[1]Omkar Mahajan
[2]Denzil Puvathingal
[3]Divyang Patel
[4]Sanket Jaiswal
[5]Dashrath Mane
[6]Prashant Kanade

YOLOv8 based fish detection and classification on fishnet dataset

Journal of Electrical Systems

Abstract: - The research examines the marine fishing sector and highlights the urgent need for an electronic monitoring system designed to meet the unique needs of fishermen. The initiative's motive is emphasized in the paper, which highlights how cutting-edge technology like object detection and tracking could revolutionize the fishing industry when integrated into an electronic monitoring framework. The research proposes an electronic monitoring system based on YOLOv8 (You Only Look Once) as a comprehensive solution to address current issues, such as out-of-date data collection methods and a lack of guiding applications. The review of the literature, which highlights gaps in the current fishing applications, is an important component of the paper. The research is geared towards training an object detection model on the fishnet dataset. The focus is on data processing created by an electronic monitoring system to rectify the current state of the fishing industry's deficiencies.

Keywords: Image Processing, YOLO, Object Detection, Electronic Monitoring, Computer Vision

I. INTRODUCTION

In the expansive realm of maritime activities, modern fishing vessels are equipped with state-of-the-art electronic monitoring systems, featuring video cameras meticulously positioned to document the bustling deck where fish are landed. Initially designed for security and safety, these camera systems hold untapped potential for identifying fish species and precisely tracking and counting the marine harvest. The urgency to automate these tasks drives our research, emphasizing the need for a real-time image processing system to ensure swift and accurate fish identification and counting.

Our research is fueled by a dual commitment—to advance technological frontiers and to foster an economically and environmentally sustainable ecosystem within the fishing industry. At its core, our primary goal is to lay the groundwork for the seamless implementation of automated counting and species identification of the fish catch. This ambition extends beyond individual fishing operations, aiming to elevate the sustainability and efficiency of marine fishing methods globally.

The fishnet dataset, a repository of annotated images extracted from video frames, serves as the backbone of our exploration. Leveraging the YOLOv8 model, a robust object detection algorithm, we delve into the intricacies of the dataset, paving the way for automation of catch reporting. While our current focus is on static images, the YOLOv8 model is poised for a dynamic transition to video feeds, contingent upon the availability of an appropriate dataset. Furthermore, our research contemplates the model and application of live video streams captured by electronic monitoring systems' cameras. The fishnet dataset, a repository of annotated images extracted from video frames, serves as the backbone of our exploration. Leveraging the YOLOv8 model, a robust object detection algorithm, we delve into the intricacies of the dataset, paving the way for automation of catch reporting. While our current focus is on static images, the YOLOv8 model is poised for a dynamic transition to video feeds, contingent upon the availability of an appropriate dataset.

In essence, our research signifies the convergence of technological prowess and environmental stewardship. By tapping into the potential of electronic monitoring systems, annotated datasets, and advanced object detection algorithms, we aspire to redefine the trajectory of marine fishing practices. The future beckons towards a realm where sustainability and efficiency harmoniously coexist, driven by the innovative solutions that emerge from our dedicated exploration of the seas.

[1] Dashrath Mane: Vivekanand Education Society's Institute Of Technology (VESIT)

II. LITERATURE REVIEW

Khokher et al discussed MLAI-based approaches for automating fishery catch and bycatch monitoring face challenges related to data quantity and operational complexity. When it comes to improving the efficiency and accuracy of data gathering in the demanding and ever-changing environment of fishing vessels, MLAI-based systems for automating fishery catch and bycatch monitoring offer an appealing answer. Obtaining annotated training datasets is a major difficulty, particularly in fisheries with changeable conditions and various catch compositions. The diversity and quality of the training data are essential for building strong models that can handle the intricacies of many species, climates, and lighting changes. In order to implement MLAI effectively, the study highlights the significance of prioritizing species according to their ecological, economic, or vulnerability status. This is because imbalanced data may result in insufficient training for specific species. To combat this bias and guarantee a more fair representation of species in training data, techniques like undersampling or creating synthetic images are recommended. Furthermore, the work highlights the significance of taxonomic categorization and suggests a multi-label, hierarchical method for MLAI models to concurrently classify captures at various taxonomic levels.[1]

