 DOI: 10.65781/actbg841

Engineering Dynamics and Applied Sciences
Open Access & Peer Reviewed Journal licensed under CC BY 4.0

Article type: Research Article

A Two-Phased Fruits Image Detection Based on Colors as an Auto-Vision Object Annotation Technique

Ankita Chaurasia[1], Mubeen Ahmed Khan[2], Akshit Harsola[3] and Kalyani Tiwari[4]

[1]Department of Computer Science and Engineering, Medicaps University, Indore, MP, 453331, India
[2]Department of Computer Science and Engineering (Cyber Security), Dayananda Sagar University, Bengaluru, Karnataka, 562112, India
[3]Department of Computer Science and Engineering, Medicaps University, Indore, MP, 453331, India
[4]Department of Computer Science and Engineering, IPS Academy Institute of Engineering, Indore, MP, 453331, India

Correspondence to: Mubeen Ahmed Khan; email: *makkhan0786@gmail.com*

(Article Info: Received: 09 Aug 2025; Accepted: 04 Nov 2025; Published online: 18 Jan 2026)
Version of record: v1

Abstract: Automatic image detection systems are computer vision solutions that use various algorithmic approaches to identify and localize objects within digital images. Current systems primarily rely on either traditional computer vision techniques or deep learning methods, presenting significant limitations in real-world applications. These systems often struggle with balancing processing speed and detection accuracy, lack adaptability across different object types, and require substantial computational resources. Additionally, existing solutions typically operate in a single mode, making them inefficient for scenarios requiring different levels of detection complexity. This paper presents AutoVision, a novel approach to a semi-automated system for detecting and annotating objects that employs an enhanced detection approach using multiple color space analysis with adaptive thresholding. Implementing a system with a comprehensive two-phase methodology which is preprocessing for image organization and quality assessment, followed by an advanced object detection algorithm utilizing combined threshold masks and morphological operations. In our experimental evaluation with a dataset of 5000 fruit images, the system achieved a remarkable processing efficiency at 757.06 images per second with an overall detection success rate of 74.68%. Also, the system maintained exceptional image quality standards, with 98.76% of processed capabilities which is 0.003 seconds per image. The research will contribute to the field of automated fruit detection by presenting a scalable, efficient solution that balances processing speed with detection accuracy.

Keywords: Object Detection; Computer Vision; Dual-Mode System; Automated Annotation; Quality Assessment; Real time systems.

1 Introduction

Computer vision (CV) and machine learning have revolutionized automated image analysis over the past decade, evolving from simple edge detection techniques to sophisticated deep learning approaches [1],[2].Traditional object detection systems initially relied on fundamental techniques such as edge detection, feature-based methods, and template matching, which while ground-breaking, faced significant limitations in processing speed and accuracy across diverse scenarios [3]. The evolution of these systems has been marked by a transition from geometric pattern recognition to more advanced approaches, including deep learning-based detection, hybrid systems, and multi-stage processing pipelines [4]-[6]. The rapid advancement of these technologies has created an unprecedented demand for annotated image datasets, particularly in machine learning applications. Current annotation systems typically fall into two distinct categories: high-speed simplified detection systems that prioritize processing speed over accuracy, and complex detection systems that offer high accuracy but require significant computational resources [7],[8]. This dichotomy has created a significant challenge in the field, where users must often choose between speed and accuracy rather than having access to a system that can provide both based on specific needs. Manual annotation systems, while offering high accuracy and flexibility, suffer from low processing speeds and high resource requirements, whereas current automated systems provide variable accuracy with limited adaptability across different scenarios [9]-[16].The computer vision community now faces critical challenges in developing annotation solutions that can effectively handle both simple and complex detection scenarios without compromising on either speed or accuracy [17]-[19]. Commercial annotation tools, open-source frameworks, and industry-specific solutions have attempted to address these challenges, but continue to face limitations in terms of processing pipelines, adaptability, and resource efficiency [20]. The growing complexity of modern applications, coupled with the need for real-time processing capabilities, has highlighted the importance of developing more flexible and efficient annotation systems that can adapt to varying levels of complexity while maintaining high quality standards.

2 Literature Review

The landscape of automated annotation systems presents a complex interplay between manual and automated approaches, each with distinct characteristics and limitations. Manual annotation systems have traditionally offered high accuracy and flexibility but suffer from significant speed limitations and resource intensiveness. Current automated systems have attempted to bridge this gap, offering improved processing speeds but often struggling with consistency and adaptability. Commercial annotation tools and open-source frameworks have emerged to address these challenges, yet they frequently fall short in providing comprehensive solutions that can handle diverse scenarios effectively [11], [12]. The field of automated object detection and annotation currently faces several critical challenges that significantly impact its practical implementation. Existing systems often demonstrate a fundamental trade-off between speed and accuracy, where improvements in one aspect typically come at the cost of the other. This limitation is particularly evident in systems that employ fixed processing pipelines, which lack the ability to adapt to varying complexity levels in different detection scenarios. Furthermore, current solutions often exhibit significant limitations in flexibility, particularly in their ability to handle different object types and environmental conditions. The resource utilization in these systems frequently proves inefficient, with high processing overhead and limited scalability capabilities [13], [14]. Quality control represents another significant challenge in current systems, where inconsistent annotation quality and limited validation methods have become persistent issues. The lack of standardized quality assessment metrics has led to difficulties in maintaining consistent output quality across different detection scenarios. Additionally, many existing systems require significant technical expertise to operate effectively, creating barriers to adoption in various practical applications. These limitations have highlighted the need for more sophisticated approaches that can maintain high quality standards while providing efficient processing capabilities [15], [16]. Research objectives in this field have evolved to address these fundamental challenges, focusing on developing more adaptive and efficient solutions. Primary objectives include the development of dual-mode detection systems that can combine high-speed and comprehensive detection capabilities, creating adaptive frameworks that can support various object types and scenarios, and implementing robust quality assessment metrics. The emphasis has also been placed on achieving real-time processing capabilities while maintaining an intuitive user interface, making these systems more accessible to a broader range of users [21]-[25].

Current research gaps reveal several areas requiring attention in the field of automated object detection and annotation. Technical limitations persist in terms of fixed processing pipelines and limited adaptability, while practical constraints continue to manifest in complex user interfaces and limited scalability. Quality assurance remains a significant concern, with inconsistent metrics and limited validation capabilities often requiring manual intervention [26]-[29]. These gaps highlight the need for more sophisticated solutions that can address both technical and practical limitations while maintaining high standards of accuracy and efficiency.

