

SAFENAVGPT: LLM and Transformer Driven Anomaly Detection for Real-Time Visual Navigation

Vaisnavi N M¹ and Dr. M. Praneesh²

UG Student, Department of Computer Science with Data Analytics¹

Assistant Professor, Department of Computer Science with Data Analytics²

Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India

Abstract: *This research enhances the “VISIONGPT” framework for safe and efficient visual navigation, targeting visually impaired individuals and autonomous systems. It incorporates transformer-based models and attention mechanisms to improve object detection, scene understanding, and real-time anomaly detection. The current system employs YOLO-World for open-vocabulary object detection, a rule-based anomaly detection module, and LLMs (GPT-3.5/4) for scene interpretation and voice-based hazard alerts. While effective, traditional CNN-based models struggle with capturing complex object relationships and adapting to dynamic environmental changes.*

To overcome these limitations, we propose an enhanced architecture featuring Vision Transformers (ViTs) and attention mechanisms. ViTs refine object recognition, especially for detecting small or distant hazards, and enable the system to predict object movements and provide anticipatory alerts. The self-attention mechanism of transformers allows the model to dynamically weigh the importance of different objects, minimizing false positives and ensuring critical hazards are prioritized.

Experimental evaluations demonstrate improved detection precision, reduced false alarms, and better latency management, making the system more practical for real-time applications. The enhanced system supports dynamic scene transitions, proactive hazard warnings, and user-personalized alert mechanisms, making it adaptable to urban, indoor, and unpredictable environments.

This work advances LLM-assisted visual navigation, contributing to AI-driven accessibility solutions. The proposed architecture is scalable and can be integrated into autonomous vehicles, robotic systems, and assistive technologies. It also pushes forward the field of vision-language models and multimodal AI, showcasing how transformer-based models can significantly enhance real-time navigation safety and accessibility.

Keywords: Vision Transformers (ViTs), Attention mechanisms, Anomaly detection, LLMs (GPT-3.5/4)

