IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

REAL TIME OPEN LID POSITION DETECTION OF BULKERS USING YOLO V7TINY, FASTER RCNN AND MASK RCNN.

[1]Dr. Sharda Chhabria, [2]Atul Ghumade, [2]Harish S. Zambani, [1]Ayushi Thakre, [1]Hritik Debnath, [1]Jayvardhan Gholap, [1]Virja Kawade
[1]Artificial Intelligence, G H Raisoni Institute of Engineering and Technology, Nagpur, India,
[2]HIXAA Technologies Pvt. Ltd. Nagpur, India

Abstract: Bulkers are parked under the silo to collect ash. Currently there are no parking systems incorporated in small scale industries to assist in the process of detecting and aligning the collector lid and attaching the ash pipe or funnel, which leads to heavy unproductivity and waste of time. In these old systems, two to three people are required to align the funnel or pipe to the open lid of the bulker. This paper proposes the incorporation of Deep Learning based approaches to automate the process of detection and alignment. The solution includes using YOLO v7 Tiny, Faster RCNN and Mask RCNN algorithms and their comparison studies on accuracy and error and deploying the best one on field.

***Index Terms* - Deep Learning, Faster RCNN, Mask RCNN, Open Lid Detection and Recognition, YOLO v7**

I. INTRODUCTION

The bulk carriers which get parked under silo to collect ash require two to three people to park and align the bulker's open lid directly under the silo ash pipe outlet. This results in the wastage of time and human resources which have the potential to be utilized elsewhere. This study on open lid position detection system is designed to assist bulker drivers for aligning the collector lid and attaching the ash pipe or funnel. The software detects whether the lid is open or closed and simultaneously detects whether the position of the bulker is in alignment with the ash pipe or funnel. The model is a camera based intelligent system which will assist drivers using tablet PCs during parking without any external manpower, thus eliminating human resourcerequirements. The study also compares the training results using Yolo V7, Faster RCNN and Mask RCNN algorithms to detect real time open lid and its position.

The software design consists of total three components which are :

1. Custom Dataset and Model Training Software
2. Detection Testing Software
3. UI Software design for Deployment

This software is assembled on a tablet accessible to the driver. The driver shall be able to control the positioning, alignment and signal the ash depositor to unload the ash into the open lid of the bulker. This will thus eliminate the need of hassle on site for manual climbing and alignment of the ash depositor pipe.

fig. 1. bulker

The bulkers (fig.1) consist of Knappco automatic hatch covers, which are to be aligned with the ash outlet and then filled. These open hatch covers are to be detected and thus help the driver in parking the vehicle without any third party human assistance. This is illustrated in detail in fig. 3.

fig. 2. silo plant and bulk carrier

In this study the automation is achieved through three deep learning algorithms which are : Faster recurrent neural network (RCNN), Mask RCNN and the YOLO v7 (You Only Look Once). The paper summarizes the advantages of each algorithm on theapplication and then presents a comparison study based on it.

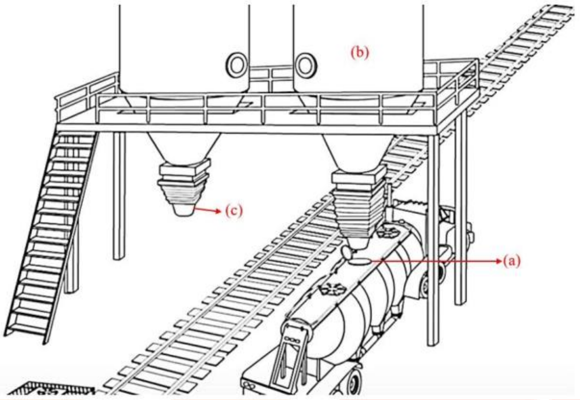

fig. 3. a) knappco automatic open hatch covers. these open hatch covers are annotated on image dataset and detected by thesoftware. b) silo ash tank. c) silo ash depositor pressure pipes

II. RELATED WORK

Previous studies on resolving the issues posed by the absence of automation for detection have been aimed at various applications like Lid Opening Detection in Manholes using RNN (Andrijašević et al.) [1]. This paper proposes the use of telecommunication manholes network equipped with smart Internet of Things (IoT) devices and sensors along with a recurrent neural network(RCNN) and LSTM (Long Short Term Memory) based approach for the manning of manholes. This has shown to reduce operational cost of manhole monitoring and increased accuracy along with effectiveness.