The evolution of automated annotation systems has demonstrated the need for solutions that can effectively balance various competing requirements [11], [13]. While progress has been made in developing more sophisticated detection and annotation techniques, significant opportunities remain for improving system adaptability, resource efficiency, and user accessibility. The field continues to evolve, with emerging technologies and methodologies offering new possibilities for addressing current limitations and advancing the capabilities of automated annotation systems.

3 Methodology

3.1 Analysis of Current Systems and Proposed Enhancements

The methodology begins with an analysis of existing annotation systems, as shown in Table 1, which compares manual annotation, current automated systems, and our proposed AutoVision approach. This analysis informed the development of our enhanced system architecture. The table presents a comparative analysis of three different approaches to image annotation: Manual Annotation, Current Automated Systems, AutoVision (proposed system)

Table 1: A comparison of AutoVision with current automated system and manual annotation

Aspect	Manual Annotation	Current Automated Systems	AutoVision
Speed	Low	Medium - High	Adaptive
Accuracy	High	Variable	Mode dependent
Flexibility	High	Limited	High
Resource Usage	High	Fixed	Optimized
Quality Control	Medium - High	Limited	Comprehensive

The table compares these three approaches across five key aspects:

- Speed: Showing that manual annotation is slow, current automated systems are medium to high speed, while the AutoVision system offers adaptive speed based on requirements.
- Accuracy: Indicating that manual annotation has high accuracy, current automated systems have variable accuracy, and AutoVision's accuracy is mode-dependent (likely adjusting based on which detection mode is being used).
- Flexibility: Showing that both manual annotation and the AutoVision system offer high flexibility, while current automated systems are limited in this regard.
- Resource Usage: Indicating that manual annotation requires high resources, current automated systems have fixed resource requirements, and AutoVision optimizes resource usage.
- Quality Control: Showing that manual annotation relies on manual quality control, current automated systems have limited quality control, while AutoVision offers comprehensive quality control mechanisms.

3.2 Technical Framework and Implementation

The technical implementation of AutoVision is structured around a layered architecture that processes images through multiple stages. Fig. 1 shows the complete system architecture with its four primary components: the input layer for initial image acquisition, preprocessing layer for quality assessment, core processing layer for detection, and output layer for result generation.

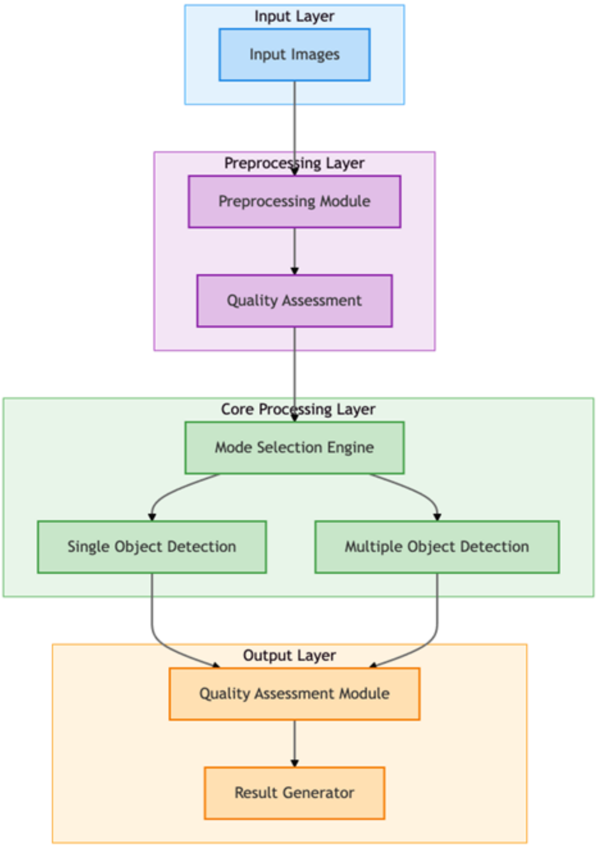

Fig. 1: Layered Architecture of the AutoVision System

This diagram illustrates the complete system architecture of AutoVision, featuring a four-layer design that processes images through sequential stages. The system begins with the Input Layer where images are initially acquired. These images then flow to the preprocessing layer, which handles organization and initial quality assessment [24]. The core processing layer contains the innovative mode selection engine that directs images to either single object detection, using enhanced canny edge detection techniques [30] or multiple object detection, utilizing YOLO [31], depending on complexity requirements. Finally, the output layer performs a second quality assessment on the detection results before generating the final output with annotations, visualizations, and quality metrics.
This architecture represents a key innovation in the project as it enables the system to adaptively process images through different detection pathways based on image complexity and requirements, combining the speed advantages of single object detection with the comprehensive capabilities of multiple object detection [32] when needed. The dual quality assessment approach (both pre- and post-processing) ensures consistently high standards in the final outputs.

3.3 Enhanced Detection Algorithm

Our detection algorithm introduces several key innovations over existing approaches. As shown in Fig. 2, the mode selection mechanism determines the optimal processing path based on image complexity and available resources, representing a critical advancement in automated detection systems.
This flowchart illustrates the innovative mode selection mechanism that forms the core of AutoVision's enhanced detection algorithm. The system begins by assessing image complexity and then follows distinct decision paths based on complexity levels and resource availability. For high-complexity images, the system checks available resources, while for low-complexity images, it evaluates speed requirements. The algorithm intelligently routes processing through either the streamlined single object mode (using enhanced Canny edge detection) when speed is critical or resources are limited, or the comprehensive multiple object mode (using YOLO-based detection) when resources are available or accuracy is paramount. The quality check mechanism provides feedback and can trigger mode switching, if necessary, with a final fallback to manual review for challenging cases that fail quality checks without a mode switch option.

