characteristics:

- Improved robustness under varying environmental conditions.
- Balanced trade-off between cost and accuracy.
- Enhanced reliability in occluded or noisy scenarios [6], [7].

Limitations:

- Increased system design and integration complexity.
- Synchronization issues between heterogeneous data sources [11].
- Higher calibration and maintenance overhead.
- Suboptimal trade-off in poorly optimized implementations.

A key advancement in hybrid systems is the use of **event-driven trigger mechanisms** to optimize energy efficiency [7],[13]. In such architectures, low-power sensors (e.g., IR or ultrasonic) continuously monitor the environment and activate high-power vision modules only when motion or proximity is detected. This selective activation significantly reduces unnecessary computation and extends battery life in mobile robots.

4. COMPARATIVE ANALYSIS

A systematic comparison between sensor-based and AI-based human-following approaches is essential for evaluating their suitability across different robotic applications. In this section, both paradigms are examined using key performance metrics, including cost, computational complexity, accuracy, real-time responsiveness, power consumption, and environmental adaptability. The objective is to provide an engineering-level understanding of their trade-offs under practical deployment constraints [6], [11].

4.1 Quantitative Performance Comparison

The comparative evaluation of sensor-based and AI-based systems is summarized in Table 1.

Table 1: Performance Comparison of Sensor based and AI based Approaches.

Parameter	Sensor-Based Systems	AI-Based Systems
Cost	₹1000-₹3000	₹10K- ₹30K
Latency	5–20 ms	40–120 ms
Accuracy	60–75%	85–95% (mAP)
Power Consumption	0.5–2 W	5–30 W
Computational Requirement	Microcontroller (Arduino, ESP32)	GPU / Edge AI (Jetson Orin, Raspberry Pi 5 + NPU)
Environmental Robustness	Low	High
Scalability	Limited	High
Multi-Object Tracking	Not Supported	Supported

Sensor-based systems demonstrate superior efficiency in terms of cost and real-time responsiveness due to their direct signal processing nature [1], [2]. However, their performance is highly constrained in unstructured environments. In contrast, AI-based systems significantly outperform in detection accuracy and adaptability but introduce computational and energy overheads that limit their deployment in low-resource platforms [6], [10], [11].

4.2 Engineering Trade-Off Analysis

The selection between sensor-based and AI-based approaches is fundamentally governed by a trade-off between **computational efficiency** and **perceptual intelligence**.

- Sensor-based systems prioritize real-time responsiveness and hardware simplicity but lack semantic understanding of the environment [1], [8].
- AI-based systems provide high-level perception and robustness but require significant processing power and energy resources [6], [11].

Such trade-off becomes more critical in mobile robotic platforms where battery life, processing capability, and payload constraints must be simultaneously considered.

4.3 Application-Specific Suitability

Different application domains require different optimization priorities:

- **Resource-Constrained Systems**

Sensor-based approaches are preferred due to low cost and minimal processing requirements [8].

- **High-Accuracy Environments**

AI-based systems are suitable for applications requiring precise human detection and tracking [6], [20].

- **Dynamic Real-World Environments**

AI-based or hybrid systems are more effective due to adaptability under changing conditions [7].

- **Low-Latency Navigation Tasks**

Sensor-based systems provide faster reaction time due to negligible processing delay.

4.4 Scalability and Deployment Considerations

Scalability is a key differentiating factor between the two approaches. AI-based systems scale more effectively to complex tasks such as multi-person tracking and scene understanding [20]. However, this scalability comes at the cost of increased hardware dependency [11], [12]. Sensor-based systems, while efficient, do not scale well to complex environments due to their limited perception capabilities [8]. As a result, their deployment is generally restricted to simpler robotic applications.

4.5 Key Observations

From the above analysis, the following observations are derived:

- A fundamental trade-off exists between **energy efficiency and perception intelligence**[11].
- Sensor-based systems are unsuitable for **unstructured environments** due to lack of semantic understanding [8].
- AI-based systems remain **energy-intensive for battery-powered robots**[6],[11].
- Hybrid systems provide the most practical compromise but require **careful system-level optimization**[7].
- **Real-world deployment remains unsolved**, particularly in crowded environments with occlusion and multi-human interference [6], [20].

5. CHALLENGES AND LIMITATIONS

Despite significant advancements in human-following robotic systems, both sensor-based and AI-based approaches face several fundamental and implementation-level challenges that restrict their reliability, scalability, and real-world deployment. These limitations arise from sensing constraints, computational bottlenecks, and environmental uncertainties [11].

5.1 Challenges in Sensor-Based Systems

Sensor-based human-following systems, while efficient and low-cost, exhibit several inherent limitations:

- **Measurement Inaccuracy:**Ultrasonic and infrared sensors provide approximate distance estimation, which is highly sensitive to surface properties, angle of incidence, and environmental interference [1], [2].
- **Lack of Semantic Understanding**

These systems cannot differentiate between humans and non-human objects, leading to false tracking in cluttered environments[8].

