Enhanced Pothole Detection using Yolov9 with Bayesian Optimizer and Segment Anything Model

Amarti Dhamele
School of Technology Management and Engineering
SVKM's NMIMS

Kharghar, India
amarti.dhamele151@nmims.in

Arohi Paigavan
School of Technology Management and Engineering
SVKM's NMIMS

Kharghar, India
arohi.paigavan083@nmims.in

Dr. Asha Rawat
School of Technology Management and Engineering
SVKM's NMIMS

Kharghar, India
asha.rawat@nmims.in

Abstract*— Pothole accidents are growing in India, and the condition is worse during the rainy season because road conditions decay very fast. Recent data shows thousands of road crashes every year due to potholes and it wounds a huge number of people, death included. This paper presents an innovative detection approach of potholes using a deep learning model YOLOv9 integrated with SAM-for better segmentation accuracy that has been optimized by Bayesian hyperparameter tuning. The model was tested on an T4 GPU, and the balance pursued on the trade-off between detection speed and accuracy is very strong. Overall accuracy: 79.2% is presented by the model, and for precision 82.9% followed by a recall of 75.5%. This highlights that an F1-Score of 79.0% along with IoU of 65.3%, Average Precision (AP) of 69.8%, mAP50 of 84.9%, and mAP50-95 of 54.7% have been achieved. This infers that the model does possess good accuracy in pothole detection efficiency in developing minimum accident rates and safe usage of roads.

***Keywords*— Pothole detection, YOLOv9, SAM, Bayesian Optimizer, Deep learning, Road Safety**

I.Introduction

Potholes are a common problem found on most roads around the world, whereby it threatens road safety and causes wear to road vehicles. However, this is a drastically increased problem in some countries with greater needs for such infrastructure, such as India. The seasonal monsoon increases the erosion of the road surfaces, which makes it weak and likely to deteriorate rapidly. In India, thousands of accidents are reported every year due to potholes. Incidents multiply many times during the rainy season as the road worsens at an alarming rate. As per recent statistics from the government, as many as 3,500 accidents have been reported in a single year resulting in significant injuries and even casualty resulting majorly during monsoons [1] . This data places a demand to develop automated systems of pothole detection that detect ahead of time and prove to be proactive in terms of hazard identification, minimization of risk chances for accidents, and road safety in general.

With greater development in computer vision and deep learning, complicated detection tasks can now be automated across multiple domains, even road infrastructure monitoring. This work utilizes the YOLOv9 model-the latest object detection framework that exemplifies the YOLO series in speed and accuracy within the real-time environment. Hence, YOLOv9's single-pass detection makes it a high-traffic scenario performer that identifies potholes under very challenging visual conditions.

To fine-tune the accuracy of detection, a Segment Anything Model [2] was added to the pipeline so that potholes can be segregated clearly from features in the road. SAM enhances the segmentation capability, providing fine-grained analysis that complements YOLOv9's detection by isolating potholes from surrounding elements, which is critical in cluttered or visually complex road environments. Additionally, Bayesian optimization is employed to fine-tune model hyperparameters, optimizing performance across diverse road conditions and ensuring the model's robustness under varying lighting and weather scenarios.

Combining YOLOv9, SAM, and Bayesian optimization, this framework is an approach to produce a comprehensive pothole detection solution tailored to the peculiar challenges in the context of monsoon-affected regions for road maintenance. This will add more safety to the roads, enable active road maintenance, and thereby bring closer proximity to more secure driving conditions for all drivers.

II. YOLOv9-E

For this project, the selected model is a high-performance, extended version, YOLOv9-E, optimized in terms of both accuracy and efficiency in parameters, also suitable for more complex detection tasks like in this pothole identification task. YOLOv9-E applies improvements to the basic architecture of YOLOv9 added to boost precision as well as detecting speed. The GELAN is a generalized efficient layer aggregation network that allows more efficient feature extraction with reduced computations. GELAN combines standard layers of convolutional with advanced gradient path planning, improving utilization of parameters without sacrificing performance for high-resolution and real-time applications.

Programmable Gradient Information is a new mechanism in YOLOv9-E that keeps the critical information within the network through stable gradient flows. Information Bottleneck Problem PGI solves because the network is generally allowed to store more accurate feature representation which results in better convergence and accuracy of the model. This will be very handy on hard object detection tasks since it minimizes loss of critical data in training processes and refines general detection performance [3].

