sciendo

Transport and Telecommunication, 2024, volume 25, no. 2, 209-217
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2024-0016

ANALYSIS OF POTHOLE DETECTION ACCURACY OF SELECTED OBJECT DETECTION MODELS UNDER ADVERSE CONDITIONS

Jaroslav Frnda[1]*, Srijita Bandyopadhyay[2], Michal Pavlicko[1], Marek Durica[1], Mihails Savrasovs[3] and Soumen Banerjee[4]

[1]Department of Quantitative Methods and Economic Informatics
Faculty of Operation and Economics of Transport and Communications
University of Zilina
Univerzitna 8215/1, 010 26 Zilina, Slovakia
jaroslav.frnda@uniza.sk
michal.pavlicko@uniza.sk
marek.durica@uniza.sk
[2]Department of Electronics and Communication Engineering
University of Engineering & Management University Area
Action Area III, New Town, West Bengal – 700160
srijita.bandopadhyay2019@uem.edu.in
[3]Transport and Telecommunication Institute
Lomonosova 1, Riga, LV-1019, Latvia
savrasovs.M@tsi.lv
[4]Narula Institute of Technology
81, Nilgunj Road, Agarpara, Kolkata, West Bengal - 700109. India
principal_nit@jisgroup.org

Potholes detection is an essential aspect of road safety and road infrastructure maintenance. Potholes, which are typically caused by a combination of heavy traffic and weather, are depressions or holes in the road surface that can cause damage to specific parts of a vehicle. Autonomous vehicles, in particular, must be capable of detecting and avoiding them. Hitting a deep or sharp-edged pothole at high speed can lead to loss of control or even an accident. This makes pothole detection all the more important. The accuracy of pothole detection systems installed in autonomous vehicles may be significantly impaired by adverse weather and bad light conditions. Therefore, the classification accuracy of selected well-known computer vision models for pothole detection under these specific conditions has been investigated. The results were then compared with state-of-the-art methods. Our findings showed that we outperformed many of them when used under adverse weather and low light situations. This paper presents valuable insights into the precision of various computer vision models for potholes detection. It may aid in selecting the optimal model for a specific application.

Keywords: Computer vision, object detection, potholes, YOLOv7, FRCNN

1. Introduction

Road infrastructure is crucial to the worldwide economy, particularly in terms of supply chain management and labour mobility. The fundamental components of road transportation include the network of roads and the vehicles that use that network. Although the quality of road materials and the number of new roads is gradually in-creasing, the volume of road traffic, in other words the number of vehicles, is growing exponentially. These circumstances are responsible for road congestion and have impacted on the need to increase national and local budgets for road maintenance. However, in low and middle-income countries, there are limitations to the road infra-structure budget, making it inevitable for the road structure to deteriorate.

Defects such as potholes, pavement cracks, and transverse or longitudinal surface roughness have significant effects on transportation systems. Deeper potholes have the potential to cause damage to tyres, wheels, steering and suspension systems, resulting in a serious traffic situation. Additionally, vehicle accidents not only cause direct damage to health or property, but they also affect the flow of traffic, so they are often a source of congestion. As evidence that poor road conditions are not only a problem of emerging economies, the report published in July 2023 (TRIP, 2023) stated that 40% of all major roads in the US are of bad quality, as well as that 7% of US bridges are in inadequate condition too.

While avoidance of road defects is critical for conventional vehicles, it is even more fundamental for autonomous vehicles (AVs). Hitting a pothole can cause a driverless car to lose control, resulting in accidents and injuries. AVs must recognize and avoid potholes to ensure the safety of their passengers and other road users. In addition, an automated pothole detection system can also be useful for road infrastructure operators. Camera systems installed in vehicles can detect cracks and potholes and send the relevant information - such as images of the potholes and GPS coordinates - to the operator's data centre. It will allow local authorities to prioritise road repairs and improve the quality/safety of roads.

Well-known AV manufacturers are trying to incorporate more and more sophisticated systems based on image processing and object recognition techniques into their cars. In order to achieve adequate pothole detection accuracy, AVs typically use a combination of sensors, such as front cameras, lidar and radar. However, there is still a lot of evidence (from real-world testing) that this is an ongoing problem, especially under challenging conditions (Kumar *et al.,* 2023; Sathya and Saleena*,* 2022 and Peraka *et al.,* 2020). AVs can quickly detect potholes under clear weather conditions. But adverse conditions like heavy rain or melting snow will cover the potholes. In such situations, many published pothole detection models have a limited ability to identify road defects properly.

