Pothole Detection and Dimension Estimation System using Deep Learning (YOLO) and Image Processing

Pranjal A. Chitale*
Student, DJSCE
Mumbai, India
pranjalchitale@gmail.com

Hrishikesh R. Shenai*
Student, DJSCE
Mumbai, India
shenaihrishikesh24@gmail.com

Jay P. Gala*
Student, DJSCE
Mumbai, India
jaygala24@gmail.com

Kaustubh Y. Kekre*
Student, DJSCE
Mumbai, India
kkekre90@gmail.com

Ruhina Karani*
Asst. Professor, DJSCE
Mumbai, India
ruhina.karani@djsce.ac.in

***Abstract* — The world is advancing towards an autonomous environment at a great pace and it has become a need of an hour, especially during the current pandemic situation. The pandemic has hindered the functioning of many sectors, one of them being Road development and maintenance. Creating a safe working environment for workers is a major concern of road maintenance during such difficult times. This can be achieved to some extent with the help of an autonomous system that will aim at reducing human dependency. In this paper, one of such systems, a pothole detection and dimension estimation, is proposed. The proposed system uses a Deep Learning based algorithm YOLO (You Only Look Once) for pothole detection. Further, an image processing based triangular similarity measure is used for pothole dimension estimation. The proposed system provides reasonably accurate results of both pothole detection and dimension estimation. The proposed system also helps in reducing the time required for road maintenance. The system uses a custom made dataset consisting of images of water-logged and dry potholes of various shapes and sizes.**

Keywords — YOLO, Deep Learning, Triangular Similarity, Image Processing, Pothole Detection, Dimension Estimation

I. INTRODUCTION

Technology has played an essential role in the development of automated systems in various sectors in the past few years. With the advent of Autonomous systems, human lives have become more convenient. Transportation and surveillance systems have significantly benefited from the incorporation of automation. Specifically for transportation, roads are of foremost importance as roads constitute the most extensive network. It is important for an autonomous system to function without compromising the safety of its users and for road transportation systems, potholes pose a great threat. According to the official data released by the Government of India, 2015 people lost their lives in 4,869 accidents caused by potholes last year. This makes road maintenance of paramount importance.

The world is suffering from COVID-19 pandemic. The lockdowns have affected various sectors, including road maintenance. This has led to a deterioration in road conditions. Thus there is a need for an autonomous system that can monitor road conditions. In this paper, a pothole detection and dimension estimation system is proposed, which uses Deep Learning and Image Processing. In recent times, a number of deep learning based object detection techniques have been developed, which use Convolutional Neural Networks for feature extraction. This paper proposes the use of YOLO (You Only Look Once) for a pothole detection system. Multiple versions of the YOLO algorithm are trained on a custom dataset, consisting of both water-logged and dry potholes of various shapes and sizes, after which results are evaluated on IoU (Intersection over Union) and mAP (mean Average Precision). The model is able to detect a variety of potholes with reasonable accuracy. Also, the proposed Image Processing based pothole dimension estimator provides fairly accurate dimensions of the detected potholes using Triangular Similarity, thus reducing the overall time required for road maintenance even further.

II. REVIEW OF LITERATURE

The system proposed by Sunil Sharma et al. [1] is an accelerometer-based approach for detecting potholes. Machine learning algorithms are applied to the accelerometer data to predict the type of pothole. The drawback of this system is that drivers tend to slow down over potholes and thus, data collected by accelerometers may not be accurate. Moreover, speed bumps would also be recorded as potholes.

978-1-7281-8579-8/20/$31.00 ©2020 IEEE

Lokeshwor Huidrom et al. [2] have proposed a system that detects road distresses like potholes, patches and cracks based on predefined threshold values of standard deviation and circularity of the object. This is done using image processing algorithms. The drawback of this system is that the same predefined thresholds cannot work for different types of road distresses as they do not have a fixed shape or size.