Zhang et al. did a study focused on improving the YOLOv5 model for the detection and analysis of Engraulis japonicus fishing vessel operations. To improve feature extraction and detection accuracy, the Squeeze-and-Excitation Network (SENet), a fusion attention mechanism, is incorporated into the proposed YOLOv5 model. The model performs better than the baseline YOLOv5 and other modified versions when measured by mean average precision (mAP), precision, recall, and loss functions. To track and count the number of fishing baskets, fishing nets, and processing vessels, the research incorporates a target identification approach that uses Kalman filtering and the Hungarian matching algorithm. In comparison to human counting, the study's automated statistical analysis of fishing vessel activity is expected to be more precise and efficient. The results show that when compared to previous models, the enhanced YOLOv5_SE model obtains a higher mAP (99.4%) as well as better precision and recall. The suggested strategy also addresses issues with statistical accuracy, including the use of threshold techniques for processing vessels and fishing nets. The report does, however, point out areas that still need work, such as the size and complexity of the model, and it makes recommendations for possible future improvements including automatic labeling and the application of sophisticated feature extraction techniques.[2]

Kay & Merrifield have provided publicly accessible data regarding camera-based electronic monitoring (EM) systems on commercial fishing vessels that was addressed by the Fishnet Open Images Database. It is the largest and most varied dataset for fish detection and classification in the context of fisheries EM, with over 86,000 photos containing 34 object classifications. The dataset comes from the increasing number of ships that have EM systems installed to apply computer vision to automate the assessment process in light of the anticipated increase in data volume. Fishnet provides a difficult benchmark for creating computer vision algorithms suited to the particular difficulties of EM images taken above water, in contrast to other datasets like ImageNet and COCO, which lack data specific to fishing.In order to obtain raw video and capture annotations, the dataset collection process entails agreements with authorities and EM service providers. Privacy is protected by means of facial blurring and the removal of identifiable vessel information. The dataset presents issues such as visual resemblance between species, skewed class distributions, and unfavorable weather circumstances. It focuses on longline tuna vessels in the western and central Pacific Ocean. It replicates real-world situations and is divided into training, validation, and test sets. This makes it an essential tool for developing computer vision applications in fisheries electronic monitoring.[3]

Qiao et al. presented a unique method for automating the detection of capture events in electronic monitoring (EM) video footage from fishing vessels using deep learning. The system uses a two-step process: it first uses frame-by-frame identification of fish and fishermen, and then it applies a temporal filter to identify catch events. Convolutional neural networks (CNNs) were the basis of the object detection framework that showed encouraging results; TensorBox, which used the ResNet 152 architecture, worked well. The catch event detection system demonstrated its potential for precise and quick analysis of massive amounts of EM data by achieving excellent recall and precision. Even with the accomplishment, there are still issues to be resolved, like the requirement for bigger and more varied training datasets to improve the model's ability to generalize across various vessels and circumstances. The suggested approach offers an alternative to the time-consuming and mistake-prone human video analysis process, addressing the growing significance of EM in fisheries management. Deep learning techniques have the potential to be widely used in fisheries monitoring as long as they continue to progress, leading to more sustainable and knowledgeable management strategies.[4]

Gilman et al. proposed the use of novel techniques such as a multi-scale fusion feature pyramid network, large-scale depthwise separable convolutions, and a cross-stage partial DWNeck backbone, the research presents CMS-YOLO, a road scene recognition method that improves real-time performance and detection accuracy. With noteworthy gains in mAP@0.5 on the BDD100K and Udacity Self-Driving datasets, CMS-YOLO shows amazing improvements over YOLOv5, and in autonomous driving scenarios, it reaches an impressive real-time detection speed of 34.5 frames per second. Turning its attention to fisheries management, the report emphasizes how EM systems—which include sensors, cameras, GPS, and data loggers—are being used more and more to address issues that human observers encounter. Situated as a cost-effective and expandable substitute, EM systems surpass traditional observer initiatives by utilizing real-time data transfer via satellite, Automatic Identification Systems (AIS), Video Monitoring Systems (VMS), and sophisticated sensors.[5]

Li et al. proposed the use of YOLOv3, YOLO-CAN which is a novel object recognition method that improves speed and accuracy, especially for small and occluded objects, by introducing innovations including an attention mechanism, CIoU loss function, Soft-NMS, and depthwise separable convolution. On the MS COCO and KAIST datasets, experimental results show that YOLO-ACN has an AP of 18.2%, a single-class mAP of over 80%, and a real-time mAP50 of 53.8%. The architecture has been improved to solve problems with small targets and occlusions. It does this by incorporating attention methods, depthwise separable convolution, and improved CIoU and Soft-NMS loss functions for accurate bounding box regression and better handling of occlusions.YOLO-ACN outperforms YOLOv3 with noticeable minor target identification improvements and is quicker and more accurate than SSD513. It also matches faster R-CNN precision. The important effect of the attention mechanism on accuracy is demonstrated by ablation studies. To sum up, YOLO-ACN provides a workable way to identify small, obscured objects in real time.[6]

