

Experimental Investigation of Computer Vision Architectures for Object Detection in Mobile Robotics

Perizat Rakhmetova
*Dept. of Robotics and technical means
of automation
Satbayev University
Almaty, Kazakhstan
p.rakhmetova@gmail.com*

Aldabergen Bektilevov
*School of transport engineering and
logistics named after M. Tynyshpayev
Satbayev University
Almaty, Kazakhstan
a.bektilevov@satbayev.university*

Lazzat Kurmangalieva
*Dept. of Robotics and technical means
of automation
Satbayev University
Almaty, Kazakhstan
lezzet@mail.ru*

Baglan Bekbossynova
*Dept. of Mechanical Engineering
Satbayev University
Almaty, Kazakhstan
0000-0002-6642-3431*

Beibit Shingissov
*Dept. of Mechanical Engineering
Satbayev University
Almaty, Kazakhstan
b.shingissov@satpayev.university*

Abstract— In recent years, computer vision has emerged as a key technology for autonomous mobile robots and transportation systems, enabling their ability to perceive the environment and make real-time decisions. One of the most critical and widely demanded tasks is object detection, which is essential for navigation, collision avoidance, and the execution of mission-specific operations. This paper examines experimental studies of computer vision algorithms used for object detection in mobile robots. The focus is on the analysis and comparison of modern methods, including classical image processing algorithms and deep learning-based neural network approaches. The experimental study was conducted in a pipeline system to detect in-pipe defects, which requires high accuracy and robustness of algorithms given the limited computing resources of mobile platforms. Experiments were conducted using various object detector architectures (YOLO, SSD, Faster R-CNN) and lightweight models, such as MobileNet, designed for resource-constrained environments. Algorithm performance was evaluated based on accuracy and processing speed. The results of the study enable the identification of optimal solutions for integrating computer vision systems into mobile robots, ensuring a balance between detection quality and operational efficiency. Furthermore, the study explores the potential adaptation of these architectures for specialized scenarios, such as pipeline inspection, where detecting complex-shaped defects within confined spaces is required.

Keywords— computer vision, object detection, mobile robotics, transportation systems, neural network, deep learning

I. INTRODUCTION

Modern mobile robots and autonomous transportation systems play a crucial role in industry, logistics, energy, and infrastructure projects. Their effectiveness directly depends on the ability to perceive the environment and make real-time decisions. Computer vision, particularly object detection methods, provides critical functionalities such as navigation, collision avoidance, object manipulation, and equipment condition monitoring. However, the integration of these technologies into mobile and transportation platforms faces several constraints: limited computational power, restricted memory capacity, and the requirement for real-time operation. These factors necessitate the optimization of

neural network architectures and the search for a balance between accuracy, speed, and energy efficiency.

Object detection is a critical function in mobile robots that allows for autonomous navigation, obstacle avoidance, and interaction with dynamic surroundings while adhering to rigorous real-time and energy restrictions [1, 2]. Early research revealed the viability of convolutional neural networks (CNNs) for navigation and obstacle detection, demonstrating higher perceptual accuracy than traditional approaches while emphasizing trade-offs between accuracy, speed, and computational cost [3]. Subsequent research extended these capabilities to more sophisticated tasks, such as automated object manipulation in construction, where vision-based systems reached sub-centimeter precision using only onboard sensors [4]. These findings highlight the critical relevance of strong perception pipelines for mobile robots [5] operating in unstructured situations.

Recent advances in deep learning have resulted in high-performance object detection architectures such as YOLO, SSD, and Faster R-CNN, which achieve cutting-edge accuracy across domains [6, 7]. Single-stage detectors, such as YOLO and SSD, are preferred for real-time applications, but two-stage versions provide improved accuracy at the cost of computational efficiency [8]. To address resource constraints on embedded platforms, lightweight backbones like MobileNet and compact detection heads (e.g., SSDLite, YOLO-Tiny) have been created, resulting in advantageous trade-offs between accuracy and performance [8, 9]. Complementary tactics, such as multi-scale feature fusion and augmentation approaches initially developed for UAV-based detection, improve robustness in demanding environments [10].