III. METHODOLOGY

The methodology we used in this study is based on SCRUM[]. SCRUM is a project management framework that was initially focused on software development, while it has now been applied to other areas such as research, sales, marketing, and cutting- edge technology. It is intended for groups with ten or fewer members who divide their work into tasks that can be finished in sprints, which are time-boxed iterations. The scrum master has significantly greater authority than a project manager when it comes to overseeing various software development processes and stages. Primarily, there are six phases in SCRUM which are as follows:

A. A product owner is initially in charge of compiling a product backlog, or a list of the tasks and specifications needed to complete the project. The product backlog is successfully prioritized by it.

B. The scrum team comes together to plan a sprint after that. The team determines where to really begin working after prioritizing the product backlog. The product backlog's subset becomes the sprint backlog at this point.

C. After that, there is a daily scrum where all of the scrum team members communicate with one another to discuss.issues and progress that were encountered during the sprint. The scrum master oversees this mostly to ensure that the team members are actually abiding by all of the scrum ideas, rules, and practices.

D. The product owner arranges a sprint review meeting when the sprint is completed. The development team presents their progress since the last sprint during this meeting. The product owner then discusses what still needs to be accomplished on the product backlog before estimating when the project will be finished.

E. After that, the entire scrum team gathers for a sprint retrospective meeting to evaluate what went well, what didn't, and how they could have done it better. This can be the result of a technical problem or a team member who is overworked. The decision-makers make plans to address these problems in the upcoming sprint after deciding how to resolve them.

F. For the remaining jobs on the product backlog, the cycle is repeated. This continues until either of the following occurs:

a) The submission deadline has passed.

b) The budget has been used up.

c) The finished product meets the needs of the product owner.

IV. DATASET

A. Data Collection

The raw video feed was recorded from the client camera from the field. For training purposes, remote access to the client enabled the collection of recorded video feeds for training purposes. After the deployment of the model on the client system, the video feeds for detection are obtained live from the camera installed on the field.

B. Data Transformation

This was converted to individual frames on which the models were trained. Initially, the raw images were of high pixel size. So, the sizes were reduced to 416x416 to decrease the computation load and increase the training and detection times. After size reduction, the images were annotated. The process of categorizing photos in a dataset in order to train a machine learning model is known as image annotation. For functional datasets, labeling images is essential because it informs the training model on the crucial components of the image (classes), allowing it to use these notes to later recognise those classes in brand-new, previously unseen images. The class labeled for the recognition task is "Open Lid".

table. 1. data representation (annotations)

label_name	bbox_x	bbox_y	bbox_width	bbox_height	image_name	image_width	image_height
open_lid	229	118	93	43	Img (1000).jpg	416	416
open_lid	243	134	85	46	Img (1001).jpg	416	416
open_lid	244	70	53	18	Img (1002).jpg	416	416
open_lid	232	121	93	38	Img (1003).jpg	416	416
open_lid	241	133	92	48	Img (1004).jpg	416	416
open_lid	243	71	58	16	Img (1005).jpg	416	416

