Yolo-based Deep Learning Techniques for Identifying Floating Bottles in Inland Water: A Comprehensive Analysis

B Saranya Devi
Department of Computer Science and Engineering,
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
b_saranya@blr.amrita.edu

Deepa Gupta*
Department of Computer Science and Engineering,
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
g_deepa@blr.amrita.edu

Rimjhim Padam Singh
Department of Computer Science and Engineering,
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
ps_rimjhim@blr.amrita.edu

2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-7024-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCCNT61001.2024.10724494

***Abstract*— Floating trash like bottles are one of the main causes of the serious environmental threat posed by the growing pollution of inland water bodies. Conservation efforts are hampered by the lack of effective and affordable techniques for locating and retrieving these bottles. Real-time object detection algorithms in conjunction with unmanned boats can offer ways to remove floating trash from rivers. A popular real-time object detection model is the You Only Look Once (Yolo) Series. Based on efficacy and accuracy rate, each YOLO model gets better than the other. The study presents an extensive analysis of YOLO designs, such as YOLOv5, YOLOv7, YOLOv8, and YOLO-NAS, for bottle identification in inland water. The dataset used is a floating trash in the river, and the performance metrics such as precision, recall and mean average precision are employed, to compare and experimentally validate each algorithm's detection efficiency and accuracy. The methodology trains each architecture for 200 epochs. Regarding the precision of detection, the experimental results shows that YOLOv8-x model has high value of 85.5% and mAP@[IoU=0.5:0.95] of 43.5%. YOLO-NAS variations have high recall, they are applicable for minimizing missed detection. YOLO-NAS m has a higher recall value of 90%.**

Keywords—Computer vision, deep learnings, object detection, floating trash, bottle detection, Flow-Img, YOLOv8, YOLONAS.

I. INTRODUCTION

Rivers, lakes, and streams are examples of inland water features that are essential to both human societies and ecosystems. Unfortunately, the amount of debris in these ecosystem is growing, creating serious ecological and esthetic problems. Trash in inland waters, especially bottles, is a serious environmental problem that has drawn attention recently. The word "trash bottle" describes used plastic bottles that wind up in inland waterways. These bottles present risks to the environment, the ecosystem, and public health when improperly disposed of or recycled. Due to their inability to break down, plastics and froth found in floating in rivers eventually through surface runoff, find their way into the ocean, where they contribute significantly to the pollution of the ocean with marine litter and microplastics [1].

While manually collecting trash from rivers and lakes is a commendable effort to combat pollution, it has several drawbacks. First of all, the procedure can be physically demanding and labor-intensive, requiring a large amount of human resources and time. Additionally, as they move through potentially dangerous situations like contaminated or sharp objects, unstable riverbanks, and polluted water, their safety may be threatened. Although manual debris collection helps with immediate cleanup, its limitations regarding size, effectiveness, and safety emphasize the need for sustainable waste management techniques and complementary approaches.

An innovative way to deal with the widespread problem of waterborne debris in rivers, lakes, and other water bodies is to use unmanned boats for trash collection. Unmanned boats have several advantages when it comes to collecting trash, one of which is their capacity to reach difficult-to-reach or challenging locations that may present challenges for human-operated vessels. These boats, which are outfitted with cameras and sensors, can identify and classify various waste materials on their own, including plastic waste and floating debris, thus streamlining the collection procedure.

Larger or irregularly shaped objects may be difficult for unmanned boats to collect efficiently. This restriction makes it difficult to carry out exhaustive cleanup operations, especially in water bodies with different types of debris. An accurate floating waste detection system that operates in real-time is essential for achieving efficient and dependable autonomous cleaning of unmanned boats.

Deep learning models are used to recognize plastic bottles once high-resolution pictures or videos of the water's surface are taken. Convolution neural networks (CNNs) are employed for object detection, image preprocessing approaches to improve visibility in different water conditions, and a reliable classification system for trash identification are the key elements of our methodology.

Challenges associated with detecting floating bottle objects are limited spatial resolution, low contrast, localization accuracy, scale variability, occlusion, reflection of light. There are several types of computer algorithms used in object detection such as the Faster RCNN model [2], which inventively generates proposal boxes and Fast RCNN [3] by utilizing the RPN network falls in two stage detectors. End-to-end detection is now possible, and the speed and accuracy of detection have significantly increased. The Yolo series algorithms [4], SSD [5], RetinaNet [6], belong to one-stage object detection algorithms.