Innovation at the system level is equally important. Edge computing frameworks, like Vivid, show how outsourcing large calculations to neighboring nodes can reduce latency and enhance navigation accuracy while maintaining privacy [11]. The integration of current detectors with control systems has been tested in outdoor mobility, with YOLOv8-based tracking enabling real-time steering of autonomous all-terrain vehicles [12]. Additional perception tasks, such as semantic segmentation, have been integrated with detection

to improve free-space reasoning [13], whilst reinforcement learning techniques use detection outputs for object-aware navigation and trajectory planning [14, 15]. These results underscore the importance of holistically evaluating detection models in closed-loop robotic systems.

Specialized applications such as pipeline inspection impose even stricter requirements. In-pipe robots must detect fine-grained defects—cracks, corrosion, weld anomalies—under poor lighting, confined spaces, and reflective surfaces. Recent studies report AI-powered inspection platforms integrating advanced locomotion and multimodal sensing (RGB, depth, thermal), achieving detection accuracies exceeding 90% in real time. Similar principles apply in forestry and agricultural robotics, where lightweight detectors like SSD-MobileNet and YOLO-Tiny deliver high AP and F1 scores at millisecond-scale inference times [9]. Comprehensive reviews emphasize that robust deployment depends on combining efficient architectures, domain-specific preprocessing, and system-level optimizations such as quantization and edge offloading [8, 10].

II. PROBLEM STATEMENT

Among the diverse applications of object detection in mobile robotics, one of the most complex and safety-critical tasks is the detection of defects inside pipeline transport systems used for the transportation of oil, gas, water, and other fluids. In-pipe defects—such as cracks, corrosion, deposits, and weld anomalies—pose significant risks to the safety and reliability of transport infrastructure. The use of mobile robots equipped with computer vision systems enables automated inspection, reducing operational costs and improving diagnostic accuracy. However, the harsh conditions inside pipelines—restricted space, low illumination, presence of moisture, contamination, and reflective surfaces—create additional challenges for object detection algorithms. These factors necessitate experimental research aimed at evaluating the performance of state-of-the-art and lightweight detection models under the specific conditions of pipeline transport systems. The objective of this study is to conduct an experimental investigation into the applicability of modern computer vision architectures for object detection in mobile robotics and transportation systems, followed by their adaptation and evaluation for in-pipe defect detection. This approach aims to assess the feasibility of transferring technologies from general-purpose object detection to specialized industrial scenarios.

III. MATERIALS AND METHODS

The experimental framework was designed to evaluate object detection architectures for mobile robotics, with a focus on lightweight models optimized for real-time performance in resource-constrained environments. Special attention was given to MobileNet-based architectures as shown in (Fig. 1) due to their proven efficiency and adaptability for embedded systems [15].

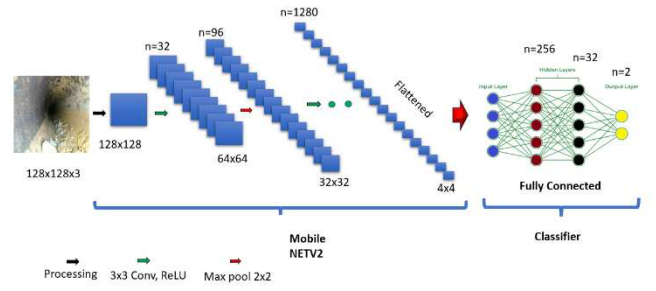


Fig. 1. Computer vision model architecture (MobileNET).

As shown in (Fig. 2), MobileNetV2—with approximately 3.4 million parameters—offers a significant computational advantage compared to heavier models, making it highly suitable for interactive and real-time applications on mobile or embedded devices.

