

# Automatic detection and counting of fisheries using fish images

Marc Momar Tall <sup>a,1,\*</sup>, Ibrahima Ngom <sup>a,2</sup>, Ousmane Sadio <sup>a,3</sup>, Adama Coulibaly <sup>a,4</sup>, Ibrahima Diagne <sup>a,5</sup>, Moustapha Ndiaye <sup>a,6</sup>

<sup>a</sup> Université Cheikh Anta Diop de Dakar / Ecole Supérieure Polytechnique, Senegal

<sup>1</sup> [marcmomar.tall@esp.sn](mailto:marcmomar.tall@esp.sn); <sup>2</sup> [ibrahima.ngom@esp.sn](mailto:ibrahima.ngom@esp.sn); <sup>3</sup> [ousmane.sadio@ucad.edu.sn](mailto:ousmane.sadio@ucad.edu.sn); <sup>4</sup> [adama.coulibaly@esp.sn](mailto:adama.coulibaly@esp.sn); <sup>5</sup> [ibrahima.diagne@esp.sn](mailto:ibrahima.diagne@esp.sn);

<sup>6</sup> [ndiayemoustapha766@gmail.com](mailto:ndiayemoustapha766@gmail.com)

\* corresponding author

## ARTICLE INFO

### Article history

Received October 6, 2023

Revised October 20, 2023

Accepted November 4, 2023

### Keywords

Fish

Fishbase

YOLO v8

Bounding Box

Detection

## ABSTRACT

In Senegal, stock recovery and fish classification are based on manual data collection, and the fish caught by the fishery are not often declared. What's more, data collection suffers from a lack of tools for monitoring and counting fish caught at fishing docks. Researchers have carried out studies on the fishery in Senegal, but data collection is almost non-existent. Moreover, there is no local fisheries database or automatic detection and counting algorithm. In this paper, a semantic segmentation algorithm is proposed using intelligent systems for the collection of fishery catches, for the formation of the local database. The data are collected by taking images of fish at the Soumbédioune fishing wharf in Senegal, and are completed with the Fishbase database. They were applied to the algorithm and resulted in a segmented dataset with masks. This constitutes our local database. The database is used with YOLO v8. The latter is very important for detecting images with bounding boxes in order to train the model. The results obtained are very promising for the proposed automatic poison detection and counting model. For example, the recall-confidence scores translate into bounding box performance with scores ranging from 0.01 to 0.75, confirming the performance of this model with bounding boxes.

This is an open access article under the [CC-BY-SA](#) license.



## 1. Introduction

Over the last forty years, Senegal's coastal marine ecosystem has undergone profound changes due to natural and, above all, anthropogenic impacts (human activities) [1]. Good fisheries management is of paramount importance in Africa, and particularly in Senegal, given the problems associated with climate change, strong fishing pressure on fishery resources, damage to marine habitats caused by poor fishing practices and conservation of the resource's components. This has led to serious imbalances in coastal ecosystems, resulting in a shortage of species fished, coastal pollution and excessive salinisation [2]. In fact, Senegal had a catch of 100,000 tonnes in 1965 [3], making it the leading producer in black Africa. Fish plays an important role in the country's diet and is the main activity of the people living in coastal areas. Domestic trade absorbs more than 90% of Senegal's total production. Since 1960, the national market has sold 80,000 tonnes of marine fish and 20,000 tonnes of freshwater fish every year. This puts Senegal in fourth place worldwide for per capita consumption [3], with an average annual consumption of 25 kg per person. What's more, this situation has an impact on the high consumption of fish dried using traditional preservation techniques [4]. This document is an extension of the work presented at the conference [5]. Given these realities, intelligent fisheries management solutions need to be proposed. Artificial intelligence models provide diverse solutions in many fields. To prevent overfishing and plan stock management policies, computer vision models are used. In order to achieve the detection goal, a suitable database is needed to automatically segment and classify captured fish. Image segmentation is one of the most important preprocessing processes to achieve recognition goals. Image recognition has been the subject of extensive research in the

literature, and many algorithms have been developed for this purpose. In [6], the algorithm used allows the separation of objects and background, the authors [7] use morphological evolution and K-means clustering to segment fish images, and several authors in turn use K-means clustering algorithm to segment fish image. Use pictures of fish. In [1], [8], species features are used for segmentation, and [9] takes a different approach from other methods by improving image preprocessing by using an adaptive histogram equalizer. All these segmentation algorithms can provide solutions according to the author's original goal, but they are not suitable for our background because photos are usually taken in environments with uneven backgrounds. This paper uses semantic segmentation algorithms to segment fish images. Most of the time, fish are caught in multi-color environments and believe in water, which explains the need for segmentation of the image and the output must be similar to the input image for predictions with the best performance. The species caught are neither reported nor collected at the docks to replenish fish stocks. To overcome the challenges of AI-powered smart fish management, segmented data is needed. Collect data by taking images of captured fish and use other resources such as FISHBASE to complete the base. All collected data then forms a library or database used to train the model. Process the data for better optimization. By using this data, it is possible to train the database, train the masks and obtain well-segmented images. In this paper, this segmented dataset is used for machine learning model training with YOLO V8. In the literature, several image processing and machine learning methods have been used to analyze visual data [10]–[12]. Detection techniques based on deep learning, as an essential branch of deep learning, have the advantage of simultaneously identifying the category of the object of interest and locating the position of the object using a bounding box. They have been widely used in various fields [13], such as facial recognition [14], [15], text detection [16], [17], pedestrian recognition [18], [19] and vehicle detection [20], [21]. The superiority of object detection techniques based on deep learning has also earned them particular attention for fish detection, particularly in the field of fish counting [22], [23], which is one of the objectives of this paper. It should be noted that many studies involve the application of deep learning in aquatic environments [24], [25], using deep learning techniques in marine environments, including convolutional neural networks (CNN) and recurrent neural networks (RNN) and generative adversarial networks (GAN) have been intensively studied. However, considering the development of all these models in our Senegalese research area, there has been no work exploring or investigating machine learning-based image recognition and fish counting techniques. This article uses the YOLO v8 detection algorithm because it is simple and fast. In the literature, some detection studies have used YOLO v4,...,v7 algorithms. Many researchers have improved the original YOLO v5 model to meet their target detection needs. Li et al. developed YOLO-FIRi, an improved YOLO v5 infrared image object detector. They added the CSP(cross-stage-partial connections) module to the feature extraction network to optimize planar features. Additionally, they used an improved attention module during rest blocks to focus more on the target and background. They also used multi-scale detection to improve the model's detection of small objects. The proposed model is evaluated on KAIST and FLIR datasets. The proposed model outperforms YOLO v4 and YOLO-FIR [26] in terms of accuracy, speed, parameter reduction, and computational cost optimization. YOLO v5-Aircraft is an improved aircraft target detection model based on YOLO v5, developed by Luo et al. proposed [27]. They combine centering and scale calibration to improve the feature extraction capabilities of the model. They replaced the residual blocks in the feature extraction backbone with CSandGlass blocks to reduce information loss. They show that the proposed model significantly improves the speed and accuracy of aircraft detection in remote sensing images, even with complex backgrounds. [28] proposed TPH-YOLO v5 for object detection in UAV detection scenarios. They added another prediction head to detect objects of different sizes. In addition, the original prediction head is replaced by the transformer prediction head. They used the CBAM attention mechanism to handle dense object scenarios. According to the experimental results, the detection performance of TPH-YOLO v5 improved by about 7% compared to the original performance of YOLO v5 [29] to improve YOLO v5 for high-quality defect detection. The authors refined the loss function and the DCGAN structure to generate high-quality defects. [30] proposes an improved original YOLO v5 method for detecting forbidden objects in baggage security. [31] improved the YOLO v5 model to detect regular and irregular cracks in tunnels. They annotated tunnel cracks using a new labeling method. [32] developed an improved YOLO v5 model for the detection of electric bicycles for optimal charging and parking management. Using YOLO v5 as a starting point, they proposed a lighter network backbone based on GhostNet. [33] integrates the GhostNet module into the YOLO v5 neck for a lightweight tea bud detection model. In addition to the performance improvements of YOLO v5, but that's not all, YOLO v8 also comes with a new backplane, a new anchor-free sensor head, and a new loss function, making it useful for a