This paper discusses literature review in section II, dataset description in section III, methodology, different YOLO models and performance metrics in section IV, and experimental analysis in section V.

II. RELATED WORK

As part of our introductory phase, few literature papers relevant to the research work on floating trash detection, small object detection was done. This provides insights into the current models used in Deep learning.

Pu et al. [7] have proposed the FloW-Img dataset that has been used to train five models: Faster RCNN, Cascade R-CNN, YOLO v3, and YOLO v5. Mean Average Precision (mAP) values were computed at various thresholds. The results showed that small-scale objects had a much higher rate of erroneous and missing detections. For small objects, YOLO v3 showed the lowest missed detection rate. For medium-sized objects, Faster RCNN had the least false detection rate.

Zhang et al. [8] proposed a hybrid attention mechanism with YOLOv5 to assist the communication among channels over a long distance while sustaining the direct correspondence between channels and their weights.

Casas, Edmundo, et al. [9] used the YOLO series to detect wildfire and smoke. YOLO-NAS variations have high recall, they are applicable for minimizing missed detection YOLOv5, YOLOv7, and YOLOv8 algorithm showed better results amongst all the metrics.

Renfei Chen et al. [10] suggests an enhanced object detection algorithm—a single shot multi-Box detector for floating targets to handle the detection task of small floating targets. Domain adaptation method and detector is combined to increase the model's detection accuracy for small floating targets in a new target domain.

This paper proposes a fusion-based approach for object detection by spatial-temporal information. This work enhances the high-resolution layers of the Single Shot Multibox Detector (SSD) network design to better accommodate the detection of small floating [11]

Cheng et al. [12] published the floating trash detection dataset in inland waters, called FloW. This consists of 2 sub – datasets: FloW-Img consist of 2000 annotated images and FloW- RI consists of 200 video sequences without annotation. 6 algorithms are used on this model: DSSD, RetinaNet, YOLOV3, Faster R-CNN, FPN and CASCADE R-CNN. mmAP average precision under IoU(Intersection over union) is calculated under different threshold values: 0.5,0.5:0.95.

Mosaic data augmentation is used by the FMA-YOLOv5 algorithm to improve small target detection during training. To ensure channel consistency, it integrates FMA layers with 1×1 convolutions and a self-attention mechanism. In the neck area, FPN and PANet are used for feature fusion. The model achieves a remarkable mAP of 79.41% and 42 FPS, outperforming YOLOv5s by 2.18% on the test dataset [13].

Zhang et al. [14] suggested an enhanced RefineDet model for detecting floating which consists of three modules: the transfer connection block, the anchor refinement module, and the object detection module.

To facilitate the removal of invasive aquatic weeds using Unmanned Floating Vehicles (UFVs) at a cheap cost, this research presents an object detection algorithm. The first phase is the accurate and real-time identification of waterlines with the K-means algorithm, which is compared to the Hough transform. After the detection of the water zone, a modified gradient-based image processing technique is useful in object detection within the water region. The approach uses the Finite Difference Central Gradient Operator instead of the Sobel Gradient Operator, which improves the accuracy of object detection in the water zone. The suggested method is to automate UFVs so they can navigate autonomously and remove weeds from water bodies with efficiency [15].

This work uses Mask R-CNN for instance segmentation and focuses on the TACO dataset for litter identification and segmentation. With the help of padding and scaling, the model employs ResNet-50 in a Feature Pyramid Network to obtain an input layer size of 1024 x 1024 pixels. The evaluation consists of two tasks: TACO-10, which is the identification and classification of 10 different litter classes, and TACO-1, which is classless litter detection. As the datsset is tiny, 4-fold cross-validation is used. Each fold is randomly divided into 80:20 ratio for training and validation respectively [16].

A popular method for background subtraction and moving object detection is the Gaussian Mixture Model (GMM). An improved GMM-based automatic segmentation method (IGASM) detects the water surface floats [[17].

A unified network is created by combining RPN and Fast R-CNN. The RPN component directs the combined network's gaze using "attention" mechanisms [18].