The system proposed by SungWon Lee et al. [3] detects potholes based on discolorations in the image. It is an image processing technique which uses a wavelet energy field to show textures in the image. Thus individual cells are checked for the presence of potholes. The drawback of this system is that discolorations on the road cannot be directly termed as potholes as discoloration can also be caused because of other reasons like wet roads, road markings and zebra crossings.

Nhat-Duc Hoang [4] has implemented a pothole detection system mainly based on Least Square Support Vector Machine which is a machine learning algorithm. The dataset used for training only consists of 200 images and thus, the model has limited application as it can predict only a specific type of potholes. Moreover, support vector machine based systems require more training time.

The system proposed by Abhishek Kumar et al. [5] is a faster region-based convolutional neural network (Faster R-CNN) for pothole detection. The dataset used is based on foreign roads. The drawback here is that R-CNN based models have longer prediction times. Also, models trained on foreign roads used for detecting potholes perform badly on Indian roads as the damage conditions are totally different.

Su-il Choi et al. [6] have implemented a pothole detection and measurement system using 2D LiDAR (Light Detection and Ranging) sensor and Camera. The images obtained from the camera are processed using image processing techniques like noise filtering and edge detection. The data from the LiDAR sensors is then used for detecting potholes and also estimating their dimensions. The drawback here is that LiDAR sensors are very expensive and installing multiple such sensors may not be cost-efficient.

The system proposed by E. N. Ukhwah et al. [7] is a YOLOv3 based pothole detection system. The images are obtained using a pavement view camera installed on the Hawkeye 2000 survey vehicle. The drawback of YOLOv3 based object detection is the results are less accurate as compared to YOLOv4. YOLOv4 gives more precise bounding boxes, thus resulting in a better IoU value.

The system proposed in this paper makes improvements over the above-mentioned systems by making use of the YOLOv4 algorithm. It is a more efficient prediction algorithm in terms of training time as well as accuracy. The YOLOv3 and YOLOv4 algorithms are trained on a custom dataset. The results obtained on the validation set are compared on the basis of mAP and IoU values for multiple iterations. The custom dataset contains pothole images from both Indian and foreign roads and under different conditions. An image processing based pothole dimension estimation system is also implemented using Triangular Similarity measure.

III. SYSTEM METHODOLOGY

A. Dataset Creation

A custom dataset is created for the model to train on. The dataset contains approximately 1300 images of both Indian and foreign roads, with multiple potholes in each image. The images are a mix of waterlogged and dry potholes of different shapes and sizes. Images were collected from Google Images, an extensive image search engine. Fig. 3.1 shows some potholes from the dataset. An open-source tool called “LabelImg” is used for annotating the images according to the YOLO standards. Annotations follow the following format - <object-class> <x_center> <y_center> <width> <height>.

Fig. 3.1. Sample Potholes from the dataset

B. System Workflow

The proposed system is divided into two stages; pothole detection and dimension estimation. Images taken from a camera are fed into the pothole detection system. The camera is kept at an elevation of 90 cm from the ground. The bounding boxes obtained as a result of applying a deep learning model on the input images are given to the dimension estimation module. This module uses the elevation of the camera from the ground and estimates the dimensions of each of the bounding boxes as output. Fig. 3.2 shows the system workflow.