variety of object detection and image segmentation tasks. Attractive choices. Additionally, it's efficient enough to run on everything from toasters to supercomputers. I hope it will include a pose estimation model like the one in YOLO v7. In short, YOLO v8 is a powerful and flexible tool that gives you the best of both worlds: the latest SOTA technology and the ability to use and compare all previous versions of YOLO. It is currently difficult to find scientific papers supporting this model. This article uses YOLO v8 for simulation, taking into account a comprehensive literature review and attention to the YOLO v8 algorithm. The rest of this paper is organized as follows: Section II presents the collection and preprocessing of images. Section III describes the formation of the database. Section IV describes the training environment. Section V describes the results and discussion. Section VI describes the balance sheet.

## 2. Method

### 2.1. Image collection and Pre-processing

- Data collection and pre-processing is one of the most important processes in forming a dataset. Fig. 1 shows the different sources used and all the processes involved in forming the dataset.

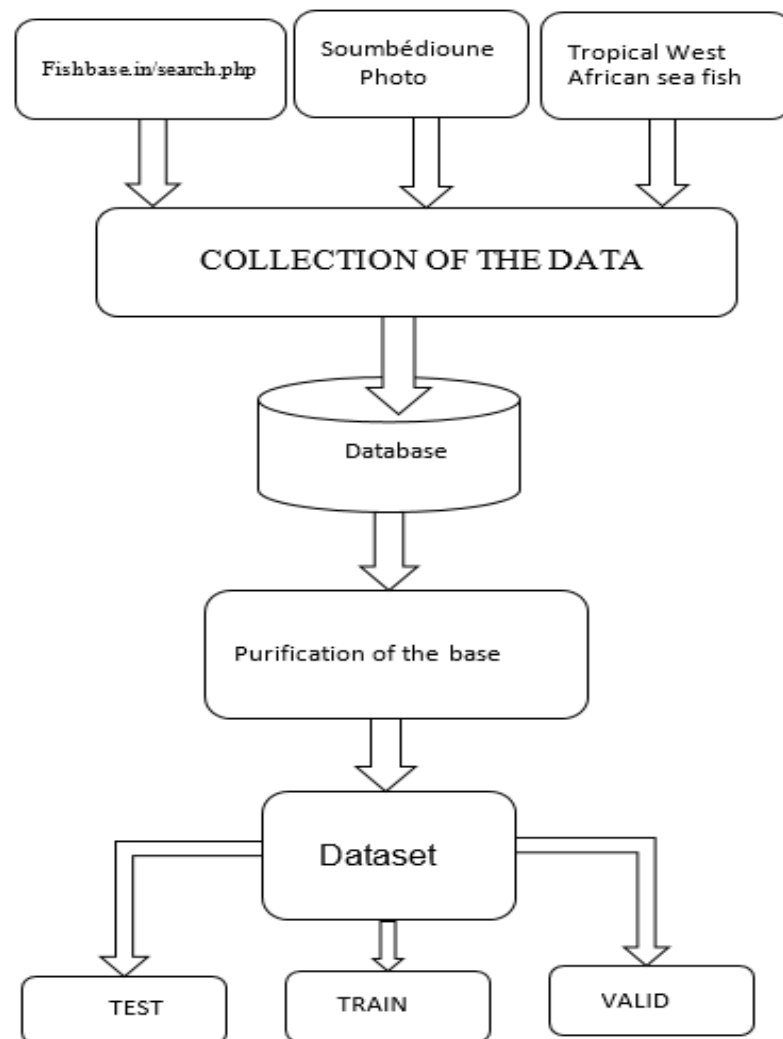


Fig. 1. Formation of datasets

Fig. 2 shows the characteristics of 825 fish species in the database. Each fish is characterised by its family name, species name, size (max and min) in millimetres and a photo of the species. The entire database is redundant. It should also be noted that the minimum size of fish is approximate and we have based it on the diameter of the fishing nets used by small-scale fishermen.