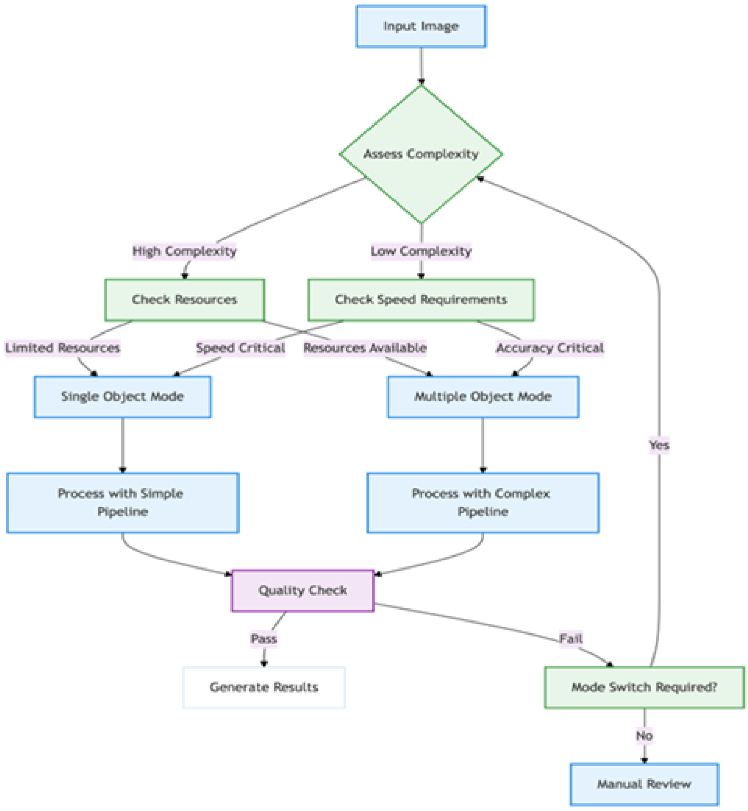

Fig. 2: Adaptive Mode Selection Algorithm and Processing Flowchart of AutoVision

Our detection algorithm introduces several key innovations over existing approaches. As shown in Fig. 2, the mode selection mechanism determines the optimal processing path based on image complexity and available resources, representing a critical advancement in automated detection systems. This dynamic routing capability enables AutoVision to achieve unprecedented levels of adaptability, ensuring optimal resource utilization while maintaining high-quality standards. The system leverages a sophisticated complexity assessment algorithm that analyzes image characteristics including contrast variations, edge density, and potential object counts to make intelligent mode selection decisions.

Unlike traditional systems that apply a fixed processing pipeline regardless of image characteristics, AutoVision's adaptive approach significantly improves both efficiency and accuracy across diverse scenarios. The dual-mode architecture combines the speed advantages of the Single Object Mode with the comprehensive capabilities of the multiple object mode, creating a versatile system that can handle varying detection requirements without compromising performance. Furthermore, the integrated quality feedback loop with mode-switching capabilities ensures robust recovery from potential detection challenges, minimizing the need for manual intervention while maximizing processing success rates.

3.4 Processing Pipeline Architecture

The processing pipeline implements a novel approach to image analysis and object detection, as illustrated in Fig. 3. This pipeline includes crucial steps such as color space conversion, noise reduction, and quality assessment, ensuring robust detection results through a series of optimized processing stages. This flowchart illustrates the comprehensive processing pipeline of AutoVision, detailing the sequential stages involved in transforming raw input images into high-quality detection results. The pipeline begins with color space conversion to optimize fruit feature extraction, followed by noise reduction to improve image quality. A preliminary quality assessment determines if the image meets minimum standards before proceeding. Images that pass continue through threshold generation, morphological operations to enhance object boundaries, and precise boundary detection. The process incorporates result validation and a second quality check before generating annotations. Images failing the initial quality assessment are flagged for manual review, ensuring all outputs—whether automatically or manually processed—meet the system's quality standards.

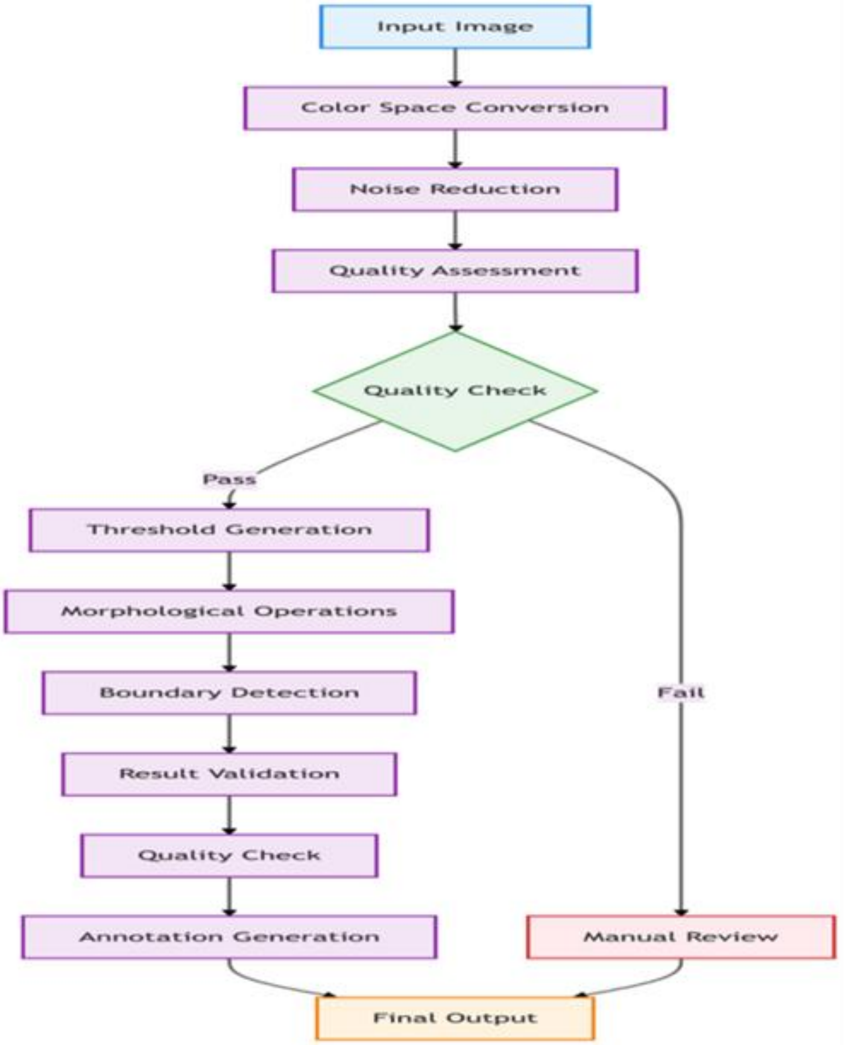

Fig. 3: Optimized Processing Pipeline for AutoVision's Image Analysis and Object Detection

The implementation of multiple color space analysis (particularly HSV transformation) significantly enhances the system's ability to identify and segment fruits with varying color characteristics, while the advanced noise reduction techniques using bilateral filtering preserve critical edge information while eliminating background interference.
The dual-stage quality assessment framework represents a critical innovation, providing early identification of problematic images and preventing wasted computational resources on images unlikely to yield acceptable results. For images meeting quality thresholds, the adaptive threshold generation process dynamically adjusts parameters based on image characteristics, enabling more accurate segmentation across diverse lighting conditions and object types. The enhanced morphological operations, utilizing optimized kernel sizes (7×7), effectively connect fragmented edges and fill gaps in object boundaries, substantially improving detection completeness. This processing pipeline culminates in a comprehensive validation system that ensures all generated annotations meet predefined quality standards, with problematic cases appropriately flagged for manual review, creating a robust end-to-end solution for automated object detection and annotation.