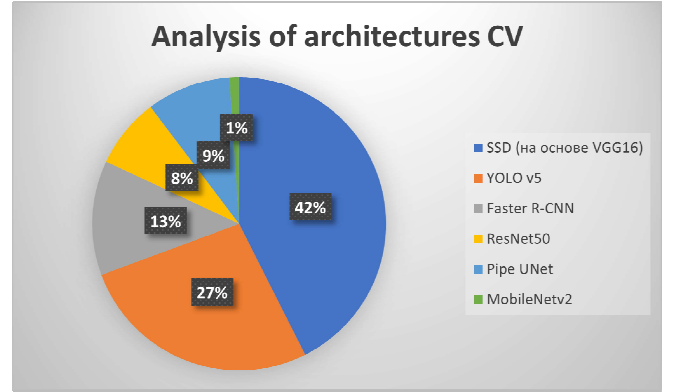


Fig. 2. Architecture of convolutional neural network in percentage by number of parameters.

Hardware Methods. The trained models were implemented on a mobile robotic platform (Fig. 3) equipped with an NVIDIA Jetson Orin development board, specifically designed for optical inspection tasks inside pipelines [5]. The system integrates an Intel RealSense RGB-D camera for image acquisition. The operational pipeline consists of continuous video capture from the internal surface of pipes, real-time frame processing, and defect localization. Detected anomalies, such as corrosion or cracks, are filtered and either stored in a database or displayed to the operator in real time.

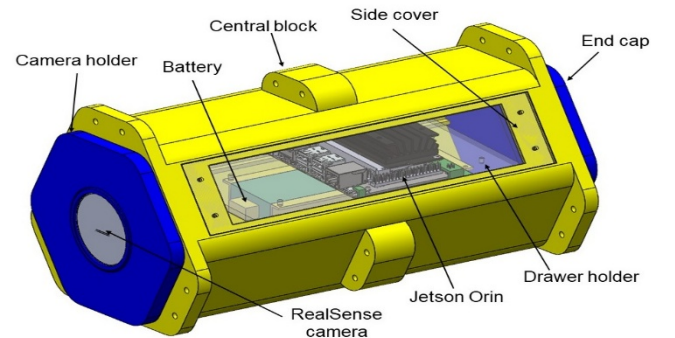


Fig. 3. An onboard system for a mobile robot equipped with a Jetson Orin microcontroller for software control of the robot.

Model Selection and Optimization. Among the evaluated architectures—YOLOv5/YOLOv8, SSD, and Faster R-CNN—MobileNet-based SSD was prioritized for deployment due to its balance of accuracy and computational

efficiency. The model was trained using TensorFlow Keras, optimized through quantization and TensorRT conversion, and deployed on Jetson Orin for real-time inference. This optimization pipeline ensures minimal latency while maintaining detection accuracy, which is critical for continuous inspection tasks.

Data Acquisition and Preprocessing. The dataset included RGB and depth images captured from pipeline interiors, annotated for defects such as cracks, corrosion, and weld anomalies, following methodologies in [3, 5]. Additional data from general mobile robotics tasks were used for benchmarking. Preprocessing involved normalization, contrast enhancement, and synthetic augmentation (illumination variation, noise injection) to simulate real-world conditions such as low lighting and reflective surfaces.

IV. RESULTS

The developed system demonstrated robust performance in detecting and localizing pipeline anomalies under real-world conditions. The system remains inactive when no anomalies are present, ensuring efficient use of computational resources. This design choice was implemented by activating the MobileNet-based detection stage only after the preliminary anomaly detection module signals a potential defect. Such a two-tier approach significantly reduces unnecessary processing and improves overall responsiveness.