FAMILY	SPECIES	MINIMUM LENGTH	MAXIMUM LENGTH	PHOTO
...	...	...	...	...
Sparidés	<i>Pagrus caeruleostictus</i>	5.00	90.00	
Sparidés	<i>Pagrus pagrus</i>	7.00	91.00	
...	...	...	...	...
Mugilidés	<i>Parachelon grandisquamis</i>	6.00	40.00	

Fig. 2. Data Base

- For pre-processing, the data is resized and rotated as shown in Table 1 by reducing the pixel size of the original images. The processed data enables fast, efficient segmentation and, above all, the information contained in the image remains constant.

Table.1 Sample Table

Auto Orient: applied	
Static Crop: 25-75% Horizontal Region, 25-75% Vertical Region	
Preprocessing	Resize: Stretch to 416x416
	Modify classes: 0 remapped, 0 dropped
	Grayscale: Applied
	Flip: Horizontal
	Corp: 0% Minimum Zoom, 30% Maximum Zoom
Increase	Rotation: Between -5° and +5
	Brightness: Between -25% and +25%
	Exposure: Between -25% and +25%.
	Noise: Up to 2% of pixels

After collecting and processing the data, we divided the dataset into three groups: the training set, the test set and the roboflow validation set. Database divisions as show in Fig. 3.

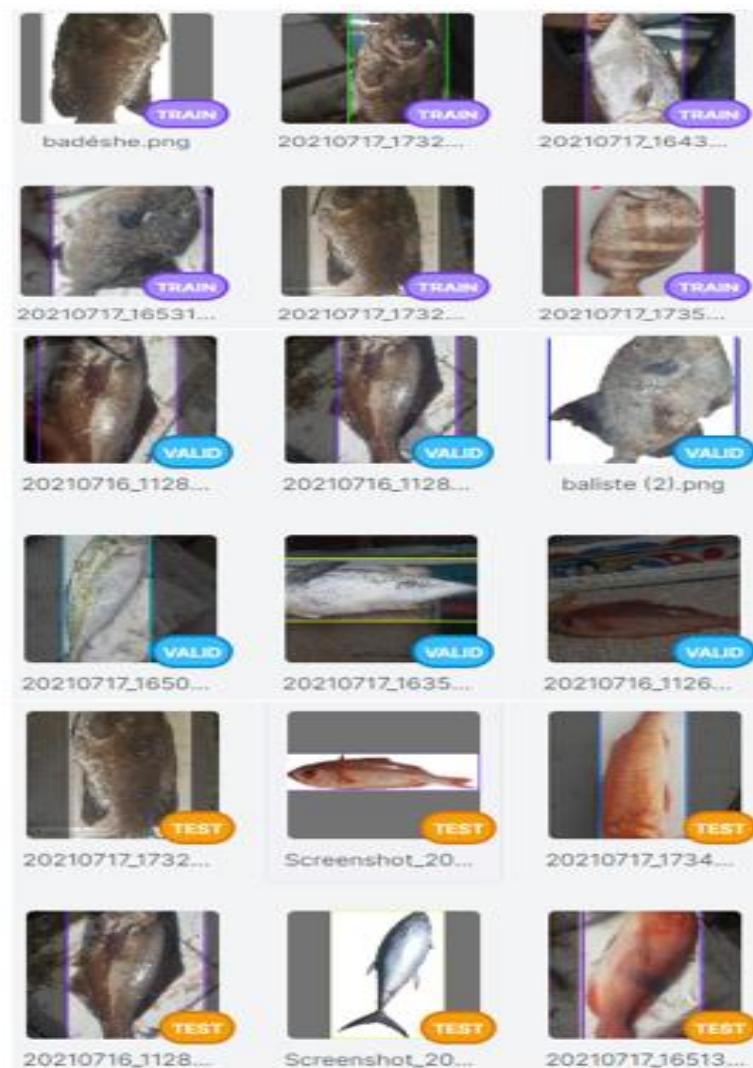


Fig. 3. Database divisions

2.2. Database Formation: Classification and Mask Formation

The segmentation algorithm is presented along with the various steps involved in running the algorithm. Table style as show in Table 2.

Table.2 Table Styles

Algorithm 1: Segmentation algorithm
Step 1: Reduce the input images to an appropriate size.
Step 2: Read the images in the subdirectory.
Step 3: Divide all images into 256x256x3 patches.
Step 4: Read each image as a BGR.
Step 5: Crop the image Image = image.resize ((SIZE_X, SIZE_Y)).
Step 6: Extract patches from each image 256 for 256 patches means no overlap.
Step 7: Carry out a pre-treatment using one of the backbones Single_patch_img = (single_patch_img.astype ('float32')) / 255.
Step 8: The same as above is repeated for masks.
Step 9: Copy images and masks to a new folder.
Step 10: Divide them into train, test and val.
Step 11: Using data generators to apply semantic segmentation semantic segmentation.

In this paper the results are obtained using python 3.7 in Jupiter. The output images are the masks of each input image.



### 2.2.1. Input Image Characteristics

The images in the database are in tiff format, and each image is characterised by its family name, species name, size (max, min), species length and photo. Input data as show in Fig. 4.

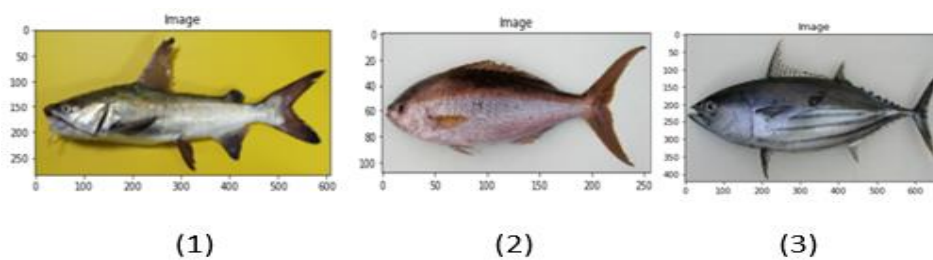


Fig. 4. Input Data

### 2.2.2. Interpretation of Masks

This part presents the formation of masks, the program developed in python 3.7 in Jupiter generates masks next to each photo of fish respectively. Using the masks in the database makes it easy for the algorithm to predict more accurately, aided by pattern recognition of the fish. Output Data Mask Images as show in Fig. 5.