3.5 User Interface and Control System

The implementation of AutoVision includes a user-friendly interface, shown in Fig. 4, which provides intuitive access to all system functionalities. This interface design focuses on simplifying complex operations while maintaining access to advanced capabilities and providing comprehensive feedback mechanisms. This screenshot displays the main interface of the AutoVision tool, featuring a clean, minimalist design that provides straightforward access to all system functionalities through a logical workflow sequence. The interface presents a dark-themed environment with high-contrast buttons for improved visibility and user experience, organizing operations in a top-to-bottom workflow that guides users through the complete detection and annotation process.

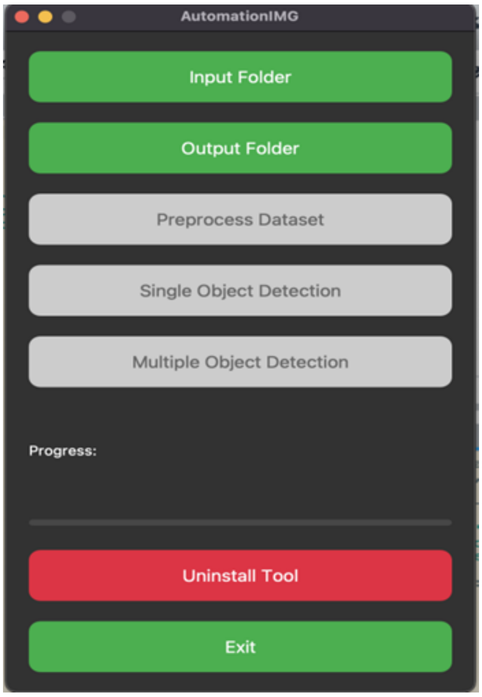

Fig. 4: AutoVision User Interface with Intuitive Control System and Workflow Management

The interface employs a carefully considered sequential layout that guides users through the logical progression of operations required for effective image processing and object detection. The system interface features several key components that facilitate efficient workflow management:

a. Input/Output Selection (Green Buttons): The prominent green "Input Folder" and "Output Folder" buttons at the top allow users to easily select source directories containing fruit images for processing and destination directories for storing results, establishing the foundation for all subsequent operations.
b. Processing Operations (Gray Buttons): The central section contains three sequential processing buttons:
   - "Preprocess Dataset" organizes and prepares images, extracting class labels and performing initial organization
   - Single Object Detection" initiates the Canny edge detection-based processing for simple, high-speed object detection
   - "Multiple Object Detection" activates the more comprehensive YOLO-based detection system for complex scenarios
c. Progress Monitoring: The interface includes a progress indicator with textual status updates and a visual progress bar, providing real-time feedback during potentially lengthy operations and ensuring users remain informed about system status
d. System Management (Bottom Buttons): The interface concludes with system management options including "Uninstall Tool" (red) for removing the application and "Exit" (green) for closing the program, maintaining consistent color coding throughout the interface.

This thoughtfully designed interface embodies AutoVision's commitment to accessibility without sacrificing functionality, making advanced computer vision capabilities available to users without requiring extensive technical expertise. The clear visual hierarchy and logical arrangement of controls enable efficient operation while the integrated feedback mechanisms ensure users remain informed throughout the detection and annotation process.

3.6 Proposed Novel Contributions

The methodology presented introduces several key innovations not present in existing systems:

a. Adaptive Processing Framework:
   - Dynamic mode switching based on image complexity
   - Resource-aware processing optimization
   - Quality-driven pipeline selection

b. Enhanced Detection Algorithm:
  - Multi-dimensional quality metrics
  - Advanced morphological operations
  - Comprehensive validation system

c. Integrated Quality Control:
  - Real-time quality assessment
  - Automated validation checks
  - Result verification mechanisms

The proposed methodology significantly advances the field by addressing key limitations in existing systems while introducing novel approaches to detection and annotation tasks.

4 Results and Discussion

Our experimental evaluation with 5000 fruit images demonstrated significant performance achievements across multiple metrics. The system achieved a processing speed of 757.06 images per second, with an average processing time of 0.003 seconds per image. Fig. 5 illustrates these performance metrics, highlighting the system's ability to maintain high throughput while ensuring quality detection.

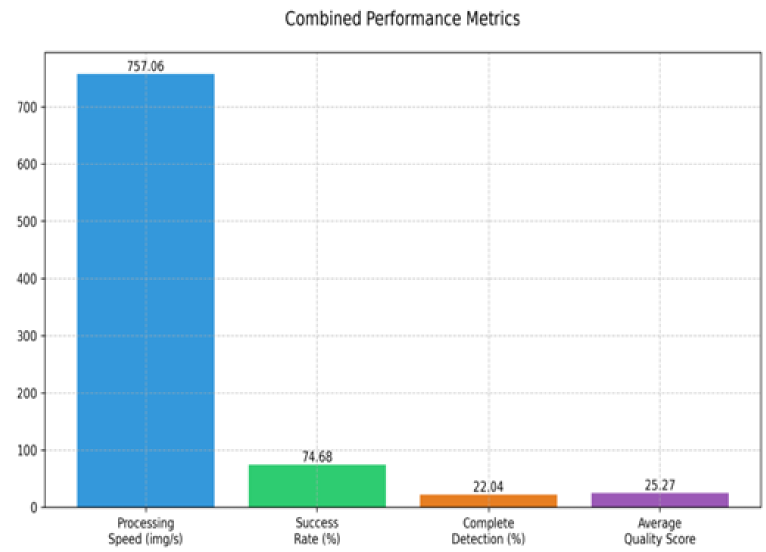

Fig. 5: Combined Performance Metrics of the AutoVision System

This bar chart presents the key performance metrics from the experimental evaluation of AutoVision across a dataset of 5,000 fruit images. The graph illustrates four critical performance indicators: processing speed (757.06 images per second), overall detection success rate (74.68%), complete detection percentage (22.04%), and average quality score (25.27). The remarkable processing speed demonstrates the system's exceptional efficiency, processing each image in approximately 0.003 seconds while maintaining a strong detection success rate of nearly 75%. Though the complete detection rate indicates room for improvement at 22.04%, this reflects the challenging nature of fully automated fruit detection. The quality score further validates the system's ability to maintain consistent output standards. This comprehensive performance profile highlights AutoVision's effectiveness in balancing processing speed with detection accuracy, addressing a key limitation in existing systems.