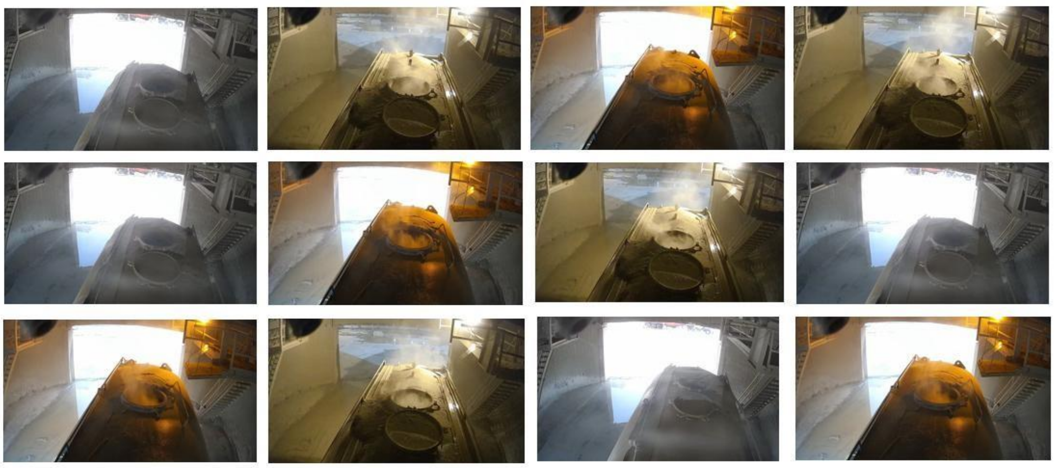

fig. 4. bulker open lid images for training

V. MASK RCNN - ARCHITECTURE IMPLEMENTATION

Kaiming et al. [2] first introduced an object segmentation mask output along with the existing bounding box and classification for object recognition task, thus extending the Faster Recurrent Neural Network (RCNN) model. The only drawback is that Mask RCNN adds a low overhead of 5fps which is negligible. Another development is the introduction of the RoIAlign which is a quantization-free layer. This fixes the misalignment in Faster RCNN which did not perform pixel-to-pixel alignment by preserving exact spatial locations. This is shown to increase the accuracy of the results by a relative 10% to 50%.

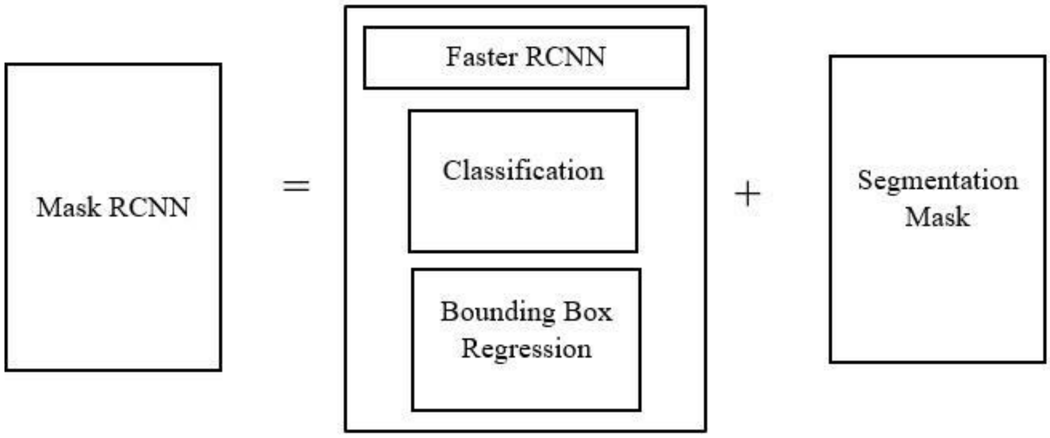

fig. 5. mask recurrent neural network (rcnn)