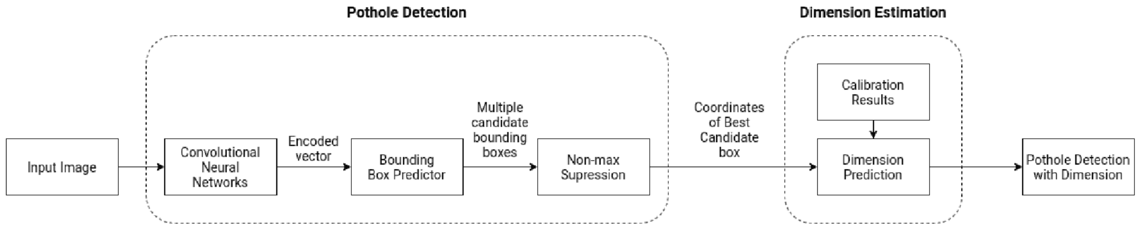

Fig. 3.2. System Workflow

C. Pothole Detection Module

The pothole detection module of the proposed system is based on the YOLO family of object detectors. YOLO [8] is an abbreviation for "You Only Look Once." As its name suggests, it is based on the principle that it is possible to detect and localize the object present in an input image at a single glance. This technique considers object detection as a regression task. These object detectors follow the one-stage architecture. Here the bounding boxes are predicted by considering the input images directly and thus, a separate region proposal step is not needed. On the other hand, in two-stage architecture based object detectors, the candidate object bounding boxes are proposed by a Region Proposal Network, which acts as the first stage of the detector. While in the second stage, the process of feature extraction is performed by the RoI Pool operation by considering every candidate box for performing further tasks of classification and bounding-box regression. The two-stage architecture based object detectors are not suitable for real-time applications due to the network complexity. Therefore, one stage architecture based object detectors - YOLOv3[9] and YOLOv4[10] have been used as the pothole detection models.

YOLOv3 compromises a bit on the speed for boosting the accuracy of the overall network. This algorithm focuses on making incremental improvements in Bounding Box Prediction, Class Prediction, Prediction Across Scales and Feature Extractor. All these improvements result in an increased mAP value and reduced localization errors. Predictions at different scales are improved as a result of using concepts similar to feature pyramid networks.

YOLOv4 aims at building a faster object detector for production systems and optimizing parallel computation. YOLOv4 boosts the performance over YOLOv3 by using strategies like Bag of Freebies and Bag of Specials. Bag of Specials causes a marginal overhead in terms of the time required in the detection phase whereas Bag of Freebies improves the performance without additional overhead in terms of time. One such strategy in the Bag of Freebies employed in YOLOv4 is the CIoU loss proposed by Zheng et al. [11]. This is a loss function that considers the overlapping area, the distance between center points and aspect ratio, thereby achieving better convergence speed and accuracy. This loss function improves the convergence accuracy because in the case of no overlap, this loss forces the network to move the predicted bounding box closer to the ground truth box. A strategy DIoU NMS proposed by Wang et al. [12] is used in the Bag of Specials. Non-max suppression filters out the bounding boxes that improperly predict the same object and retains the one with the highest score. DIoU is a factor in NMS. DIoU considers IoU and an object's distance from the center. The CIoU and DIoU loss used in YOLOv4 help attain significant performance improvement in terms of IoU. This improvement is critical in the case of the proposed model, as the dimension of potholes are estimated by considering the size of the bounding box. Thus YOLOv4 gives a better estimation of the pothole dimensions.

Encoded image representation of the input image is obtained after passing it through the Convolutional Neural Network. Bounding Box Predictor predicts multiple candidate bounding boxes along with different anchors. Non-max suppression is applied to these bounding boxes, which yields the best bounding boxes with the highest IoU. Both the models are trained on the custom dataset for multiple iterations (3000, 4000, 5000 and 6000). The batch size is 64 and the number of subdivisions are 8 throughout the training process.

D. Dimension Estimation Module

The dimension estimation module of the proposed system is an image processing based solution to estimate the dimensions of the detected potholes. The module uses triangle similarity property: Consider we have an object of width 'W.' This object is placed at a distance of 'D' from the camera. By taking a picture of this object from a distance 'D,' we get the apparent width in pixels 'P.' Now, the perceived focal length 'F' of the used camera is derived as shown in (1).

$$F = (P * D) / W \quad (1)$$

This perceived focal length maintains the relation of inversely proportional pixel length and the camera distance. It is important to note that the pixel length is dependent upon the pixel density of the image measured in PPI (pixels per inches). The pixel length of an object will be larger for an image having higher PPI than for an image having lower PPI, even when both the images are taken from the same distance. Hence, it is recommended to apply preprocessing so as to fix the PPI of the input image for maintaining consistency. In the proposed set-up, since the same camera is used for calibration as well as for testing, PPI conversion is not required.