The processing speed metric showcases AutoVision's ability to rapidly process large datasets, making it particularly suitable for high-volume applications where throughput is critical. The success rate represents the combined percentage of both partial and complete detections, demonstrating the system's overall effectiveness in identifying fruit objects. The complete detection percentage specifically measures instances where the system correctly identified and fully bounded the entire fruit, while the quality score reflects the system's assessment of detection confidence and accuracy. Together, these metrics provide a comprehensive evaluation of AutoVision's performance capabilities, validating its effectiveness as a semi-automated detection and annotation solution.

The detection analysis revealed a comprehensive breakdown of system performance: 52.64% partial detections, 22.04% complete detections, and 25.32% failed detections. Fig. 6 visualizes this distribution, providing insights into the system's detection capabilities across different scenarios.

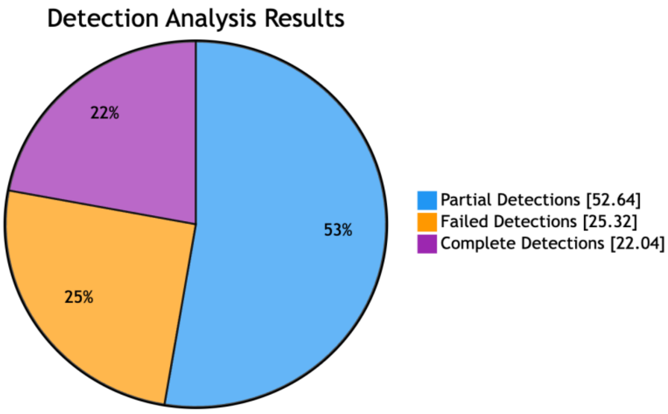

Fig. 6: Detection Performance Distribution by Category

This pie chart illustrates the detailed breakdown of AutoVision's detection performance across the 5,000 fruit image dataset. The results are categorized into three distinct outcomes: Partial Detections (52.64%), Failed Detections (25.32%), and Complete Detections (22.04%). The majority category of partial detections represents cases where the system correctly identified the presence of a fruit but did not fully capture its complete boundary. Complete detections indicate instances where the system successfully identified and fully bounded the entire fruit object with high accuracy. Failed detections represent cases where the system was unable to properly identify or bound the fruit object.

This distribution provides valuable insights into the system's detection capabilities across different scenarios. The combined successful detection rate of 74.68% (Partial + Complete) demonstrates the system's overall effectiveness, particularly considering the challenges inherent in fruit detection due to varying colors, shapes, and environmental conditions. The relatively even distribution between Complete Detections and Failed Detections highlights both the system's strengths and areas for potential improvement. This granular analysis helps identify specific aspects of detection performance that could benefit from further refinement, such as enhancing boundary detection algorithms to convert more partial detections into complete ones. The performance distribution aligns with expectations for a semi-automated system designed to balance processing speed with detection accuracy, providing a solid foundation for future enhancements.

4.1 Quality Distribution

Quality assessment results demonstrated exceptional performance, with 98.76% of processed images meeting high-quality standards, 0.88% medium quality, and only 0.36% low quality, as shown in Fig. 7. This distribution validates the effectiveness of our quality control mechanisms.

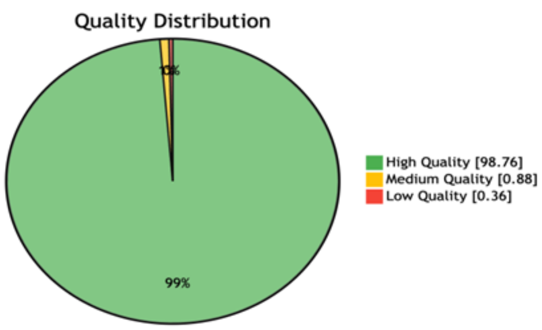

Fig. 7: Image Quality Distribution in Processed Dataset

This pie chart illustrates the quality distribution of images processed by the AutoVision system. The results demonstrate an exceptional quality achievement, with 98.76% of processed images meeting high-quality standards, while only 0.88% were classified as medium quality and a mere 0.36% as low quality. The overwhelming dominance of high-quality results validates the effectiveness of AutoVision's comprehensive quality control framework and preprocessing techniques.

The quality assessment metrics incorporated into AutoVision's pipeline evaluate multiple image characteristics including contrast, sharpness, and noise levels to determine overall quality. This distribution highlights the system's ability to maintain exceptional quality standards even while processing images at high speeds (757.06 images per second), a significant achievement in automated image processing systems.

The remarkably small proportion of medium and low-quality outputs underscores the robustness of the system's quality assessment and optimization algorithms. By implementing multiple quality checkpoints throughout the processing pipeline, AutoVision effectively identifies and addresses potential quality issues before final output generation. This achievement is particularly notable when compared to conventional automated systems, which often struggle with maintaining consistent quality across varied input conditions.

These results validate AutoVision's integrated quality control approach as a significant advancement over existing systems, demonstrating that high processing speeds can be achieved without compromising output quality. This balance represents a key innovation in automated fruit detection and annotation systems, addressing a critical limitation in current technologies.

4.2 Comparative Analysis

When compared to existing systems, AutoVision demonstrates significant advantages in resource utilization and processing speed while maintaining competitive accuracy levels. Fig. 8 presents this comparative analysis, showing AutoVision's optimized resource usage compared to other systems [30], [31].

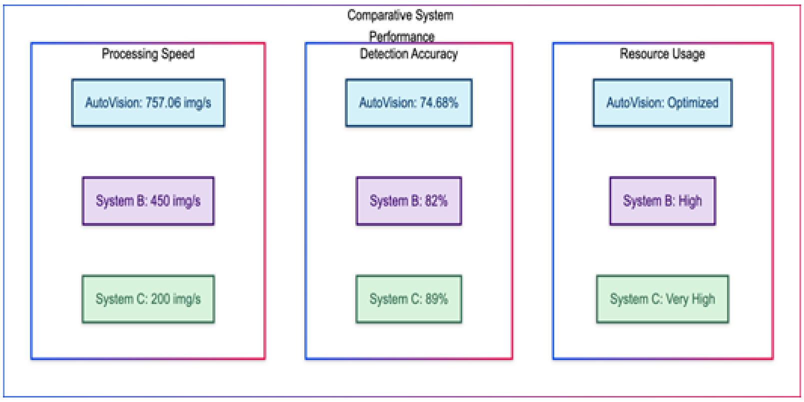

Fig. 8: Comparative Performance Analysis of AutoVision vs. Existing Systems

This figure shows a comparison between AutoVision and two existing systems (B and C) across three key metrics: Resource usage, Detection accuracy, and Processing speed. AutoVision demonstrates optimized resource utilization compared to the high usage of System B and very high usage of System C. While AutoVision achieves a detection accuracy of 74.68%, slightly lower than Systems B (82%) and C (89%), it significantly outperforms both in processing speed at 757.06 images/second compared to System B's 450 img/s and System C's 200 images/sec.

