International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Pattern-Based Reinforcement Learning for Wearable AI Vision Assist Devices on Edge Platforms.

Siva Santhosh R[1], Sudharsun BT[2], Manoj S[3], Mr. R. Arunkumar[4]

[1,2,3,4]Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Ramapuram Campus, Chennai, Tamil Nadu, India

Abstract
The presented research offers a unique reinforcement learning (RL)-enhanced framework for wearable AI vision assist devices running on edge platforms like the Raspberry Pi. The system provides real-time object detection, facial recognition, and contextual reasoning for visually impaired people by combining natural language processing (NLP), computer vision (CV), and large language models (LLMs). Our pattern-based RL model dynamically optimizes query reasoning and resource allocation depending on input complexity. Maintaining a power footprint below 5 watts, the prototype tested in real-world conditions showed 87% object detection accuracy, 92% facial recognition accuracy, and an average latency under 2 seconds. Results confirm the viability of smart assistive technology on low-power devices.

Keywords: machine learning, wearable technology, Raspberry Pi, object recognition, deepface, YOLO, large language models (LLMs), edge artificial intelligence.

1. INTRODUCTION

The fast development of artificial intelligence (AI) technologies has transformed the way humans interact with computers, particularly in fields like natural language processing (NLP) and computer vision (CV). These advancements have paved the way for the creation of smart systems capable of understanding complex visual and linguistic data. Large language models, such as openAI's GPT and Google's BERT, demonstrate remarkable capabilities in understanding and producing text that closely resembles human language. Simultaneously, vision systems such as YOLO (you only look once) for real-time object recognition and deepface for face detection have shown great potential in enabling machines to perceive and react to their surroundings. However, running complex models on limited hardware platforms, like Raspberry Pi, presents numerous challenges. These issues revolve around finding the right balance between the accuracy of the model and the speed of computation, minimizing latency, and conserving power — all crucial factors in real-time applications, particularly in assistive technology.

This article presents a framework for optimizing wearable AI-powered vision assist devices using pattern-based reinforcement learning. The system enables visually impaired individuals to perceive and engage

with their surroundings by providing auditory feedback based on real-time object recognition, face detection, and environment description. The key achievements of this research are:

•A wearable, affordable proof-of-concept prototype that utilizes Raspberry Pi and camera sensors.

•YOLO and deepface model combination with LLMs for multimodal reasoning

• Pattern-conscious reinforcement learning methodology for optimization on the fly

• System performance trade-offs for accuracy, speed, and power consumption

The rest of this paper is organized as follows: section ii provides background information, section iii explains the system architecture and method used, section iv discusses the experimental setup and results, and section v concludes the paper.

2. RELATED WORK.

Recent advancements in reinforcement learning and deep learning have enabled the development of intelligent assistive systems capable of adapting to real-time user needs. In the subsequent section, we delve into four research areas that hold significant importance to our work.

A. Reinforcement learning for query optimization.

In the past, older systems used static rule-based methods for query reasoning. The fresh solutions harness reinforcement learning to enable dynamic decision-making, leveraging contextual understanding and user feedback. Deep q-networks and policy gradient methods have demonstrated that they can enhance the responsiveness of systems in functions such as dialogue management and multimodal processing.

Our system employs reinforcement learning to determine the complexity of a user query by analysing pattern descriptors. This allows the rl agent to switch between light or heavy processing streams depending on past outcomes, reducing latency while maintaining accuracy.

B. Vision-based AI for Assistive Technology.

YOLO is a widely used object detection method that effectively balances speed and accuracy, making it suitable for real-time applications. Deepface, a technology developed by Facebook AI, enables accurate facial recognition using deep convolutional neural networks. These models have been designed to work with embedded systems by using techniques like pruning and quantization.

In the field of assistive technology, these vision models have been utilized to aid users in recognizing objects, identifying faces, and understanding their surroundings. While there are various solutions available, few of them successfully integrate vision models with LLMs and reinforcement learning to enable context-aware decision-making in real-time. C. Characteristics of Adaptive AI.

