RESEARCH ARTICLE OPEN ACCESS

Real-Time Pothole Detection for Visually Impaired Navigation Using YOLOv8 and Transformer Architecture

Arul Selvan R*, R. Kalaichelvan**
*(Department Of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India
Email: arulselvan104@gmail.com)
** (Department Of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India
Email: Kalaichelvan.r@drngpasc.ac.in)

----------**********----------

Abstract:

Visually impaired individuals face serious challenges while navigating roads due to the inability to visually identify surface hazards such as potholes and uneven road conditions. Existing assistive tools like white canes provide only close-range detection and cannot deliver advance warnings about road surface defects ahead. This paper proposes a Transformer-based real-time pothole detection system to support safe and independent navigation for visually impaired individuals using live camera input and deep learning techniques. The system captures live road scenes using a camera and applies YOLOv8 integrated with Transformer-based attention mechanisms to detect potholes accurately in real time. When a pothole is identified, an audio alert is generated immediately to warn the user before they reach the hazard. The dataset used for training is sourced from Mendeley Data and Roboflow, containing annotated road images exported in YOLO format. The proposed system enhances safety, mobility, and independent movement for visually impaired users in real-world urban and rural road environments..

Keywords: ***Pothole Detection, Visually Impaired Navigation, YOLOv8, Transformer Architecture, Real-Time Object Detection, Assistive Technology, Deep Learning***

----------**********----------

I. INTRODUCTION

The ability to navigate independently is fundamental to the quality of life of visually impaired individuals. Roads present numerous hazards that are impossible to detect without vision, including potholes, broken surfaces, and uneven terrain. These hazards pose significant risks of falls, injuries, and reduced mobility. In developing countries, the prevalence of poorly maintained roads makes the problem even more severe, limiting the independence and safety of visually impaired pedestrians.

Conventional assistive devices such as white canes are effective for detecting nearby obstacles but are incapable of identifying road surface defects in advance. Electronic assistive technologies such as ultrasonic sensors and GPS-based navigation systems have been explored, but they often lack the precision and real-time responsiveness needed for detecting small localized road defects like potholes.

Recent advances in deep learning, particularly in object detection and Transformer-based architectures, have opened new possibilities for real-time visual hazard detection. This paper proposes a Transformer-based real-time pothole detection system that uses a live camera feed to detect potholes and generate immediate audio alerts. The system integrates YOLOv8 with Transformer attention mechanisms to combine the speed of convolutional detection with the contextual understanding of Transformers, enabling accurate and reliable pothole identification in diverse real-world conditions.

II. LITERATURE SURVEY

Several image processing and computer vision approaches have been explored for road

damage detection. Early methods relied on traditional image processing techniques such as edge detection, thresholding, and morphological operations to identify road surface anomalies. These methods were effective under controlled conditions but struggled with varying lighting, shadows, and road textures encountered in real-world environments.

Machine learning approaches using Support Vector Machines (SVM) and Random Forests were subsequently applied to road defect classification with moderate success. The introduction of Convolutional Neural Networks (CNN) significantly improved detection accuracy by enabling automatic feature learning from images. YOLO-based architectures further advanced real-time detection by performing detection in a single forward pass, making them suitable for live video analysis.

Transformer architectures, originally developed for natural language processing, were adapted for computer vision through the Vision Transformer (ViT) model proposed by Dosovitskiy et al. (2020). ViT-based models demonstrated superior performance for global context understanding by treating image patches as sequences and applying multi-head self-attention. Integrating Transformer modules into YOLO-based detection systems has shown improved accuracy for objects at varying distances and orientations, making them well-suited for road hazard detection in complex outdoor environments.

III. PROBLEM STATEMENT

Visually impaired individuals face significant difficulties while navigating roads due to the inability to visually identify hazards such as potholes, uneven surfaces, and damaged roads. Existing assistive tools like white canes can detect obstacles only at very close range and cannot provide advance warning about road surface defects. Current road monitoring systems are mostly manual, time-consuming, and not designed for real-time personal navigation assistance.

There is a critical need for an automated, real-time system that can continuously monitor road conditions using camera input, detect potholes and road hazards accurately at safe stopping distances, and immediately alert visually impaired users through audio notifications to enable timely course correction and safe navigation.