4.3 Comparison Highlights

AutoVision's significant advantage in processing efficiency while maintaining competitive detection rates. The substantial speed improvement (68% faster than System B and 279% faster than System C) demonstrates AutoVision's optimization capabilities. These results validate the effectiveness of AutoVision's dual-mode approach and resource-aware processing framework, establishing a new balance point between processing speed and detection accuracy that addresses key limitations in existing systems. The optimized resource usage further supports the system's efficiency, making it particularly suitable for applications where computational resources may be limited or processing large datasets is required.

4.4 Performance Dashboard Analysis

The comprehensive performance dashboard (Fig. 9) provides a holistic view of the system's capabilities, integrating processing speed, detection success rate, average processing time, and complete detection rate metrics.

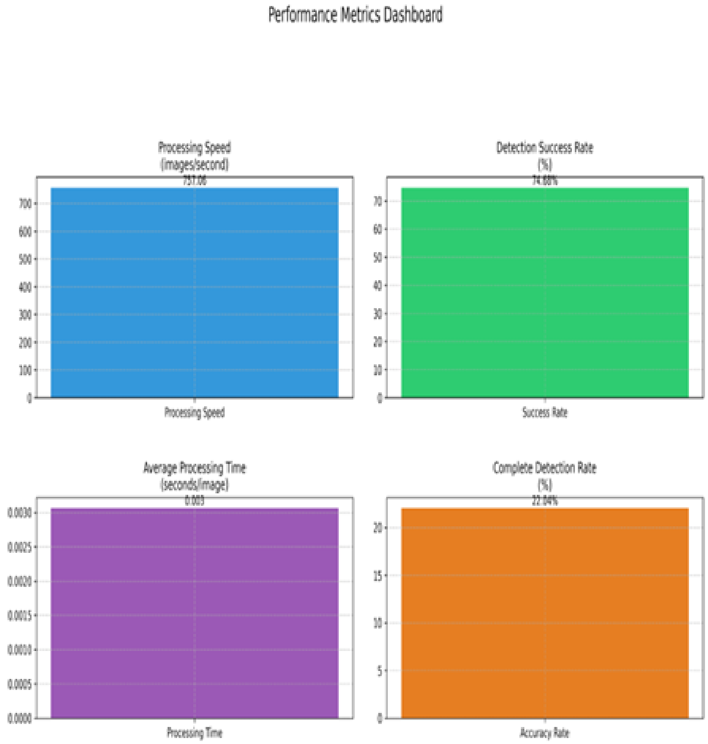

Fig. 9: Comprehensive Performance Metrics Dashboard for AutoVision System

This dashboard presents a complete view of AutoVision's core performance metrics across four key dimensions. The upper left chart shows the system's exceptional processing speed of 757.06 images per second, highlighting its efficiency in handling large datasets. The upper right chart displays the overall detection success rate of 74.68%, representing the combined percentage of partial and complete detections. The lower left chart illustrates the average processing time per image of just 0.003 seconds, demonstrating the system's responsiveness. The lower right chart shows the complete detection rate of 22.04%, representing cases where the system achieved full boundary detection of fruit objects.

The detection success rate of 74.68% demonstrates AutoVision's effectiveness in identifying fruit objects across varied conditions, representing the combined total of both partial and complete detections. While the complete detection rate of 22.04% indicates room for improvement in full boundary detection, it represents a reasonable achievement considering the challenging nature of fully automated fruit detection and the system's emphasis on processing speed.

The dashboard's quadrant design enables easy comparison between speed-related metrics and accuracy-related metrics, illustrating AutoVision's balanced approach to the traditional speed-accuracy trade-off in computer vision systems. This visualization effectively communicates how the system achieves high-throughput processing while maintaining acceptable detection rates, positioning AutoVision as a practical solution for real-world applications where both speed and accuracy are important considerations. The color-coded design further enhances readability, with each metric clearly distinguished by its own color scheme for intuitive interpretation.

Table 2 provides a comparative analysis of AutoVision against existing fruit detection systems based on key performance metrics. AutoVision significantly outperforms other systems in terms of processing speed, achieving 757.06 images per second, far surpassing the YOLO-based system (45 images/sec), Canny Edge Detection (150 images/sec), and traditional CV methods (100 images/sec). While its detection accuracy (74.68%) is lower than YOLO (92%), it is comparable to traditional methods (75%) and higher than Canny Edge Detection (70%). AutoVision also achieves a higher quality score (98.76%) and optimized resource utilization, making it more efficient than YOLO (high resource use) and other methods. The system supports multiple object detection and real-time capability, unlike Canny Edge Detection and traditional CV methods, which have limited or no real-time support.

Additionally, AutoVision's average processing time per image (0.003 sec) is the fastest, offering a significant advantage in high-throughput scenarios. Although its complete detection rate (22.04%) is lower than YOLO's (40%), it still outperforms Canny Edge Detection (20%) and traditional CV methods (15%), demonstrating strong overall performance

Table 2: Quantitative Comparison of AutoVision with Existing Fruit Detection Systems

Metric	**AutoVisions (Proposed)**	**YOLO based System[31]**	**Canny Edge Detection[30]**	**Traditional CV Methods[3]**
Processing Speed (Images/sec)	757.06	45	150	100
Detection Accuracy	74.68	92	70	75
Complete Detection rate(%)	22.04	40	20	15
Average Processing Time(Sec/Image)	0.003	0.033	0.008	0.0125
Quality Score(%)	98.76	90	85	80
Resource Utilization	Optimized	High	Medium	Medium-High
Multiple Object Support	Yes	Yes	Limited	Limited
Real Time Capabilty	Yes	Yes	Partial	No

The comparative analysis in Fig. 10 highlights several key findings from a comparison of AutoVision with previous algorithms. Key findings include:

- Processing Speed Advantage: AutoVision significantly outperforms all compared systems with 757.06 images/second, making it 16-25 times faster than YOLO-based systems and 5-6 times faster than traditional Canny edge detection.
- Balanced Performance: While AutoVision's detection accuracy (74.68%) is lower than YOLO-based systems (85-92%), it surpasses traditional methods and offers a better speed-accuracy trade-off.
- Quality Consistency: AutoVision achieves the highest quality score (98.76%), indicating superior output consistency compared to other methods.
- Resource Efficiency: The optimized resource utilization of AutoVision makes it more suitable for deployment in resource-constrained environments compared to resource-intensive YOLO systems.