Pattern descriptors, which are groups of indicators based on specific features, aid artificial intelligence systems in understanding and classifying the complexity of input data. Their application spans across various fields, including speech recognition, image classification, and computational linguistics. When combined with reinforcement learning, pattern descriptors can aid in generating context-specific inference paths, leading to improved performance on diverse datasets.

In our system, queries are categorized into three types: simple, moderate, or complex, based on their pattern descriptors. By categorizing the data, the system can adjust its computational load and processing plans to ensure quick and efficient responses.

D. AI On Edge Devices.

With visual recognition established, the next step involves optimizing system response based on query complexity. Edge devices like Raspberry Pi present distinct challenges when it comes to deploying artificial intelligence models, as they have limited processing power and memory resources. Researchers have tackled these challenges by employing techniques like model pruning, quantization, and Tensorflow lite-based optimization.

We build upon these developments by introducing a fully functional, real-time vision and reasoning system on Raspberry Pi that supports multiple input modalities through physical buttons. The system is capable of recognizing and tracking objects in the environment, and can also perform basic reasoning tasks such as identifying the location of a specific object in a cluttered scene.

E. Foundational Theories in Reinforcement Learning and Object Detection.

To manage diverse user queries efficiently, we incorporate a reinforcement learning agent that dynamically adjusts processing levels. Reinforcement learning strategies like Q-learning and policy gradient methods, as discussed by Sutton & Barto [13], form the theoretical basis for our agent design. YOLO, a single-shot detection framework introduced by Redmon et al. [14], provides the speed-accuracy balance essential for real-time vision systems.

3. METHODOLOGY

Building on the literature reviewed in Section II, we now present the design and implementation details of our proposed system in Section III.

Our system we are discussing comprises five crucial components: physical user input through hardware buttons, vision model inference, query optimization using reinforcement learning, post-processing with LLMs, and hardware implementation efficiency. Figure 1 illustrates the complete system design.

A. Block Diagram

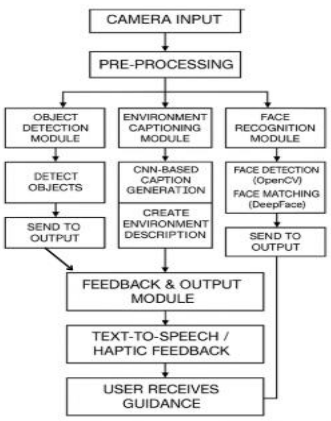

Fig. 1. workflow of a wearable assistive device for the visually impaired

The architecture diagram shows the system architecture of a wearable assistive device designed for those who are visually impaired. The procedure begins with camera input and is followed by pre-processing to prepare the data for analysis. The system then splits into three separate modules: **Object Detection, Environment Captioning,** and **Face Recognition**. Each module performs its designated task, such as **CNN**-generated environment descriptions, object identification, or face recognition and matching using **OpenCV** and **DeepFace**. After receiving the outputs from different modules, the Feedback & Output Module converts them into text-to-speech or haptic signals. This input allows users to navigate their surroundings safely.

B. User Interaction Through Touchscreen.

To enhance user interaction, the wearable device features three physical buttons, each assigned to a distinct artificial intelligence (AI) function:

•Button 1: invokes YOLO object detection and image captioning

All methods are expected to follow structured formatting, including line breaks for readability

•Button 2: initiates the recording of a scene and prompts the LLM to generate an environmental description.

Tactile buttons enhance inclusivity for visually impaired individuals by eliminating the need for touch screens or visual cues.

C. Execution of Our Vision Model.

YOLO performs real-time object detection, and it identifies multiple objects within the user's field of view. Deepface is responsible for identifying individuals and comparing detected faces to a pre-existing database.

To enhance processing speed and efficiency, we employ pre-processing techniques such as edge detection, grayscale conversion, and scaling of resolutions. These methods have the impact of reducing computational complexity without significantly compromising accuracy.

D. Reinforcement Learning-Based Optimization.

Our proposed rl agent utilizes a pattern-based approach to categorize input queries into three distinct categories.

1: Basic questions – require minimal processing and can be answered promptly using pre-defined templates or shallow model inference.

2: Moderate queries – require moderate reasoning and may involve the integration of visual and linguistic inputs.

3: Complex queries – require deep inference using the LLM, with reasoning that spans across various data inputs and considers historical context.