In conclusion, small objects are easily missed or misclassified, detecting them is an important task in computer vision. Though there are numerous deep learning algorithms, but due to the data's size, shape, and other issues like reflection and illumination, 10–20% of it is always missed while detection. Hence the work is carried out using the recent object detection algorithm YOLO for detecting the floating objects in the inland water.

III. DATASET DESCRIPTION

The first ever dataset for the perspective of water surface floating garbage detection from an unmanned ship Flow[12] is used for the experiment. This dataset consists of 2 sub datasets namely FloW_Img and FloW_RI. FloW_Img dataset is a gathering of images of plastic bottles and aluminium can float in inland water which is being captured from various directions, angles, and on different lighting and wave conditions. The dataset consists of 2000 images, where 1200 are used for training and 400 used for validation and 400 used as test images.

TABLE 1. Analysis of the Flow_img dataset based on the size.

Images	***Small***	***Medium***	***large***	***Total***
Training	1797	1263	188	3248
Validation	600	350	57	1012
Test	599	361	57	1012

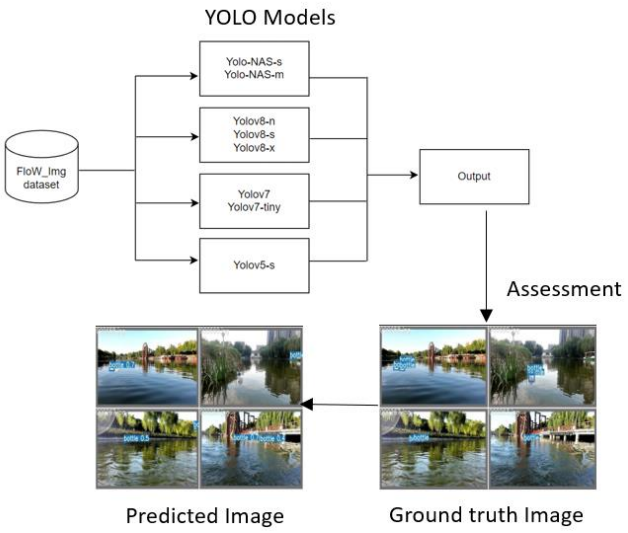


Figure 1 . Flow Diagram for bottle detection

There are totally 5272 labels on 2000 images, these labelled objects are further divided into three subcategories based on the area. Table 1 shows the analytical statistics of FloW_Img dataset grounded on the three subcategories.

Small targets(small) of area lesser than 32x32, medium targets(medium) of area between 32x32 and 96x96 as, and large targets(large) of area greater than 96x96 as.

IV.METHODOLOGY

The methodolgy for small object detection involves a systematic approach to address the challenges associated with detecting objects of small size in image or video. Figure 1 shows the flow diagram used for floating bottles in inland water using the FloW_img datset. The dataset is trained on several Yolo models, first assessment is done on test dataset which and second assessment is done on custom dataset with confidence score greater than 0.50.

A. Object detection algorithm

Figure 2 shows the YOLO algorithm's[19] network structure, an input image is taken and resized to 448 x 448 before passing it through the convolution network. By suppressing the less likely bounding boxes, non-max suppression (NMS) chooses the most suitable bounding box and provides the result.

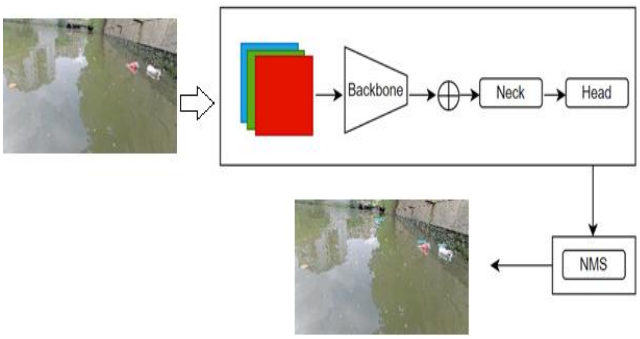


Figure 2. YOLO algorithm network structure.

1) Yolo v5 algorithm

The backbone is CSPDarkNet53. This structure stacks multiple CBS(convolution +Batch normalization+ SiLU) modules and component of 3 convolution layer (C3) is the bottleneck modules and one SPPF module is connected. Neck of YOLOv5 [20] uses two methods: Feature pyramid Network (FPN) and Path aggregation network (PANet) boosts the information flow. FPN upsamples the output feature map generated by multiple convolutions down sampling operations from the feature extraction network [21].