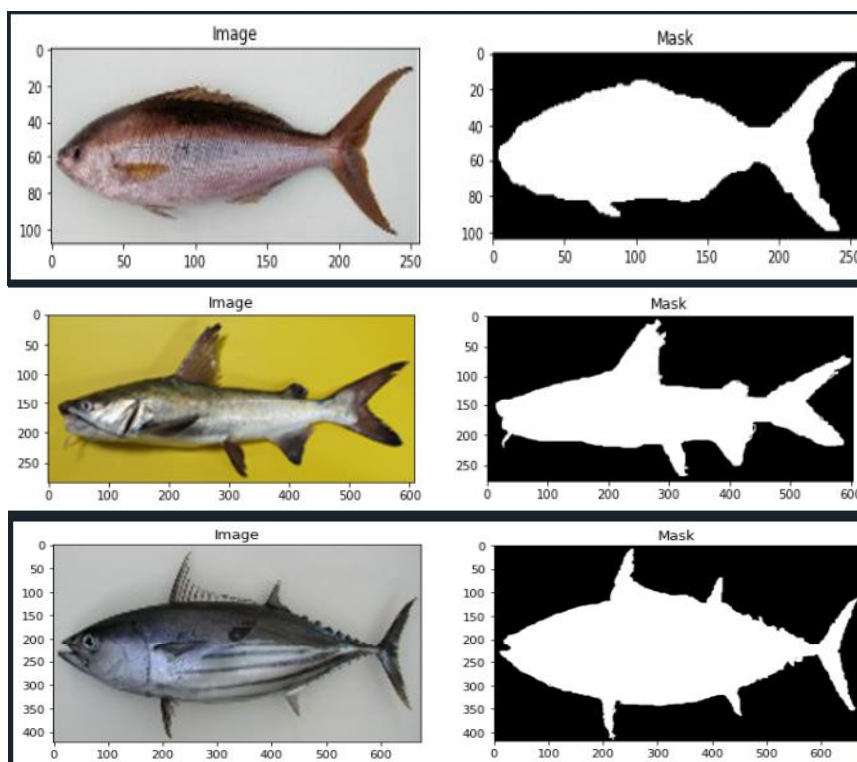


Fig. 5. Output Data Mask Images

## 3. Results and Discussion

This paper uses the colab GPU and the latest detector based on the YOLO architecture, YOLO v8 [21]. This is a one-step anchorless object detector. The main reasons for choosing this model are its flexibility, high accuracy, ease of training and deployment, and speed of detection. The new version adds new features and enhancements to improve flexibility and performance. Key innovations include a new backbone network, a tetherless detection head and a new loss function that can be run on CPUs or GPUs. Depending on the scaling factor, YOLO v8 comes in different size models to meet the needs of different scenarios. Unlike YOLO v5, the C2f structure with a richer gradient flow replaces the C3 structure of YOLO v5 in the backbone and neck area, and different channel numbers are modified for the various scaling models.

Image detection algorithms typically use non-maximal suppression during prediction to discard redundant detection bounding boxes, as shown in Fig. 6. The principle is to retain the highest scoring predicted bounding box after traversing all detection bounding boxes.

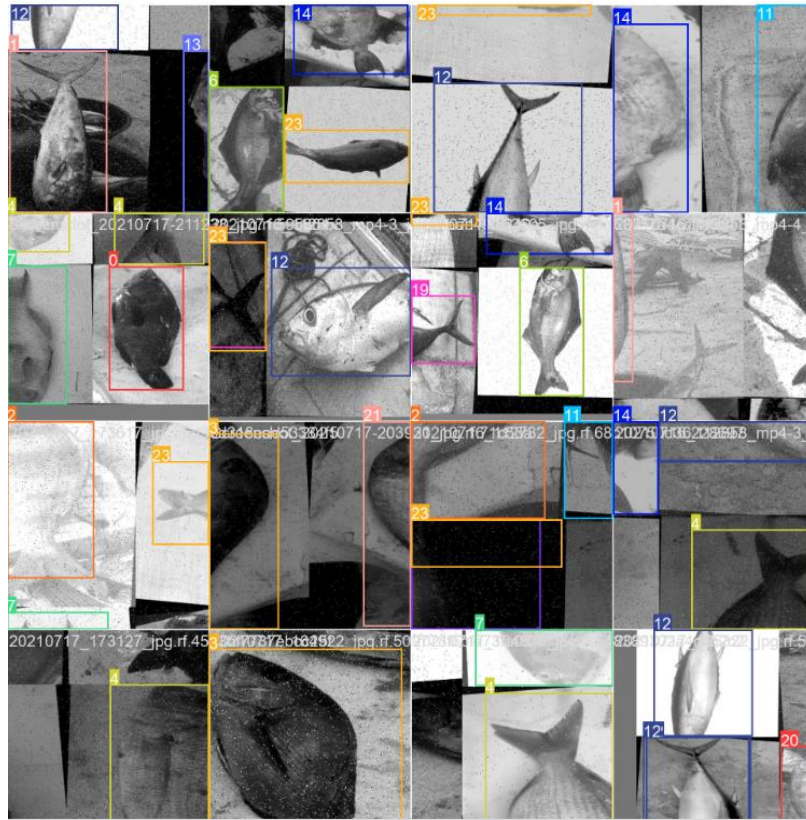


Fig. 6. Bounding Boxes

These datasets are used to train detection and counting models. Use the YOLO v8 algorithm for training, and the data used is imported from the Roboflow library. Confidence curve and Precision confidence curve as show in Fig. 7 and Fig. 8.