Thus, this paper examines a pothole detection models able to operate in real-time which are available as open-source, namely YOLO (You Only Look Once) and FRCNN (Faster Region-Based Convolutional Neural Network). We also investigated whether image pre-processing techniques, such as grayscale conversion and edge detectors, could improve the accuracy of detecting small potholes. We tested open-source algorithms because they are cost-effective and easy to implement in real-world applications.

2. Motivation and related work

Several research papers have been published that deal with automatic road pothole detection systems. The authors have used different approaches, ranging from relatively simple methods to very sophisticated algorithms with significant hardware or software requirements. We can say that there are two main research directions.

The first is the so-called vibration-based technique which is based on the evaluation of pavement quality using various sensors in cooperation with GPS system. Data from sensors such as accelerometers or gyroscopes are collected and sent to a data centre where the data is post-processed (Wang *et al.,* 2015). Nowadays, due to the increase in computing power of mobile devices, several papers have shown that smartphones or tablets can identify important patterns from the data (collected by their internal sensors) and detect potholes or cracks. These devices can also transmit the exact location of potholes which can then be displayed on an online map (Wang *et al.,* 2015; Sattar *et al.,* 2021; Harikrishnan *et al.,* 2017 and Agrawal *et al.,* 2021). A disadvantage of this technique is the need to drive through the pavement defect in order to record it. This can result in additional costs, e.g. for tyres, due to hitting the pothole. In addition, the vibration-based technique allows us to determine the depth of the pothole, however, it lacks information on the area and shape of the pothole. On the other hand, this system is independent of visibility conditions, although it needs peers for the continuous generation of sensor data.

Terrestrial laser scanning or stereo vision systems can reconstruct a 3D model of the road with all the dimensions of a pothole with high accuracy, but it can suffer from high computational complexity. Moreover, motion sensitivity or camera misalignment can affect the output quality and, most importantly, it is expensive to equip the vehicle with laser scanners (Ahmed, 2021). In general, the 3D reconstruction approach is considered to be the most expensive method.

Automatic vision-based pothole detection is the second approach we would like to mention here. It is the most common detection approach and is also considered to be the most cost-effective. There is a variety of approaches to identify potholes from im-ages captured by front cameras, but they all require a computer system to process the image input in real-time. Authors in (Guo *et al.,* 2022) proposed a pavement distresses identification method. Their method based on K-means algorithm and pretrained You Only Look Once algorithm allows to identify the type of road damage, such as crack, potholes or damaged manhole covers. Work (Ouma and Hahn, 2017) used integrated wavelet filtering and fuzzy c-means algorithm for image clustering of road defects and non-defects. While there are other different vision-based methods, the main direction of research today is focused on machine learning and neural networks, in particular convolutional neural networks (CNN). After 2012, when the SuperVision team's AlexNet won the ImageNet Large Scale Visual Recognition Challenge (Krizhevsky *et al.,* 2017), CNN has become more popular in 2D vision-based classification. Today, several CNN-based algorithms exist. Region-based CNN (R-CNN) and You Only Look Once (YOLO) are examples of widely used methods. R-CNN and its later successors, Fast and Faster CNN, combine two key elements and are therefore commonly referred to as two-stage detectors.

In R-CNN, the image is divided into the proposal regions which are sliced from the input image and then classified by the CNN. In the next step, a support vector machine (SVM) is used to determine the bounding boxes using the features of the CNN (Girshick *et al.,* 2014). In contrast, the Fast R-CNN detector works on the whole image. It introduces the region of interest pooling layer to reduce the computations for the overlapping regions (Girshick, 2015). The Faster R-CNN presented by (Ren *et al.,* 2017) proposed the model containing two cooperating CNNs, namely Fast R-CNN and Region Proposal Network (RPN). Faster computation time is achieved by sharing convolutional features between these two networks. On the contrary, YOLO is called a one-stage detector because it does not use a special region proposal system. The first version of the YOLO algorithm was presented by (Redmon *et al*., 2016). The main difference from previous approaches is that the image is divided into predefined grid cells. For each cell, a confidence score is calculated.

