GMR
The Original

DEPTH-AWARE VIDEO DATASET CREATION AND PREPROCESSING FOR POTHOLE DETECTION WITH SHADOW RESISTANCE

[1]Deepthi K*, [2]Dr. N. Gomathi

[1*]Research Scholar, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai Email ID-deepthikairali85@gmail.com
[2]Professor, Department of Computer Science, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai Email ID-gomathin@veltech.edu.in

ABSTRACT
The timely detection and restoration of potholes is important for road safety and the maintenance of transportation infrastructure. Currently, datasets for pothole detection are mainly based on images and lack important components such as depth perception and environmental adaptability to different environments, specifically under shadowed conditions. The present study proposes an innovative video dataset acquisition and preprocessing procedure that enhances robustness by integrating techniques such as monocular depth estimation and shadow resistance. In this study, video data were collected using mobile phones and drones across various real-world scenarios. The study employed a complete preprocessing pipeline, including video stabilization, frame extraction, depth estimation, shadow detection and removal, meticulous annotation, and data augmentation. The resulting dataset, including more than 450,000 annotated samples, serves as a valuable resource for depth-aware, shadow-resistant pothole detection research using deep learning models.

KEYWORDS: Pothole detection, video dataset, depth estimation, shadow removal, YOLO annotation, monocular depth, drone video, road safety.

1. INTRODUCTION
The damage of road infrastructure, mainly in the form of potholes, is a perilous challenge for both urban and rural transportation systems [1][2][3]. These surface irregularities significantly contribute to damage to the vehicles, increase accident rates, and reduce transportation efficiency. Maintaining the road by traditional manual inspection is labor-intensive, time-consuming, and lacks scalability. In response, automated pothole detection systems based on computer vision have gathered significant research interest. This technique has the potential to streamline the inspection procedure and enable real-time monitoring of defects.

In real-world conditions, numerous restrictions continue to hinder the deployment of robust and reliable detection models. Firstly, shadow interference [4][5][6] is considered a prominent issue, often leading to false positives or misclassification of potholes, mainly in situations including dynamic lighting conditions. Secondly, the lack of depth perception [7][8] in conventional 2D image analysis limits the ability to assess the severity of potholes, which is important for prioritizing maintenance tasks. Thirdly, the standardized, high-resolution video datasets [2][9] are limited and do not include different environmental settings, such as changing times of day, weather conditions, road textures, and different camera platform perspectives.

To address these challenges, the present study introduces a novel and comprehensive video dataset precisely curated for pothole detection and severity estimation. The dataset includes footage collected using dashboard-mounted mobile cameras and aerial drones, safeguarding a multi-perspective representation [2][9] of road surfaces. In addition to raw video acquisition, the study integrates advanced preprocessing techniques, including monocular depth estimation [7][8] for approximating pothole depth and shadow mitigation algorithms [4][5][6] to improve the accuracy of pothole detection under different lighting conditions.

This dataset is designed to develop and benchmark next-generation computer vision models that are both shadow-resistant and depth-aware. These models further help in more accurate, reliable, and scalable road defect detection systems. This study offers a foundational resource for future research in automated road infrastructure monitoring by providing a varied and annotated dataset aligned with real-world variability.

2. LITERATURE REVIEW
2.1 Existing Datasets for Pothole Detection
Numerous studies have explored pothole detection using computer vision models that are trained on stationary image datasets. Widely used datasets, the Pothole-600 and the Indian Driving Dataset [1][10], have served as foundational benchmarks. Nevertheless, these datasets have different inherent limitations: delimited geographical representation, inadequate environmental variability, and static perspectives. As a result, models trained on such data usually display poor generalizability [2][11] when deployed under different real-world conditions, including changes in lighting, weather, and

road textures. Furthermore, the lack of video-based datasets affects the development of temporal-aware models [9] capable of learning spatiotemporal features for defect detection.