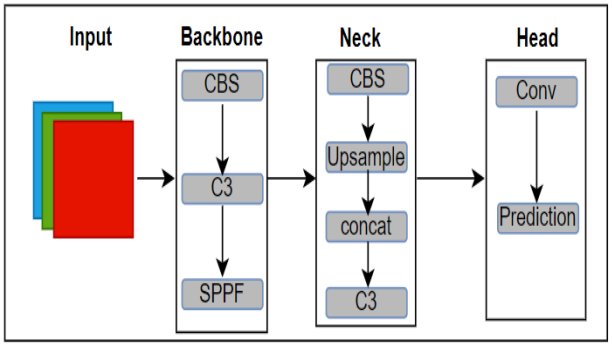


Figure 3 . YOLOv5 architecture

2) Yolo v7 algorithm

Yolo architecture is a fully connected neural network. There are three main components in Yolo framework: backbone, head and neck. Essential features of an image is extracted in the backbone which are of different sizes, these extracted features are fused to the head through the neck. Neck creates the feature pyramids. Head consists of output layers and has final detections. By several architectural reforms, YOLOv7 [22] enhances both the speed and accuracy compared to the previous versions. The architectural changes are E-ELAN (Extended Efficient Layer Aggregation Network) used as a backbone, SPP CSPC is a Cross stage partial network (CSPNet) with spatial pyramid pooling layer (SPP) block. CSPNet separates the feature map of the base layer. There are seven models: YOLOv7, YOLOv7-Tiny, YOLOv7-X, YOLOv7-W6, YOLOv7-E6, YOLOv7-D6, YOLOv7-E6E.

YOLOv7 and YOLOv7-tiny model is used for training the dataset.

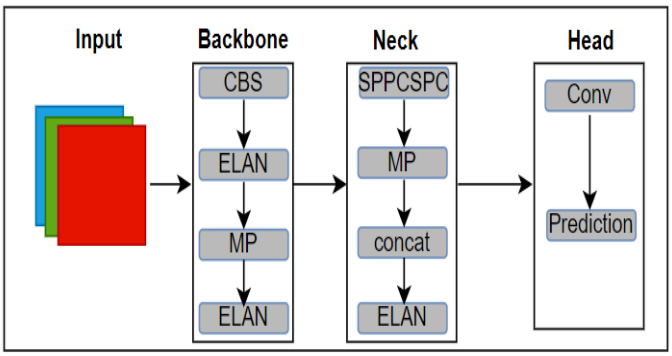


Figure 3 . YOLOv7 architecture

3) *Yolo v8 algorithm*

YOLOv8 follows three primary architectural elements backbone, neck and head that make up the YOLO framework. As shown in Figure 4, these parts function together, with the CPSDarknet53 Backbone extracting features from image, which is followed by the Course to fine (C2f) module instead of the traditional Neck architecture, decoupled head performing the prediction of object locations and classes separately thereby increasing the performance.

Like its predecessors, Yolov8 is available in five different versions: extra large (x), large (l), medium (m), small (s), and nano (n). Yolov8-n and Yolov8-s and Yolov8-x versions are used in this work.

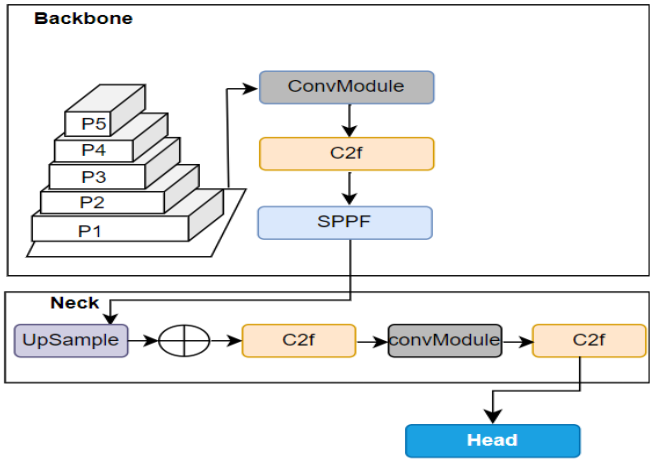


Figure 4. YOLOv8 architecture

4) *YOLO - NAS*

Deci.ai's creative real-time object detection model, YOLO-NAS, is the recent deep learning technology. In YOLO-NAS, "NAS" stands for "Neural Architecture Search. Finding the ideal balance between complexity in computation, size of the model and accuracy is the main objective of NAS. Figure 5 shows the high-level overview of the YOLO-NAS architecture. YOLO-NAS [23] comes in three variations: large (l), medium (m), and small (s).

