Walking Assistant Robot for Blind People

Mir Fahad Abdullah
2121572642
North South University
Dhaka, Bangladesh
fahad.abdullah@northsouth.edu

Sairat Jamin Shefa
2122172042
North South University
Dhaka, Bangladesh
shirat.shefa@northsouth.edu

Most Shirajum Munira Ahmad
2022354042
North South University
Dhaka, Bangladesh
shirajum.munira@northsouth.edu

Abstract— This paper presents the design and development of a low-cost, autonomous walking assistant robot intended to aid visually impaired individuals in navigating complex and resource-constrained environments, particularly in Bangladesh To provide dependable obstacle detection and responsive navigation, the system combines real-time sensor data fusion, natural language processing for Bangla voice assistance, and lightweight object recognition using Deep Learning. The robot uses effective voice synthesis models that are adapted for Bangladesh to improve user communication. Stability and terrain adaptation, including stair traversal, are made possible via a rocker-bogie suspension system. By providing a locally adaptive, computationally efficient, and linguistically inclusive mobility aid, the suggested approach overcomes significant drawbacks of traditional assistive technologies, including their high cost, English-only interaction, and not being fully autonomous. The robot is capable of providing precise, real-time guidance that is appropriate for daily use in Bangladeshi cities.

Keywords— Assistive robotics, object detection, Bangla NLP, MobileNet-SSD, Tiny-YOLOv4, EfficientDet-Lite, Jetson Orin Nano, visual impairment, voice navigation, real-time system.

I. Introduction

According to the 2020 Nationwide Blindness Survey in Bangladesh, about 1% of people over the age of 30 are visually impaired. For many of them, navigating daily life without assistance can be difficult and, at times, even dangerous. While canes and other traditional mobility aids are useful but limited because they are unable to detect impediments above ground, such as tree branches or hanging signs. The majority of individuals in Bangladesh cannot afford guide dogs, despite the fact that they provide a superior alternative.

The reality on the ground in Bangladesh is quite harsh. Overcrowded sidewalks, unpredictable road conditions, and noisy environments make it extremely challenging for visually impaired individuals to move around safely. Although artificial intelligence (AI) and natural language processing (NLP) technologies have demonstrated promise in helping blind people around the world, the majority of these solutions depend on expensive hardware or reliable internet access, neither of which are always available or inexpensive in our environment.

AI-powered robotics and assistance navigation have a strong foundation thanks to earlier research. To enable autonomous service robots, for example, Ekvall et al. showed how object detection and mapping with SLAM may work in tandem [4]. Their technology, however, was tested in controlled indoor circumstances and relied on the availability of a robust robotic platform with pan-tilt-zoom cameras and laser scanners—hardware that is far too costly and sophisticated for everyday usage in low-resource contexts. Similarly, while the work of Saini and Joseph [5]. discusses the role of NLP in robotics and even touches on emotional intelligence and sentiment analysis, it remains largely conceptual and lacks practical, deployable implementations in real-world assistive scenarios.

Furthermore, these studies usually focus on English-based interactions while ignoring the linguistic and cultural contexts of non-Western countries. This is a big problem in places like Bangladesh, where a lot of people feel more comfortable speaking Bangla. If language localization is not used, even the most advanced assistive technology could end up being ineffective for the very people it is meant to help.

We provide a practical and cost-effective solution to fill these gaps: a completely autonomous walking assistant robot designed for Bangladesh. Compared to current systems, our robot can navigate in Bangla by speech and is built to work offline. It uses computationally efficient, lightweight object detection techniques that work well with embedded systems, making it affordable and widely available. Our ultimate objective is to provide visually impaired people with a tool that respects their surroundings while simultaneously improving their mobility.

II. Literature Review

A key component of robotic assistance systems for the blind and visually impaired is object detection. Traditional CNN-based techniques, such as R-CNN, Faster R-CNN, YOLO, and Mask R-CNN, are still in use because they can provide real-time performance with minimal accuracy loss. Particularly noteworthy for striking the appropriate balance between speed and accuracy, YOLO is perfect for mobile robotics [1]. Notwithstanding these advantages, there are still several obstacles to overcome for real-world deployment in settings like Bangladeshi city streets, including occlusions, conflicting perspectives, dim illumination, and congested scenes [1].