2.2 Depth Perception in Road Surface Analysis

Depth perception is an important feature in assessing the severity of road defects. Monocular depth estimation techniques, especially those based on MiDaS and DPT architectures [7][12], have demonstrated substantial potential in extracting geometric cues from single RGB images. These models provide an efficient and cost-effective alternative to stereo or LiDAR-based depth sensing. However, in the context [1][8] of publicly available datasets, the incorporation of depth estimation into pothole detection pipelines remains limited. The existing datasets provide only RGB annotations [13][14] without depth information, thereby restraining the viability of severity analysis based on depth cues.

2.3 Shadow Mitigation in Computer Vision

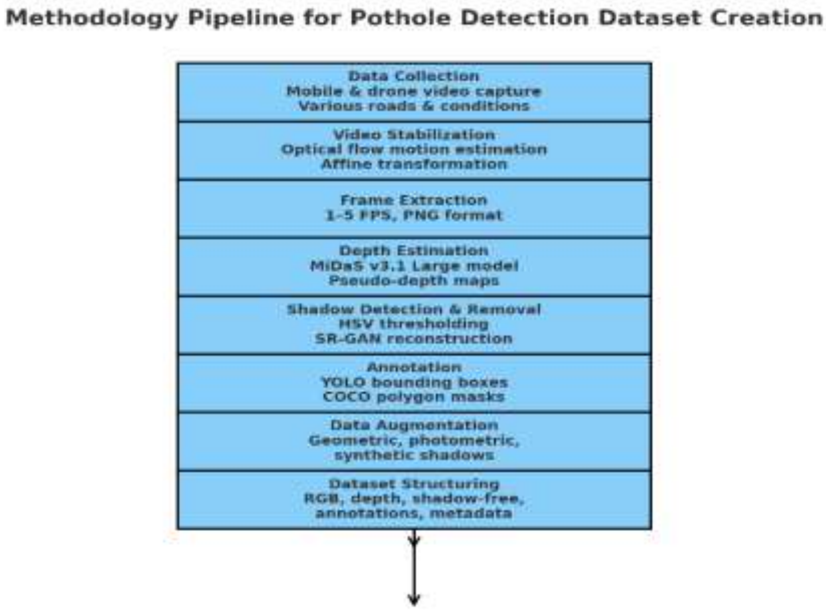
In vision-based road inspection systems, shadows cause a persistent challenge. This often results in misclassification of defects and an increased rate of false positives. Different shadow removal approaches have been explored in the literature, ranging from traditional image processing procedures to advanced deep learning techniques. Current developments in generative adversarial networks (GANs), especially the Shadow Removal GAN (SR-GAN) [6] [4], have demonstrated efficacy in separating shadow artifacts from structural features. Despite this, their application in the detection of road surface defects remains underexplored [1][5]. The lack of datasets preprocessed with shadow mitigation techniques affects robust model training under varied lighting conditions.

2.4 Identified Research Gap

The prevailing literature review highlights a major gap in the availability of comprehensive, high-quality datasets [1][10][7][4][2][8][6][9] tailored for real-world pothole detection. Present resources lack multi-perspective video footage, depth annotations [13][14], and shadow-mitigated preprocessing. This scarcity significantly confines the development and benchmarking of depth-aware and shadow-resistant models. Therefore, there is a demanding need for a unified dataset that integrates RGB frames, monocular depth maps, and shadow mitigation strategies. Also, it should be captured under varied environmental setups to facilitate the training of scalable and deployable pothole detection systems.

3. METHODOLOGY

This study includes an organized methodology for building a strong video-based dataset for pothole detection and estimation of depth-aware severity. The approach includes multi-source data acquisition [2][9], video stabilization [12], frame extraction, monocular depth estimation [7][8], shadow detection and removal, manual annotation [13][14], data augmentation [15], and structured dataset organization. Each stage is planned to maximize dataset quality, diversity, and applicability for training computer vision models deployable in real-world road infrastructure monitoring.