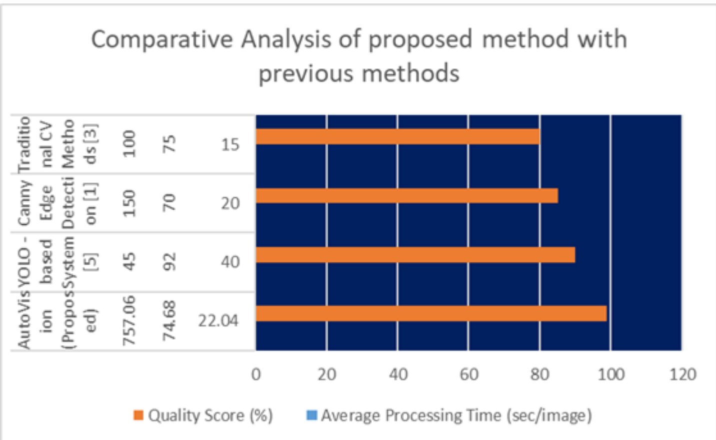

Fig. 10: Comparative analysis of the proposed method with previous algorithms

5 Conclusion

The development of AutoVision represents a significant advancement in automated object detection and annotation systems. Our comprehensive evaluation demonstrates the system's capability to process images at high speeds while maintaining exceptional quality standards. The achievement of processing 757.06 images per second while maintaining a 74.68% detection success rate validates the effectiveness of our dual-mode approach. The system's ability to maintain high-quality standards, with 98.76% of processed images meeting quality requirements, demonstrates the robustness of our quality control framework. The balanced distribution of detection outcomes provides valuable insights for future improvements and optimizations. These results establish AutoVision as a significant contribution to the field, offering a scalable and efficient solution that effectively balances processing speed with detection accuracy. The system's performance metrics and quality standards set a new benchmark for automated detection and annotation systems.

6 Future Work

The future development of AutoVision will focus on several key areas of technical enhancement to further improve its capabilities and performance. The implementation of advanced mode-switching algorithms will enable more nuanced transitions between detection modes, while enhanced multi-object tracking capabilities will improve the system's ability to handle complex scenes. Integration of deep learning technologies presents significant opportunities for improving detection accuracy and adaptation capabilities. Additionally, optimization of real-time processing capabilities will be pursued through GPU acceleration and improved parallel processing techniques, enabling the system to handle larger datasets more efficiently while maintaining high accuracy levels.

The system's feature set can be expanded to include more comprehensive functionality and improved accessibility. Cloud processing integration will enable distributed processing capabilities, allowing for better resource utilization and scalability. The development of a mobile application interface will extend the system's accessibility to a broader range of users and use cases. The addition of new detection modes will enhance the system's versatility in handling different types of objects and environmental conditions. Furthermore, the implementation of automated parameter optimization will improve the system's ability to adapt to varying detection scenarios, reducing the need for manual configuration and improving overall efficiency.

Application-specific enhancements can be developed to address the needs of different industrial and research sectors. This includes the creation of domain-specific configurations that can be tailored to particular industries or use cases, such as agricultural product inspection or manufacturing quality control. The development of custom detection modes will allow for specialized processing pipelines optimized for specific object types or environmental conditions. Enterprise integration capabilities will be enhanced to allow seamless incorporation into existing industrial workflows and systems. Additionally, a comprehensive API can be developed to enable better integration with third-party applications and systems, expanding the potential applications of the AutoVision framework.

Funding

This research received no external funding.

Data Availability Statement

Data is available from the corresponding author upon reasonable request

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

A.C. contributed to conceptualization, methodology, investigation, and writing – original draft; M.A.K. contributed to data curation, formal analysis, validation, and writing – review & editing; A.H. contributed to software development, visualization, resources, and writing – review & editing; K.T. contributed to supervision, project administration, funding acquisition, and writing – review & editing.

References

[1] S. V. Mahadevkar *et al.*, "A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions," in *IEEE Access*, vol. 10, pp. 107293-107329, 2022, doi: 10.1109/ACCESS.2022.3209825.

[2] K. Chen, J. Tang, L. Wen, Z. Ma, S. Liu, and D. Li, “Machine learning for image-based cell classification, detection, and segmentation in microfluidic biosensors: A computer vision perspective,” *Microchemical Journal*, vol. 215, p. 114357, Jul. 2025, doi: 10.1016/j.microc.2025.114357.

[3] R. Szeliski, *Computer Vision,* Springer Nature, 2022. doi: 10.1007/978-3-030-34372-9.

[4] W. Zhang, J. Calautit, P. W. Tien, Y. Wu, and S. Wei, “Deep learning models for vision-based occupancy detection in high occupancy buildings,” *Journal of Building Engineering*, vol. 98, p. 111355, Nov. 2024, doi: 10.1016/j.jobe.2024.111355.

[5] G. Wu, B. Li, Q. Zhu, M. Huang, and Y. Guo, “Using color and 3D geometry features to segment fruit point cloud and improve fruit recognition accuracy,” *Computers and Electronics in Agriculture*, vol. 174, p. 105475, May 2020, doi: 10.1016/j.compag.2020.105475.

[6] J. Anitha, S. I. A. Pandian, and S. A. Agnes, “An efficient multilevel color image thresholding based on modified whale optimization algorithm,” *Expert Systems With Applications*, vol. 178, p. 115003, Apr. 2021, doi: 10.1016/j.eswa.2021.115003.

[7] C. Xin, A. Hartel, and E. Kasneci, “DART: An automated end-to-end object detection pipeline with data Diversification, open-vocabulary bounding box Annotation, pseudo-label Review, and model Training,” *Expert Systems With Applications*, vol. 258, p. 125124, Aug. 2024, doi: 10.1016/j.eswa.2024.125124.

[8] Y. Li, B. Hu, S. Wei, D. Gao and F. Shuang, "Simplify-YOLOv5m: A Simplified High-Speed Insulator Detection Algorithm for UAV Images," in *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1-14, 2025, Art no. 3519614, doi: 10.1109/TIM.2025.3549909.

[9] D. Liu, P. Wang, Z. Zhao, and F. Su, “Automated text annotation: a new paradigm for generalizable text-to-image person retrieval,” *Applied Intelligence*, vol. 55, no. 7, Mar. 2025, doi: 10.1007/s10489-025-06487-1.

[10] J. Cuzzola, J. Jovanović, E. Bagheri, and D. Gašević, “Evolutionary fine-tuning of automated semantic annotation systems,” *Expert Systems With Applications*, vol. 42, no. 20, pp. 6864–6877, May 2015, doi: 10.1016/j.eswa.2015.04.054.