For the calibration process, an object with a sharp edge like a ruler-scale, whose length is known, is taken and multiple images are captured from various heights like 15cm, 30cm, etc. Edges of the object are detected by applying a Canny edge detector algorithm to get the pixel length of the object. The perceived focal lengths are calculated using the formula specified in (1) for each of the images with their respective parameters. The average of these focal lengths gives the final perceived focal length.

The pothole detection stage provides the bounding box coordinates, from which pixel length and pixel width is obtained. Using the fixed camera distance of 90 cm and the calculated perceived focal length, the actual length is calculated using the formula specified in (1).

IV. RESULTS AND DISCUSSIONS

The model is trained on the custom dataset for multiple iterations and the mAP and the IoU values are recorded. Validation accuracy is monitored throughout the training process to avoid overfitting. Early stopping regularization technique is used to prevent overfitting. The early stopping technique states that the training of the model should be continued as long as the validation score on the considered metric keeps improving over every iteration. The training process should be ceased at the point where the validation score degrades over the previous iteration.

In the case of the proposed system, the indicator used is mAP@0.50. Fig. 4.1 shows the plot of the mAP@0.50 score on the validation set. It is observed that the validation performance on the mAP@0.50 metric starts degrading post 4000 iterations. Therefore 4000 iterations should be considered as the stopping point in this case.

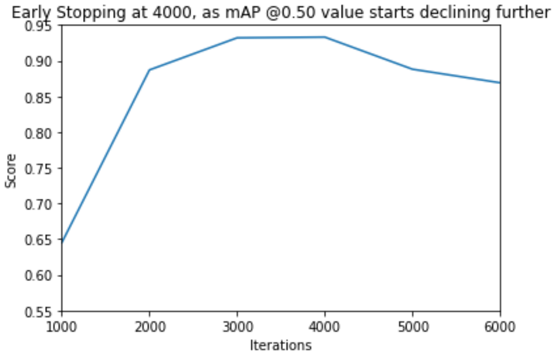

Fig. 4.1. mAP scores v/s Iterations

Table I shows the results obtained while training both the YOLOv3 and YOLOv4 based models. From the table, it can be inferred that the results peaked at 4000 iterations. The maximum mAP and IoU values obtained are 0.889 and 0.635 for the YOLOv3 based model and 0.933 and 0.741 for the YOLOv4 based model.

TABLE I. MODEL RESULT COMPARISON

Model	Iterations	mAP @ 0.50	IoU
YOLOv3	3000	0.860	0.592
	4000	**0.889**	**0.635**
	5000	0.875	0.612
	6000	0.861	0.592
YOLOv4	3000	0.932	0.727
	4000	**0.933**	**0.741**
	5000	0.888	0.717
	6000	0.869	0.716

Fig. 4.2 is the graphical representation of result metrics obtained for the YOLOv3 and YOLOv4 based models. IoU and mAP are used as metrics for performance evaluation. IoU is the ratio between the intersection and the union of the ground truth bounding boxes and the predicted bounding boxes. The mAP metric is a combined metric that takes into account both precision as well as recall.

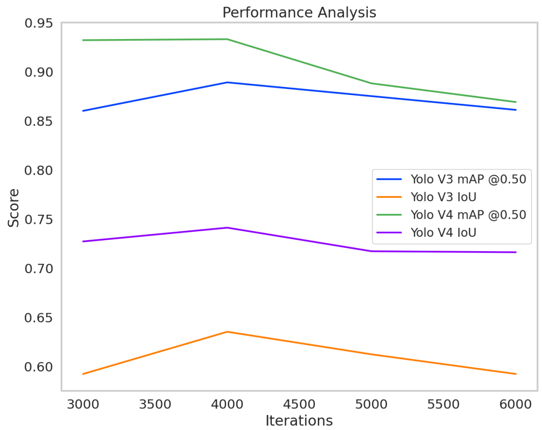

Fig. 4.2. YOLOv3 and YOLOv4 Performance Analysis

It can be inferred from the above graph that YOLOv4 gives better results as compared to YOLOv3. This behavior can be attributed to the strategies like Bag of Freebies and Bag of Specials used in YOLOv4. For instance, the CIoU and DIoU losses are used in YOLOv4. These loss functions consider overlapping area, distance between center points and aspect ratio thereby achieving better convergence speed and help attain significant performance improvement in terms of IoU.