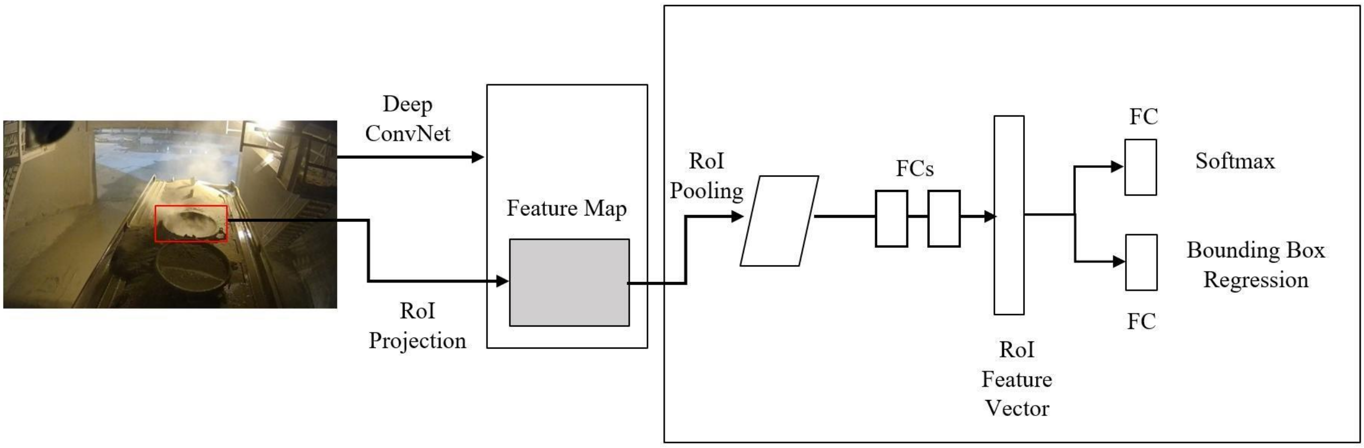

fig. 6. you only look once (yolo) architecture

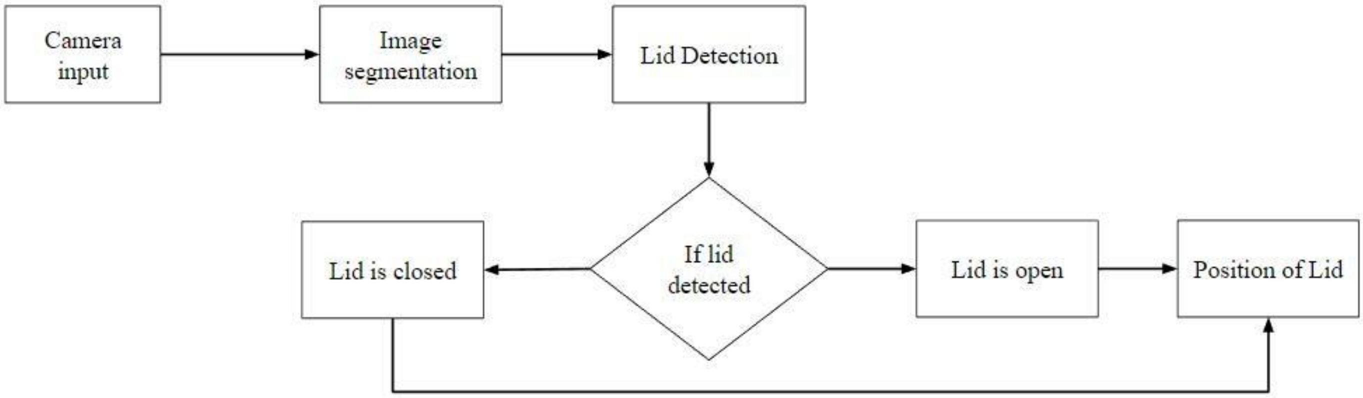

fig. 7. block diagram of the approach employed by this paper

VI. FASTER RCNN - ARCHITECTURE AND IMPLEMENTATION

The Faster RCNN was first proposed by Ren et al. [3]. This is one of the most widely used versions of the recurrent neural network group. The Faster RCNN consists of four basic components :

A. Region Proposal Network (RPN) - This generates bounding boxes for the region that the RCNN suspects containing the object.