Hartill et al. pointed out that the underuse of digital cameras for tracking recreational fishing efforts is highlighted in the literature, which draws on early adopter research conducted in Germany, Australia, and New Zealand. Conventional approaches, characterized by their labor-intensive and irregular nature, impede a thorough comprehension of dynamic recreational fisheries. The writers support affordable digital camera monitoring systems, highlighting their capacity to track trends of fishing efforts over time. Different jurisdictions have different ethical standards, including privacy rights. It is advised to follow privacy regulations and use low-resolution photographs. For monitoring to be effective, representative monitoring locations, infrastructure requirements, and technological factors like lens selection and data storage are essential. Over time, the program's cost-effectiveness and efficacy are influenced by strategic choices on monitoring methodologies, ongoing attention to system components, and resolution of possible problems.[7]

III. Existing System

Electronic monitoring (EM) technologies for fisheries management have made significant strides, but there are still several restrictions and knowledge gaps. Although improving recall and precision, the suggested YOLOv5-based method for Engraulis japonicus fishing recognition might not be scalable or universal. It has successfully cut down on storage and review time, but the automated catch event detection framework still needs more validation to be more flexible. Though useful, the Fishnet Open Images Database might not accurately depict EM difficulties in the real world. Although the approach for automatic species identification utilizing optical tracking and convolutional neural networks shows promise, issues with species variety and environmental variability require more research. For EM systems to be fully and successfully used in fisheries management, several deficiencies must be filled.

IV. PROPOSED SYSTEM

A. Introduction to YOLO

You Only Look Once (YOLO) operates on the principle of unified object detection through a grid-based approach, exhibiting notable efficacy in real-time computer vision applications. YOLO's core mechanism involves dividing input images into a grid, and each grid cell is responsible for predicting bounding boxes and class probabilities concurrently. This unique design facilitates the expeditious and comprehensive identification of multiple objects within the image. Within each grid cell, YOLO predicts bounding box coordinates relative to the cell's spatial dimensions, along with associated class probabilities. The model employs a single convolutional neural network (CNN) to process the entire image, enabling end-to-end predictions. This streamlined architecture significantly accelerates inference speed while maintaining competitive accuracy, addressing the imperative of real-time applications.

YOLO integrates non-maximum suppression to refine the output by eliminating redundant bounding boxes and enhancing localization precision. The model is trained through a comprehensive loss function, encompassing localization, confidence scores, and class predictions, optimizing parameters through backpropagation. The YOLOv8 model is the latest version of the YOLO model developed by ultralytics company.

Keys features of YOLOv8:

Anchor-Free Detection: YOLOv8 diverges significantly from earlier models by adopting an anchor-free approach in object detection. This entails direct prediction of the object's center, eliminating the need for offsets from predefined anchor boxes. Anchor boxes, a historical challenge in earlier YOLO models, often did not align with the distribution of custom datasets. The shift to anchor-free detection in YOLOv8 addresses the complexities associated with anchor boxes, reducing computational load and enhancing adaptability to custom datasets. This modification significantly impacts the number of box predictions, consequently expediting Non-Maximum Suppression (NMS), a critical post-processing step that refines candidate detections post-inference. This streamlined approach improves the model's efficiency without compromising its real-world applicability.

Closing the Mosaic Augmentation: In deep learning research, while model architecture often takes the spotlight, the training routine is crucial for the success of models like YOLOv5 and YOLOv8. YOLOv8 employs online image augmentation during training, exposing the model to slightly varied images in each epoch. A significant technique is mosaic augmentation, where four images are stitched together to challenge the model with new object locations, partial occlusion, and different surroundings. However, empirical evidence suggests that continuous use of mosaic augmentation throughout training can degrade performance. To address this, it is beneficial to disable it for the final ten training epochs. This strategic adjustment exemplifies the meticulous attention given to refining YOLO modeling, evident in the YOLOv5 repository and ongoing YOLOv8 research.