Comparative analysis reveals that YOLOv9-E outperformed other variants in terms of accuracy and reduced parameters and computational cost, making it an excellent choice for high-traffic roads pothole identification due to real-time-intensive applications. Through GELAN and PGI, YOLOv9-E manages to balance efficiency with the ability to detect objects, and the paper holds it as a cutting-edge approach in the direction towards scalable automatic object detection within practically high-demand settings [3].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

III. BAYESIAN OPTIMIZER

We use Bayesian optimization in this paper to search for the hyperparameters of the YOLOv9-E model. Also, we fine-tune it towards its higher accuracy and efficiency. Bayesian optimization differs from traditional optimization techniques since it constructs a probabilistic model of the objective function. The most widely used method of Bayesian optimization is the Gaussian Process (GP). It allows performing multi-objective optimization, in such a way that accuracy and efficiency are balanced within the same process of optimization. GP is defined with the mean function $\mu(x)$ and covariance function $k(x,x')$, which estimate the relationship between hyperparameters and results of the model:

$$f(x) \sim GP(\mu(x), k(x,x'))$$

where x represents a set of hyperparameters. A surrogate model allows Bayesian optimization to look into promising regimes in the hyperparameter space as guided by previously observed data.

A major component of this approach is the acquisition function, which serves as a guide for the search between exploration and exploitation. Probably the most popular acquisition function is Expected Improvement (EI), defined as:

$$EI(x)=E\ [max\ (0, f(x)-f\ (xbest\ ))]$$

where $f(xbest)$ is the best observed value. By starting from the points that maximize the expected improvement, EI directs the optimizer to the configurations that will be expected to yield the largest gains.

Bayesian optimization in the project has efficiently tuned learning rate, batch size, and confidence threshold hyperparameters optimizing model performance within fewer evaluations. It is a multi-objective probabilistic approach that reduces computations involved and maintains requirements for real-time YOLOv9-E pothole detection applications. [4]

IV. SEGMENT ANYTHING MODEL (SAM)

Meta AI Research developed a versatile foundation model for image segmentation called the Segment Anything Model [2]. It generalizes over almost any visual task with prompts. Unlike most models that require very specific training per task, SAM achieves great adaptability through its all-encompassing training on the extensive SA-1B dataset. This dataset may cover more than 1 billion masks and 11 million images. Such a dataset is enough for the strong zero-shot capabilities of SAM over diversified segmentation applications.

The architecture comprises three elements: high-resolution feature extraction through an image encoder based on a Vision Transformer, which is designed as the ViT; prompt encoding, with elements comprising points, bounding boxes, and text prompts, allowing for the specification of targets for segmentation; and mask decoding to generate real-time segmentation masks. This prompt-oriented architecture enables SAM to generate accurate masks within milliseconds, being sensitive to and flexible in its response to any prompt that may induce ambiguity, thereby producing multiple masks.

In this experiment, SAM is combined with YOLOv9-E to improve segmentation accuracy; it is specifically focused on complex and irregularly shaped pothole areas. With zero-shot capability, SAM is able to delineate the boundaries of potholes so that false detection is minimized. Combining YOLOv9-E for object detection and more precise segmentation through SAM results in a better, more detailed detection framework for the real-time identification of potholes on various road surfaces [2].

V. LITERATURE REVIEW

A. Automated Road Damage Detection

High-level computer vision and machine learning have taken over the automation of increased detecting of road damage. For instance, YOLOv9 can depict high-accuracy detection of road damage, that is potholes and cracks, from training on large data. The finer details are still challenging in this network. Again, these techniques hint at the role of advanced augmentation techniques being able to realize improvements in detection across a variety of damage types [5] .

B. Automatic Pothole Detection Using YOLO

Recent approaches have been optimizing pothole detection for a better balance between accuracy and computational efficiency by variants of YOLOX and YOLOv8. For example, YOLOX-Nano performs well under limited resource conditions, while MS-YOLOv8 enhances the multiscale features that are very important for road maintenance [6] [7] [8] . The modified YOLOv3 and YOLOv5, amongst deep learning models, did an amazing job in resource-constrained setup pothole detection and thus proved to be quite promising for developing nations. Performing very high in real-time monitoring of roads [9] .