The initial anomaly screening relied on a combination of Canny edge detection and DBSCAN clustering, which effectively identified suspicious regions in real time. As illustrated in (Fig. 4), the system successfully detected and clustered various pipeline anomalies using RGB visualization. These algorithms enabled multiple stages of analysis, including the identification of candidate regions, bounding-box annotation of anomalies, and detection of surface changes such as corrosion or deposits. The RGB-based visualization adapted well to typical pipeline lighting conditions, reliably highlighting high-contrast debris and corrosion defects. For example,

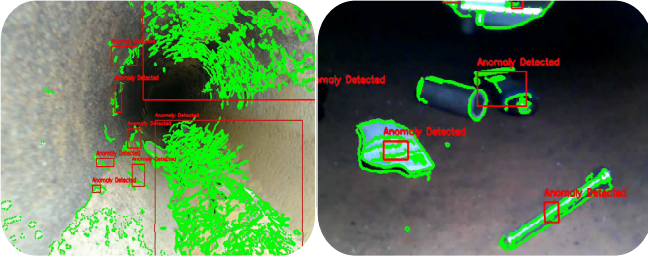


Fig. 4. Detection of various anomalies and defects inside the pipe.

Fig. 4 (right) shows several objects localized on the pipe floor using combined color and texture analysis, while (Fig. 4 left) highlights corrosion areas marked in green, providing clear visual cues for structural assessment.

To evaluate real-time suitability, inference times were measured for each stage (Table I). The edge-based pre-filtering achieved an average processing time of 12 ms per frame, while MobileNet-based SSD required 48 ms per frame after TensorRT optimization. This hybrid approach ensures that MobileNet is only activated when necessary, maintaining overall throughput at 20 FPS, which meets real-time requirements for pipeline inspection.

TABLE I. COMPARATIVE TIMING ANALYSIS OF DETECTION STAGES.

Stage	Average time per frame	Frames per second
Canny Edge+DBSCAN	12 ms	83
MobileNet	48 ms	20

The MobileNet-based detector provided accurate classification and localization of confirmed anomalies during the secondary detection stage. Although its inference time was higher than that of the edge-based pre-filtering approach, MobileNet delivered superior precision in distinguishing defect types, including fine cracks and weld anomalies, which are critical for predictive maintenance. The combined pipeline thus achieved a practical compromise: rapid anomaly screening followed by high-confidence classification, enabling real-time decision-making for maintenance operations. Furthermore, the modular design allows adaptation to different pipe diameters, materials, and lighting conditions, and can be extended with additional sensor modalities (e.g., depth or infrared) to further enhance detection robustness. This adaptability ensures that the proposed architecture remains scalable and effective across diverse operational environments, reinforcing its suitability for embedded systems in mobile robotics.

V. DISCUSSION

The experimental results confirm that the proposed two-stage detection pipeline—combining Canny edge detection with DBSCAN clustering for anomaly pre-filtering and MobileNet for secondary classification—offers a practical balance between computational efficiency and detection accuracy. The pre-filtering stage significantly reduces the computational burden by activating the MobileNet model only when anomalies are detected, as illustrated in Fig.4. This design choice is particularly advantageous for embedded platforms such as Jetson Orin, where energy efficiency and real-time responsiveness are critical.

The RGB-based visualization approach demonstrated strong adaptability to typical pipeline lighting conditions, reliably identifying corrosion, deposits, and structural irregularities in real time (Fig. 3). The ability to localize multiple anomalies and highlight surface changes with clear visual cues enhances the interpretability of the system for operators. Furthermore, the modular nature of the pipeline allows parameter tuning for different pipe diameters, materials, and environmental conditions, while the integration of additional sensor modalities (e.g., depth or infrared) could further improve robustness in challenging scenarios.

Comparative analysis revealed that while MobileNet delivers superior classification accuracy for fine-grained defects, its inference time is higher than that of the edge-based pre-filtering stage. However, by limiting MobileNet activation to confirmed anomaly cases, the system achieves real-time performance without sacrificing detection reliability. This hybrid approach aligns with findings in prior research emphasizing the importance of lightweight architectures and system-level optimizations for mobile robotics [7, 8, 9].