Building on this, Ekvall et al. suggested a more comprehensive framework for robotic perception by combining Simultaneous Localization and Mapping (SLAM) with object identification to produce semantically rich maps of indoor spaces [4]. In order to facilitate autonomous navigation and job planning, their system employed Receptive Field Co-occurrence Histograms (RFCHs) to recognize objects and combine those detections with spatial data. With the use of this method, robots may not only find themselves but also recall the location of particular objects. For example, they can recall the location of the kitchen by looking for typical objects like cups or rice bags. Human-robot interaction is greatly enhanced and more sophisticated autonomy is supported by the synergy between geometric mapping and semantic comprehension.

In the meantime, how robots and humans communicate is also changing as a result of the convergence of computer vision and natural language processing (NLP). The integration of CV and NLP enables robots to not only "see" but also "describe" and "explain" what they see, as noted by Wiriyathammabhum et al [2]. This is crucial for applications such as visual question answering and scene captioning. One of the main factors facilitating more natural human-robot interactions is the capacity to include contextual semantics, such as object connections and qualities, into robotic systems.In the meantime, how robots and humans communicate is also changing as a result of the convergence of computer vision and natural language processing (NLP). The integration of CV and NLP enables robots to not only "see" but also "describe" and "explain" what they see, as noted by Wiriyathammabhum et al [2]. This is crucial for applications such as visual question answering and scene captioning. One of the main factors facilitating more natural human-robot interactions is the capacity to include contextual semantics, such as object connections and qualities, into robotic systems.

Extending this notion, Saini and Joseph [5] examined how NLP can be leveraged in robotic systems for emotional and contextual understanding. Their work highlights the potential of NLP techniques such as sentiment analysis, stemming, and TF-IDF in training robots to interpret human speech beyond basic commands. For visually impaired users, this means a robot could not only guide them through a hallway but also respond empathetically to voice tone and phrasing. The authors emphasize that machine learning allows robots to learn from user interactions, reducing the need for continuous reprogramming and improving adaptability over time.

Mazzei et al.'s bibliometric assessment of NLP's function in assistive systems in the social robotics area revealed a discernible shift toward "soft" human-robot interaction, where language and emotional engagement are valued more highly than physical reactions [3]. Their results support the worry that existing systems are less usable in linguistically diverse areas because they are primarily designed for English-speaking users. The use and usefulness of such assistive technologies are restricted, especially in Bangladesh, by the absence of support for Bangla-language natural language processing.

All things considered, even if there has been a lot of advancement, there is still a long way to go before these technologies are accessible, flexible, and useful in non-Western, low-resource settings. In order to effectively meet the demands of visually impaired users in Bangladesh and other comparable contexts, a low-cost, autonomous robotic assistant that combines lightweight object identification with Bangla voice command navigation has been proposed.

III. Methodology

Walking Assistant Robot for Visually Impaired Users tackles the issue of learning with a modular and iterative methodology toward unifying deep learning, natural language processing, and embedded systems. The principal components of the methodology are hardware configuration, object detection, Bangla voice navigation, system calibration, and finally integration with all ends focused toward keeping real-time capability on low-power hardware.

IV. Hardware Requirements

The hardware configuration of the Walking Assistant Robot is specifically designed to be cost-effective, light in weight, and having real-time obstacle detection and navigation features. The system makes it portable and energy-efficient so that it is simple to operate for visually impaired individuals. The system includes the essential components of environment awareness sensors, a processing unit for performing object detection algorithms, and an auditory output system for the delivery of voice instruction. This compact yet effective design renders the robot operational in various environments without being expensive and inaccessible.