B. Feature Generation Stage - The features of the object recognized using the RPN are generated. This is usually performed using a convolutional neural network (CNN).

C. Classification Layer - This classifies the detected objects into predefined categories.

D. Regression Layer - This is deployed to make the coordinates of the bounding boxes more precise and accurate.

VII. YOLO v7 - ARCHITECTURE IMPLEMENTATION

You Only Look Once (YOLO) is a single stage object detector. Single stage object detectors view the object detection task as a simple regression problem. They skip the region proposal network (RPN) stage, which is generally found in two stage detectors. Therefore the detection is done directly upon the feature map instead of passing the feature map through the RPN layer. Thedetailed architecture of YOLO is depicted in fig.6. YOLO has over 5 versions and has been used in various applicationsthroughout almost every industry. YOLOv1 was the first single-stage object detector approach that addressed detection as a regression problem, and it represented a significant advancement in the field of object identification proposed by Redmon et al.[5]. YOLO v2 was proposed by Redmon et al.[6] in 2017. Although it is similar to the YOLO v1 model, the only difference is in the training strategies. It was trained on Pascal VOC and MS COCO datasets predicting more than 9000 different categories. YOLO v3 was introduced by Redmon et al.[7] and it introduced a new architecture called the Darknet-53. It is as accurate as Single-Shot Detector (SSD321) and three times faster. Bochkovskiy et al. proposed YOLO v4 in 2020 [8]. It raises YOLOv3's mAP and FPS by 10% and 12% respectively. The YOLOv4 model merged various features with others to create “Bag of Freebies” for better model training and “Bag-of-Specials” for better object detection accuracy.