Vision-based systems are naturally limited in their ability to determine or estimate pothole depth. This is caused by them using two-dimensional input. In addition, these systems are affected by the weather and the light conditions. In (Bucko *et al.,* 2022) authors proved that YOLOv3 has a higher success rate under good visibility conditions, while Sparse R-CNN achieves better results under poor visibility and weather conditions. However, the size of the Sparse R-CNN model was three times larger and it required more computation time. Paper (Wang *et al.,* 2022) provides the methodology that significantly enhances YOLOv3 through various image adjustments, transformations and clustering. The pothole samples were generated by random occlusion and added noise. The aim was to increase the robustness of the model. Rainy and sunny conditions were taken into account to ensure the correct detection of the pothole in different weather conditions. Authors (Du and Jiao, 2022) proposed a different road defect detection model based on the 5th version of YOLO algorithm and Bidirectional Feature Pyramid Network (BIFPN) to improve feature extraction, solve sample imbalance problem and reduce misdetection. Research team led by Jakubec (Jakubec *et al.,* 2023) compared several CNN-based models for pothole detection under different weather conditions. Published results together with work (Bucko *et al.,* 2022), served as the baseline for our experiments. In paper (Aparna *et al.,* 2022) authors proposed a completely different and novel method using CNN based on the Residual Networks (ResNet) model. They build the model on thermal images instead of detecting from standard images. After collecting the appropriate number of thermal images of potholes in different weather and daylight situations, and training the model, an average accuracy of 95% has been achieved using ResNet models.

As can be seen from the aforementioned works, pothole detection is an important research topic, especially for autonomous vehicles and local road maintenance services. In this paper, we tested several of most common open-source frameworks focusing on detection accuracy in bad light/weather. The major contributions of this work are following:

- We investigated the detection accuracy of two mainstream object detection models and compared the results with published works. We focused our attention on real images of adverse weather conditions. We performed many experiments to find the best hyperparameters setting. This includes adjusting parameters such as learning rate, batch size, or regularisation techniques.
- We tested the AutoML framework Roboflow. This framework advertises itself as a tool for building and deploying object recognition models without coding knowledge or hyperparameters tuning (Dwyer *et al.*, 2020).
- We analysed the impact of converting input images to grayscale and applying an edge detector as a pre-processing step. Grayscale images can be less affected by variations in lighting and shadows in comparison to colour images. This can make it easier to detect objects in different lighting conditions.
- We extended the dataset (Bucko *et al.,* 2022) by annotation using the PASCAL format.

3. Methodology

The presented work is based on a comprehensive methodology for pothole detection, leveraging the synergy of Deep Learning and Computer Vision techniques. The study highlights two different experimentations. Figure 1 explains the two different steps used in the work which are illustrated below:

1. The first method involves the use of a framework in which the dataset (Bucko *et al.,* 2022) is first annotated, and then it is split amongst test and train dataset. Different object detection architectures like YOLOv7, FRCNN and Roboflow are selected. After the validation, the unseen test dataset is employed to evaluate the chosen model. Finally, the detected pothole images are obtained along with different scores.

2. The second method uses an extra image processing step before the annotation. The Sobel filter is used to highlight edges in images by calculating the gradient of the image intensity and identifying the areas where there is a significant change in intensity. To achieve this, we had to use a pre-processing technique called Pixel Brightness Transformation (PBT). PBT is responsible for converting the image to a grey scale. Since there is no colour information to be represented for a grey scale image, there is less information that needs to be stored for each individual pixel. The aim of the use of the Sobel filter is to extract valuable information from the images, such as edges or curves. During image segmentation, it is not trivial to incorporate small potholes into a suitable region of pixels. Therefore, we wanted to try and see if enhancing the edges could lead to better accuracy. The identification of edges is crucial for finding the contours of objects. In order to obtain the best final result, we tested different thresholds (limits for which pixel orientations are considered as edges). Performance indicators are generated and compared with the other method.