YOLO Network in Road Damage Detection: From YOLOv1 to YOLOv8, innovations like SPD-Conv modules and multiscale feature extraction have brought significant improvements in speed and accuracy for road damage detection, enabling real-time applications. However, issues like small object detection still need further development [10] [11] .

Variants for Improved Defect Detection in Roads: Enhanced versions, such as ML-YOLO and YOLOv8-CM, include advanced feature extraction techniques with attention modules, allowing for precise defect detection, making them valuable for complex road monitoring tasks [12] [13] .

Nighttime Pothole Detection Challenges: Detecting Road damage at night requires specifically designed datasets, like the Nighttime Pothole Dataset. Training a modified YOLOv8 on this dataset has shown improved nighttime performance, underscoring the need for low-light data and algorithms [14].

C. Hyperparameter Optimization in Machine Learning

One of the most important aspects in machine learning is hyperparameter tuning, which can be accomplished in various ways, including Bayesian Optimization. Bayesian Optimization models the relationship between hyperparameters and performance using Gaussian processes, making it far more efficient than brute-force methods for this task. It thus becomes the best fit for complex models like neural networks and tree ensembles. Practical Bayesian Optimization further improves this by considering the cost variability of experiments, optimizing both efficiency and performance [15] [4].

D. Bayesian Optimization for Hyperparameter Tuning

Bayesian Optimization simplifies hyperparameter tuning by predicting the optimal set of parameters, reducing the need for exhaustive searches. It has been shown to outperform manual tuning methods in models like CNNs and SVMs, with substantial evidence of its effectiveness in searching for better hyperparameters and improving model performance [16] .

E. Recent Advances in Hyperparameter Optimization

The growing complexity of modern machine learning models has led researchers to explore efficient ways to optimize hyperparameter settings. For instance, black-box and multi-fidelity optimizations have advanced computational efficiency for tasks like deep learning and reinforcement learning, where rapid scalability is crucial [17] .

F. Segment Anything Model

Segment Anything Model for Crater and Pothole Detection: SAM is a flexible segmentation model used for detecting craters, showing potential for pothole detection. This makes it versatile due to its zero-shot segmentation capability [18] [2] .

VI. METHODOLOGY

A. Setting up the Environment

The code was executed on the Google Colab environment, which leveraged the GPU to speed train/infer faster. Use of Google Drive was possible to handle the dataset and model weights efficiently; the colab checked for the availability of a GPU in case of heavy computation.

B. YOLOv9 Model and Dataset Preparation

YOLOv9 is used in particular due to its outstanding real-time object detection capability. The YOLOv9 repository was cloned, and its dependencies installed, along with the pre-trained weights for transferring learning towards optimizing the efficiency of pothole detection. Besides, the dataset was created by combining two open datasets; images containing annotated potholes were derived from the Roboflow Platform. There are a total of 3703 images, with divisions into training, validation, and testing sets. The training set includes 2727 images. The YOLOv9 configuration model was modified for a single-class model, pothole detection only.

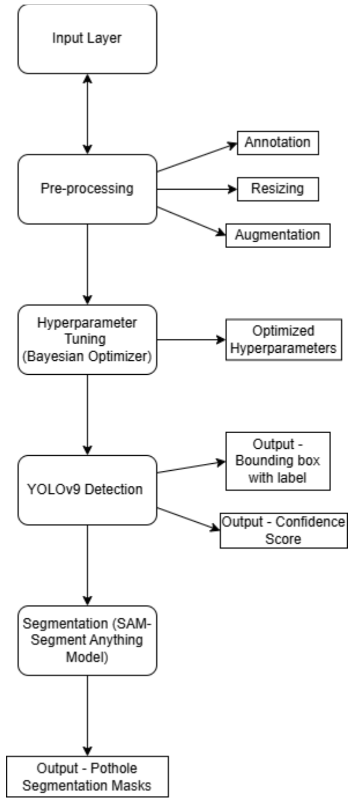

Fig.1 System Architecture of the Proposed Model

C. Training Model and Hyperparameter Tuning

The performance of the first model was not satisfactory enough, so, for hyperparameter fine-tuning, Bayesian optimization was used. Bayesian optimization was used due to its capability for multi-objective optimization between accuracy of the model and computationally efficient usage. All the above was applied to hyperparameter tune the significant parameters. The items displayed are- Learning Rate, (lr0), Final Learning Rate (lrf), Momentum, Weight Decay, Warmup Epochs, Warmup Momentum, Warmup Bias Learning Rate, Box Loss Gain, Classification Loss Gain, Distribution Focal Loss Gain (dfl),Hue (hsv_h), Saturation (hsv_s), Value (hsv_v), Image Scaling (scale), Image Flip (fliplr), Mosaic, Mix, Mixup, Copy-Paste.