The platform includes the following major components:

- Nvidia Jetson Orin Nano (4GB) – Handles both AI processing and motor/sensor control
- USB Camera (Logitech C270)
- VL53L1X ToF Sensors (2x or more)
- HC-SR04 Ultrasonic Sensors (2x)
- MPU6050 IMU Sensor
- L298N or TB6612FNG Motor Driver
- DC Gear Motors with Wheels (4x or 6x)
- USB Speaker or I2S Audio Module (MAX98357A)
- Power Bank (5V, 3A or higher)
- Li-ion Battery Pack
- Buck Converter (MP2307)

A. Robot Architecture

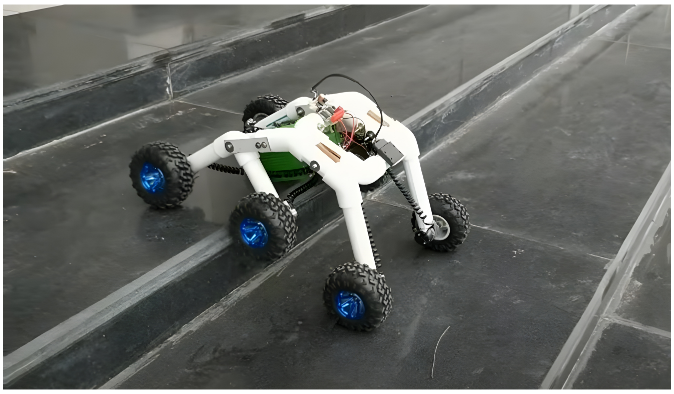

Figure I: Walking Assistant Robot's Design

The rocker-bogie suspension system, which was inspired by the Mars Rover, enables each wheel to independently articulate and adapt to the terrain's shape, enabling the robot to climb stairs. When working with irregular surfaces, such as stair steps, this configuration helps preserve optimum ground contact and stability. For adequate traction on risers, which are usually angled between 30° and 35°, the robot uses high-torque DC motors in conjunction with big, deep-tread rubberized wheels. The lightweight, long-lasting PVC used to make the chassis is inexpensive, easy to fabricate, and provides good structural strength for academic prototyping. By placing the hardware and power supply close to the base, the low center of gravity enhances balance and lessens the possibility of tipping.

B. System Diagram

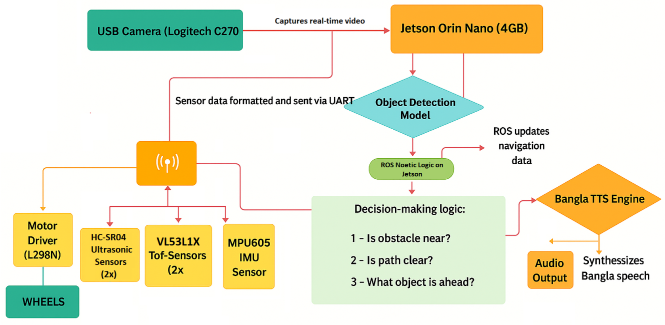

Figure II: System Diagram

The Walking Assistant Robot assists blind people to walk independently with a Jetson Orin Nano (4GB) module. The Jetson Orin Nano executes high-level AI operations (object detection using a USB camera) and also manages low-level motor control and sensor data acquisition simultaneously. It receives obstacle and orientation data from MPU6050, VL53L1X, and ultrasonic sensors and processes it locally.

DC gear motors are controlled through motor driver modules, and Bangla voice instructions are generated using onboard text-to-speech synthesis. Powered through a Li-ion battery and buck converter, the integrated setup presents a small-sized, affordable voice-guided navigation system optimized for visually impaired individuals in resource-poor settings.

V. DL in Action (Object Detection Models)

Object detection is a fundamental computer vision and image processing capability that enables detection and localization of objects such as people, obstacles, or vehicles in digital images and video streams. In our project, the Walking Assistant Robot for Blind People, object detection is a very critical component by enabling the system to perceive the world and detect potential obstacles in real time, hence enabling safe and effective movement.

There are several Deep Learning models that are optimized and runs well on low-powered processing units. Some of them include:

- **Tiny-YOLOv3 / Tiny-YOLOv4:**

  Tiny-YOLOv4 (or v3) provides a great trade-off between speed and accuracy, enabling real-time detection on Jetson Orin Nano (4GB) without using external accelerators. It enables support for several object classes and can be used with Tensorflow-Lite, TensorRT or converted into ONNX for GPU inference at high efficiency.