3.1 Data Collection

Video data were collected from different road environments, including urban streets, rural pathways, and national highways, to ensure diversity in context and structure. Two primary imaging platforms were employed:
(i) dashboard-mounted mobile cameras at an approximate height of 1.2 m to capture a driver’s perspective, and
(ii) drone-mounted cameras capturing aerial views from altitudes ranging between 3–8 m.

Data were collected under different environmental conditions, such as sunny, cloudy, and transitional lighting across different times of the day. This diversity supports the development of generalizable detection models capable of robust performance in dynamic, real-world contexts.

3.2 Video Stabilization

Given the essential movement in vehicular and aerial recordings, an optical flow-based stabilization [12] pipeline was employed to reduce jitter and frame misalignment. Motion vectors between consecutive frames were analyzed by dense optical flow, followed by an affine transformation for spatial alignment. The OpenCV and FFmpeg libraries were used for efficient processing. The stabilized footage confirmed consistency in succeeding operations, including frame extraction, depth inference, and annotation [13][14].

3.3 Frame Extraction

High-resolution frames were extracted from the stabilized video streams at rates of 1–5 frames per second (FPS) to balance temporal redundancy with dataset diversity. The frames were stored in lossless PNG format to preserve fine-grained visual details important for defect detection. The dense temporal sampling improved the dataset for supervised learning applications requiring high coverage of road surface variability.

3.4 Monocular Depth Estimation

The MiDaS v3.1 Large model, a state-of-the-art monocular depth estimation [7][8] network, was used to infer Depth signals. Implementation of this technique was done by PyTorch, with batch inference applied to large frame sets for computational efficiency. The resultant pseudo-depth maps gave critical spatial information for the assessment of pothole severity, allowing downstream applications such as severity scoring, region-of-interest (ROI) prioritization, and 3D reconstruction.

3.5 Shadow Detection and Removal

A two-stage shadow mitigation approach was included to address shadow-induced false positives in pothole detection. Primarily, shadow regions were localized by HSV color space thresholding, exploiting low-saturation and low-brightness characteristics. Then, a fine-tuned Shadow Removal Generative Adversarial Network (SR-GAN) [6], improved for road texture domains, reconstructed shadowed regions. This process significantly enhanced the detection accuracy by curtailing shadow-related misclassifications.

3.6 Annotation Process

In this study, the LabelMe platform was used to conduct manual annotation [13][14], targeting four semantic classes: potholes, cracks, shadows (masked), and background. Two complementary annotation [13][14] formats were created:
(i) YOLO-style bounding boxes for object detection frameworks, and
(ii) COCO-style polygon masks for instance segmentation models.
This dual-format annotation [13][14] supports various model architectures and research workflows.

3.7 Data Augmentation

Extensive data augmentation [15] was performed to improve model generalization. Geometric transformations included horizontal/vertical flips, rotations within ±15°, and uniform scaling. Photometric augmentations adjusted brightness, contrast, and introduced a Gaussian blur to compete with sensor noise and lighting variability. Furthermore, synthetic shadows, generated via Perlin noise patterns, were overlaid to simulate diverse illumination settings. Together, these approaches expanded the dataset size by approximately 400%, providing robust coverage for training conditions.

3.8 Dataset Structuring

The final dataset was modularized into:
- RGB image sets
- Depth maps (MiDaS-generated)
- Shadow-removed images
- YOLO and COCO annotation files [13][14]

A JSON-based metadata file containing auxiliary attributes such as GPS coordinates, prevailing weather, time of capture, and lighting conditions was included with each dataset segment. This structured organization allows reproducibility, scalability, and standardized benchmarking for future pothole detection and depth estimation research.