Non-optimized parameters like the IoU, training threshold, anchors per layer, and certain augmentations (rotation and perspective) were kept at default values. A confidence threshold of 0.1 was set to enhance the model's sensitivity to smaller or partially occluded potholes. After fine-tuning, the final setup was configured with a Batch Size of 8, an Image Size of 640x640, and 100 Epochs. Augmentation data techniques applied—flipping, scaling, and HSV color space modifications—are applied to enhance the robustness of the model in diversified environmental conditions.

D. Inference and Conclusion

With the optimized hyperparameters, the model is tested over some new images, and the accuracy of pothole detection increases with the optimized settings. The requirement for inference on the CPU was in order to not exhaust the GPU, hence this model validates its capability of working on a different hardware environment.

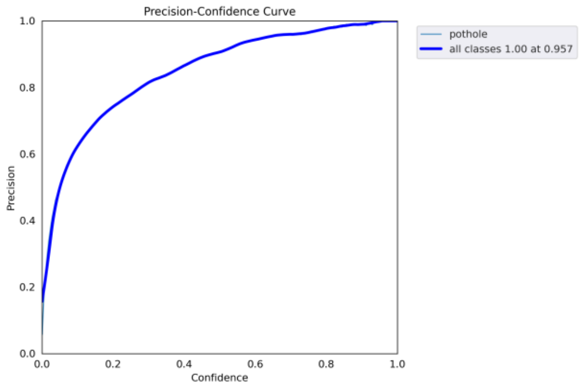

Fig.2 Precision-Confidence Curve

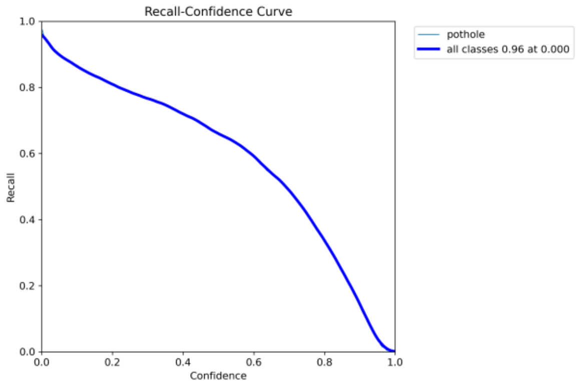

Fig.3 Recall-Confidence Curve

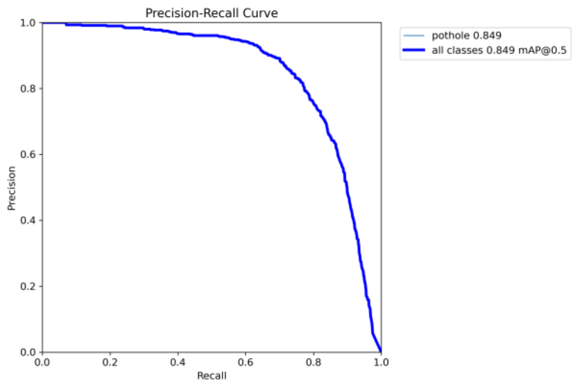

Fig.4 Precision-Recall Curve

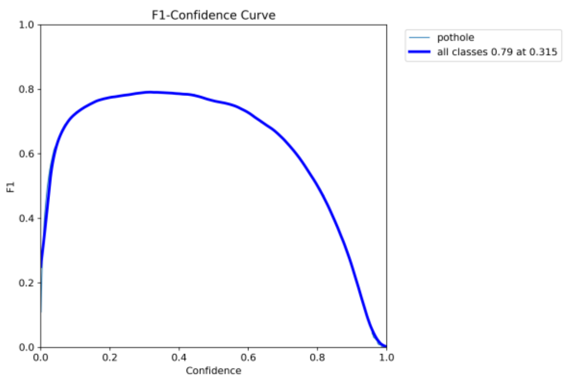

Fig.5 F1- Confidence Curve

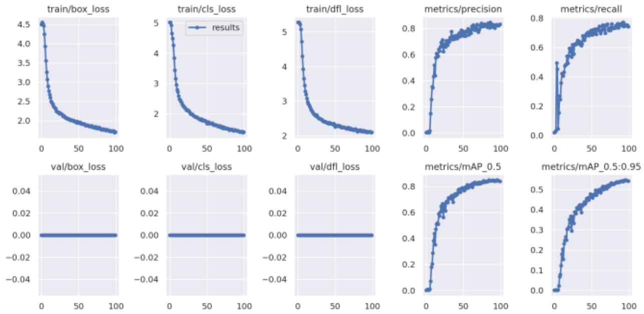

Fig.6 Training Performance

E. Segmentation and Mapping

For boundary delineation of the segmented potholes, the Segment Anything Model [2] was used with YOLOv9. Further, these segmented outputs can be mapped onto geographic coordinates for the real-time monitoring and safety applications of roads.

VII. RESULTS

The model's performance was evaluated on the test dataset, and the following metrics were obtained:

Metric	Value
Precision	0.829
Recall	0.755
mAP50	0.849
mAP50-95	0.547
Accuracy	0.792
F1-score	0.790
Average Precision	0.698
IoU (Intersection over Union)	0.653

Table.1 Performance Metrics of the Model

With an accuracy of 79.2% and a precision rate of 82.9%, the model effectively distinguishes potholes, minimizing false detections. The F1-score is 79.0%, and the mAP50 stands at 84.9%, indicating a balanced performance in both identifying and localizing potholes. Bayesian optimization was applied to fine-tune optimal hyperparameters for the YOLOv9-E model, including learning rate, batch size, and confidence threshold. This optimization enhanced the model's ability to detect and locate potholes efficiently with minimal false positives.

Fig.7 Input Images

Fig.8 Potholes Detection by YOLOv9

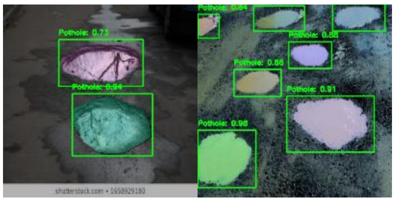

Fig.9 Potholes Segmentation by SAM