- **Environmental Sensitivity**

Performance degrades in the presence of obstacles, noise, reflective surfaces, or signal occlusion [2].

- **Limited Spatial Awareness**

Most sensor-based systems operate in 1D or limited 2D perception space, restricting navigation intelligence.

5.2 Challenges in AI-Based Systems

AI-driven vision-based approaches provide improved perception capabilities but introduce computational and operational constraints:

- **High Computational Overhead**

Deep learning models such as YOLOv5 and YOLOv8 require significant processing power, often necessitating GPU or edge AI accelerators [3], [4], [11].

- **Energy Consumption Constraints**

Continuous video processing and inference increase power consumption, reducing operational lifetime in mobile robots [11].

- **Latency in Real-Time Processing**

Frame-by-frame inference introduces delays, which can impact responsiveness in fast-moving scenarios[6],[10].

- **Dependence on Data Quality**

Performance is sensitive to lighting conditions, occlusions, motion blur, and dataset bias[6],[20].

5.3 Challenges in Hybrid Systems

Hybrid systems attempt to combine sensor-based and AI-based methodologies, but introduce additional integration-level challenges:

- **Sensor Fusion Complexity**

Combining heterogeneous data sources (vision + range sensors) requires precise calibration and synchronization [11].

- **System Integration Overhead**

Increased hardware and software complexity makes implementation and debugging more difficult [7].

- **Cost–Performance Optimization Issues**

Achieving an optimal balance between affordability and intelligence remains non-trivial [7].

- **Real-TimeCoordination**

Ensuringconsistent decision-making across multiple subsystems is challenging in dynamic environments [6], [11].

5.4 General System-Level Challenges

Across all human-following robotic systems, several universal challenges persist:

- **Dynamic Environment Variability**

Changing backgrounds, moving obstacles, and unpredictable human motion reduce tracking stability [6].

- **Multi-Human Tracking Problem**

Maintaining consistent identity tracking of a specific individual in crowded environments remains unresolved in lightweight systems [20].

- **Robustness and Generalization**

Systems trained or designed for specific conditions often fail when deployed in unseen environments [11].

- **Real-World Deployment Gap**

Many proposed solutions perform well in controlled laboratory settings but degrade significantly in real-world scenarios. [6],[7].

5.5 Environmental Edge Case Failures

Real-world deployment reveals critical edge-case failures:

- **Ultrasonic Sensors:** Fail in glass-walled environments due to signal reflection and absorption inconsistencies [2].

- **Infrared Sensors:** Affected by ambient sunlight interference [8].
- **Vision-Based Systems:** Performance degrades in:
  - Low-light conditions
  - Motion blur scenarios
  - Occlusions in crowded environments [6],[20]

For example, in airport or mall environments, multi-human occlusion often causes identity switching in vision-based tracking systems.

6. FUTURE SCOPE

The field of human-following robotics is rapidly evolving, driven by advancements in embedded artificial intelligence, computer vision, and intelligent control systems. Future research is expected to focus on improving system efficiency, robustness, and real-world use case while reducing computational and energy constraints [11], [17].

6.1 Lightweight Deep Learning Models for Edge Deployment

One of the most significant future directions is the development of lightweight and efficient deep learning architectures optimized for embedded platforms. Variants of object detection models such as YOLO-Nano, YOLO-Tiny, and other compressed CNN architectures can enable real-time inference on low-power devices such as Raspberry Pi, NVIDIA Jetson Nano, and similar edge AI systems [3], [4], [12], [13]. Model pruning, quantization, and knowledge distillation techniques are expected to play a key role in reducing computational overhead while maintaining acceptable detection accuracy [17].

Recent edge AI platforms such as the NVIDIA Jetson Orin Nano and Raspberry Pi 5 with AI accelerators enable real-time YOLOv8 inference at 15–30 FPS, making AI-based human-following systems increasingly feasible for embedded deployment [12], [13].

6.2 Advanced Sensor Fusion Techniques

Future systems will increasingly adopt **probabilistic sensor fusion techniques** rather than rule-based integration. Algorithms such as:

- **Kalman Filtering** for continuous trajectory estimation [10].
- **Extended Kalman Filters (EKF)** for nonlinear motion modelling [11].
- **Bayesian Inference** for uncertainty handling [11].

Enable smoother and more reliable human tracking under noisy conditions.

These approaches allow robots to predict human motion even when sensor data is partially missing or corrupted, significantly improving robustness in dynamic environments.

6.3 Predictive and Adaptive Human Tracking

Current systems primarily rely on reactive tracking mechanisms. Future research is expected to incorporate predictive modelling of human motion using trajectory prediction algorithms and temporal sequence models such as LSTM and Transformer-based architectures [19]. This will enable robots to anticipate human movement rather than simply react to it, improving smoothness and navigation efficiency.