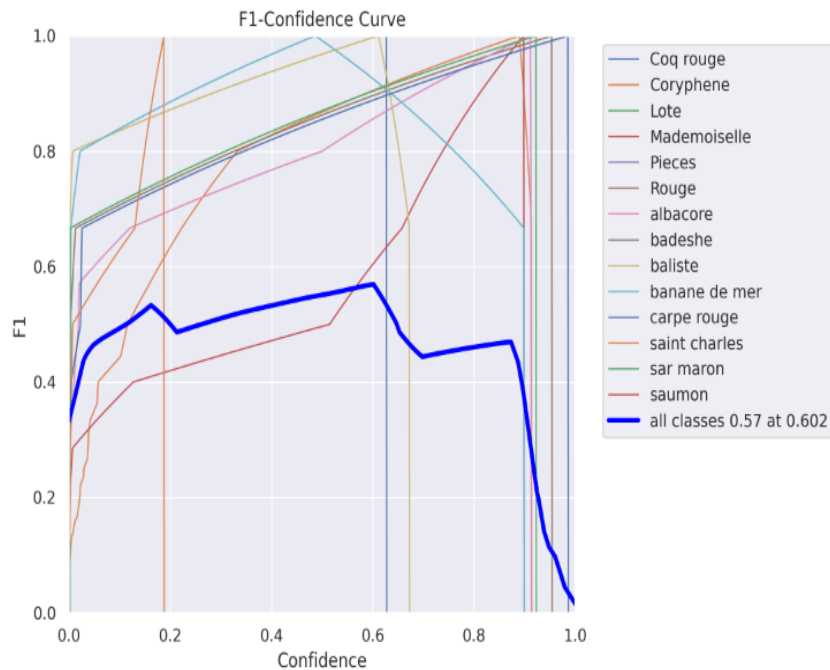
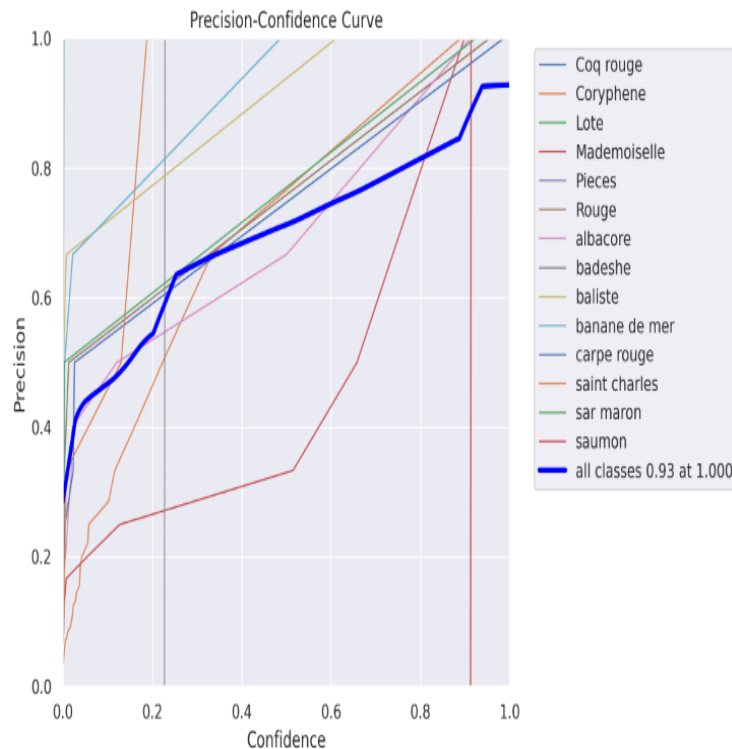
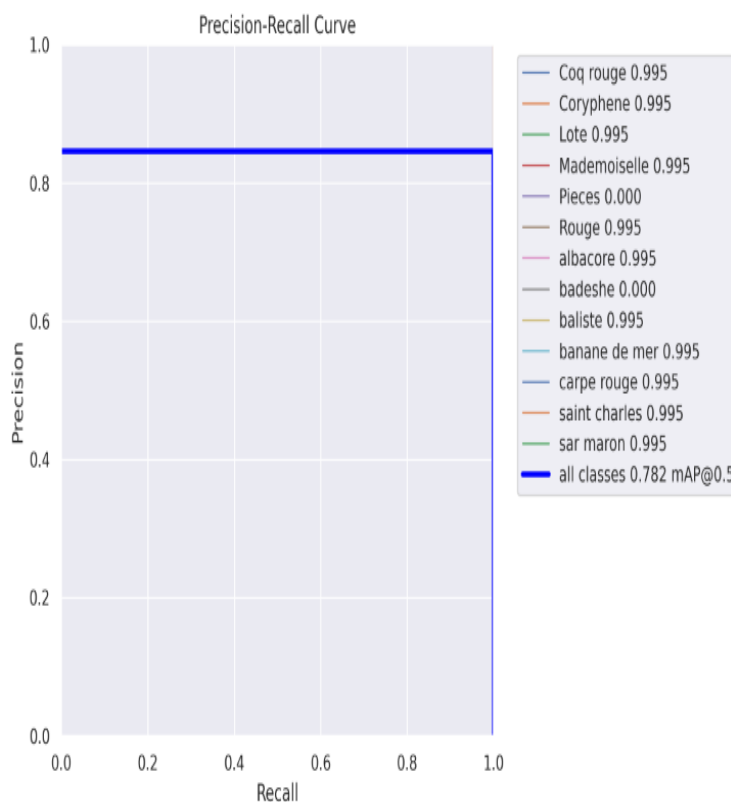


Fig. 7. Confidence Curve



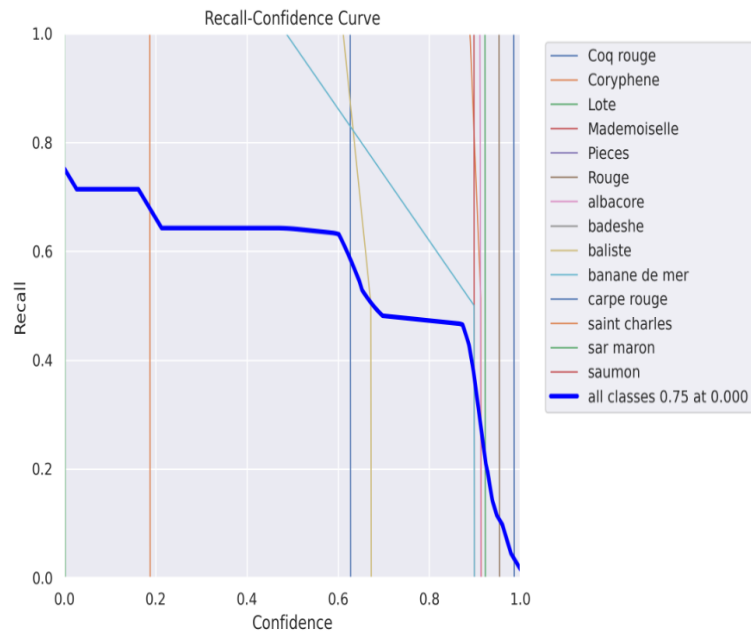
**Fig. 8.** Precision-Confidence Curve

Fig. 9 and Fig. 10 shows the precision-recall curve of the YOLO v8 model. The precision-recall curve shows the trade-off between precision and recall at different thresholds. The larger the area under the curve, the higher the recall and precision rates. For example, the model achieved an average mAP@0.5 value of 0.782 across all categories, justifying the choice of YOLO v8.