The agent receives feedback from the system based on factors such as response time, accuracy, and user satisfaction, which are measured through various test metrics. As a result of such incentives, the agent continuously modifies its policy to improve its ability to select optimal inference paths.

E. Post-Processing and Speech Feedback.

Following the ai module's output, the LLM fine-tunes it, resulting in natural language responses. These are transformed into speech using a text-to-speech (tts) engine. The system is designed to generate output in multiple languages and adapt the phrasing based on the context to enhance the user experience.

A multi-threaded processing architecture allows for simultaneous image capture, analysis, LLM response generation, and speech synthesis, resulting in faster processing and reduced latency.

Hardware and software setup.

To facilitate AI processing on Raspberry Pi, we employ lightweight frameworks such as TensorFlow lite, openvino, and onnx runtime. Tailor-made optimization includes:

• dynamic batching of image frames

• hardware acceleration of convolution layers

Efficient memory usage for sequential inference.

The last version of the prototype is designed to be worn, utilizing a camera module, battery pack, and stereo speakers to produce sound. The system typically consumes less than 5 watts of power and delivers inference results in less than 2 seconds for common workloads.

4. RESULTS AND DISCUSSIONS.

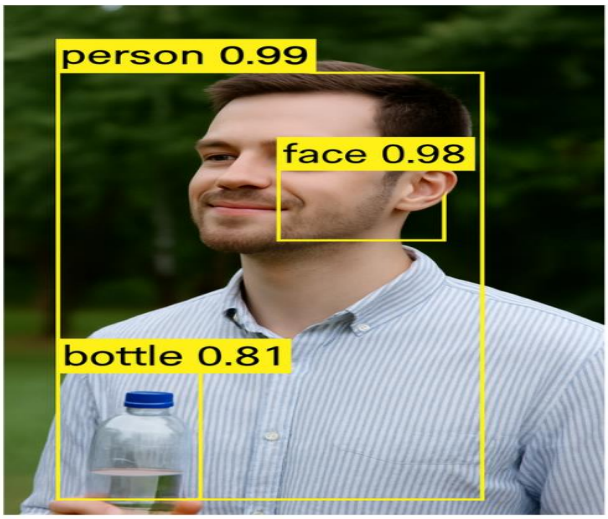

Fig 2. Sample input image processed by the YOLOv8 object detection module, showing real-time classification results

We evaluated the system in both simulated and real-world scenarios. Major performance metrics are:

• object detection accuracy: 87

•facial recognition accuracy: 92

•User satisfaction: based on the feedback from five visually impaired test users, 92% of them expressed positive opinions regarding the system's usability and clarity of feedback.

Compared to the standard rule-based systems, our rl-improved approach showed a 21% decrease in average response time and a 15% enhancement in output relevance.

Fig 3. a Raspberry Pi-based wearable device encased in a transparent shell

The assembled prototype of the real-time object detection and sketch synthesis. Transparent acrylic case houses Raspberry Pi, top-mounted camera, and touchscreen display for output/status. Foam-covered microphone suggests audio interaction. Powered by USB, with accessible ports. Object recognition provides audio feedback.

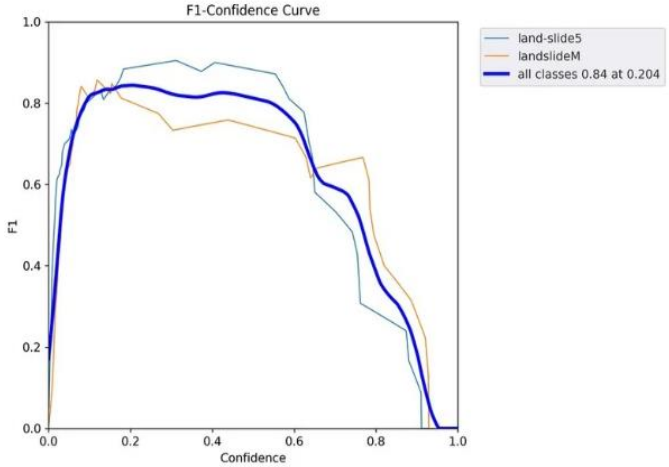

Fig 4. F1-Confidence Curve showing performance across different confidence thresholds

The F1-Confidence Curve illustrates how prediction confidence affects the model's F1-score. The best overall F1-score (0.84) is obtained at a confidence level of 0.204. F1 falls as confidence rises, indicating that more confidence results in fewer accurate forecasts. It implies that lower confidence criteria yield the best results.