VI. CONCLUSION

This study demonstrates that combining classical computer vision techniques with deep learning models can effectively address the dual challenges of accuracy and efficiency in mobile robotic inspection tasks. The implementation on Jetson Orin confirms the feasibility of deploying advanced vision systems in real-world industrial environments, enabling real-time decision-making and reducing operational costs.

Future work will focus on expanding the dataset to include a wider range of defect types and environmental conditions, integrating multimodal sensing (depth and thermal imaging), and exploring advanced optimization techniques such as pruning and quantization-aware training to further enhance performance. Additionally, incorporating uncertainty estimation and predictive maintenance analytics could transform the system into a comprehensive diagnostic tool for critical infrastructure monitoring.

FUNDING INFORMATION

This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP22685781 Development of a manipulator control system using computer vision to enhance flexibility and automate processes)

REFERENCES

- [1] Raj R, Kos A. A Comprehensive Study of Mobile Robot: History, Developments, Applications, and Future Research Perspectives, 2022.
- [2] Q. Bai, S. Li, J. Yang, Q. Song, Z. Li and X. Zhang, "Object Detection Recognition and Robot Grasping Based on Machine Learning: A Survey," pp. 181855-181879, 2020.
- [3] Verbitsky, Nikita S., Eugene V. Chepin, and Alexander A. Gridnev. "Experimental studies of a convolutional neural network for application in the navigation system of a mobile robot.", pp. 611-616, 2018.
- [4] Asadi, K., Haritsa, V. R., Han, K., & Ore, J. P. Automated object manipulation using vision-based mobile robotic system for construction applications. 2021.
- [5] Kenzhekhan, A., Bakytzhanova, A., Omirbayev, S., Tuieubayev, Y., Daniyalov, M., Yeshmukhametov, A. Design and Development of an In-Pipe Mobile Robot for Pipeline Inspection with AI Defect Detection System. pp. 579–584, 2023.
- [6] S. Bogdan, U. Alisher and M. Togmanov, "Development of System using Computer Vision to Detect Personal Protective Equipment Violations," pp. 1-4, 2024.
- [7] Terven J, Córdova-Esparza D-M, Romero-González J-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. 2023.
- [8] Manakitsa N, Maraslidis GS, Moysis L, Fragulis GF. A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision. 2024.
- [9] da Silva DQ, dos Santos FN, Sousa AJ, Filipe V. Visible and Thermal Image-Based Trunk Detection with Deep Learning for Forestry Mobile Robotics. 2021.
- [10] Cazzato D, Cimorelli C, Sanchez-Lopez JL, Voos H, Leo M. A Survey of Computer Vision Methods for 2D Object Detection from Unmanned Aerial Vehicles. 2020.
- [11] W. Zhao, L. Xu, B. Qi, J. Hu, T. Wang and T. Runge, "Vivid: Augmenting Vision-Based Indoor Navigation System With Edge Computing," pp. 42909-42923, 2020.
- [12] Dang T-V, Bui N-T. Multi-Scale Fully Convolutional Network-Based Semantic Segmentation for Mobile Robot Navigation. Electronics. 2023.
- [13] A. Staroverov, D. A. Yudin, I. Belkin, V. Adeshkin, Y. K. Solomentsev and A. I. Panov, "Real-Time Object Navigation With Deep Neural Networks and Hierarchical Reinforcement Learning," in pp. 195608-195621, 2020.
- [14] R. Chai, H. Niu, J. Carrasco, F. Arvin, H. Yin and B. Lennox, "Design and Experimental Validation of Deep Reinforcement Learning-Based Fast Trajectory Planning and Control for Mobile Robot in Unknown Environment," pp. 5778-5792, 2024.
- [15] Rakhmetova, P., Sergazin, G., Altay, Y., Dauletiya, D., & Kurmangaliyeva, L. (2025). Development of in-pipe defects detection and classification system. Eastern-European, pp. 80–89, 2025.