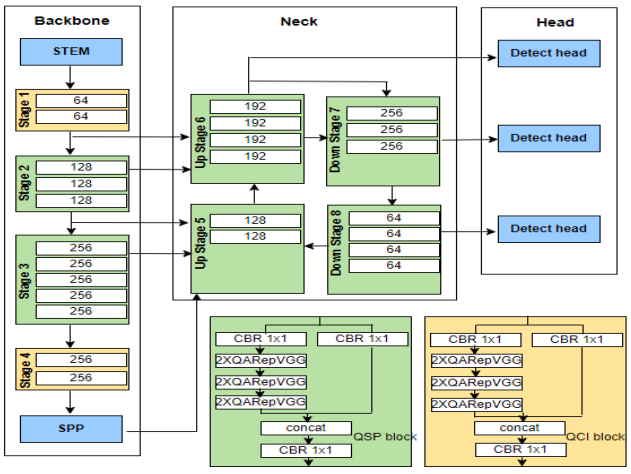


Figure 5. YoloNAS architecture

B. *Perfromance Metrics*

Precision(Pr), Recall(Re), Mean Average Precision(mAP), and loss are the metrics considered to detect the object based on the literature review.

Loss is defined as the difference between a model's expected and actual outputs. When it comes to object detection, the loss is used to assess how well the model detects objects in an image in relation to their actual locations and labels. Box loss, objectness loss, and classification loss are used in YOLOv5 and YOLOv7. YOLOv8 makes use of distributional focal lass (dfl), box loss, and classification (cls) loss. Finally, IoU loss, cls loss, and dfl are used in YOLO-NAS. Below, we give a brief explanation of each of these losses.

a. Box loss: The difference between the ground truth box coordinates and the predicted bounding box coordinates is measured by this loss function. Usually, it makes use of metrics like smooth L1 loss or mean squared error (MSE).
b. Objectness (obj) loss: This calculates the error in detecting whether an object is present inside the grid or not.
c. Classification (cls) loss: The difference between true class labels and predicted class probabilities. Categorical cross-entropy loss is used.
d. Distributional Focal Loss (dfl): This function addresses class imbalance in object detection and is an addition of the focal loss. It gives tougher examples that are more difficult to correctly classify a higher weight.
e. Intersection over Union (IoU) loss: The consistency between the ground truth boxes and the predicted bounding boxes is assessed by the IoU loss using the IoU metric.

V. EXPERIMENTAL RESULTS

The evaluation metrics mean average precision (mAP)@0.50 [24], recall, mAP@0.50:95 [24], and precision are used to test the model. The formula to calculate the mAP is shown in Equation 1. The average precision of each class is averaged, to obtain the mAP value.

$$\text{mAP} = \frac{\sum_{i=1}^{C} APi}{C} \qquad (1)$$

Eight models- Yolov5-s, Yolov7, Yolov7-tiny, Yolov8-n, Yolov8-s, Yolov8-x, Yolo-NAS-s, and Yolo-NAS-m—are used to train the dataset. Each of the eight models has undergone 200 training cycles.

Table 2. Performance metrics of YOLO-NAS for bottle detection

Model	*Pr@0.50*	*Re@0.50*	*mAP@0.50*	*mAP@0.5:0.95*
YOLO-NAS s	0.102	0.895	0.796	0.367
YOLO-NAS m	0.111	0.91	0.796	0.376