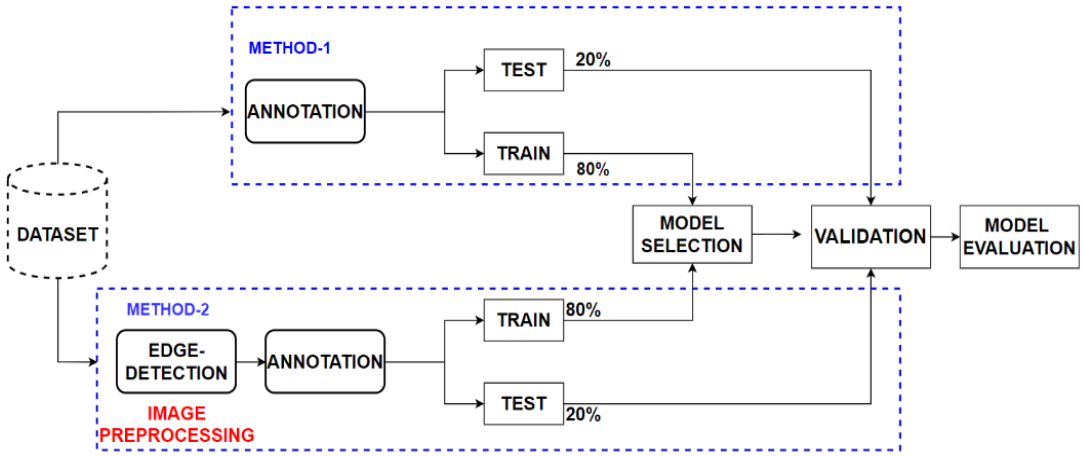


Figure 1. Workflow diagram

3.1. Dataset

The dataset (Bucko *et al.,* 2022) used in this study is meticulously organized and categorized into five distinct weather situations, namely Clear, Rainy, Night, Evening and Sunset, reflecting real-world driving conditions. The dataset's images (examples are depicted in Figure 2) are 640 x 640 pixels in size, comprising a total of 2 099 images. Because this dataset suffers from data imbalance (class Clear makes up almost half of all the images), Generative Adversarial Network (GAN) was used to enhance the quality of the dataset. GAN allowed us to generate synthetic data for training models, so after this step, the total number of images for training and testing was 5 000. This step also helped us to achieve better class balance, which can be crucial for the model's performance and generalization. Selected synthetic images are depicted in Figure 3. There are different types of GANs for different purposes. Based on our positive experience with a combination of FRCNN and CycleGAN in (Roy *et al.,* 2023), we have chosen CycleGAN again. Following a standard 80:20 ratio for training and testing data, we randomly selected 4000 images for training and 1000 for testing. For the test set, our aim was to ensure a balanced set of data, and we repeated the random selection until all classes were almost equally represented.

We used two different models for object detection: YOLOv7 and Faster RCNN (FRCNN). YOLOv7 was built on the Darknet framework as the backbone model (responsible for extraction of high-level features from the input image), while FRCNN utilized Resnet152v2 as its backbone architecture. For the YOLO model, the dataset needs annotating for potholes in the YOLO format, while for the FRCNN training PASCAL format is applied, and this involves converting XML files to TensorFlow record files. In addition, we explored the feasibility of using AutoML tools for model building, which does not require any coding or hyperparameter fine-tuning. After careful consideration, we have selected Roboflow. As of 2023, Roboflow has over 250,000 active users. Its AutoML tool is named Roboflow Train (Dwyer et al., 2020).

The dataset is well structured with different weather conditions and detailed annotations. This provides a solid basis for training and evaluation of the object recognition models performance while detecting and locating potholes in different environmental conditions.