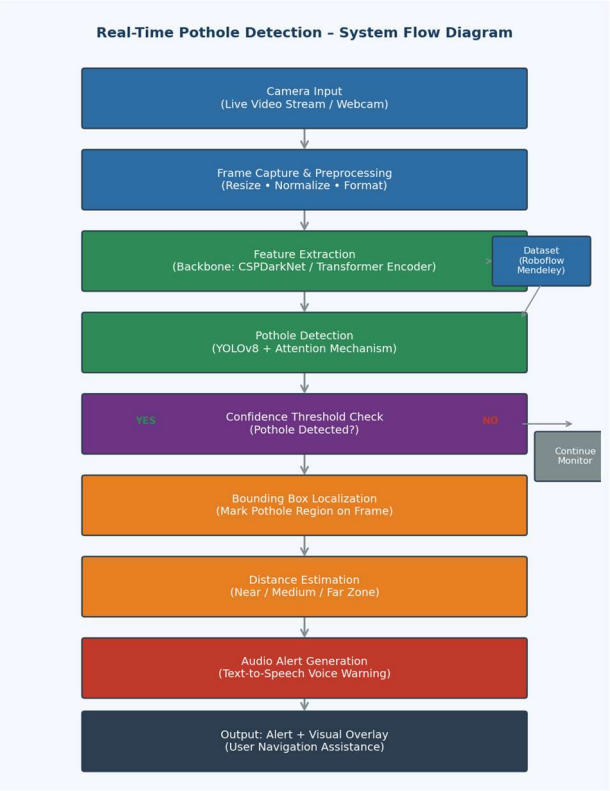

Fig. 1: Real-Time Pothole Detection – System Flow Diagram

IV. PROPOSED SYSTEM

Real-Time Pothole Detection for Visually Impaired Navigation is a deep learning-based assistive system designed to continuously monitor road surfaces through live camera input and provide immediate audio alerts when potholes or road hazards are detected. The system integrates YOLOv8 with Transformer-based attention mechanisms to achieve accurate real-time detection across diverse road and lighting conditions.

The system captures live video frames using a webcam or mobile phone camera and passes each frame through a preprocessing pipeline that resizes and normalizes the input for the detection model. The YOLOv8 backbone extracts multi-scale spatial features from the input frame, while the integrated Transformer encoder applies multi-head self-

attention to capture long-range contextual dependencies and improve detection accuracy for partial or distant potholes.

When the detection confidence exceeds the defined threshold, the system localizes the pothole using a bounding box, estimates the proximity of the hazard based on bounding box size, and triggers an audio alert through a text-to-speech engine. The alert message is calibrated to the estimated distance, providing early warning for far hazards and urgent warnings for near hazards, giving users sufficient time to navigate safely around the identified road defect.

The training dataset is sourced from Mendeley Data and Roboflow, containing road images collected under varied lighting, weather, and road surface conditions. All images are manually annotated with bounding boxes indicating pothole locations and exported in YOLO format. The model is trained using PyTorch on this dataset to achieve reliable detection performance across diverse real-world scenarios.

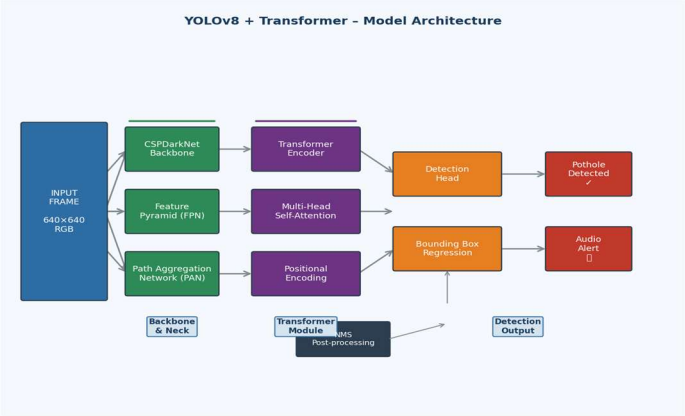

Fig. 2: YOLOv8 + Transformer Model Architecture

V. SYSTEM ARCHITECTURE

The architecture of the Real-Time Pothole Detection system consists of four primary components: the input processing layer, the detection backbone with Transformer module, the detection head, and the alert generation module. The input processing layer captures frames from the camera feed and applies preprocessing operations including resizing to 640x640 pixels, normalization, and format conversion for model compatibility.

