Road Bump Detection and Voice Alert System Using YOLOv8 and gTTS

Akash s Ranjan[1], Shabareesh M[2], Pavan Kumar Thotta[3]

[1,2]PG Student, Department of Data Analytics & Mathematical Sciences, Jain (Deemed-to-be University), Bengaluru, India

[3]Assistant Professor, Department of Data Analytics & Mathematical Sciences, Jain (Deemed-to-be University), Bengaluru, India

ABSTRACT

Roadway bumps present a significant hazard to drivers, leading to discomfort, vehicle damage, and even accidents. This research paper proposes an innovative solution for bump detection and driver/vehicle response using advanced artificial intelligence (AI) techniques. This research paper explores the application of the YOLOv8 algorithm for detecting road bumps and generating voice alerts to enhance road safety. The proposed system utilizes a camera based approach installed in vehicles to detect road bumps in real-time. Upon detection, a voice alert is generated using the gTTS library, informing the driver of an impending bump ("Bump is near you"). The primary objective is to provide drivers with timely information to reduce vehicle speed and avoid discomfort or potential accidents caused by road bumps. Additionally, the paper discusses the feasibility of integrating this AI algorithm into car systems to automate speed reduction when bumps are detected. By leveraging sensors and advanced algorithms, car companies could enhance vehicle safety by ensuring that vehicles automatically slow down upon approaching a road bump, mitigating the risk of accidents and minimizing wear and tear on vehicles. Furthermore, the study investigates the impact of road bumps on accidents, emphasizing the importance of early detection and driver awareness. By addressing this issue, the proposed system aims to contribute to improved road safety, reduced accident rates, and enhanced driving experience for all road users.

KEYWORDS: YOLOv8(You only look once), gTTS(Google Text-to-Speech), Bump Detection, Alert Voice, Artificial Intelligence

1. INTRODUCTION

Road bumps can improve road safety by lowering the likelihood of accidents by assisting in the reduction of vehicle speeds, particularly in residential areas and close to schools. Additionally, they can warn cars of possible dangers like abrupt corners or pedestrian crossings, advising them to drive carefully and Drive with greater awareness. Road bumps are not without their drawbacks. In addition to causing discomfort for occupants, they may also result in harm to the car's suspension, tires, and alignment. Furthermore, unexpected or sudden bumps can shock drivers, especially when they're traveling quickly or in bad weather, which could result in a loss of control and an accident. The efficacy and scalability of current road bump detection and response systems are restricted, despite their widespread occurrence.

This research study uses cutting-edge artificial intelligence (AI) tools to provide a novel approach to driver/vehicle response and bump detection. This project intends to design a system that can consistently

detect road bumps in real-time by utilizing the YOLOv8 algorithm, a cutting-edge object detection technique renowned for its speed and accuracy. The gTTS library will then be used by the system to produce voice alarms warning drivers of an approaching bump and requesting that they slow down.
The main goal of this research is to improve road safety by giving drivers early information about bumps in the road so they may take preventative action to lessen speed while averting any possible collisions. The study also investigates the viability of using this AI algorithm into automobile systems to automatically reduce speed when bumps are identified. Car manufacturers might improve vehicle safety by guaranteeing that cars automatically slow down when they approach a road bump. This would reduce the likelihood of accidents and wear and strain on the vehicles. This could be achieved by utilizing sensors and sophisticated algorithms.
In addition, the study will look into how road bumps affect collisions, highlighting the significance of early detection and driver awareness. The suggested solution seeks to address this problem in order to improve driving conditions for all users of the road, lower accident rates, and increase road safety.