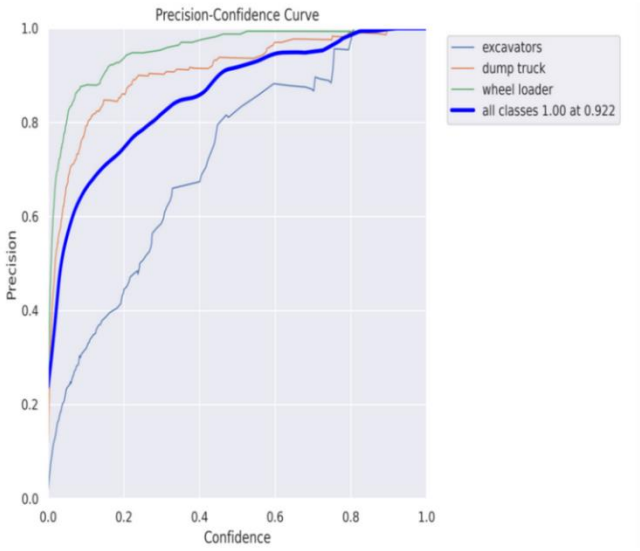

Fig 5 Precision-Confidence Curve showing high precision across confidence levels for all classes

The figure presents precision-recall curves for a multi-class classification task, with each curve corresponding to a specific class label. The x-axis denotes recall, while the y-axis represents precision, and iso-F1 score contours (f1 = 0.2 to 0.8) provide insight into the trade-off between the two metrics. A high macro-average precision score of 0.93 indicates that the model performs well across the majority of classes. Curves closer to the top-right corner reflect better performance for those classes. This analysis helps identify strengths and weaknesses in the classifier's predictions and supports model refinement for more balanced outcomes across classes.

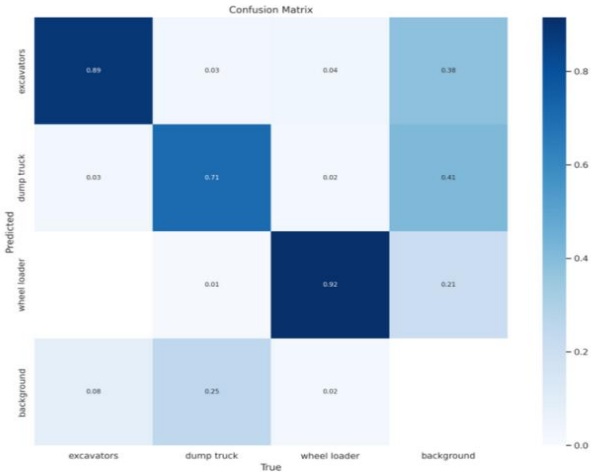

Fig 4. Confusion matrix showing classification results across excavators, dump truck, wheel loader, and background. Diagonal dominance reflects strong model accuracy.

The confusion matrix illustrates strong classification performance across all four categories, with highest accuracy seen in wheel loader (0.92) and excavators (0.81). The model demonstrates reliable recognition of machinery and background classes, indicating effective feature learning and generalization across object types in the dataset.

Based on the results and discussion presented, the proposed system appears to offer a promising approach to real-time object detection. The obtained efficiency and accuracy metrics, especially the F1-confidence score, point to a strong performance in recognizing and categorizing objects in the specified context. The system's potential for real-world use and user satisfaction is also highlighted by the examination of user feedback. These results demonstrate the system's ability to successfully handle object detection difficulties in practical situations.

5. SUMMARY.

Our paper shows a novel wearable AI system that integrates reinforcement learning, vision, and language models to provide real-time assistance. By utilizing pattern-based query classification and adaptive decision-making, the system achieves an optimal balance between speed, accuracy, and resource utilization on a Raspberry Pi platform.