The detection backbone uses CSPDarkNet as the primary feature extractor, supplemented by a Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) for multi-scale feature fusion. The Transformer encoder module applies multi-head self-attention with positional encoding to the extracted feature maps, enabling the model to capture global context and improve detection of partially occluded or distant potholes. The detection head performs bounding box regression and classification, followed by Non-Maximum Suppression (NMS) post-processing to eliminate redundant detections. The audio alert module uses a text-to-speech engine to generate distance-calibrated spoken warnings delivered through a speaker or headphones.

VI. DATASET DESCRIPTION

The dataset used for training the pothole detection model is obtained from publicly available sources hosted on Mendeley Data and Roboflow platforms. Mendeley Data provides curated academic datasets containing road and pothole images captured under real-world conditions across various locations. Roboflow is used to import, organize, annotate, and preprocess the pothole images for model training.

The dataset includes images collected from urban and rural roads, captured using mobile and vehicle-mounted cameras under different lighting conditions, weather conditions, and road surface types. All images are manually annotated with bounding boxes indicating pothole locations. The final dataset is exported in YOLO format, consisting of image files and corresponding annotation files. Standard data augmentation techniques including flipping, rotation, brightness adjustment, and mosaic augmentation are applied during training to improve model robustness and generalization.

VII. RESULTS AND DISCUSSION

The Real-Time Pothole Detection system was evaluated using test images and live video streams from the trained YOLOv8 model with Transformer integration. The model demonstrated

accurate pothole detection across varied road conditions including different lighting levels, road surface textures, and partial occlusion scenarios. Bounding box localization was precise, and the confidence scores were consistent across diverse input conditions.

The audio alert system functioned reliably, generating timely spoken warnings when pothole detections exceeded the confidence threshold. Distance zone estimation based on bounding box area provided appropriate differentiation between near, medium, and far hazard alerts, enabling users to respond with sufficient advance notice. The integration of Transformer attention mechanisms improved detection performance for distant and partially visible potholes compared to standard CNN-based baselines, demonstrating the benefit of global contextual understanding for road hazard detection in complex real-world environments.

VIII. CONCLUSION

This paper presents a Transformer-Based Real-Time Pothole Detection System aimed at enhancing the safety and independence of visually impaired individuals. By integrating YOLOv8 with Transformer-based attention mechanisms, the system overcomes key limitations of conventional CNN-based approaches, particularly in long-distance detection and contextual understanding of road surface conditions. The proposed system successfully processes real-time video streams and provides timely audio alerts to enable safe navigation. The combination of a high-quality annotated dataset, an optimized detection model, and an accessible audio interface makes the system a practical and effective assistive technology solution for visually impaired pedestrians in real-world road environments.

IX. FUTURE SCOPE

Future enhancements include integrating the system with wearable devices such as smart glasses or chest-mounted cameras for hands-free operation. Incorporating GPS-based pothole mapping would enable community reporting and alert generation based on previously identified hazard locations. Extending the detection capability to identify multiple types of road hazards including speed bumps, open drains, and construction obstacles would improve overall navigation safety. Edge deployment on low-power embedded devices such as Raspberry Pi would enable fully mobile and self-contained assistive navigation units accessible to a wider population.

X. REFERENCES

[1] Dosovitskiy, A. et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929.

[2] Redmon, J. et al. (2016). You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788.

[3] Ultralytics. (2023). YOLOv8 Object Detection Framework Documentation. Available at: https://docs.ultralytics.com

[4] Mendeley Data. Pothole and Road Damage Image Dataset for Academic Research. Available at: https://data.mendeley.com

[5] Roboflow. Dataset Annotation, Preprocessing and Export Platform. Available at: https://roboflow.com

[6] Fan, R. et al. (2019). Pothole Detection Based on Disparity Transformation and Road Surface Modeling. IEEE Transactions on Image Processing, 29, 897-908.

[7] Maeda, H. et al. (2018). Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. Computer-Aided Civil and Infrastructure Engineering, 33(12), 1127-1141.

[8] Wang, C. Y. et al. (2022). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. arXiv:2207.02696.