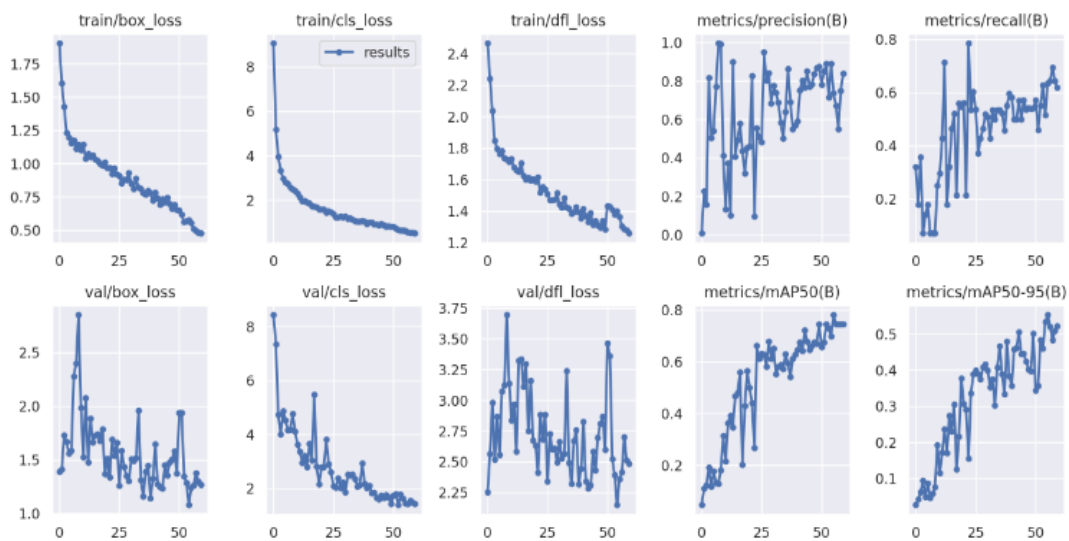
**Fig. 9.** Precision-Recall Curve





**Fig. 10.** Precision-Recall Curve

Fig. 11 shows the F1 curve of the YOLO v8 model. The horizontal axis represents the confidence, and the vertical axis represents the F1 score, which is located in the interval  $[0, 1]$ . Fig. 10 shows the trade-off between precision and recall. The results assume that precision and recall are equally important. The overall F1 score is 0.57 and the confidence level is 0.93. At this confidence level, the best precision and recall values are achieved.



**Fig. 11.** YOLO v8 Evaluation Measure

The development curve is shown in Fig. 11. Object detection, precision, recall, and mAP curves are commonly used to evaluate model performance. The accuracy curve illustrates the percentage of bounding box predictions that are accurate. The recall curve, on the other hand, shows the percentage of ground-truth bounding boxes that were accurately predicted. It is important to note that the higher the precision and recall values, the better the classification performance of the model.

The curves shown in Fig. 11 indicate an improvement in the stability of the loss curve as the number of epochs increased from 20 to 60. However, the reduction in the value of the loss function and the increase in mAP were slower thereafter. The object detection loss function including classification loss, location loss and confidence loss was used to evaluate the model. The bounding

box category evaluates the ability of a model to accurately locate the center point of an object and determine whether the bounding box accurately matches the detection targets. A lower value of this loss indicates a more accurately predicted bounding box. The confidence loss is used to determine whether an object exists in the predicted box. The lower the value of the confidence loss, the greater the accuracy. The classification loss indicates the model's ability to correctly predict a given object class. The lower the value of the loss, the better the classification.

Fig. 10 shows the recall-confidence scores, which translates into bounding box performance with scores of 0.01 - 0.75 and confirms the model detection performance with bounding boxes. Fig. 12 shows the predicted bounding boxes compared with the boxes produced using the YOLO v8 model.

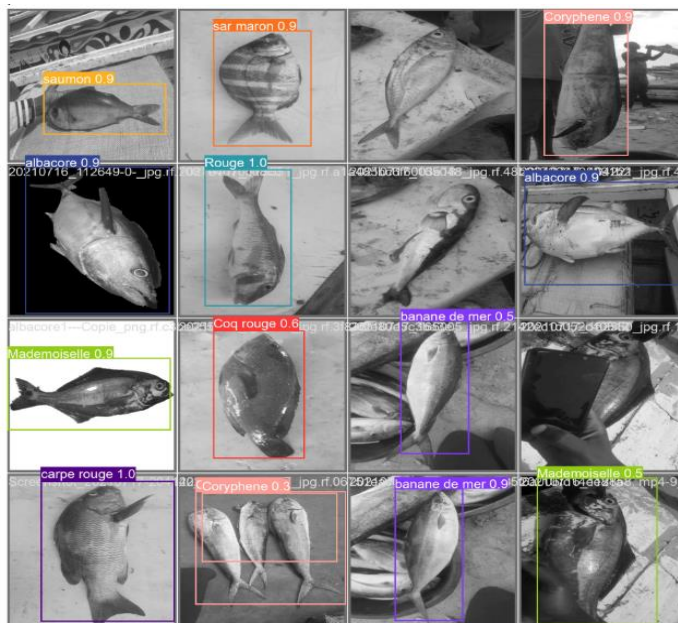


Fig. 12. Detecting Poisons

Fig. 13 shows the confusion matrix of the YOLO v8 model for the classification of poisons. The actual classes are on the horizontal axis, while the predicted classes are on the vertical axis. The diagonal line in the middle of the confusion matrix indicates the significance of the predicted results. The horizontal line represents false negatives and the vertical line false positives. The YOLO v8 models can correctly classify most large defect classes