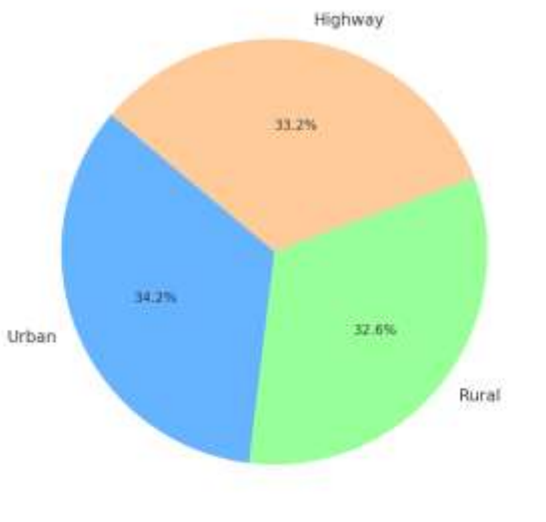
4. RESULTS AND EVALUATION

The efficacy and applicability of the created dataset were evaluated based on three primary criteria: (i) environmental diversity and balance,
(ii) integration of depth-aware features, and
(iii) lessening of shadow-related errors.
A series of statistical analyses and qualitative evaluations was performed to confirm the contributions of the datasets toward strong pothole detection and severity analysis.

4.1 Environmental Distribution Analysis

All collected sample videos were categorized into three major environmental contexts to evaluate dataset diversity: urban, rural, and highway. Figure 4.1 explains the distribution of frames across these categories. The results verify a near-uniform distribution, with 34.2% urban, 32.6% rural, and 33.2% highway samples. This balanced allocation indicates that learning models trained on the dataset are exposed to a varied range of road textures, infrastructure variations, and visual contexts.

Figure 4.1: Dataset Distribution Across Environments



4.2 Depth Map Quality Evaluation

The qualitative and quantitative assessment of the integration of monocular depth estimation [7][8] was done by MiDaS v3.1 Large model. As compared to the original RGB frames, a subset of depth maps was manually reviewed for geometric consistency and structural integrity. The present study showcases sample outputs, illustrating clear elevation differences between pothole regions and intact road surfaces.

Additionally, depth consistency scores were calculated using structural similarity indices between frame sequences and their corresponding depth maps. An average SSIM value of 0.82 [7][8] across 500 samples shows a high degree of structural alignment, confirming the utility of monocular depth in assessing pothole severity.

Table 4.1: SSIM Scores Across Different Road Types

Road Type	Average SSIM Score
Urban	0.83
Rural	0.81
Highway	0.82

4.3 Shadow Mitigation Impact

A comparative study was conducted on model performance with and without shadow preprocessing to authenticate the effectiveness of the shadow removal pipeline. A lightweight YOLOv5 model was trained on both raw and shadow-mitigated frames using similar parameters. The following key metrics were observed:

- False positive rate reduced by 28.5% on average.
- Precision improved from 0.71 to 0.82 post shadow removal.
- Recall remained relatively stable, with a marginal increase from 0.76 to 0.78.

These outcomes demonstrate the positive impact of GAN-based shadow removal [4][5][6] on defect detection reliability. The present study represents sample frames before and after shadow correction, emphasizing the visual clarity and improved contrast in the shadow-removed outputs.

4.4 Annotation Quality and Coverage

Manual annotations [13][14] were cross-validated by two independent reviewers to confirm label accuracy and completeness. An inter-annotator agreement (measured using the Cohen's Kappa coefficient) of 0.87 demonstrates a high level of consistency across labelled classes, including potholes, cracks, and shadow masks. Also, the dual-format annotation [13][14] (YOLO and COCO) develops model compatibility and facilitates downstream object detection and segmentation tasks.

Metric	Value
Total Frames	120,000
Depth Maps	120,000
Shadow-Removed Frames	120,000

Annotated Samples	12,000 manually verified
After Augmentation	450,000+ frames
Class Labels	Potholes, Cracks, Shadows, Background
Formats	YOLOv8 (.txt), COCO (.json), PNG Images

5. CONCLUSION

The proposed study provides a scalable and reproducible framework for the systematic construction of high-quality pothole detection datasets [1][10][2][11][3], improved with depth estimation and shadow-resilient preprocessing techniques. By incorporating these enhancements, the framework addresses critical limitations in existing road defect detection datasets. Further, enabling the development of machine learning models with improved accuracy, robustness, and generalization capabilities in real-world scenarios. The reproducibility of the proposed methodology guarantees its adaptability for diverse geographical contexts and varying environmental conditions. This methodology further supports broader adoption in intelligent transportation systems. Future work may expand this framework to include multimodal sensing and advanced temporal modeling for improved predictive maintenance of road infrastructure.