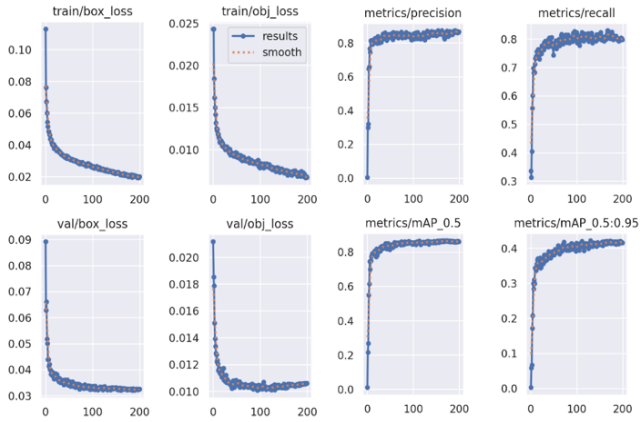
Table 2 presents performance metrics that provides insight of the behavior showed by the various YOLO-NAS variants. The recall @0.50 is high, above 89% which indicates that most actual positive cases are detected. On the other hand, the Precision rate is very low, which indicates that there is a high

rate of detecting the false positives. The mAP@0.50 for both the models is 79.6%. Yolo-NAS outperformance only for the recall value at mAP@0.50.

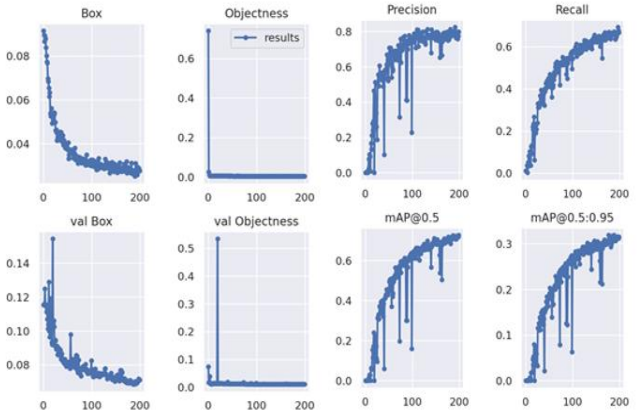
Table 3 represents the performance metrics for all the eight models, on the test data. Yolov8x has the highest mAP value at IoU[0.5-0.95] at 43.5%, and Yolov5 has the highest recall of 78.8% and mAP value at IoU[0.5] at 83.9%.

Table 3. Performance metrics of Yolo models for bottle detection

Model	***Pr***	***Re***	***mAP@0.50***	***mAP@0.5:0.95***
YOLOv5s	0.842	0.788	0.839	0.41
YOLOv7	0.821	0.7	0.811	0.385
YOLOV7-tiny	0.773	0.61	0.66	0.263
YOLOv8n	0.848	0.75	0.82	0.404
YOLOv8s	0.852	0.775	0.836	0.427
YOLOv8x	0.855	0.765	0.828	0.435