- **MobileNet-SSD:**

  Light and agile, MobileNet-SSD performs well on Jetson Orin Nano with low GPU usage and attains decent accuracy without consuming much memory relative to the full-precision YOLO models.

- **EfficientDet-Lite:**

  Optimized for edge devices like Jetson Orin Nano, EfficientDet-Lite0 and Lite1 are low-latency inference with reasonable accuracy for detecting small objects. TensorRT-acceleration and model quantization can be used with them to improve their performance.

A. Schematic Diagram

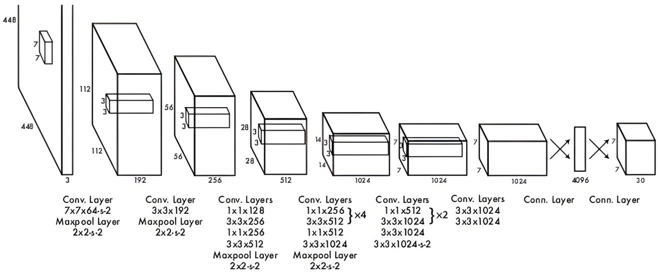

Figure III: Schematic Diagram (Tiny YOLOv4)

B. Model Architecture (YOLOv4)

- **Input Layer:** Accepts the input image and normalized to be of uniform size and proper pixel value ranges for processing.
- **13 Convolutional Layers:** Use Conv2D operations to extract visual features, with Batch Normalization to prevent training instability and LeakyReLU to maintain gradient flow for negative values. These layers learn incrementally from simple edges and textures to complex object structures.
- **6 MaxPooling Layers:** Do downsampling by selecting maximum values within regions, reducing spatial dimensions to reduce computation and highlight significant features while achieving spatial invariance.
- **2 Fully Connected Layers:** Convert extracted features into predictions—bounding box coordinates, objectness scores, and class probabilities. Detection is performed at two spatial scales to improve accuracy for small and large objects.
- **Output Layer:** Ends with two or more output layers. Normally the output layers are determined by the number of classes of the dataset or classification problem of the model.

C. Why Tiny or Low-Powered Models?

Suppose: for Tiny-YOLOv4

- Input channels: 32
- Filter: 3×3
- Number of Filters: 64

Then:

- Conv2D parameters: $(3 \times 3 \times 32) \times 64 = 18{,}432$
- BatchNorm parameters: $4 \times 64 = 256$
- Total for this layer: **18,688 parameters**

Total number of parameters in Tiny-YOLOv4 is approximately: 6.06 million parameters.

This includes all convolutional + batch normalization layers.

On the other hand, **for YOLOv4**:

Total number of parameters in YOLOv4 is approximately: 64 million!!!!

The computational requirements of a deep learning model are in direct proportion to its architectural complexity and the number of parameters. Models with lots of parameters, such as YOLOv4 or YOLOv7, are highly accurate but expensive in terms of memory consumption, time taken for training, inference time, and power requirements. Models like these usually require high-end computing setups with GPUs or TPUs supporting parallel processing capabilities, high memory bandwidth, and thermal management.

But in real-world applications, especially when dealing with embedded systems or edge devices such heavy resource models are not practicable. Instead, light models like MobileNet or Tiny-YOLOv4 are utilized. These models are designed in such a way to offer a compromise between performance and efficiency by reducing the number of parameters and using techniques like depthwise separable convolutions or quantization.

Tiny-YOLOv4, for instance, is a lightweight version of the YOLO object detection model with significantly less computational burden yet maintaining good accuracy. Its minimalistic architecture allows it to run in real-time on low-processing power and power-constrained devices. It is thus ideal for real-time use cases such as surveillance, visually impaired individuals' assistance technology, robotics, or autonomous navigation where rapid decisions based on visual input are critical.

