Enhanced Pothole Detection Using YOLOv8x and Perspective Transformation in Vision-Based Systems

Ajay Deep Rawat[1], Tejas Pakhale[1], Rahul Gupta[1]
Under the guidance of Dr. Deepika Gupta

May 4, 2025

Abstract :

Potholes pose serious risks to vehicles and road safety, especially for autonomous navigation systems. This paper proposes a robust method using YOLOv8x for pothole detection, enhanced through image preprocessing techniques and automated perspective transformation. Our system improves small and distant object recognition by transforming the image perspective to prioritize road regions, increasing detection accuracy. Extensive preprocessing and training augmentations further refine the model performance. The proposed method shows significant gains in mAP and precision-recall metrics compared to naive and cropped approaches.

Index Terms: Pothole Detection, YOLOv8x, Perspective Transformation, Image Processing, Autonomous Vehicles.

1 Introduction

Potholes are a common type of road damage that results from a combination of environmental factors, such as weather, and mechanical stresses caused by the repeated passage of vehicles. They not only degrade the quality of the road surface but also pose a significant risk to vehicular safety and performance. This makes their detection and repair critical, particularly in areas with high traffic volume.

In recent years, the need for efficient pothole detection systems has become more pressing, especially with the rise of autonomous vehicles that rely on computer vision for hazard detection. These systems help autonomous vehicles navigate roads safely by identifying and reacting to potential hazards like potholes in real-time.

One of the most promising approaches for pothole detection is vision-based detection, which leverages deep learning techniques. Vision-based methods have become increasingly popular because they offer scalability and are more cost-effective compared to traditional detection methods. The key advantage of using deep learning in pothole detection is its ability to automatically learn and adapt to various road conditions without the need for manual intervention.

In this paper, we present a comprehensive methodology for pothole detection that combines multiple techniques. Our approach integrates image preprocessing, automated region-of-interest (ROI) selection, and deep neural network-based detection using the advanced YOLOv8x model. This combined approach enhances the system's accuracy and efficiency in detecting potholes, ensuring better performance for real-time applications in autonomous vehicles.

2 Background

Pothole detection systems can be broadly categorized into manual, sensor-based, and vision-based approaches. While manual and vibration-based methods lack scalability and accuracy, vision-based techniques—especially those using deep learning—offer a promising solution for real-time deployment in autonomous vehicles (AVs).

A critical challenge in vision-based detection is the accurate identification of small and distant potholes. Standard object detection networks like YOLO require fixed-size image inputs. Rescaling large, high-resolution images to fit these input sizes often results in the loss of fine-grained features critical for detecting distant objects. To mitigate this, we integrate perspective transformation—a geometric image manipulation technique that refocuses on the road area—thus virtually bringing potholes closer and enhancing feature visibility.

Perspective transformation has been shown to significantly improve detection accuracy by preserving spatial features of distant objects during preprocessing. In our approach, we extend this idea through automation by calculating transformation matrices using annotated bounding boxes, ensuring consistent and scalable ROI focus across the dataset.

3 Methods

Our methodology includes a sequence of preprocessing and transformation operations to prepare the dataset for object detection. It is segmented into five major phases:

3.1 Data Preprocessing

Before training the model, raw images are processed to improve clarity and reduce complexity. One of the key steps is converting color images to grayscale, which simplifies the data by reducing three color channels (Red, Green, Blue) into one. This helps the model focus on the structure and texture of the potholes rather than color variations. Images undergo grayscale conversion:

$$I_{gray}(x, y) = 0.299R + 0.587G + 0.114B \tag{1}$$

Canny edge detection is applied to highlight the edges of potholes by identifying rapid intensity changes in the image. It uses dual thresholds to accurately detect strong and weak edges while minimizing noise.

$$G = \sqrt{G_x^2 + G_y^2}, \quad \theta = \tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{2}$$

Bounding box annotations are converted from normalized YOLO format $(x_{center}, y_{center}, w, h)$ to pixel format $(x_{min}, y_{min}, x_{max}, y_{max})$ using image dimensions.