Result-1: Fig. 4.3 compares the pothole detection results produced by the YOLOv3 and YOLOv4 based models. It can be seen in the results that YOLOv4 gives more precise bounding boxes thus better IoU value as compared to YOLOv3.


Fig. 4.3. Result comparison between YOLOv3 (left) and YOLOv4 (right)

Result-2: Once the pothole detection model provides predicted results, the coordinates of the bounding box are forwarded to the dimension estimation module. This module estimates the dimensions and prints a textual output. Fig.s 4.4, 4.5, 4.6 and 4.7 show the output of the Dimension Estimation Module. Ground-Truth values are mentioned at the bottom of the Fig.s and are written from left to right according to order of occurrence. The error rate in prediction of area based on estimated dimensions is calculated for all the images. The mean of this error is considered as the error rate of the Dimension Estimation Module. The error rate of the module is 5.868%.

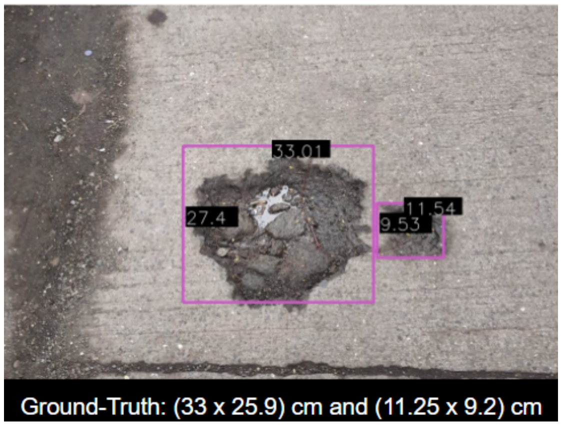

Fig. 4.4. Output of Dimension Estimation Module

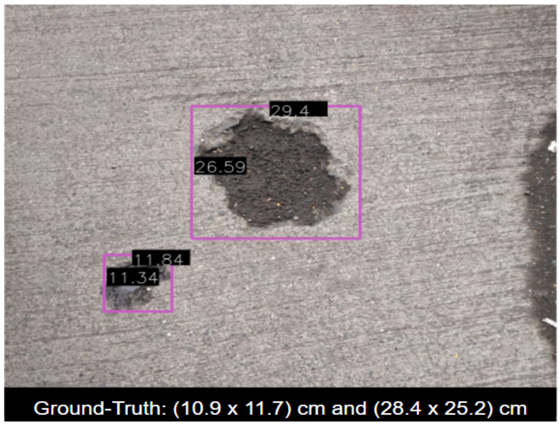

Fig. 4.5. Output of Dimension Estimation Module

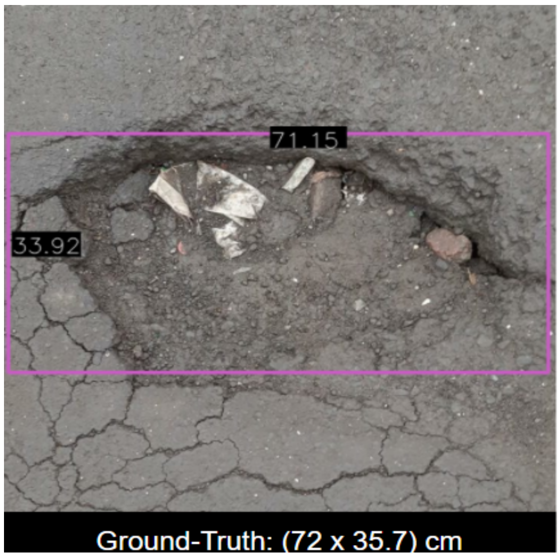

Fig. 4.6. Output of Dimension Estimation Module

Fig. 4.7. Output of Dimension Estimation Module

V. CONCLUSION AND FUTURE WORK

The system proposed in this paper will help reduce the dependency of human workers for the task of road maintenance, especially during the pandemic.