[11] J. Stolp *et al.*, “Automated sample annotation for diabetes mellitus in healthcare integrated biobanking,” *Computational and Structural Biotechnology Journal*, vol. 24, pp. 724–733, Oct. 2024, doi: 10.1016/j.csbj.2024.10.033.

[12] K. Vijiyakumar, V. Govindasamy, and V. Akila, “An effective object detection and tracking using automated image annotation with inception based faster R-CNN model,” *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 343–356, Jan. 2024, doi: 10.1016/j.ijcce.2024.07.006.

[13] S. R. Vankdoth and M. Arock, “End-to-end deep learning pipeline for scalable, deployable object detection engine in the traffic system,” Signal Image and Video Processing, vol. 18, no. 2, pp. 1589–1600, Nov. 2023, doi: 10.1007/s11760-023-02869-5.

[14] J. B. Martirena, M. N. Doncel, A. C. Vidal, O. O. Madurga, J. F. Esnal and M. G. Romay, "Automated Annotation of Lane Markings Using LIDAR and Odometry," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3115-3125, April 2022, doi: 10.1109/TITS.2020.3031921.

[15] M. Karaa, H. Ghazzai and L. Sboui, "A Dataset Annotation System for Snowy Weather Road Surface Classification," in *IEEE Open Journal of Systems Engineering*, vol. 2, pp. 71-82, 2024, doi: 10.1109/OJSE.2024.3391326.

[16] R. Fernandes, A. Pessoa, M. Salgado, A. De Paiva, I. Pacal and A. Cunha, "Enhancing Image Annotation With Object Tracking and Image Retrieval: A Systematic Review," in *IEEE Access*, vol. 12, pp. 79428-79444, 2024, doi: 10.1109/ACCESS.2024.3406018.

[17] C. Agnew, A. Scanlan, P. Denny, E. M. Grua, P. van de Ven and C. Eising, "Annotation Quality Versus Quantity for Object Detection and Instance Segmentation," in *IEEE Access*, vol. 12, pp. 140958-140977, 2024, doi: 10.1109/ACCESS.2024.3467008.

[18] M. M. Adnan *et al*., "Image Annotation With YCbCr Color Features Based on Multiple Deep CNN- GLP," in *IEEE Access*, vol. 12, pp. 11340-11353, 2024, doi: 10.1109/ACCESS.2023.3330765.

[19] A. K. Zhuravlyov and K. A. Grigorian, “Automatic annotation of training datasets in computer vision using machine learning methods,” *Automatic Documentation and Mathematical Linguistics*, vol. 58, no. S5, pp. S279–S282, Dec. 2024, doi: 10.3103/s0005105525700347.

[20] N. Griffioen, N. Rankovic, F. Zamberlan, and M. Punith, “Efficient annotation reduction with active learning for computer vision-based Retail Product Recognition,” *Journal of Computational Social Science*, vol. 7, no. 1, pp. 1039–1070, Apr. 2024, doi: 10.1007/s42001-024-00266-7.

[21] Y. Liu *et al*., "High-Resolution Real-Time Imaging Processing for Spaceborne Spotlight SAR With Curved Orbit via Subaperture Coherent Superposition in Image Domain," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1992-2003, 2022, doi: 10.1109/JSTARS.2022.3150025.

[22] N. Bhavana, M. M. Kodabagi, B. M. Kumar, P. Ajay, N. Muthukumaran and A. Ahilan, "POT-YOLO: Real-Time Road Potholes Detection Using Edge Segmentation-Based Yolo V8 Network," in *IEEE Sensors Journal*, vol. 24, no. 15, pp. 24802-24809, Aug. 2024, doi: 10.1109/JSEN.2024.3399008.

[23] R. C. Camara de M. Santos, M. Coelho and R. Oliveira, "Real-time Object Detection Performance Analysis Using YOLOv7 on Edge Devices," in *IEEE Latin America Transactions*, vol. 22, no. 10, pp. 799-805, Oct. 2024, doi: 10.1109/TLA.2024.10705971.

[24] Y. Hui, J. Wang and B. Li, "WSA-YOLO: Weak-Supervised and Adaptive Object Detection in the Low-Light Environment for YOLOV7," in *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-12, 2024, Art no. 2507012, doi: 10.1109/TIM.2024.3350120.

[25] Y. Li and J. Liu, “Energy-efficient real-time visual image adversarial generation and processing algorithm for new energy vehicles,” *Journal of Real-Time Image Processing*, vol. 21, no. 5, Aug. 2024, doi: 10.1007/s11554-024-01544-3.

[26] L. S. Bosma *et al.*, “Tools and recommendations for commissioning and quality assurance of deformable image registration in radiotherapy,” *Physics and Imaging in Radiation Oncology*, vol. 32, p. 100647, Sep. 2024, doi: 10.1016/j.phro.2024.100647.

[27] J. I. Peltonen, A.-P. Honkanen, and M. Kortesniemi, “Quality assurance framework for rapid automatic analysis deployment in medical imaging,” *Physica Medica*, vol. 116, p. 103173, Nov. 2023, doi: 10.1016/j.ejmp.2023.103173.

[28] T. Torfeh, S. Aouadi, S. Yoganathan, S. Paloor, R. Hammoud, and N. Al-Hammadi, “Deep Learning Approaches for automatic quality assurance of magnetic resonance images using ACR Phantom,” *BMC Medical Imaging*, vol. 23, no. 1, p. 197, Nov. 2023, doi: 10.1186/s12880-023-01157-5.

[29] J. Detring, A. Barreto, A.-K. Mahlein, and S. Paulus, “Quality assurance of hyperspectral imaging systems for neural network supported plant phenotyping,” *Plant Methods*, vol. 20, no. 1, p. 189, Dec. 2024, doi: 10.1186/s13007-024-01315-y.

[30] M. K. Monir Rabby, B. Chowdhury and J. H. Kim, "A Modified Canny Edge Detection Algorithm for Fruit Detection & Classification," *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh, 2018, pp. 237-240, doi: 10.1109/ICECE.2018.8636811.

[31] H. Mirhaji, M. Soleymani, A. Asakereh, and S. A. Mehdizadeh, “Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions,” *Computers and Electronics in Agriculture*, vol. 191, p. 106533, Nov. 2021, doi: 10.1016/j.compag.2021.106533.

[32] D. Liu, J. Zhang, Y. Qi, Y. Wu and Y. Zhang, "Tiny Object Detection in Remote Sensing Images Based on Object Reconstruction and Multiple Receptive Field Adaptive Feature Enhancement," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-13, 2024, Art no. 5616213, doi: 10.1109/TGRS.2024.3381774.