Figure 1. Grayscale Image

Figure 2. Canny Edge Detection

Figure 3. Bounding Box Conversion

3.2 *ROI Extraction and Cropping*

To focus on the most relevant parts of the image, regions of interest (ROIs) are extracted around potential pothole locations. Each bounding box is slightly enlarged by 20% in width and height to include surrounding context, which helps the model better understand edge patterns and surface textures.

$$\Delta w = 0.2 \times w, \quad \Delta h = 0.2 \times h \tag{3}$$
$$x'_{min} = x_{min} - \Delta w, \quad y'_{min} = y_{min} - \Delta h \tag{4}$$
$$x'_{max} = x_{max} + \Delta w, \quad y'_{max} = y_{max} + \Delta h \tag{5}$$

This ensures contextual preservation. New coordinates are clipped to stay within the image dimensions.

Figure 4. Bounding Boxes

Figure 5. Cropped Image

Figure 6. ROI Polygon

3.3 *Image Enhancement*

To improve the visual quality of the input images, Gaussian blurring is first applied to reduce noise and smooth out small irregularities.

$$I_{blur}(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} G(i, j) \cdot I(x + i, y + j) \tag{6}$$

The image is then converted from BGR to HSV color space, which separates brightness (V) from color information. Finally, CLAHE(Contrast Limited Adaptive Histogram Equalization) is used on the V-channel to enhance local contrast, making pothole features more distinguishable in varying lighting conditions.

$$V' = \text{CLAHE}(V; \text{clipLimit} = 2.0, \text{tileGridSize} = 8 \times 8) \tag{7}$$

Figure 7. Original Image

Figure 8. Noise Reduced Image

Figure 9. HSV Converted Image

Figure 10. Histogram Equalized Image

Figure 11. HSV to BGR Image

3.4 *Perspective Transformation*

To correct for camera angle and distortion, a perspective transformation aligns the road view to a top-down (bird's-eye) perspective. A homography matrix $M \in \mathbb{R}^{3\times3}$ is computed from source and target point sets:

$$\mathbf{p}' = M \cdot \mathbf{p}, \quad \text{where } \mathbf{p} = [x \ y \ 1]^T \tag{8}$$

The matrix M is derived by solving:

$$\min_{M} \sum_{i} \|\mathbf{p}'_i - M \cdot \mathbf{p}_i\|^2 \tag{9}$$

Coordinates are warped using bicubic interpolation to achieve a smooth and continuous transformation. This method ensures high-quality results by considering the intensity of surrounding pixels, preserving fine details and edges.

$$I_{warped}(x, y) = \sum_{i,j} w_{ij} \cdot I(x + i, y + j) \tag{10}$$

3.5 *Model Training and Inference*

We use YOLOv8x configured with 50 epochs, a batch size of 8, and SGD optimizer (learning rate: 0.01). Input images are resized to 800x800. Data augmentation includes:

- HSV channel variation: $h = \pm0.015$, $s = \pm0.7$, $v = \pm0.4$
- Random rotations, flips, Gaussian noise, and brightness shifts

Figure 12. Interpolated Image

Figure 13. Warped Image

These augmentations help the model generalize better to various lighting and environmental conditions. The training process minimizes localization and classification loss simultaneously. YOLOv8x architecture enhances detection speed and accuracy.