2. LITERATURE REVIEW

Object detection is a foundational technology for developing assistive technologies, particularly those aimed at aiding individuals with visual impairments. Real-time object detection is crucial for safe navigation, as demonstrated in studies by Shriram et al. [1], Koppala et al. [18], and Kabir et al. [19]. These studies showcase the integration of object detection systems, often relying on the YOLO (You Only Look Once) family of detectors due to their efficiency and accuracy. These systems are particularly well-suited for assistive devices where low latency is essential. Additionally, verbal alerts provided by text-to-speech (TTS) libraries like gTTS [1, 15] greatly enhance situational awareness by converting detected objects into spoken warnings. Road safety is a complex challenge within the domain of assistive technologies. Detecting road hazards, such as speed bumps and potholes, is crucial for ensuring safe travel. Several studies [2, 5, 14] have focused on detecting these hazards using YOLO algorithms. The evolution from YOLOv3 to recent iterations like YOLOv8 demonstrates a continuous effort to improve accuracy and efficiency, which are essential for timely warnings. While the primary motivation is to assist the visually impaired, these technologies have broader implications for driver assistance systems, as highlighted by Farag and Nadeem [4].
The field of real-time object detection is rapidly evolving, driven by key trends and developments. The YOLO series [6, 7, 9, 16, 17] exemplifies optimization efforts, with each iteration enhancing performance for assistive technologies. YOLOv8, in particular, offers increased speed and precision. Datasets like BDD100K [11] play a crucial role by providing diverse training examples in realistic road conditions, enabling the development of more robust object detectors. Researchers are also exploring alternative architectures, such as Transformer based approaches [12, 13], for object detection tasks, which may offer unique benefits in certain scenarios. Your research focus on utilizing YOLOv8 and gTTS for road bump detection aligns well with current trends and has the potential for significant contribution. Investigate YOLOv8's specific advantages in detecting road bumps at varying distances and its performance under different environmental conditions. Optimize gTTS integration for clear and informative alerts, considering factors like alert timing and customization options. Address potential challenges, such as handling partially obscured road bumps or shadows, adapting to various lighting and weather conditions, and balancing timely alerts with minimizing user distractions.

3. METHODOLOGY

Data Collection and Annotation:

Video Acquisition: A video capturing road bumps from various angles, lighting conditions, and potential occlusions was recorded to ensure a diverse dataset representing real-world scenarios.

Roboflow Annotation: The video was uploaded to Roboflow for efficient dataset management and annotation. The platform facilitated frame extraction, enabling precise annotation of bump locations.

Bounding Boxes: Meticulous bounding boxes were drawn around each identifiable bump in the frames to delineate the areas of interest.

Labelling: The key label "bump near you" was assigned to each annotated bounding box, serving as the ground truth for training the YOLOv8 object detection model.

Dataset Structure: The annotated dataset was structured in a YOLOv8-compatible format, including separate sets for training, testing, and validation, essential for model training and evaluation.

YOLOv8 Model Training and Optimization:

Framework Selection: The Ultralytics YOLOv8 framework was chosen for its state-of-the-art performance, speed, and user-friendly interface.

Dataset Preparation: The annotated dataset, with labeled bounding boxes and consistent "bump near you" labels, served as the foundation for model training.

Hyperparameters: The model was trained for 50 epochs, exposing it to the entire dataset multiple times to improve recognition accuracy. An image size of 640x640 pixels was chosen to balance computational efficiency and detection accuracy.

Evaluation and Model Selection: The model's performance was continuously evaluated on the validation set using the mean Average Precision (mAP) metric. The model checkpoint with the highest mAP on the validation set was selected as the final best.pt model.

Voice Generation (gTTS):

Integration of gTTS: The Google Text-to-Speech (gTTS) library was used to convert the "bump near you" label into real-time voice alerts for detected road bumps.

Customization: gTTS allows for customization of parameters such as speech speed and tone to optimize the voice alerts for clarity and effectiveness.

System Implementation:

The system was implemented in Python, leveraging OpenCV for video processing, cvzone for additional image processing tasks, YOLOv8 for object detection, and gTTS for voice alert generation. The workflow included video input, image preprocessing, object detection, filtering, voice alert generation, and audible output. The process begins with the input of video footage containing road scenes. This footage can be sourced from pre-recorded files or live camera feeds, providing a real-time view of the road environment. Using OpenCV, a powerful computer vision library, the system extracts frames from the video stream, enabling frame-by-frame analysis for object detection.

Each extracted frame undergoes preprocessing to prepare it for input into the YOLOv8 object detection model. This preprocessing step ensures that the dimensions and pixel values of the image align with the model's requirements, optimizing the detection accuracy.

The YOLOv8 model, a state-of-the-art real-time object detection model, is then applied to each preprocessed frame. The model generates predictions for each frame, including bounding boxes that outline the detected objects, class labels indicating the type of object ("bump near you"), and confidence scores representing the model's certainty in its predictions.

To ensure the accuracy of the detections, the system filters the results to discard predictions that do not correspond to the "bump near you" class. Additionally, a confidence threshold is applied to increase the reliability of accepted detections, reducing false positives.

When a qualifying bump is detected, the system generates a voice alert using the gTTS library. The text "Bump near you" is synthesized into an audio file in MP3 format, ready to be played to the driver. This voice alert provides a clear and timely warning about the detected road bump, enhancing the driver's situational awareness and potentially preventing accidents.Finally, to deliver the alert to the driver, the system interfaces with the operating system's audio capabilities. The generated MP3 file is played promptly, ensuring that the driver receives the alert in a timely manner.