Moreover, reducing model size not only improves responsiveness but also decreases heat generation and power consumption, which is very important for battery-powered systems. In all practical scenarios, model optimization of such constraints through techniques like pruning, quantization, or model distillation is essential to deploy deep learning applications in the field.

VI. NLP in Action:

ONatural Language Processing (NLP) functions as an important translation service which converts AI system data along with sensor information into understandable Bangla voice commands for visually impaired users. The communication system becomes more effective through these features which allow individuals to make proper decisions during their navigation journeys. The principal functions of the NLP component embrace:

- The NLP component converts both object detection results together with navigation directions into easily understandable and direct Bangla language instructions.
- The system needs to create voice instructions which match the situation and sound naturally and exhibit emotional elements to build trust between the device and users.
- The system meets the requirement of delivering real-time processing that qualifies for deployment in wearable assistive systems using minimal power.

A. Speech Synthesis Models

- A system uses real-time Bangla voice instruction production for navigation texts through rule-based and deep learning programming techniques.
- Data-driven models outperform rule-based models

regarding natural sound but rule-based systems provide better speed performance.

- The system runs in real-time by implementing model compression together with quantization while maintaining short and direct outputs.
- The developed system provides distinct and reactive voice guidance in Bangla language for secure navigation purposes.

B. *Speech Challenges and Solutions*

- The production of robot voice using Bangla speech entailed overcoming numerous obstacles. The system required pronunciation adjustments because different Bangla speech patterns existed across regions and we addressed this issue by implementing data collection which modified the system to achieve broader user understanding.
- Premade systems presented either artificial speech along with insufficient real-time performance capabilities. The team trained better responsive and clear lightweight models to address the issue.
- The system performance required model compression and short speech delivery to run effectively on available hardware. The pitch and vocal tone received modification to create a singing voice which facilitated better understanding for listeners.

C. *Model Optimization*

Organizing AI models for execution on Jetson Orin Nano needs specific optimization steps to keep the system operating at fast speeds and using minimal resources yet maintaining effective performance. The model optimization process consisted of size reduction along with speed enhancement combined with memory optimization while keeping accurate outputs.

- **Quantization**:

  The quantization enabled both small model size and high operating speed. Quantization of 32-bit floating point numbers to 8-bit integers enabled faster operation of the deep learning engine on Jetson Orin Nano (4GB) systems without sacrificing the accuracy level too much.

- **Pruning:**

  The model received pruning treatment which eliminated nonessential parameters. The removal of superfluous weights in Tacotron 2 through this method reduced both computational load during inference time and system memory requirements.

- **Lightweight Models:**

  We can accelerate the system performance by implementing these simple lightweight models.

  - Tiny DeepSpeech for speech-to-text capabilities.
  - Festival TTS and Google TTS API for straightforward Bangla voice generation.
  - FastSpeech for low-latency, expressive TTS.

- **Vocabulary Limiting:**

  The system performance improves when navigation terms from a limited vocabulary are enforced for both speech processing and synthesizing functions. The system became both simpler and faster and more predictable due to these restrictions.

- **Frame Skipping and Buffering:**

  The system uses a frame skipping function and buffering technique to manage CPU performance and prevent audio problem occurrences. The system bypassed unneeded audio frames which enabled smooth continuous voice instructions to flow uninterrupted to users.

D. *System Integration Challenges*

- **Real-Time Synchronization:** The integration of multiple real-time subsystems including perception navigation along with communication became a difficulty to manage because of hardware restrictions.
- **Modular Design Approach:** A modular framework served to enable different subsystems when operating separately yet remaining coordinated. This improved overall system reliability and maintainability.
- **Performance Optimization:** The optimization process minimized model memory demands and processing requirements to achieve better performance on minimal hardware.
- **Efficient Communication:** The system developed a speedy communication protocol that enabled dependable data transfers among its different components.
- **Sensor Reliability:** The system used filtering methods together with cross-validation techniques to manage inconsistent and noisy sensor information by improving its accuracy and stability levels.