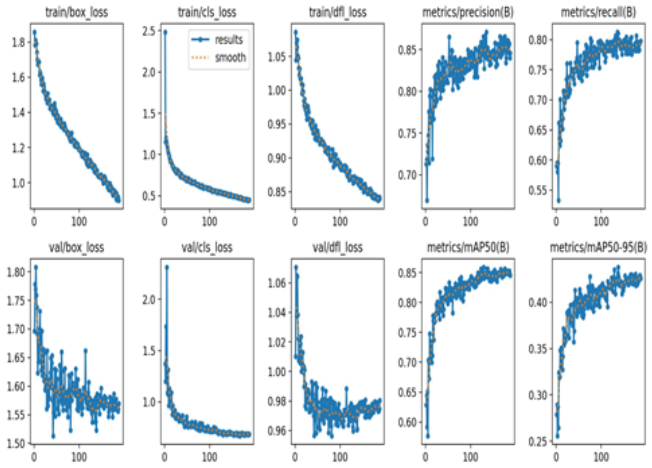


A: YOLOv5-s

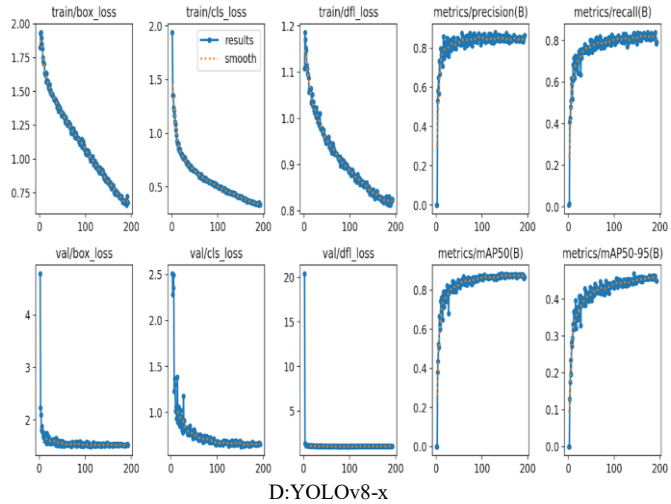


B:YOLOv7

As we move towards the in-depth analysis of evaluation metrics. Figure 6. A, B and C and D shows the graphical representation of the training and validation loss, and other metrics such as precision, recall, mAP@0.50 and mAP@0.50:0.95 of YOLOv5, YOLOv7 and YOLOv8s, YOLOv8n model respectively. Box loss and obj_loss shows the number of epochs in the horizontal axis and error percentage on the vertical axis. We can infer that the loss value is reducing during the training process. mAP@0.5:0.95 for YOLOv5(A) is ranging between 38% to 42% from 90th epoch. For YOLOv7(B) is ranging between 30% to 33% from epoch 180. YOLOv8s(C) is ranging between 40-44% from 150th epoch. And YOLOv8n is ranging between 38-40% from 100th epoch. Therefore, YOLOv8s is giving better mAP value compared to the other models.



C:YOLOv8-s



D:YOLOv8-x

Figure 6. A, B, C and D is the graphical visualization of training loss, validation loss and other performance metrics of Yolov5-s, YOLOv7, YOLOv8-s and YOLOv8-x respectively on training data.

Figure 7 shows the prediction of bottles from each model in the test data from four different scenes. In scene A, YOLOv7 is detecting floating objects with high accuracy rate. In scene B, Yolov8n is incorrectly predicting the image. In scene C, all the objects are identified correctly, YOLOv7 identifies the object with high threshold. Scene D is one of the challenging images, YOLO-NAS missed to predict the

small object due to the reflection in water, whereas YOLOv8-x detects at high threshold. In scene D, YOLO-NAS, YOLOv5, YOLOv7 and YOLOv8-s predicts the image incorrectly, whereas YOLOv8-x predicts the image correctly.

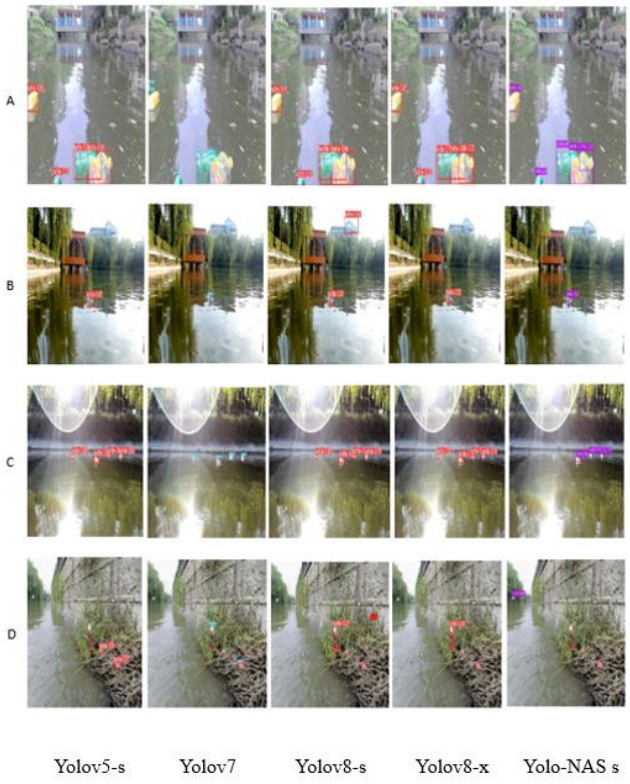


Figure 7. The detection results of A, B, C, D are the predicted images from each model in four complex scenes.

VI. Conclusion and Future Work

The thorough performance evaluation of several YOLO architectures, for the detection of floating bottles in inland water is concluded in this paper. The objective is to access the small objects. To achieve this, we used the FloW_Img dataset which consists of 2000 images with 5727 labels. Various metrics were used to evaluate the performance that includes Recall, Precision, mAP@0.50 and mAP@0.50:0.95. Every model underwent 200 training epochs, with YOLOv8-x with more effective balance across all metrics. Among all the models, YOLO-NAS variations had the highest recall scores, making them stand out for critical applications where high recall is required to minimize false negatives. Limitation in the model Yolov8-x is that the number of parameter and the computational unit required to train is high. Future work could be to reduce the amount of analysis of the challenges in the dataset like bottle submerge, reflection of water, half visibility of the bottle in the image into different classes and perform the analysis.