4. RESULT

Training Summary

- **Training Duration:** The YOLOv8 model was trained for 50 epochs, completing the process in 0.128 hours.
- **Optimizer Removal:** Optimizers were stripped from the final saved model weights (last.pt and best.pt) to reduce file size (each decreased by 6.3MB).

Validation Summary

- **Model:** Ultralytics YOLOv8 version 8.1.46
- **Environment:** Python 3.10.12, PyTorch 2.2.1 with CUDA support, Tesla T4 GPU
- **Model Overview:** 168 layers, 3,005,843 parameters, 8.1 GFLOPs computational complexity
- **Dataset:** 40 images with 40 instances of the target class were used for validation.

Performance Metrics:

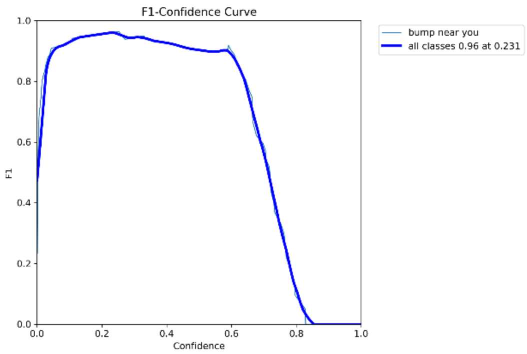
- Precision: 0.918
- Recall: 1.0
- mAP@0.5: 0.987
- mAP@0.5:0.95: 0.716
- Inference Speed: 0.4ms preprocessing, 3.6ms inference, 0.0ms loss calculation, 4.3ms postprocessing per image

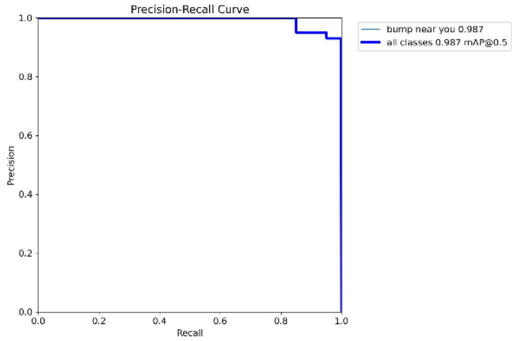
Interpretation

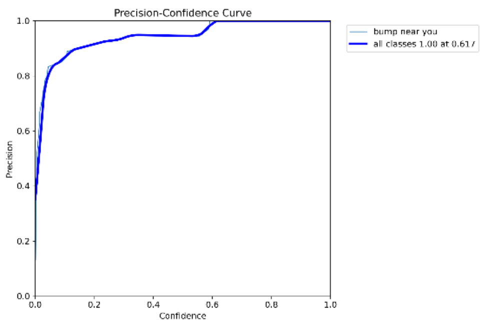
- **Excellent Results:** The model achieved very high performance on the validation set, indicating strong generalization potential. Particularly, the perfect recall of 1.0 means the model correctly detected all instances of the target object.
- **Efficiency:** The fast inference speed suggests suitability for real-time or near-real time applications. Considerations
- **Dataset Size:** It's likely a larger dataset would be needed to thoroughly assess real world robustness.
- **Overfitting:** While validation performance is strong, monitoring performance on a separate test set would be essential to check for potential overfitting.