VII. Process of Work

The process flow includes the following major steps:

- **Hardware & System design:** Putting all the hardware components to initiate the process of constructing the robot. May face robot structural and adequate weight distribution issues here.
- **Object detection:** Through Data preprocessing, Data augmentation and then training different DL models with different large datasets and comparing

the performance in order to choose the optimized model.

- **Bangla Voice Guidance using NLP:** Utilize Bangla TTS to provide clear, real-time navigation instructions on low-power devices.
- **Sensor based issue evaluation:** Resolving user tracking and climbing stairs sensor-based issues.
- **Testing & optimization:** Optimizing the AI models followed by testing the robot.

This system places a greater emphasis on a balance between performance and computational efficiency, thus being ideal for low-power aiding hardware.

VIII. Future Scope

The developed system creates a cost-effective platform designed specifically to assist visually impaired users with their mobility needs. Future developments will concentrate on implementing quality enhancements for both system functionality and user interface improvement:

- Users would benefit from basic voice command functionality that enables them to control the robot system using their spoken commands.
- The device would incorporate an SOS emergency alert system which enables contact notifications to emergency contacts during unsafe situations and emergency accidents.
- An upgrade to the robot will enable automatic user tracking and movement adaptation for following the user.
- The robot would get outdoor navigation capabilities thanks to GPS integration through which users could receive guided travel in outdoor environments.
- Enhancing the obstacle detection system to detect several types of obstacles would lead to more precise guidance.
- Custom Bangla language models should be developed for better natural and dialect-aware speech output.
- Future development will concentrate on creating a portable small version of the device alongside making it easily wearable.

IX. Conclusion

The primary work of this project involved developing user-robot interface communication tools by using TTS technologies and Bangla NLP systems. To achieve the requirements the voice interface needed both natural accuracy alongside quick reliable performance. Our solution meets the needs of blind users in Bangladesh thanks to optimized models and tailored speech training and considerate implementation methods. Through its development the project achieves two main goals: solving technological difficulties as well as improving accessibility safety in public spaces for blind users.

References

[1] G. Xu, A. S. Khan, A. J. Moshayedi, X. Zhang, and Y. Shuxin, “The Object Detection, Perspective and Obstacles In Robotic: A Review,” EAI Endorsed Transactions on AI and Robotics, vol. 1, no. 1, Oct. 2022, doi: 10.4108/airo.v1i1.2709.

[2] P. Wiriyathammabhum, D. Summers-Stay, C. Fermüller, and Y. Aloimonos, “Computer Vision and Natural Language Processing: Recent Approaches in Multimedia and Robotics,” ACM Computing Surveys, vol. 49, no. 4, Art. 71, Dec. 2016, doi: 10.1145/3009906.

[3] D. Mazzei, F. Chiarello, and G. Fantoni, “Analyzing Social Robotics Research with Natural Language Processing Techniques,” Cognitive Computation, vol. 13, pp. 308–321, Jan. 2021, doi: 10.1007/s12559-020-09799-1.

[4] S. Ekvall, D. Kragic, and P. Jensfelt, “Object Detection and Mapping for Service Robot Tasks,” Robotica, vol. 25, pp. 175–187, 2007, doi: 10.1017/S0263574706003237.

[5] V. Saini and N. Joseph, “Artificial Intelligence in Robotics Using NLP,” Preprint, Dec. 2022. [Online]. Available: https://www.researchgate.net/publication/366313247

[6] E. Garcia, M. A. Jimenez, P. G. De Santos and M. Armada, ”The evolution of robotics research,” in IEEE Robotics & Automation Magazine, vol. 14, no. 1, pp. 90-103, March 2007, doi: 10.1109/MRA.2007.339608.

[7] Y. Yang, C. L. Teo, C. Fermüller and Y. Aloimonos, ”Robots with language: Multi-label visual recognition using NLP,” 2013 IEEE International Conference on Robotics and Automation, 2013, pp. 4256-4262, doi: 10.1109/ICRA.2013.6631179.