The paper proves that the YOLOv4 based model outperforms the YOLOv3 based model in detecting potholes accurately. The dimensions of the detected potholes are estimated with good accuracy and a considerably lower error rate. As YOLOv4 attains a better IoU, it provides a precise estimation of the dimensions of the potholes.

Future work involves extending the system by fitting it under surveillance vehicles, so the condition of the road can be precisely monitored automatically. These surveillance vehicles would also be fitted with a GPS module, so the exact location of the potholes can be recorded. The calculated pothole dimensions would help in estimating the extent of the road damage and would also help to estimate the raw materials required to fill the potholes. Thus most of the inspection and planning can be done remotely.

REFERENCES

[1] Sunil Kumar Sharma and Rakesh Chandmal Sharma, “Pothole Detection and Warning System for Indian Roads,” Advances in Interdisciplinary Engineering, 2019, pp. 511-519, doi:10.1007/978-981-13-6577-5_48, Springer.

[2] Lokeshwor Huidrom, Lalit Kumar Das and S.K. Sud, “Method for automated assessment of potholes, cracks and patches from road surface video clips,” Conference of Transportation Research Group of India (CTRG), 2013, pp. 312-321, doi:10.1016/j.sbspro.2013.11.124, Elsevier.

[3] SungWon Lee, SeokJin Kim, Kwang Eun An, Seung-Ki Ryu and Dongmahn Seo, “Image Processing-based Pothole Detecting System for Driving Environment,” International Conference on Consumer Electronics (ICCE), 2018, pp. 1-2, doi:10.1109/ICCE.2018.8326141, IEEE.

[4] Nhat-Duc Hoang, “An Artificial Intelligence Method for Asphalt Pavement Pothole Detection Using Least Squares Support Vector Machine and Neural Network with Steerable Filter-Based Feature Extraction,” Advances in Civil Engineering, vol. 2018, 2018, pp. 1-12, doi:10.1155/2018/7419058, Hindawi.

[5] Kumar, A., Chakrapani, Kalita, D. J. and Singh, V. P. , ”A Modern Pothole Detection technique using Deep Learning,” International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-5, doi:10.1109/idea49133.2020.9170705, IEEE.

[6] Byeong-ho Kang and Su-il Choi, “Pothole detection system using 2D LiDAR and camera”, International Conference on Ubiquitous and Future Networks (ICUFN), 2017, pp. 744-746, doi:10.1109/ICUFN.2017.7993890, IEEE.

[7] Ukhwah, E. N., Yuniarno, E. M. and Suprapto, Y. K. , “Asphalt Pavement Pothole Detection using Deep learning method based on YOLO Neural Network,” International Seminar on Intelligent Technology and Its Applications (ISITIA), 2019, pp. 35-40, doi:10.1109/isitia.2019.8937176, IEEE.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi. ”You Only Look Once: Unified, Real-Time Object Detection,” Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi:0.1109/CVPR.2016.91, IEEE.

[9] Joseph Redmon and Ali Farhadi, “YOLOv3: An incremental improvement,” 2018, preprint arXiv:1804.02767, arXiv.

[10] Alexey Bochkovskiy, Chien-Yao Wang and Hong-Yuan Mark Liao, ”Yolov4: Optimal speed and accuracy of object detection,” 2020, preprint arXiv:2004.10934, arXiv.

[11] Zhaohui Zheng, Ping Wang, Dongwei Ren, Wei Liu, Rongguang Ye, Qinghua Hu and Wangmeng Zuo, “Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation,” 2020, preprint arXiv:2005.03572, arXiv.

[12] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye and Dongwei Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” 2019, preprint arXiv:1911.08287, arXiv.