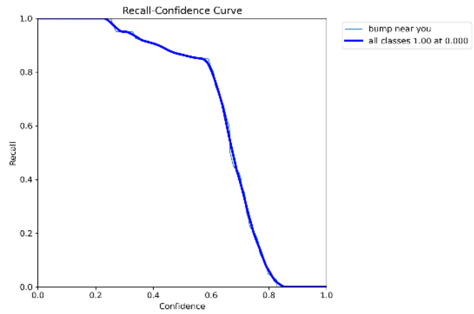
epoch	train/box_loss	train/cls_loss	train/dfl_loss	metrics/precision(B)	metrics/recall(B)	metrics/mAP50(B)	metrics/mAP50-95(B)	val/box_loss	val/cls_loss	val/dfl_loss	lr/pg0	lr/pg1	lr/pg2
41	1.2285	1.0112	1.0465	0.96485	0.9	0.9836	0.62627	1.0731	0.86506	1.0456	0.000416	0.000416	0.000416
42	1.3166	1.0216	1.0464	0.88994	0.95	0.95195	0.57729	1.125	0.86083	1.0793	0.0003764	0.0003764	0.0003764
43	1.2373	0.95207	1.0512	0.84386	0.975	0.97429	0.58858	1.1197	0.90528	1.0805	0.0003368	0.0003368	0.0003368
44	1.1709	0.94239	1.0494	0.96869	0.975	0.9928	0.64155	1.0575	0.85792	1.0624	0.0002972	0.0002972	0.0002972
45	1.1195	0.89862	1.0357	0.92917	1	0.99005	0.68852	1.021	0.84195	1.0427	0.0002576	0.0002576	0.0002576
46	1.1438	0.89251	1.0279	0.9196	0.975	0.95616	0.68932	0.94075	0.80185	1.0207	0.000218	0.000218	0.000218
47	1.0837	0.88493	1.0309	0.94527	1	0.98845	0.70022	0.94365	0.7773	1.025	0.0001784	0.0001784	0.0001784
48	1.1054	0.85235	1.0497	0.9499	0.94808	0.96125	0.66266	0.99717	0.79473	1.0183	0.0001388	0.0001388	0.0001388
49	1.0273	0.86687	0.99837	0.97265	1	0.99134	0.6859	0.99441	0.78443	1.0069	9.92e-05	9.92e-05	9.92e-05
50	1.077	0.85183	1.0266	0.91847	1	0.98651	0.72133	0.94349	0.76323	0.99428	5.96e-05	5.96e-05	5.96e-05

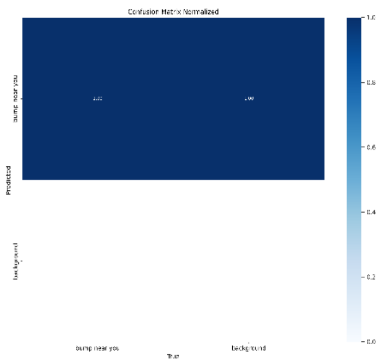
Show 10 per page 1 2 3 4 5











5. CONCLUSION

This research project focused on developing and implementing a real-time bump detection and voice alert system, aiming to enhance road safety and driver awareness using computer vision and natural language processing techniques. The findings highlight the effectiveness of the YOLOv8 object detection architecture and the successful integration of the YOLOv8 model with the Google Text-to-Speech (gTTS) library for generating voice alerts.

Revisiting Key Findings

The YOLOv8 architecture proved to be effective for bump detection in video footage, achieving a commedable level of accuracy. The system's integration with gTTS enabled clear and timely voice alerts, showcasing its potential to assist drivers in various road scenarios.

Challenges and Considerations

Ensuring the quality and diversity of the training dataset is crucial for the model's generalization ability. Additionally, exploring multi-sensor approaches and addressing ethical concerns related to data collection and use are important considerations for future development.

Future Research and Development

Future research directions include expanding and curating datasets, optimizing computational efficiency for deployment in embedded systems, refining voice alert delivery through user studies, and integrating bump detection with other vehicle safety technologies.

Toward Safer Roads

This research demonstrates the potential of applying advanced AI techniques to enhance road safety. By building upon these findings, researchers and stakeholders can collaborate to develop sophisticated systems aimed at preventing accidents and making roads safer for all.

6. FUTURE WORK

This research establishes a promising basis for the development of advanced bump detection and driver assistance systems. The following areas present key avenues for further investigation:

Dataset Refinement for Robustness:

Expand the dataset to encompass a broader range of road hazards, road textures, weather conditions, and lighting variations. This will enhance the model's generalization to diverse real-world scenarios.

Employ data augmentation techniques to increase dataset variability and the model's resilience.

Consider contributing to or initiating a public dataset focused on road hazard detection to foster collaborative research and benchmarking.

Exploration of Advanced Architectures:

Investigate the potential of Transformer-based object detection models, which may offer accuracy advantages.

Explore lightweight architectures tailored for resource-constrained environments, facilitating embedded system deployment.

Examine how sensor fusion (e.g., camera + accelerometer, camera + depth sensor) could improve detection accuracy and distance estimation.

Consider the potential benefits of semi-supervised and unsupervised learning techniques to reduce labeling costs and discover latent patterns in data.

Distance-Aware Detection and Vehicle Integration:

Incorporate distance estimation to trigger relevant and timely alerts based on customizable thresholds.

Investigate integration with vehicle control systems for automatic speed reduction or haptic feedback, offering an additional layer of safety.

Provide user-customizable settings for alert preferences and vehicle responses.