fig 8. results on the test set in mask rcnn

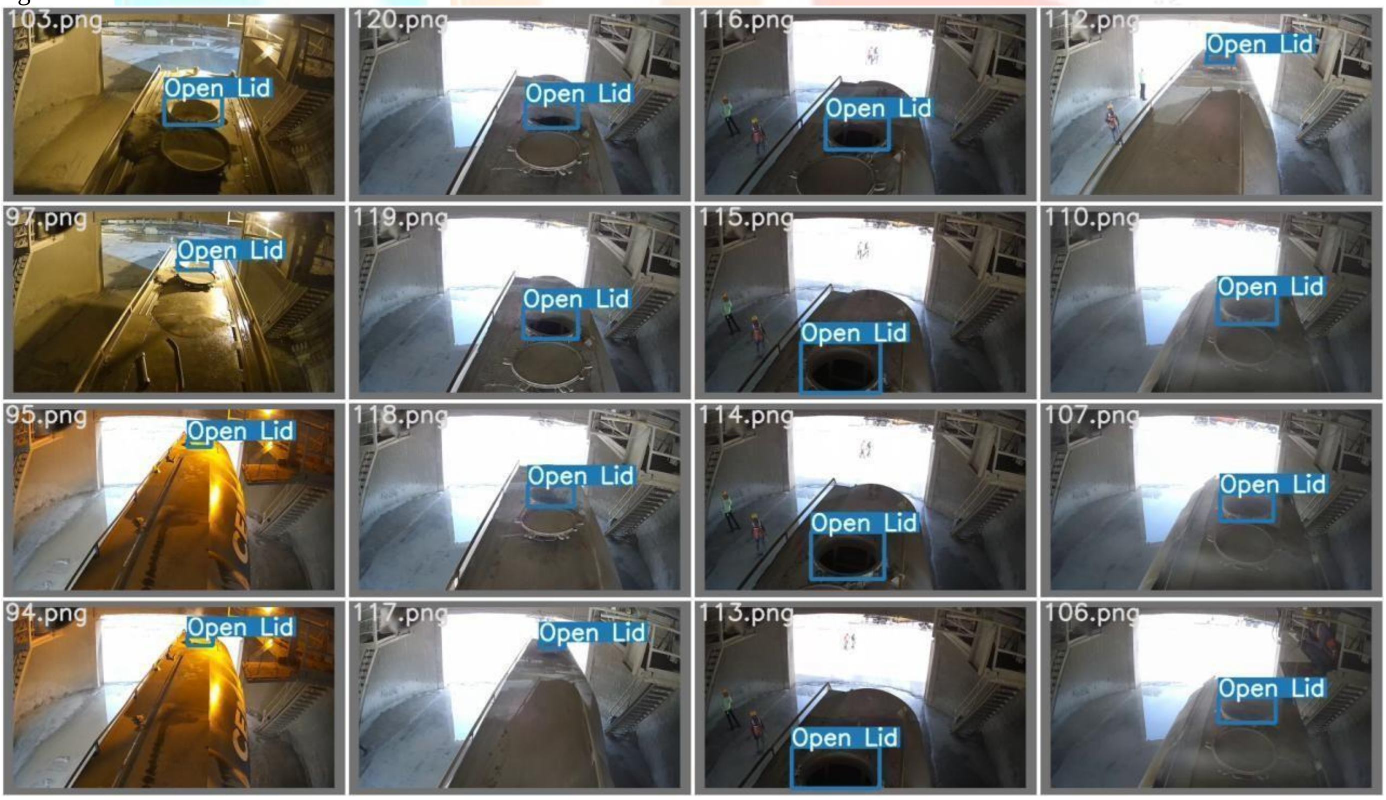

fig. 9. results on the test set in yolo v7

To determine the location of the items and their class labels, the detection architecture simply took a single look at the image. While Faster RCNN and similar algorithms use the Region Proposal Network to identify potential regions of interest before performing recognition on each of those regions independently, YOLO v7 tiny conducts all of its predictions using a single fully connected layer. Modern real-time object detectors may effectively have their parameters and computations reduced by roughly 40% and 50%, respectively, by using the method that is suggested by Wang et al. [4]. It also has a faster inference speed and a higher detection accuracy. A new real-time object detector architecture and associated model scaling strategy are also proposed.[4] To improve the accuracy of object detection, the study conducted by Wang et al. [4] suggests using the trainable bag-of-freebies method. On the basis of the aforementioned, it created the YOLOv7 tiny series of object detection systems, which produces cutting-edge results. The test results which show considerably high accuracy and inference speed are the reasons for the selection of this model for the real time deployment in industry.

VIII. COMPARISON OF RESULTS OF THE STATE-OF-THE-ART MODELS

A) Faster RCNN Results

Fig. 7. shows the results of test images passed to the Faster RCNN detector.