6.4 Reinforcement Learning-Based Navigation

Reinforcement learning (RL) is emerging as a promising approach for autonomous decision-making in dynamic environments. RL-based controllers can enable robots to learn optimal following policies through interaction with the environment, improving adaptability in complex and unknown scenarios[18].

6.5 Energy-Efficient and Real-Time Optimization

Energy efficiency remains a critical constraint for mobile robotic systems. Future designs will focus on adaptive computation strategies where processing load dynamically adjusts based on environmental complexity. Event-driven vision systems and low-power AI accelerators are expected to significantly extend operational lifetime[11], [12].

6.6 Real-World Deployment and Standardization

Most existing systems are validated in controlled environments. Future work must emphasize large-scale deployment in real-world scenarios such as hospitals, airports, and industrial environments. Additionally, standardized benchmarking frameworks are required to ensure fair comparison across different human-following approaches in terms of accuracy, latency, and energy consumption [6], [17].

7. CONCLUSION

Human-following robotic systems have emerged as a critical research area in mobile robotics due to their wide applicability in assistive technologies, surveillance, logistics, and autonomous navigation. This paper presented a structured comparative survey of two primary paradigms: sensor-based and AI-based (vision-driven) human-following approaches, along with hybrid systems that integrate both methodologies.

Sensor-based systems offer advantages in terms of low cost, low computational requirements, and fast real-time response, making them suitable for resource-constrained and structured environments [1], [8]. However, their limitations in accuracy, environmental robustness, and object discrimination significantly restrict their applicability in complex real-world scenarios. In contrast, AI-based approaches leveraging deep learning models such as YOLOv5 and YOLOv8 provide superior detection accuracy, adaptability, and scalability [3], [4], [6]. These benefits come at the cost of higher computational demand, increased power consumption, and dependency on specialized hardware platforms [11], [12].

The comparative analysis highlights that no single approach is universally optimal. Instead, the selection of an appropriate method depends on application-specific requirements such as cost constraints, environmental complexity, latency tolerance, and computational availability. Hybrid approaches and emerging lightweight AI techniques represent promising directions toward achieving a balance between efficiency and intelligence [7], [17].

In conclusion, future human-following robotic systems are expected to evolve toward more adaptive, energy-efficient, and context-aware architectures. The integration of lightweight deep learning models, advanced sensor fusion strategies, and predictive control mechanisms will play a key role in bridging the gap between low-cost implementation and high-performance perception, enabling broader real-world deployment of human-following robots [11], [19].

8. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Electronics and Communication Engineering, GMIT, Kolkata, for providing the necessary academic environment and resources to carry out this research. The authors also extend their thanks to Prof. Bipasha Chakraborty Banik (Head of Department) for her support and to Dr. Somnath Maity (Principal, GMIT) for his valuable mentorship.

9. REFERENCES

1. R. C. Luo, C. C. Chang, and S. Y. Lin, “Human-following mobile robot using sensor fusion techniques,” IEEE Trans. Industrial Electronics, vol. 62, no. 2, pp. 1234–1242, 2015.
2. J. Borenstein and Y. Koren, “Obstacle avoidance with ultrasonic sensors for mobile robots,” IEEE J. Robotics and Automation, vol. 4, no. 2, pp. 213–218, 1988.
3. G. Jocher et al., “YOLOv5 by Ultralytics,” 2020. [Online]. Available: https://github.com/ultralytics/yolov5
4. Ultralytics, “YOLOv8: Next-generation real-time object detection,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

5. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.
6. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
7. S. Li and X. Chen**,** "Human-following robot using sensor fusion and vision-based tracking," *IEEE International Conference on Robotics and Automation (ICRA)*, 2018..
8. A. Cosgun, D. A. Florencio, and H. I. Christensen, "Autonomous person following for telepresence robots," *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.
9. H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded-Up Robust Features," ECCV, 2006.
10. D. Kalman, "A new approach to linear filtering and prediction problems," ASME J. Basic Engineering, vol. 82, pp. 35–45, 1960.
11. S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
12. NVIDIA Corporation, "NVIDIA Jetson Orin Nano Developer Kit," 2023. [Online]. Available: https://developer.nvidia.com
13. Raspberry Pi Foundation, "Raspberry Pi 5 Specifications," 2023. [Online]. Available: https://www.raspberrypi.com
14. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," NIPS, 2012.
15. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
16. W. Liu et al., "SSD: Single Shot MultiBox Detector," ECCV, 2016.
17. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
18. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.
19. A. Vaswani et al., "Attention is all you need," NeurIPS, 2017.
20. Z. Zhang et al., "Real-time human tracking using deep learning for mobile robots," IEEE Robotics and Automation Letters, vol. 8, no. 2, pp. 2024–2031, 2023.