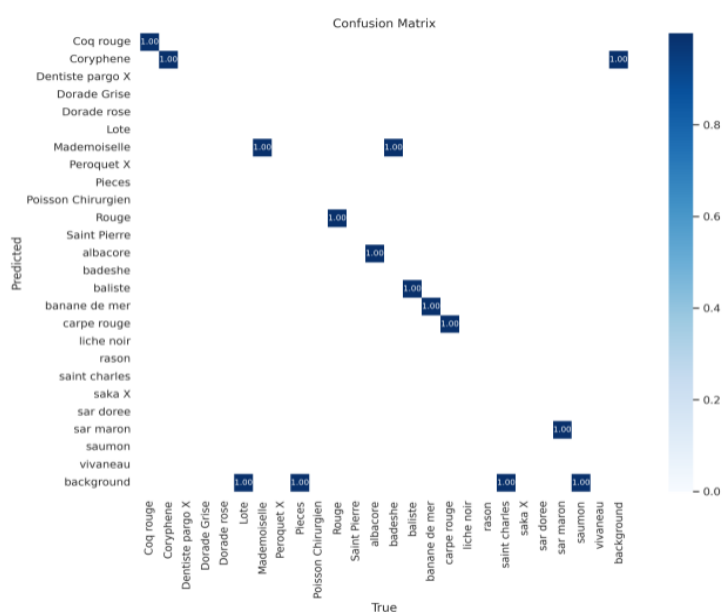


Fig. 13. Confusion matrix

#### 4. Conclusion

The use of the YOLO v8 algorithm with bounding boxes to provide a solution for stock recovery and the classification of fish caught on the Senegalese coast gives the desired performance to the proposed image detection model. The results obtained promise a transition from manual to automated management for stock recovery and classification of fish caught on the Senegalese coast. The data collected by taking photos at the Soumbédioune quay and at Fishbase are processed and analysed to form a redundant, well-segmented database, using semantic segmentation. Use of the database built using this model has produced promising results for improving data collection techniques and fisheries management in Senegal.

#### References

- [1] J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, and X. Si, "Fish species classification by color, texture and multi-class support vector machine using computer vision," *Comput. Electron. Agric.*, vol. 88, pp. 133–140, Oct. 2012, doi: [10.1016/j.compag.2012.07.008](https://doi.org/10.1016/j.compag.2012.07.008).
- [2] A. Newton *et al.*, "Anthropogenic, Direct Pressures on Coastal Wetlands," *Front. Ecol. Evol.*, vol. 8, p. 512636, Jul. 2020, doi: [10.3389/fevo.2020.00144](https://doi.org/10.3389/fevo.2020.00144).
- [3] R. Priyadharsini and T. S. Sharmila, "Object Detection In Underwater Acoustic Images Using Edge Based Segmentation Method," *Procedia Comput. Sci.*, vol. 165, pp. 759–765, Jan. 2019, doi: [10.1016/j.procs.2020.01.015](https://doi.org/10.1016/j.procs.2020.01.015).
- [4] S. K. Amit, M. M. Uddin, R. Rahman, S. M. R. Islam, and M. S. Khan, "A review on mechanisms and commercial aspects of food preservation and processing," *Agric. Food Secur.*, vol. 6, no. 1, p. 51, Dec. 2017, doi: [10.1186/s40066-017-0130-8](https://doi.org/10.1186/s40066-017-0130-8).
- [5] M. M. Tall, I. Ngom, O. Sadio, and I. Diadne, "Proposal for a Local Database: Segmentation and Classification Algorithm," in *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, Apr. 2023, pp. 846–849, doi: [10.1109/CSNT57126.2023.10134645](https://doi.org/10.1109/CSNT57126.2023.10134645).
- [6] B. Meissa and D. Gascuel, "Overfishing of marine resources: some lessons from the assessment of demersal stocks off Mauritania," *ICES J. Mar. Sci.*, vol. 72, no. 2, pp. 414–427, Jan. 2015, doi: [10.1093/icesjms/fsu144](https://doi.org/10.1093/icesjms/fsu144).
- [7] R. Bonnardel, "Les problèmes de la pêche maritime au Sénégal," *Ann. Geogr.*, vol. 78, no. 425, pp. 25–56, 1969, doi: [10.3406/geo.1969.14498](https://doi.org/10.3406/geo.1969.14498).
- [8] E. H. Balla Dieye, A. Tahirou Diaw, T. Sané, and N. Ndour, "Dynamique de la mangrove de l'estuaire du Saloum (Sénégal) entre 1972 et 2010," *Cybergeogeo*, vol. 2013, p. 27, Jan. 2013, doi: [10.4000/cybergeogeo.25671](https://doi.org/10.4000/cybergeogeo.25671).
- [9] H. Qin, W. Zhou, Y. Yao, and W. Wang, "Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data," *Remote Sens. Environ.*, vol. 280, p. 113143, Oct. 2022, doi: [10.1016/j.rse.2022.113143](https://doi.org/10.1016/j.rse.2022.113143).
- [10] Y. Kutlu, B. Iscimen, and C. Turan, "Multi-stage fish classification system using morphometry," *Fresenius Environ. Bull.*, vol. 26, no. 3, pp. 1911–1917, 2017, [Online]. Available at: <https://www.researchgate.net/profile/Yakup-Kutlu/publication/314284234>.
- [11] Y. Zhao, Z.-Y. Sun, H. Du, C.-W. Bi, J. Meng, and Y. Cheng, "A novel centerline extraction method for overlapping fish body length measurement in aquaculture images," *Aquac. Eng.*, vol. 99, p. 102302, Nov. 2022, doi: [10.1016/j.aquaeng.2022.102302](https://doi.org/10.1016/j.aquaeng.2022.102302).
- [12] H. Mohammadi Lalabadi, M. Sadeghi, and S. A. Mireei, "Fish freshness categorization from eyes and gills color features using multi-class artificial neural network and support vector machines," *Aquac. Eng.*, vol. 90, p. 102076, Aug. 2020, doi: [10.1016/j.aquaeng.2020.102076](https://doi.org/10.1016/j.aquaeng.2020.102076).
- [13] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: [10.1109/TNNLS.2018.2876865](https://doi.org/10.1109/TNNLS.2018.2876865).