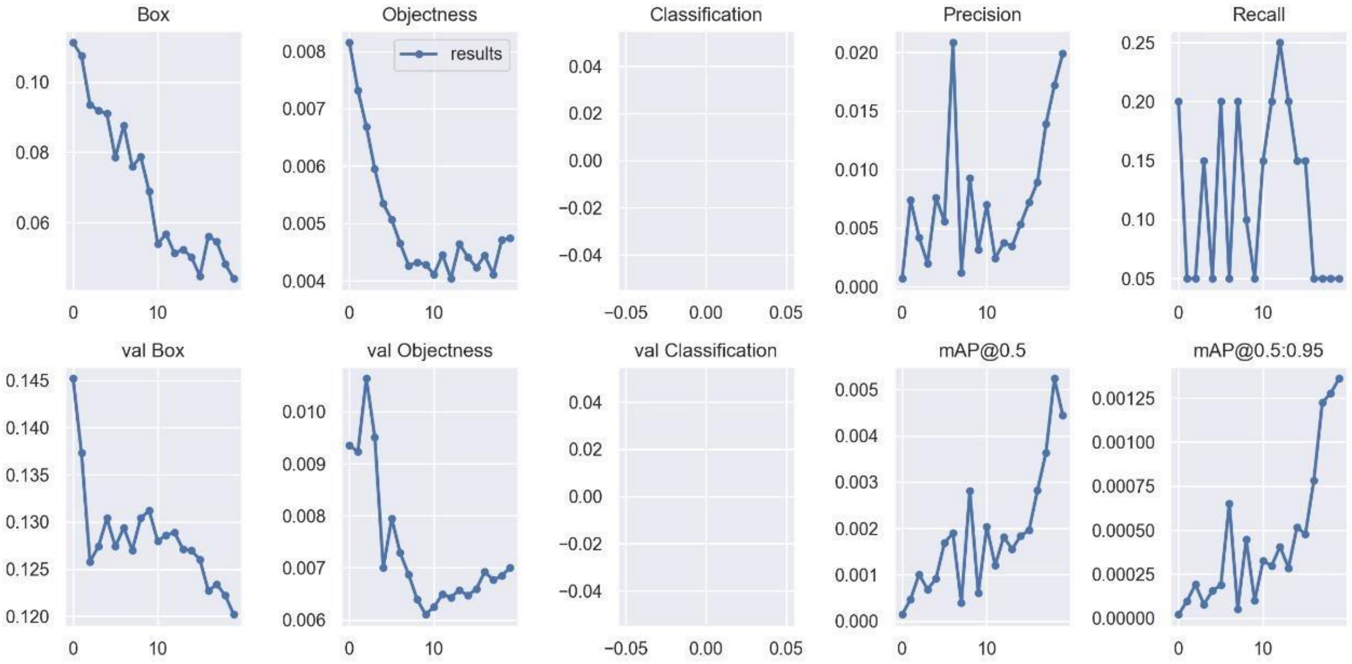

fig. 10. precision, recall and map (mean average precision) curves of faster rcnn on the test dataset

B) Mask RCNN Results

Fig. 8. shows the results of test images passed to the Faster RCNN detector. The only difference in the results of the Mask RCNN model is the presence of segmentation masks in the output images and video feeds. Although the accuracy is worthwhile, the speed of the output and processing is high which is the reason that Mask RCNN was not used for deployment purposes.

C) YOLO v7 tiny Results

Fig. 9. shows the results of test images passed to the YOLO v7 tiny detector. The YOLO v7 tiny performed better than the other two models when tested on recorded video feeds and images in several categories like the presence of lesser number of parameters which contributes significantly to the faster inference speed of YOLO v7 tiny. It is also better at detecting far-off or small objects than Faster RCNN and Mask RCNN. Thus, for real time deployment on the client system, YOLO v7 tiny has been used as it is both accurate and fast.