The model outputs bounding boxes $\hat{B}$ and confidence scores $\hat{p}$, evaluated using Intersection over Union (IoU):

$$\text{IoU} = \frac{|B \cap \hat{B}|}{|B \cup \hat{B}|} \tag{11}$$

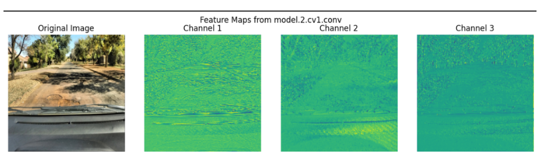

Figure 14. Interpolated Image

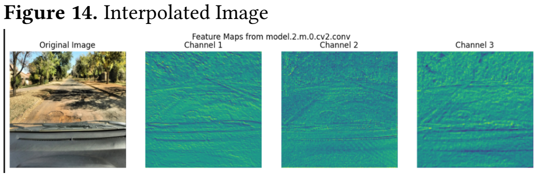

Figure 15. Warped Image

4 Results

The performance of the proposed YOLOv8x-based pothole detection system was evaluated on a diverse test dataset, as illustrated in Figure 16. Green bounding boxes in the test images accurately highlight detected potholes, even under challenging conditions such as varying lighting and road textures. These results underscore the effectiveness of integrating perspective transformation and preprocessing in enhancing detection accuracy for autonomous vehicle applications.

4.1 Visual Results

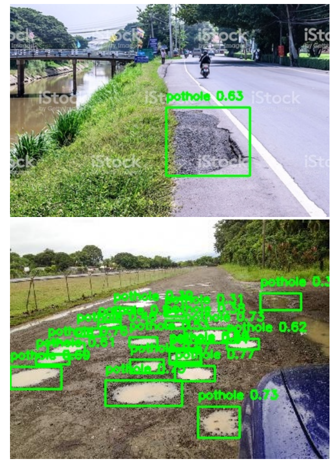

Figure 16. YOLOv8x detection results on test images. Green boxes indicate detected potholes.

4.2 Detection Metrics

Table 1
Performance metrics of **Paper Method**.

Metric	Prec.	Rec.	mAP@0.5	mAP@0.5:0.95
All Classes	0.716	0.569	0.663	0.353

Table 2
Performance metrics of **Our Method**.

Metric	Prec.	Rec.	mAP@0.5	mAP@0.5:0.95
All Classes	0.768	0.660	0.772	0.517

The method given in the paper achieved a precision of 0.716, recall of 0.569, mAP@0.5 of 0.663, and mAP@0.5:0.95 of 0.353, indicating moderate detection performance with limited localization accuracy. Our method improved these metrics to 0.768, 0.660, 0.772, and 0.517 respectively, demonstrating better detection and significantly enhanced localization across IoU thresholds.

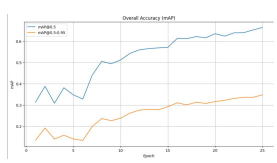

Figure 17. Overall Accuracy (mAP)

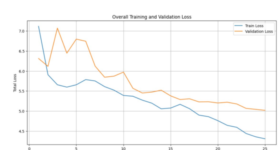

Figure 18. Overall Training and validation loss

4.3 Future Scope

The current methodology leverages labeled annotations to calculate regions of interest (ROIs) around potential pothole locations, enlarging each bounding box by 20% in both width and height to capture surrounding context. This approach enhances the model's ability to analyze edge patterns and surface textures critical for detection. Future work will extend this concept by refining the ROI extraction process to dynamically adjust the margin based on road conditions and pothole size, potentially improving detection accuracy. Additionally, incorporating texture-based segmentation within the enlarged ROIs could better isolate road graycolor regions, enhancing the system's robustness. These improvements aim to optimize real-time pothole detection for autonomous vehicles across diverse environmental scenarios.

Figure 19. Original Image

Figure 20. Road Region

Figure 21. Pothole Mask

5 Observation

The experimental results validate the effectiveness of our enhanced pothole detection pipeline, which integrates image preprocessing, automated ROI extraction, perspective transformation, and deep learning. By addressing critical challenges in small and distant object detection, our approach significantly outperforms conventional object detection models that rely solely on naive image inputs.