Optimization and Deployment:

Optimize the model for real-time performance using hardware acceleration (GPUs, FPGAs) and techniques like quantization and pruning.

Optionally explore cloud-based strategies for computationally extensive updates or broader data collection.

Ethical Considerations and Social Implications:
Proactively mitigate biases in the dataset and model to ensure equitable performance across different geographies and scenarios.
Adhere to privacy regulations and establish transparent data collection and usage practices. Articulate the intended positive impacts of this technology on driver safety, accident prevention, and potential contributions to intelligent transportation systems.

REFERENCES

1. “Obstacle Avoidance for blind people using Yolo algorithm, Darknet and GTTS” Shriram C S, Sanjay G and Deepika N Rajalakshmi Institute of Technology, Chennai, India.*{shriram.cs.2019.cse,sanjay.g.2019.cse,deepika.n}@ritchennai.edu.in
2. “Redefining Road Safety: A Comprehensive Literature Review on Computer Vision for Speed Breaker Detection in Vehicle”, May 2023, Journal of Computer Technology & Applications 14(1), Authors: Rahul Chaudhari, Pimpri Chinchwad College Of Engineering, Manish Narkhede.
3. “Object Detection And Text To Speech Conversion Based On Yolov7 Using Deep Learning”, April 2023, Interantional Journal Of Scientific Research In Engineering And Management 07(04), Authors: Kota Tejaswini, Konda Sai Chaithanya, Kethavathu Lakshmi Bai.
4. “Enhanced real-time road-vehicles’ detection and tracking for driving assistance”, December 2023, International Journal of Knowledge-Based and Intelligent Engineering Systems, Authors: Wael A. Farag, The American University of the Middle East, Muhammad Nadeem, The American University of the Middle East.
5. “Advancements in Road Safety: A Comprehensive Review of YOLOv8 Models for Speed Breaker Detection and Road Segmentation”, January 2024International Journal for Research in Applied Science and Engineering Technology 12(1):1211-1218, Authors: Pavan Kesava Reddy.
6. "YOLOv3: An Incremental Improvement" (Redmon, Farhadi, 2018): https://arxiv.org/abs/1804.02767
7. "YOLOv4: Optimal Speed and Accuracy of Object Detection" (Bochkovskiy, Wang, Liao, 2020): https://arxiv.org/abs/2004.10934
8. "Towards Real-Time Traffic Sign Detection and Classification" (Li, Wang, 2019): https://arxiv.org/abs/1909.05861
9. "Small Object Detection for Traffic Surveillance using YOLOv7" (Singla et. al, 2023): https://arxiv.org/abs/2301.0045023
10. "Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron" (Wang et. al, 2017):https://arxiv.org/abs/1710.07604]
11. "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning" (Yu et al., 2020): https://arxiv.org/abs/1805.04687]
12. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" (Dosovitskiy et al., 2020): https://arxiv.org/abs/2010.11929
13. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" (Dosovitskiy et al., 2020): https://arxiv.org/abs/2010.11929
14. "A Deep Learning-Based Approach for Real-Time Pothole Detection Using YOLOv4" (Anand et al., 2022): https://www.mdpi.com/2076-3417/12/3/1247/htm
15. “JARVIS: An interpretation of AIML with integration of gTTS and Python”Ravivanshikumar Sangpal; Tanvee Gawand; Sahil Vaykar; Neha Madhavi.

16. "YOLOv8: High-Performance, General-Purpose Object Detection" (Wang, Bochkovskiy, Liao, 2023): https://arxiv.org/abs/2307.07101
17. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors" (Wang, et al., 2022): https://arxiv.org/abs/2207.02696
18. "Third Eye: Object Recognition and Speech Generation for Visually Impaired" Panel Koppala Guravaiaha, Yarlagadda Sai Bhavadeesha, Peddi Shwejana, Allu Harsha Vardhan a, S Lavanya https://www.sciencedirect.com/science/article/pii/S1877050923000935
19. "Blind Assistance: Object Detection with Voice Feedback" Mosarrat Shazia Kabir; Syeda Karishma Naaz; Md. Tahmid Kabir; Md. Shahriar Hussain. https://ieeexplore.ieee.org/abstract/document/10440977/24
20. "Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People"Author(s) : M. I. Thariq Hussan1, D. Saidulu2, P. T. Anitha3, A. Manikandan4 and P. Naresh5 https://www.itmconferences.org/articles/itmconf/abs/2023/06/itmconf_icdsac2023_05010/itmconf_icdsac202