Reasonable performance is attained in "Advanced Automated Road Damage Detection Using YOLOv9," although the accuracy and recall are only 0.643 and 0.626, respectively, suggesting possible problems with false positives and missed detections [5]. With a mAP@50 of 0.774, "Improved YOLOv8 Road Defect Detection" outperforms YOLOv9 in road fault recognition while striking a balance between precision and recall, achieving greater precision of 0.803 and recall of 0.675 [10].

However, with a precision of 0.829, recall of 0.755, and a remarkable mAP@50 of 0.849, our model performs better than both YOLOv9 and Improved YOLOv8. Furthermore, as shown in Table 1, it obtains an F1 score of 0.790 and an IoU of 0.653, demonstrating that the model is excellent at both identifying and precisely localizing road faults on the road surface. With these enhanced metrics, the new model outperforms both YOLOv9 and YOLOv8, making it the most reliable and effective choice for road defect detection applications.

VIII. CONCLUSION

Using the YOLOv9-E model, which was refined by Bayesian hyperparameter tuning and supplemented by the Segment Anything Model [2] for accurate segmentation, this work presented a robust pothole detection system. The model demonstrated its deploy ability for real-time applications by showcasing high-performance detection and segmentation in the main evaluation metrics. Using Bayesian optimization to optimize the essential hyperparameters significantly increased the model's accuracy, precision, and stability.

The combination of YOLOv9-E and SAM presents a viable strategy that might be applied in real-world scenarios, enhancing road safety through proactive maintenance and prompt pothole detection. The model's strength will be further increased in subsequent stages by exposing it to different lighting conditions and probably implement the application on the edge devices to achieve real-time in-field application.

IX. LIMITATIONS AND FUTURE SCOPE

There are still certain restrictions even though this approach has a lot of promise for identifying and classifying potholes. The scale and effectiveness of model training and evaluation were impacted by the principal limitation, which was the restricted availability of GPU resources. The framework may handle data more quickly and optimise it further with more processing power. Furthermore, changes in weather, road conditions, and lighting might affect the accuracy of identification, indicating the necessity for ongoing model improvement to guarantee reliable performance in a range of scenarios.