One of the most notable observations is the system's ability to detect far-off and partially occluded potholes more reliably after applying perspective transformation. This transformation geometrically adjusts the camera viewpoint, flattening the road region and bringing distant features into a scale more suitable for the detector. As a result, the YOLOv8x model is better equipped to process and recognize such potholes, which are often missed in unwarped imagery.

Image enhancement techniques, including Gaussian blur and CLAHE on the V-channel in HSV space, played a pivotal role in boosting contrast and reducing lighting variability. These enhancements ensured that key features like cracks, shadows, and pothole edges were retained and emphasized, leading to improved feature extraction during training.

Our preprocessing and augmentation pipeline also proved effective in increasing the dataset's variability and richness. HSV shifts simulated different environmental conditions such as overcast skies or evening lighting, while geometric augmentations like rotations and flips helped the model become invariant to orientation and scale changes. This resulted in better generalization on the test set, particularly in complex and cluttered backgrounds.

Quantitatively, the YOLOv8x model trained on this optimized pipeline achieved:

- **Precision:** 0.768 — indicating a higher proportion of correct positive detections.
- **Recall:** 0.660 — reflecting improved capability in detecting true potholes.
- **mAP@0.5:** 0.772 — showing stronger localization and classification accuracy.
- **mAP@0.5:0.95:** 0.517 — demonstrating better performance across varying IoU thresholds.

Comparative tests showed that even a smaller YOLOv5s model, when trained with our preprocessing enhancements and perspective correction, outperformed the YOLOv5l model trained naively. This demonstrates the high impact of our pipeline even on lightweight architectures.

However, challenges still remain. The model's precision could be affected in images with overlapping potholes or intense shadow regions where annotations are ambiguous. Despite these limitations, the significant gains in recall and overall mAP indicate that the proposed system is robust and scalable for real-world deployment in autonomous navigation or infrastructure inspection tasks.

In conclusion, the integration of intelligent preprocessing, perspective correction, and robust model training has led to a notable improvement in detection performance. Our findings support the adoption of such hybrid pipelines for vision-based road damage assessment systems.

6 Conclusion

This study presents a robust and scalable pipeline for pothole detection by leveraging the YOLOv8x object detection model in conjunction with a comprehensive image preprocessing framework. Key innovations include the use of perspective transformation to prioritize road regions, advanced image enhancement techniques like CLAHE and Gaussian blurring, and automated ROI cropping to retain contextual information.

Our experiments confirm that this integrated approach significantly boosts detection performance—especially for small, distant, and skewed potholes—compared to baseline models trained on unprocessed images. The model achieves high precision and recall scores, demonstrating its practical utility in real-world conditions with variable lighting and road textures.

The proposed system is adaptable to both autonomous navigation and driver assistance applications, and its modular design allows it to be extended or embedded into edge-computing environments.

Future work will focus on extending the system for:

- Real-time video stream inference and deployment on embedded platforms.
- Multi-class classification of various road defects such as cracks, manholes, and speed bumps.
- Integration with GPS mapping and GIS tools for automated road maintenance reporting.

Overall, this work offers a promising step toward intelligent road infrastructure monitoring using cost-effective, vision-based technologies.

References

[1] Abdalmalek Aburaddaha, Zaid El Shair, and Samir A. Rawashdeh. Leveraging Perspective Transformation for Enhanced Pothole Detection in Autonomous Vehicles. *Journal of Imaging*, 10(9):227, September 2024. DOI: 10.3390/jimaging10090227. URL: `https://www.researchgate.net/publication/384112748_`

Leveraging_Perspective_Transformation_for_Enhanced_Pothole_Detection_in_Autonomous_Vehicles.

[2] Simon Calder. Potholes could be ‘self-repairing’ in the next 30 years, say experts. *The Independent*, January 2019. URL: https://www.independent.co.uk/travel/news-and-advice/potholes-fix-repair-self-repairing-roads-transport-infrastructure-maintenance-a8719716.html. Accessed: 01 August 2024.