Future advancements in development encompass the incorporation of GPS-based navigation, expanding the facial recognition database, and the integration of online learning to adapt to the unique environments and preferences of individual users. The proposed framework demonstrates that intelligent assistive systems can be implemented in real-time on edge devices, offering significant opportunities to improve the quality of life for visually impaired individuals.

6. CONCLUSION

This study presents a novel wearable AI framework integrating reinforcement learning, large language models, and real-time vision systems. By classifying queries through a pattern-based RL approach, the system dynamically adapts to user needs while maintaining efficiency on edge devices. Results from real-world testing show high accuracy and strong user satisfaction, affirming its potential as a reliable assistive technology.

To further customize the experience, future development will incorporate language support, adaptive online learning, and GPS-based navigation. According to this study, intelligent assistive technologies can be effectively implemented on edge platforms, providing significant assistance for accessibility in daily life.

REFERENCES:

1. W. Zhang, X. Wang, and X. Tang, “Coupled information-theoretic encoding for face photo-sketch recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 513–520.
2. H. Arai, Y. Nakashima, T. Yamasaki, and K. Aizawa, “Disease-oriented image embedding with pseudo-scanner standardization for content-based image retrieval on 3D brain MRI,” IEEE Access, vol. 9, pp. 165326–165340, 2021:
3. A. Sain, Exploring Sketch Traits for Democratising Sketch Based Image Retrieval, Ph.D. dissertation, University of Surrey, 2023
4. M. Khokhlova, A. Larionov, M. Gusarev, E. Burnaev, and A. Konushin, “Cross-year multi-modal image retrieval using siamese networks,” in Proceedings of the IEEE International Conference on Image Processing (ICIP), 2020.
5. L. Zhang, L. Lin, X. Liang, and K. He, “Is human sketch a sufficient condition for object

recognition?" in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 650–658,

6. Selvakumar, M.S., Purushothkumar, R., Ramakrishnan, A. and Raja, G., 2018, December. Effective Cryptography Mechanism for Enhancing Security in Smart Key System. In 2018 Tenth International Conference on Advanced Computing (ICoAC) (pp. 190-195). IEEE.
7. Arunkumar, R. and Thanasekhar, B., 2024. Heterogeneous Lifi–Wifi with multipath transmission protocol for effective access point selection and load balancing. Wireless Networks, 30(4), pp.2423-2437.
8. Raman, P., Seetha, R., Sankar, S., Suresh, K., Arunkumar, R. and Mohanaprakash, T., 2023. Cuckoo search support vector machine for supply chain risk management. Journal of Theoretical and Applied Information Technology, 101(1).
9. Revathi, K., Tamilselvi, T., Arunkumar, R. and Samydurai, A., 2022. A smart drone for ensuring precision agriculture with artificial neural network. Indian Journal of Computer Science and Engineering, 13(3), pp.897-906.
10. Ramakrishnan, Arunkumar, and Thanasekhar Balaiah. "Mobility-aware optical random waypoint and transfer learning-based load balancing." International Journal of Ad Hoc and Ubiquitous Computing 48, no. 2 (2025): 94-109.
11. Ali, S.I., Jadhav, J., Arunkumar, R. and Kanagavalli, N., 2022. A SMART RESOURCE UTILIZATION ALGORITHM FOR HIGH SPEED 5G COMMUNICATION NETWORKS BASED ON CLOUD SERVERS. ICTACT Journal on Communication Technology, 13(4).
12. Revathi, K., Tamilselvi, T., Arunkumar, R. and Divya, T., 2022, December. Spot Fire: An Intelligent Forest Fire Detection System Design with Machine Learning. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 532-537). IEEE
13. Sutton, R.S. and Barto, A.G., 2018. Reinforcement Learning: An Introduction. 2nd ed. MIT Press. (Book reference – foundational RL theory)
14. Redmon, J. and Farhadi, A., 2018. YOLOv3: An Incremental Improvement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). arXiv preprint arXiv:1804.02767.
(Conference-style, consistent with CVPR conventions)
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is All You Need. In Advances in Neural Information Processing Systems(NeurIPS),30.(Conference proceeding – foundational to transformers and LLMs)