There are numerous chances for future improvement using this framework. Measurements of pothole attributes such depth, size, length, and width would be a useful feature. Combining information from accelerometer and gyroscope sensors could greatly enhance depth estimate and provide a more thorough picture of pothole severity to help with prioritizing repairs.

Also, if the model is implemented in a mobile application, pothole detection may become more widely available, allowing people to promptly discover and report road problems in real time. In addition to supporting proactive road repair and improving road safety for all drivers, this would expedite reporting for authorities.

REFRENCES

[1] "Road Crashes Caused by Potholes in India."

[2] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, W.-Y. Lo, and P. Dollár, "Segment Anything," *arXiv preprint arXiv:2304.02643*, Apr. 2023.

[3] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.13616

[4] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization," Journal of Electronic Science and Technology, vol. 17, no. 1, pp. 26-40, 2019. [Online]. Available: https://doi.org/10.11989/JEST.1674-862X.80904120

[5] J. Estilong, "Advanced Automated Road Damage Detection Usingyolov9: Leveraging Machine Learning and Computer Vision for Enhanced Infrastructure Maintenance," 2024.

[6] M. P. B. and S. K. C., "Enhanced Pothole Detection System Using YOLOX Algorithm," Autonomous Intelligent Systems, vol. 2, no. 22, 2022. [Online]. Available: https://doi.org/10.1007/s43684-022-00037-z

[7] Y. M. Kim, Y. G. Kim, S. Y. Son, S. Y. Lim, B. Y. Choi, and D. H. Choi, "Review of Recent Automated Pothole-Detection Methods," Jun. 01, 2022, *MDPI*. doi: 10.3390/app12115320.

[8] M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," Preprints, vol. 2024, no. 2024101785, 2024. [Online]. Available: https://doi.org/10.20944/preprints202410.1785.v1

[9] S. Chougule and A. Barhatte, "Smart Pothole Detection System using Deep Learning Algorithms," *International Journal of Intelligent Transportation Systems Research*, vol. 21, no. 3, pp. 483–492, Dec. 2023, doi: 10.1007/s13177-023-00363-3.

[10] Z. Sun, L. Zhu, S. Qin, Y. Yu, R. Ju, and Q. Li, "Road Surface Defect Detection Algorithm Based on YOLOv8," *Electronics*, vol. 13, no. 12, p. 2413, 2024. [Online]. Available: https://doi.org/10.3390/electronics13122413

[11] Z. Han, Y. Cai, A. Liu, Y. Zhao, and C. Lin, "MS-YOLOv8-Based Object Detection Method for Pavement Diseases," *Sensors*, vol. 24, no. 14, Jul. 2024, doi: 10.3390/s24144569.

[12] T. Li and G. Li, "Road Defect Identification and Location Method Based on an Improved ML-YOLO Algorithm," Sensors, vol. 24, no. 21, p. 6783, 2024. [Online]. Available: https://doi.org/10.3390/s24216783.

[13] C. Zhang, X. Chen, P. Liu, B. He, W. Li, and T. Song, "Automated detection and segmentation of tunnel defects and objects using YOLOv8-CM," *Tunnelling and Underground Space Technology*, vol. 150, Aug. 2024, doi: 10.1016/j.tust.2024.105857.

[14] M. Ling, Q. Shi, X. Zhao, W. Chen, W. Wei, K. Xiao, Z. Yang, H. Zhang, S. Li, C. Lu, et al., "Nighttime Pothole Detection: A Benchmark," Electronics, vol. 13, no. 19, p. 3790, 2024. [Online]. Available: https://doi.org/10.3390/electronics13193790

[15] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, and S. H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, Mar. 2019, doi: 10.11989/JEST.1674-862X.80904120.

[16] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," in Neural Information Processing Systems, 2012.

[17] F. Hutter, L. Kotthoff, and J. Vanschoren, "The Springer Series on Challenges in Machine Learning Automated Machine Learning Methods, Systems, Challenges." [Online]. Available: http://www.springer.com/series/15602

[18] I. Giannakis, A. Bhardwaj, L. Sam, and G. Leontidis, "A Flexible Deep Learning Crater Detection Scheme Using Segment Anything Model (SAM)," Icarus, vol. 408, p. 115797, 2024. [Online]. Available: https://doi.org/10.1016/j.icarus.2023.115797