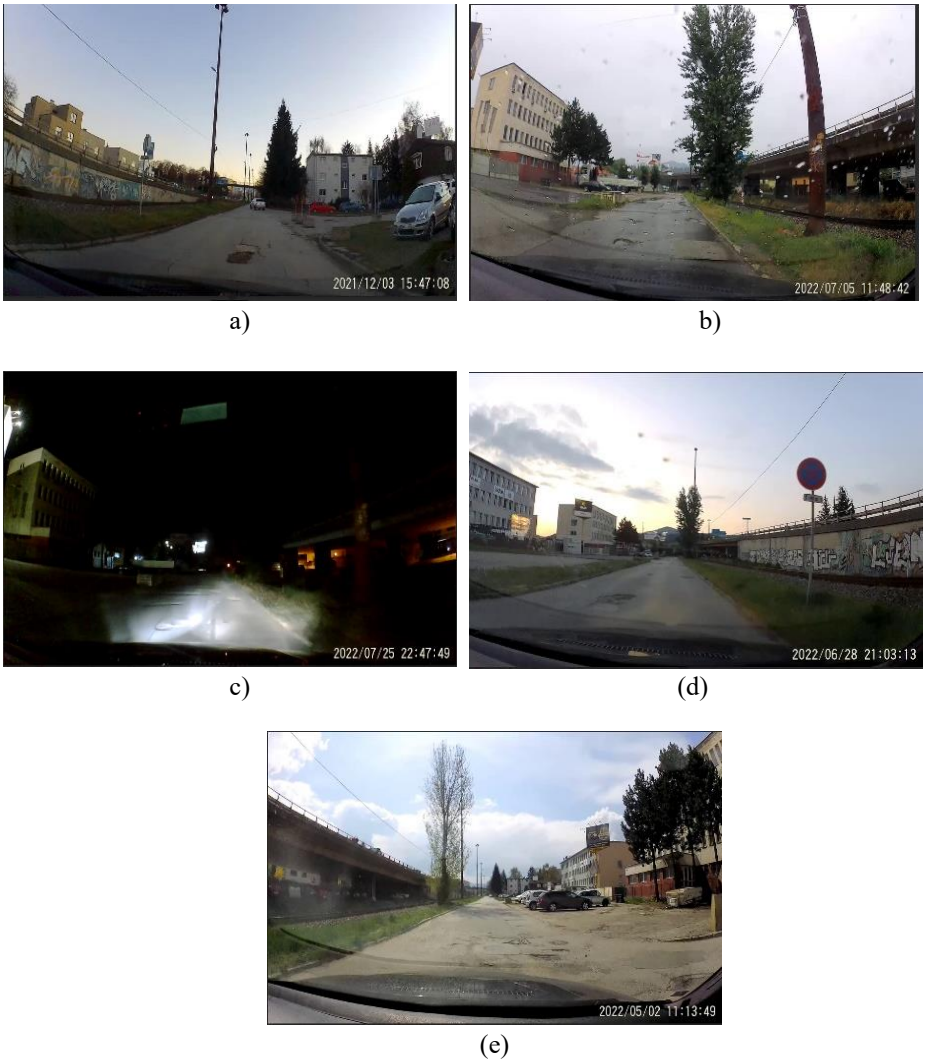


Figure 2. Some samples of Pothole (a) Sunset (b) Rain (c) Night (d) Evening (e) Clear



Figure 3. Examples of synthetic images

3.2. Evaluation Metrics

Evaluation metrics in the context of object detection are quantitative measures used to assess the performance and effectiveness of a model during identification and localization of objects within an image. These metrics provide insights into various aspects of the model capabilities, including accuracy, precision, recall, and efficiency. The evaluation metrics used in the presented work are calculated based on the values obtained from the confusion matrix which consists of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

Precision: Precision measures the accuracy of positive predictions by calculating the ratio of correctly predicted positive instances to the total predicted positive instances. It helps to assess the model's ability to avoid false positives.

$$Precision\ (PPV) = \frac{TP}{TP+FP}. \quad (1)$$

Recall (Sensitivity): Recall measures the ability of the model to capture all actual positive instances. It calculates the ratio of correctly predicted positive instances to the total actual positive instances.

$$Recall\ (Sensitivity) = \frac{TP}{TP+FN}. \quad (2)$$

Mean Average Precision (mAP): It is a crucial evaluation metric in object detection, providing a comprehensive measure of the model performance across multiple object classes. It combines precision and recall values at various confidence thresholds, and it computes the average precision for each class. The precision-recall trade-off is assessed by varying the confidence threshold for positive predictions, generating a precision-recall curve. In the context of object detection, mAP is often computed at different Intersection over Union (IoU) thresholds, such as 0.5 (mAP@0.5) or a range from 0.5 to 0.95 (mAP@[0.5:0.95]), providing insights into the ability of the model to accurately localize objects at varying levels of overlap.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i. \quad (3)$$

4. Results

At first, we wanted to examine the classification accuracy (one-class detection) of Roboflow Auto ML tool. In this platform, all the hyperparameters were set automatically. As depicted in Figure 4, we deployed a model with low prediction accuracy.

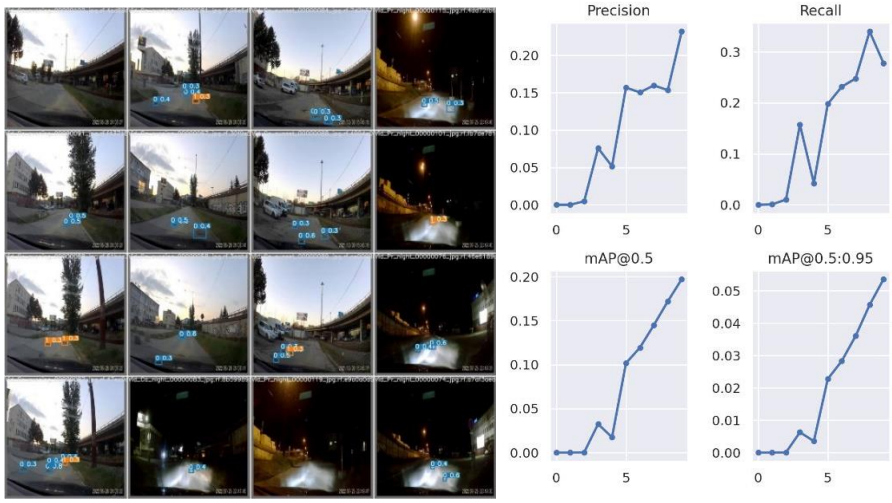


Figure 4. Performance analysis of Roboflow AutoML framework

Second group of experiments were dedicated to YOLO and FRCNN models. We received the best results with the following settings:

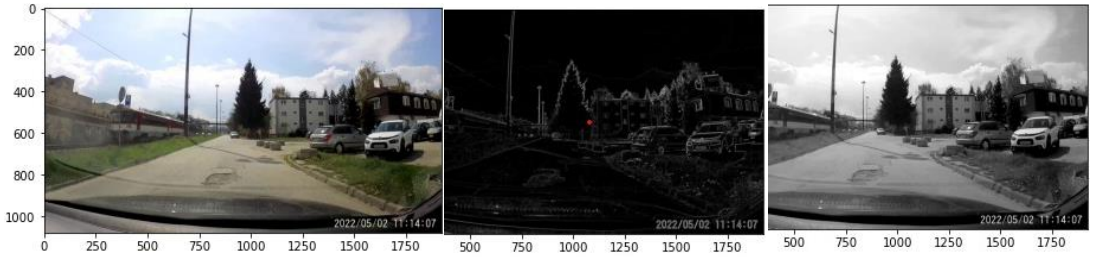
- Number of epochs: (early stopping function was implemented) – up to 100 epochs.
- Batch size: 64 and initial learning rate: 0.01
- Pre-trained weights: yes
- Number of classes: 1 and number of workers: 4

"Workers" represent the maximum number of dataloader workers. This parameter is calculated as ratio between CPUs and GPUs (e.g., 8 CPUs and 4 GPUs are available, so each GPU uses 2 CPU = worker is set to 2).

The obtained results of potholes recognition pointed out that models selected by us overperformed the state of the arts methods. Results shown in Table 1 verified our assumption about FRCNN model that it is currently a very promising framework for objects detection. In-depth analysis of detection results under

adverse conditions revealed several oddities. Small cracks or dark regions on the road can be mistaken for potholes in images with limited visibility. Similarly, in rainy images, reflections from objects on the wet road or rain droplets collected on the vehicle windscreen could be incorrectly classified as potholes. Compared to the current state-of-the-art methods, the detection accuracy of the FRCNN model is significantly higher, especially at night. As it can be seen from the Table 1, correct pothole detection at night was a major challenge for these models. For this type of image, we also observed the most significant difference in accuracy between FRCNN and YOLOv7.

Image pre-processing by applying Sobel filter did not overcome YOLOv7 in its basic settings but the detection ability was very similar. One of the potential reasons is that edge detectors are sensitive to noise. Especially the environment introduces white noise to image, which reduces effectiveness of edge detectors. For this purpose, several authors paid their attention to creation of rain noise removal methods (Ran *et al.,* 2023; Matsui and Ikehara, 2020). However, additional techniques for noise reduction increase the computing power required for real-time detection. Images captured in dark environment are very similar to their grayscale copies, thus there is no significant difference between grayscale and night images. On the other hand, grayscale image can be processed more quickly because the image consists of one single channel instead of the three channels in RGB colour model (e.g. Tesla autopilot processes images converted to grayscale plus red channel). In our experiments, grayscale image combined with Sobel filter reduced training time by more or less 5%. Both YOLO and FRCNN require RGB format, so we can simply copy this channel three times. The PBT process is depicted in Figure 5.



a) Original image b) Grayscale transformation c) Sobel filter

Figure 5. Scheme of converting image to grayscale and applying Sobel filter

Following table shows performance analysis of selected algorithms under various weather conditions. The best values are bolted. Italics are used for models proposed by ours.