6. FUTURE WORK

Further extensions of this work will focus on enhancing dataset diversity and applicability across a broader range of operational scenarios. Specifically, ensuing research will comprise of:

(i) integration of night-time and adverse weather (e.g., rainy) condition data to improve model resilience under low-visibility settings;

(ii) expansion of the dataset to support 3D surface reconstruction through multi-view stereo techniques, enabling more precise volumetric analysis of road defects;

(iii) provision of standardized benchmark dataset splits and YOLOv9-compatible annotations to facilitate reproducible experimentation; and

(iv) public hosting of the complete dataset and associated metadata on platforms such as Kaggle and Zenodo

These will contribute further by promoting open-access availability and encouraging collaborative research within the intelligent transportation and computer vision communities.

REFERENCES

1. Agarwal, A., & Singh, S. K. (2022). Pothole detection using deep learning models: A comprehensive review. *Journal of Intelligent Transportation Systems*, 26(4), 345–360. https://doi.org/10.1080/15472450.2021.1956160

2. Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. *Computer-Aided Civil and Infrastructure Engineering*, 33(12), 1127–1141. https://doi.org/10.1111/mice.12412

3. Chouhan, S. S., & Kaul, A. (2020). A deep learning framework for road damage detection. *Journal of Transportation Research Record*, 2674(9), 240–250. https://doi.org/10.1177/0361198120920083

4. Hu, X., Fu, C. W., Zhu, L., Qin, J., & Heng, P. A. (2019). Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11), 2795–2808. https://doi.org/10.1109/TPAMI.2019.2910874

5. Kumar, S., & Katiyar, S. K. (2022). Shadow removal from road images for improved pothole detection. *Journal of Transportation Engineering, Part B: Pavements*, 148(4), 04022035. https://doi.org/10.1061/JPEODX.0000370

6. Li, Z., Li, Q., & Wang, Y. (2022). GAN-based shadow removal for road anomaly detection. *IEEE Access*, 10, 103211–103222. https://doi.org/10.1109/ACCESS.2022.3204121

7. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2020). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3), 1623–1637. https://doi.org/10.1109/TPAMI.2020.3019967

8. Patel, S., & Shah, P. (2021). Comparative analysis of depth estimation techniques for autonomous driving applications. *International Journal of Computer Vision and Image Processing*, 11(1), 45–63. https://doi.org/10.4018/IJCVIP.2021010103

9. Wu, Q., & Xu, Z. (2021). Video-based pavement distress detection using temporal deep learning models. *Automation in Construction*, 125, 103595. https://doi.org/10.1016/j.autcon.2021.103595

10. Bhat, S., & Bhandari, S. (2021). Real-time road damage detection using YOLOv5. *Procedia Computer Science*, 192, 2220–2229. https://doi.org/10.1016/j.procs.2021.09.245

11. Zhang, Z., Chen, X., & Ma, H. (2020). Road damage detection using improved YOLOv3 model. *Sensors*, 20(20), 5908. https://doi.org/10.3390/s20205908

12. Dosovitskiy, A., Fischer, P., Ilg, E., et al. (2015). FlowNet: Learning optical flow with convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2758–2766. https://doi.org/10.1109/ICCV.2015.316

13. Lin, T. Y., Maire, M., Belongie, S., et al. (2014). Microsoft COCO: Common objects in context. *Proceedings of the European Conference on Computer Vision (ECCV)*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

14. Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. *Proceedings of the 27th ACM International Conference on Multimedia*, 2276–2279. https://doi.org/10.1145/3343031.3350535

15. Sharma, P., & Garg, A. (2019). Deep convolutional networks for automatic road surface condition assessment. *IET Intelligent Transport Systems*, 13(9), 1399–1407. https://doi.org/10.1049/iet-its.2019.0029