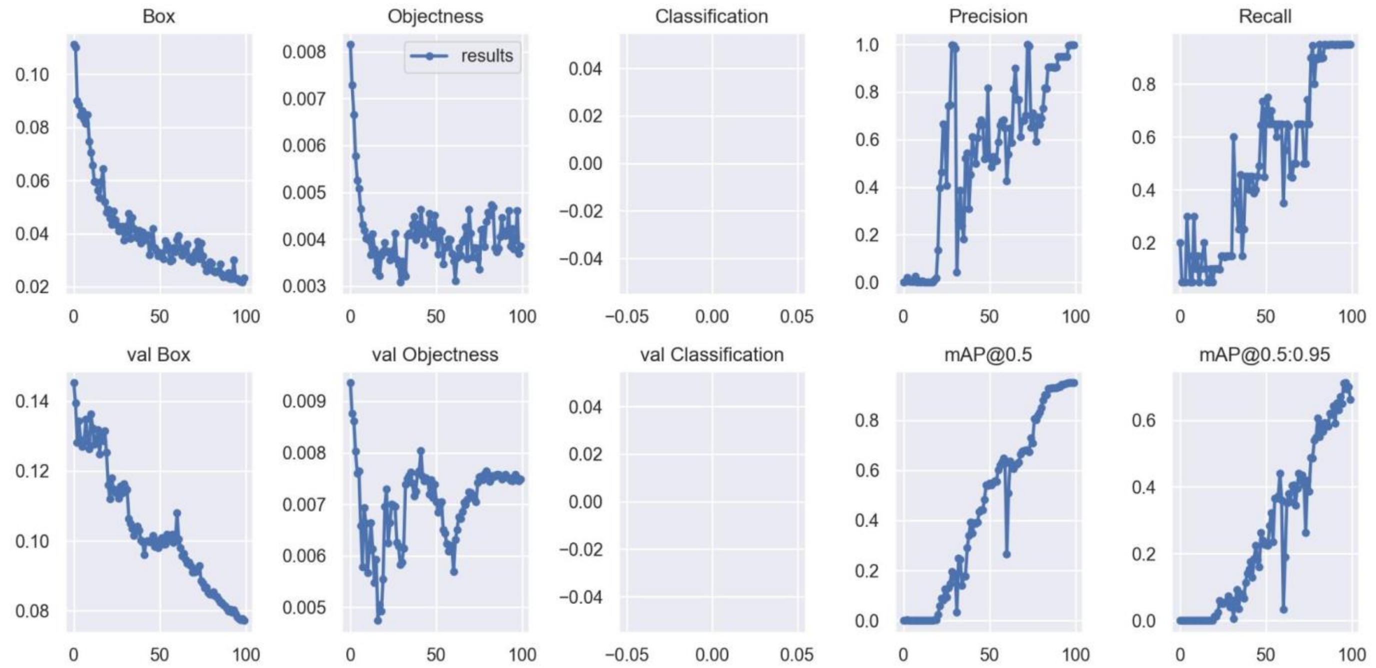

fig. 11. precision, recall and map (mean average precision) curves of yolo on the test datasettable. 2. comparison of

faster rcnn, mask rcnn and yolov7 tiny models

Model	Parameters	mAP	Inference Speed
Faster RCNN	53 M	82.8 %	7 FPS
Mask RCNN	44 M	78.1 %	3 FPS
YOLO v7 tiny	6.2 M	77.8 %	37 FPS

IX. CONCLUSION

This paper makes use of three algorithms - Faster RCNN, Mask RCNN and YOLO v7 tiny models for open lid detection and positioning. This shall reduce wastage of human resources which can be utilized at a more productive task. The costs of employing manpower for this task shall also be reduced thus saving a considerable amount of money for the institution utilizing the software. Time is also a factor that would be optimized after the successful installation of the software. The paper presented a concise comparison study of the YOLO family and the main differences between them after experimentation on the working of each. Implementation and comparison of results of the three models led to the inference that deployment of YOLO v7 tiny would be the better choice amongst the three strategies.

X. Acknowledgment

The authors wish to thank HIXAA Technologies Pvt. Ltd. and Thermal Power Plant for providing live video feeds of bulkers for compilation of the dataset used for this study and remote access to the on-site camera for live testing. They also want to acknowledge G. H. Raisoni Institute of Engineering and Technology, Nagpur, for providing access to NVIDIA GPU for faster and optimized training of the algorithms.

REFERENCES

[1] Andrijašević, Uroš, Jelena Kocić, and Vladan Nešić. 2020."Lid Opening Detection in Manholes using RNN." 2020 28th Telecommunications Forum (TELFOR). IEEE, pp. 1-4

[2] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. 2017. “Mask R-CNN”, Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961-2969

[3] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. 2015. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Advances I n Neural Information Processing Systems 28

[4] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2017."YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv preprint arXiv:2207.02696

[5] Redmon, Joseph, et al. 2016. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788

[6] Redmon, Joseph, and Ali Farhadi. 2017. "YOLO9000: better, faster, stronger." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271

[7] Redmon, Joseph, and Ali Farhadi. 2018. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767
[8] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020."Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934