Table 1. Overall classification performance of selected models

Model	Situation	Precision	Recall	mAP@0.5	mAP@[0.5:0.95]
YOLOv7	Clear	0.91	**0.91**	**0.91**	0.63
YOLOv7 Sobel		0.89	0.89	0.89	0.61
FRCNN		**0.92**	**0.91**	**0.91**	**0.79**
Optimized Deep Recurrent Neural Network (Sathya and Saleena, 2022)		0.90	0.88	-	-
YOLOv5 (Ahmed, 2021)		0.87	0.90	-	0.62
FRCNN (ResNet50 backbone) (Ahmed, 2021)		**0.92**	-	-	0.64
Sparse RCNN (Bucko *et al.,* 2022)		0.69	0.54	0.73	0.32
YOLOv3 (Bucko *et al.,* 2022)		0.78	0.77	0.77	0.33
Cascade RCNN (Jakubec *et al.,* 2023)		0.71	0.55	0.74	0.34
*SPFPN-YOLOv4 (Heo *et al.,* 2023		0.89	0.84	0.8	-
YOLOv5 (Ahmed, 2021)		0.87	-	-	0.62
Faster RCNN (ResNet50) (Ahmed, 2021)		**0.92**	-	-	0.64
YOLOv7	Night	0.58	0.57	0.57	0.39
YOLOv7 Sobel		0.51	0.51	0.51	0.32
FRCNN		**0.86**	**0.86**	**0.86**	**0.65**
YOLOv3 (Bucko *et al.,* 2022)		0.36	0.16	0.18	0.06
Cascade RCNN (Jakubec *et al.,* 2023)		0.35	0.34	0.33	0.11
ROBOFLOW		0.24	0.29	0.2	0.06
YOLOv7	Sunset	**0.89**	**0.89**	**0.89**	**0.63**
YOLOv7 Sobel		0.86	0.86	0.74	0.59
FRCNN		**0.89**	**0.89**	**0.89**	0.61
YOLOv3 (Bucko *et al.,* 2022)		0.69	0.5	0.53	0.19
Cascade RCNN (Jakubec *et al.,* 2023)		0.57	0.48	0.52	0.2

Continuation of Table 1

Model	Situation	Precision	Recall	mAP@0.5	mAP@[0.5:0.95]
YOLOv7	Evening	0.87	0.87	0.87	0.57
YOLOv7 Sobel		0.83	0.83	0.73	0.57
FRCNN		**0.89**	**0.89**	**0.89**	**0.60**
YOLOv3 (Bucko *et al.*, 2022)		0.74	0.48	0.47	0.18
Cascade RCNN (Jakubec *et al.*, 2023)		0.52	0.46	0.48	0.14
YOLOv7	Rainy	0.8	0.8	0.8	0.51
YOLOv7 Sobel		0.72	0.72	0.64	0.45
FRCNN		**0.82**	**0.82**	**0.82**	**0.54**
YOLOv3 (Bucko *et al.*, 2022)		0.61	0.52	0.51	0.2
Cascade RCNN (Jakubec *et al.*, 2023)		0.56	0.47	0.46	0.17

*Note: SPFPN combines Spatial pyramid pooling (SPP) and feature pyramid network (FPN).

5. Conclusions

Our experiments have shown that current computer vision methods have improved on the ability of their predecessors to correctly detect potholes on the road, particularly in adverse weather conditions and when visibility is poor. Two established models, YOLOv7 and FRCNN, were included in our tests. We also investigated the effect of converting images to greyscale and applying a Sobel filter during the data pre-processing phase. In contrast to unbalanced dataset used in the original paper (Bucko *et al.,* 2022), our approach involved increasing the balance of data through synthetic data generated by the GAN technique. This resulted in doubling of the dataset size, having positive effects not only on the training phase, which benefited from more training samples, but also on the classification accuracy. We found out that using FRCNN produced the best results. Especially for dark images, FRCNN significantly improved the ability to detect potholes in comparison to the results published in previous papers. The results also support the conclusion that YOLOv7 is a strong competitor, with the exception of dark scenes, due to its faster prediction time and model size. Model size plays a crucial role in memory consumption and the need for computational resources. The size of YOLOv7 model was about six times smaller (approximately 36 MB) than the size of the FRCNN Resnet152v2 model (232 MB). However, we were surprised by the poor performance of the Roboflow AutoML tool at night. The deployed model achieved accuracy similar to the relatively old and computationally undemanding YOLOv3 model (released in 2018). As a result, we did not continue with model validation for other image types.

This work has highlighted that pothole detection under adverse conditions is a relatively new topic with limited up-to-date research results, providing many opportunities for further research. Future research will propose expansion of the road damage dataset to include other forms of adverse conditions and conduct additional tests using recently released object detectors, such as YOLOv8. In order to compare the performance of selected models, we will also focus on running models on different hardware configurations.

Acknowledgements

This research was funded by Slovak Grant Agency for Science (VEGA) under Grant No. 1/0509/24.