References

[1] An L. H., Li H., Wang F. F., et al. (2022) International governance process and countermeasures of marine plastic pollution[J]. Environmental Science Research, 35(06): 1334-1340

[2] Ren S., He K., Girshick R, et al. (2017) Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(6): 1137-1149.

[3] R. Girshick, Fast R-CNN[C]. (2015) IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448.

[4] Redmon J, Farhadi A. (2018) YOLOv3: An Incremental Improvement[J]. arXiv e-prints

[5] Liu W, Anguelov D, Erhan D, et al. (2015) SSD: Single Shot MultiBox Detector[P]

[6] Lin T Y, Goyal P, Girshick R, et al. (2017) Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, PP(99): 2999-3007.

[7] Z. Pu, X. Geng, D. Sun, H. Feng, J. Chen and J. Jiang, "Comparison and Simulation of Deep Learning Detection Algorithms for Floating Objects on the Water Surface," 2023 4th International Conference on Computer Engineering and Application (ICCEA), Hangzhou, China, 2023, pp. 814-820.

[8] Zhang, X.; Min, C.; Luo, J.; Li, Z. YOLOv5-FF: Detecting Floating Objects on the Surface of Fresh Water Environments. Appl. Sci. 2023, 13, 7367.

[9] Casas, Edmundo, et al. "Assessing the Effectiveness of YOLO Architectures for Smoke and Wildfire Detection." IEEE Access (2023).

[10] Renfei Chen, Jian Wu, Yong Peng, Zhongwen Li, Hua Shang, "Solving floating pollution with deep learning: A novel SSD for floating objects based on continual unsupervised domain adaptation", Engineering Applications of Artificial Intelligence,Volume120,2023,105857,ISSN0952-1976.

[11] Chen Renfei, Wu Jian, Peng Yong, Li Zhongwen, Shang Hua, "Detection and tracking of floating objects based on spatial-temporal information fusion", Expert Systems with Applications, Volume 225, 2023, 120185, ISSN 0957-4174.

[12] Y. Cheng et al., "FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 10933-10942

[13] Lin, F.; Hou, T.; Jin, Q.; You, A. Improved YOLO Based Detection Algorithm for Floating Debris in Waterway. Entropy 2021, 23, 1111.

[14] L. Zhang, Y. Wei, H. Wang, Y. Shao and J. Shen, "Real-Time Detection of River Surface Floating Object Based on Improved RefineDet," in IEEE Access, vol. 9, pp. 81147-81160, 2021.

[15] Sravanthi, R., Sarma, A.S.V. "Efficient image-based object detection for floating weed collection with low cost unmanned floating vehicles". Soft Compute 25, 13093–13101 (2021).

[16] Pedro F Proença and Pedro Simões, " TACO: Trash Annotations in Context for Litter Detection,"2020.

[17] X. Jin, P. Niu and L. Liu, "A GMM-Based Segmentation Method for the Detection of Water Surface Floats," in IEEE Access, vol. 7, pp. 119018-119025, 2019.

[18] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[19] Lohit, G. A. Vishnu and Nalini Sampath. "Multiple Object Detection Mechanism Using YOLO." IEEE International Conference on Data Engineering (2020).

[20] Kailash, A. Siva, et al. "Deep Learning based Detection of Mobility Aids using YOLOv5." 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP). IEEE, 2023.

[21] S. M. Tarachandy and A. J, "Enhanced Local Features using Ridgelet Filters for Traffic Sign Detection and Recognition," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1150-1156.

[22] S. S, D. S and S. N. Reddy, "Efficient Real-time Breed Classification using YOLOv7 Object Detection Algorithm," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6.

[23] Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr. 2023, 5, 1680-1716.

[24] O. E. Olorunshola, M. E. Irhebhude, and A. E. Evwiekpaefe, "A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms", JCSI, vol. 2, no. 1, pp. 1-12, Feb. 2023.