- 
- [14] W. Liu, I. Hasan, and S. Liao, "Center and Scale Prediction: Anchor-free Approach for Pedestrian and Face Detection," *Pattern Recognit.*, vol. 135, p. 109071, Mar. 2023, doi: [10.1016/j.patcog.2022.109071](https://doi.org/10.1016/j.patcog.2022.109071).
- [15] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019, doi: [10.1109/TPAMI.2017.2781233](https://doi.org/10.1109/TPAMI.2017.2781233).
- [16] Y. Xu *et al.*, "Gliding Vertex on the Horizontal Bounding Box for Multi-Oriented Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021, doi: [10.1109/TPAMI.2020.2974745](https://doi.org/10.1109/TPAMI.2020.2974745).
- [17] J. Ma *et al.*, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals," *IEEE Trans. Multimed.*, vol. 20, no. 11, pp. 3111–3122, Nov. 2018, doi: [10.1109/TMM.2018.2818020](https://doi.org/10.1109/TMM.2018.2818020).
- [18] M. M. Islam, A. A. R. Newaz, and A. Karimoddini, "Pedestrian Detection for Autonomous Cars: Inference Fusion of Deep Neural Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 23358–23368, Dec. 2022, doi: [10.1109/TITS.2022.3210186](https://doi.org/10.1109/TITS.2022.3210186).
- [19] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware Fast R-CNN for Pedestrian Detection," *IEEE Trans. Multimed.*, vol. 20, no. 4, pp. 1–1, Apr. 2017, doi: [10.1109/TMM.2017.2759508](https://doi.org/10.1109/TMM.2017.2759508).
- [20] G. Li, Z. Ji, and X. Qu, "Stepwise Domain Adaptation (SDA) for Object Detection in Autonomous Vehicles Using an Adaptive CenterNet," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 17729–17743, Oct. 2022, doi: [10.1109/TITS.2022.3164407](https://doi.org/10.1109/TITS.2022.3164407).
- [21] H. Wang, Y. Yu, Y. Cai, X. Chen, L. Chen, and Q. Liu, "A Comparative Study of State-of-the-Art Deep Learning Algorithms for Vehicle Detection," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 2, pp. 82–95, 2019, doi: [10.1109/MITS.2019.2903518](https://doi.org/10.1109/MITS.2019.2903518).
- [22] A. Ben Tamou, A. Benzinou, and K. Nasreddine, "Multi-stream fish detection in unconstrained underwater videos by the fusion of two convolutional neural network detectors," *Appl. Intell.*, vol. 51, no. 8, pp. 5809–5821, Aug. 2021, doi: [10.1007/s10489-020-02155-8](https://doi.org/10.1007/s10489-020-02155-8).
- [23] T. Liu, P. Li, H. Liu, X. Deng, H. Liu, and F. Zhai, "Multi-class fish stock statistics technology based on object classification and tracking algorithm," *Ecol. Inform.*, vol. 63, p. 101240, Jul. 2021, doi: [10.1016/j.ecoinf.2021.101240](https://doi.org/10.1016/j.ecoinf.2021.101240).
- [24] X. Yang, S. Zhang, J. Liu, Q. Gao, S. Dong, and C. Zhou, "Deep learning for smart fish farming: applications, opportunities and challenges," *Rev. Aquac.*, vol. 13, no. 1, pp. 66–90, Jan. 2021, doi: [10.1111/raq.12464](https://doi.org/10.1111/raq.12464).
- [25] D. Li and L. Du, "Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish," *Artif. Intell. Rev.*, vol. 55, no. 5, pp. 4077–4116, Jun. 2022, doi: [10.1007/s10462-021-10102-3](https://doi.org/10.1007/s10462-021-10102-3).
- [26] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, "YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection," *IEEE Access*, vol. 9, pp. 141861–141875, 2021, doi: [10.1109/ACCESS.2021.3120870](https://doi.org/10.1109/ACCESS.2021.3120870).
- [27] S. Luo, J. Yu, Y. Xi, and X. Liao, "Aircraft Target Detection in Remote Sensing Images Based on Improved YOLOv5," *IEEE Access*, vol. 10, pp. 5184–5192, 2022, doi: [10.1109/ACCESS.2022.3140876](https://doi.org/10.1109/ACCESS.2022.3140876).
- [28] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, vol. 2021-Octob, pp. 2778–2788, doi: [10.1109/ICCVW54120.2021.00312](https://doi.org/10.1109/ICCVW54120.2021.00312).
- [29] Y. Jin, H. Gao, X. Fan, H. Khan, and Y. Chen, "Defect Identification of Adhesive Structure Based on DCGAN and YOLOv5," *IEEE Access*, vol. 10, pp. 79913–79924, 2022, doi: [10.1109/ACCESS.2022.3193775](https://doi.org/10.1109/ACCESS.2022.3193775).
- [30] Z. Wang, H. Zhang, Z. Lin, X. Tan, and B. Zhou, "Prohibited Items Detection in Baggage Security Based on Improved YOLOv5," in *2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI)*, Jun. 2022, pp. 20–25, doi: [10.1109/SEAI55746.2022.9832407](https://doi.org/10.1109/SEAI55746.2022.9832407).
-

- 
- [31] Y. Li, R. Cheng, C. Zhang, M. Chen, J. Ma, and X. Shi, "Sign language letters recognition model based on improved YOLOv5," in *2022 9th International Conference on Digital Home (ICDH)*, Oct. 2022, pp. 188–193, doi: [10.1109/ICDH57206.2022.00036](https://doi.org/10.1109/ICDH57206.2022.00036).
  - [32] C. Zhang, A. Xiong, X. Luo, C. Zhou, and J. Liang, "Electric Bicycle Detection Based on Improved YOLOv5," in *2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*, Apr. 2022, pp. 1–5, doi: [10.1109/CTISC54888.2022.9849750](https://doi.org/10.1109/CTISC54888.2022.9849750).
  - [33] M. Cao, H. Fu, J. Zhu, and C. Cai, "Lightweight tea bud recognition network integrating GhostNet and YOLOv5," *Math. Biosci. Eng.*, vol. 19, no. 12, pp. 12897–12914, 2022, doi: [10.3934/mbe.2022602](https://doi.org/10.3934/mbe.2022602).