References

1. Agrawal, H., Gupta, A., Sharma, A., Singh, P. (2021) Road pothole detection mechanism using mobile sensors. In: *2021 International Conference on Technological Advancements and Innovations (ICTAI)*.
2. Ahmed, K.R. (2021) Smart Pothole Detection Using Deep Learning Based on Dilated Convolution. *Sensors*, 21, 8406. https://doi.org/10.3390/s21248406.
3. Aparna, Bhatia, Y., Rai, R., Gupta, V., Aggarwal, N., Akula, A. (2022) Convolutional neural networks based potholes detection using thermal imaging. *Journal of King Saud University - Computer and Information Sciences* 34, 578–588.
4. Bucko, B., Lieskovská, E., Zábovská, K., Zábovský, M. (2022) Computer Vision Based Pothole Detection under Challenging Conditions. *Sensors*, 22, 8878. https://doi.org/10.3390/s22228878.
5. Du, F.J., Jiao, S.J. (2022) Improvement of Lightweight Convolutional Neural Network Model Based on YOLO Algorithm and Its Research in Pavement Defect Detection. *Sensors*, 22, 3537. https://doi.org/10.3390/s22093537.
6. Dwyer, B., Nelson, J., Solawetz, J. *et al*. Roboflow (Version 1.0) [Software]. Available online: https://roboflow.com. (accessed on 29 December 2023).

7. Girshick, R. (2015) Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV).* 1440-1448.
8. Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 580-587.
9. Guo, K., He, C., Yang, M., Wang, S. (2022) A pavement distresses identification method optimized for YOLOv5s. *Sci Rep*, 12, 3542, https://doi.org/10.1038/s41598-022-07527-3
10. Harikrishnan, P. M., Gopi, V. P. (2017) Vehicle vibration signal processing for road surface monitoring. *IEEE Sensors Journal*, 17, 5192–5197.
11. Heo, D.-H., Choi, J.-Y., Kim, S.-B., Tak, T.-O., Zhang, S.-P. (2023) Image-Based Pothole Detection Using Multi-Scale Feature Network and Risk Assessment. *Electronics*, 12, 826. https://doi.org/10.3390/electronics12040826.
12. Jakubec, M., Lieskovská, E., Bučko, B., Zábovská, K. (2023) Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation. *Appl. Sci.*, 13, 5810. https://doi.org/10.3390/app13095810.
13. Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017) ImageNet classification with deep convolutional Neural Networks. *Communications of the ACM*, 60, 84–90.
14. Kumar, P., Pooja, Chauhan, N., Chaurasia, N. (2023) Vision-based pothole detection using CNN model. *SN Computer Science*, 4.
15. Matsui, T., Ikehara, M. (2020) Gan-based rain noise removal from single-image considering rain composite models. *IEEE Access*, 8, 40892–40900.
16. National Transportation Research Nonprofit. Key facts about the U.S. surface transportation system July 2023. Available online: https://tripnet.org/wp-content/uploads/2020/04/TRIP_Fact_Sheet_NATL.pdf (accessed on 29 December 2023).
17. Ouma, Y. O., Hahn, M. (2017) Pothole detection on asphalt pavements from 2D-colour pothole images using fuzzy c-means clustering and morphological reconstruction. *Automation in Construction*, 83, 196–211.
18. Peraka, N.S., Biligiri, K.P. (2020) Pavement Asset Management Systems and technologies: A Review. *Automation in Construction*, 119, 103336.
19. Ran, W., Yang, B., Ma, P., Lu, H. (2023) TRNR: Task-driven image rain and noise removal with a few images based on patch analysis. *IEEE Transactions on Image Processing*, 32, 721–736.
20. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016) You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
21. Ren, S., He, K., Girshick, R., Sun, J. (2017) Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 1137–1149.
22. Roy, K., Chaudhuri, S.S., Frnda, J., Bandopadhyay, S., Ray, I. J., Banerjee, S., Nedoma, J. (2023) Detection of tomato leaf diseases for agro-based industries using novel PCA deepnet. *IEEE Access*, 11, 14983–15001.
23. Sathya, R., Saleena, B. (2022) A framework for designing unsupervised pothole detection by integrating feature extraction using deep recurrent neural network. *Wireless Personal Communications*, 126, 1241–1271.
24. Sattar, S., Li, S., Chapman, M. (2021) Developing a near real-time road surface anomaly detection approach for road surface monitoring. *Measurement*, 185, 109990.
25. Wang, D., Liu, Z., Gu, X., Wu, W., Chen, Y., Wang, L. (2022) Automatic Detection of Pothole Distress in Asphalt Pavement Using Improved Convolutional Neural Networks. *Remote Sens.*, 14, 3892. https://doi.org/10.3390/rs14163892.
26. Wang, H.-W., Chen, C.H., Cheng, D.Y., Lin, C.H., Lo, C.C. (2015) A real-time pothole detection approach for Intelligent Transportation System. *Mathematical Problems in Engineering*, 2015, 1–7.