[8] Nitin Madnani, “Getting Started on Natural Language Processing with Python” ed:Crossroads 13(4):5,September 2007

[9] Jost Schatzmann, Karl Weilhammer, Matt StuttleE, Steve Young “A Survey Statistical User Simulation Techniques for Reinforcement-Learning of Dialogue Management Strategies,”ed:The Knowledge Engineering Review, Vol. 00:0, 1–24. c 2006, Cambridge University Press

[10] K. S. Jones and J. Galliers. Evaluating Natural Language Processing Systems: An Analysis and Review. Springer Verlag, 1996.

[11] R. Klinkenberg and I. Renz. Adaptive information filtering: learning in the presence of concept drift. In AAAI/ICML-98 Workshop on Learning for Text Cat-

egorization, Technical Report WS-98-05, 1998. Madison, Wisconsin.

[12] Smarr, C. A., Prakash, A., Beer, J. M., Mitzner, T. L., Kemp, C. C., & Rogers, W. A. (2012, Sep). Older Adults' Preferences for and Acceptance of Robot Assistance for Everyday Living Tasks. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Boston, MA: HFES.

[13] Zhang, L., Jiang, M., Farid, D., & Hossain, M. A. (2013, October). Intelligent Facial Emotion Recognition and Semantic-Based Topic Detection for a Humanoid Robot. Expert Systems with Applications, 40(13), 5160–5168. doi:10.1016/j.eswa.2013.03.016

[14] Lars Kunze , Nick Hawes , Tom Duckett , Marc Hanheide , and Toma ́s Krajnık "Artificial Intelligence for Long-Term Robot Autonomy: A Survey", IEEE Robotics and Automation Letters, VOL. 3, NO. 4, October 2018

[15] Jim Torresen "A Review of Future and Ethical Perspectives of Robotics and AI" Front. Robot. AI, 15 January 2018 Sec. Robot Learning and Evolution https://doi.org/10.3389/frobt.2017.00075

[16] W. Zhang, S. Wang, S. Thachan, J. Chen and Y. Qian, "Deconv R-CNN for Small Object Detection on Remote Sensing Images," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 2483-2486, DOI: 10.1109/IGARSS.2018.8517436.

[17] D. Kumar and X. Zhang, "Improving More Instance Segmentation and Better Object Detection in Remote Sensing Imagery Based on Cascade Mask R-CNN," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 4672-4675, DOI: 10.1109/IGARSS47720.2021.9554512.

[18] C. Gao, Y. Zhai and X. Guo, "Visual Object Detection and Tracking System Design based on MobileNet-SSD," 2021 7th International Conference on Computer and Communications (ICCC), 2021, pp. 589-593, DOI: 10.1109/ICCC54389.2021.9674450.

[19] B. Wu, A. Wan, F. Iandola, P. H. Jin and K. Keutzer, "SqueezeDet: Unified, Small, Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 446-454, DOI: 10.1109/CVPRW.2017.60.

[20] A. Sarda, S. Dixit and A. Bhan, "Object Detection for Autonomous Driving using YOLO [You Only Look Once] algorithm," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1370-1374, DOI:10.1109/ICICV50876.2021.9388577.

[21] J. Fan, J. Lee, I. Jung and Y. Lee, "Improvement of Object Detection Based on Faster R-CNN and YOLO," 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 2021, pp. 1-4, DOI: 10.1109/ITC-CSCC52171.2021.9501480.

[22] S. T. Blue and M. Brindha, "Edge detection based boundary box construction algorithm for improving the precision of object detection in YOLOv3," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-5, DOI: 10.1109/ICCCNT45670.2019.8944852.

[23] S. Ali, A. Siddique, H. F. Ateş and B. K. Güntürk, "Improved YOLOv4 for Aerial Object Detection," 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021, pp. 1-4, DOI: 10.1109/SIU53274.2021.9478027.

[24] M. Sharma et al., "YOLOrs: Object Detection in Multimodal Remote Sensing Imagery," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 1497-1508, 2021, DOI: 10.1109/JSTARS.2020.3041316.

[25] S. Mane and S. Mangale, "Moving Object Detection and Tracking Using Convolutional Neural Networks," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1809-1813, DOI: 10.1109/ICCONS.2018.8662921.