B. Dataset characteristics

Images from longline tuna vessels in the western and central Pacific are included in the fishnet dataset. Four visually similar tuna species (albacore, yellowfin, skipjack, and bigeye) are represented by almost 85% of the included fish annotations. The "L1" label collection comprises 25 more species from which the other fish annotations are derived. The FAO ASFIS List of Species for Fishery Statistics Purposes is another source of 12 coarser classes that group related species and create the "L2" label set. "Unknown" L1 classes and those with fewer than 1000 labels are included in the L2 "OTH" (other) class; sharks are excluded for conservation purposes. Indicators of fishing activity include annotations made by humans. Due to uneven catch among boats and skewed species distribution, both the L1 and L2 class distributions are long-tailed, as shown in the dataset.

C. Data Split

The training, validation, and testing sets are designed to emulate real-world operating situations for EM algorithms. Mimicking both visible (current EM program members) and invisible (new EM program members) vessels is critical. The Fishnet validation and test sets include equal amounts of imagery from "seen" and "unseen" cameras, which is consistent with datasets that originate photographs from many distinct places. The final split included 59,497 training images, 13,648 validation images, and 12,891 test images. Figure 1 depicts the class distribution. Class Frequency Distribution

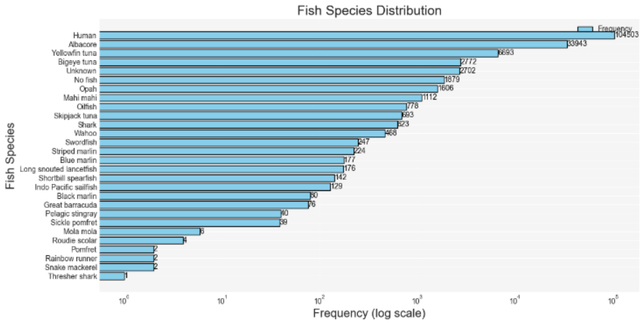

Fig. 2. Class Frequency Distribution

The dataset is highly unbalanced as some classes have little representation, which is unsuitable for a good result. As a result, only the most common classes were chosen for training: "Albacore", "Bigeye tuna", "Yellowfin tuna", and "Unknown". A subset of the dataset was isolated such that each class would have 2700 instances spread across 6618 images.

D. Evaluation Metrics

The following metrics are used to evaluate the model performance[9]:

mAP - Mean Average Precision (mAP) is the primary performance measure of computer vision models. mAP is equal to the average of the Average Precision metric across all classes in a model. You can use mAP to compare both different models on the same task and different versions of the same model. mAP is measured between 0 and 1

Precision - Precision is a metric in a confusion matrix measuring the accuracy of positive predictions. It is calculated as the ratio of true positives to the sum of true positives and false positives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall - Recall, also known as sensitivity or true positive rate, is a metric in a confusion matrix measuring the model's ability to correctly identify all positive instances.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Experimental Environment:

CPU: Intel(R) Xeon(R) CPU @ 2.00GHz
GPU: Tesla P100-PCIE-16GB
Driver Version: 535.129.03
CUDA Version: 12.2
Python version” 3.10.12

Hyperparameters:

Model type: yolov8s
Epochs: 100
Image Size: 640x640
Batch Size: 16
Optimizer: SGD
IOU: 0.7
Momentum: 0.937
Weight_decay: 0.001

V. Results and Evaluations

The model was trained with a mAP50 of 0.90351 and a mAP50-95 of 0.69167. The Precision-Recall curve below further illustrates the mAP of distinct classes.

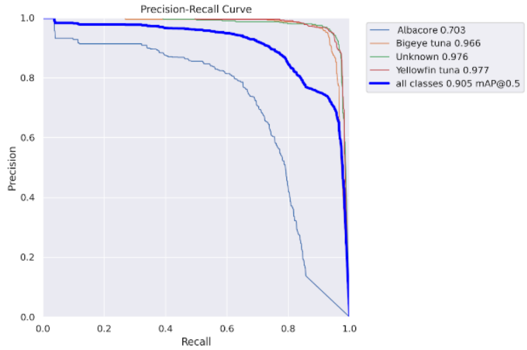

Fig. 2. Precision-Recall Curve

The model's extensive training routine produced outstanding results, with a mean Average Precision at IoU (Intersection over Union) 50 (mAP50) of 0.90351 and a mAP50-95 of 0.69167. These metrics are strong indicators of the model's ability to detect and localize fish in the dataset. The Precision-Recall curve in Figure 2 provides a detailed representation of the mean Average Precision for each class.

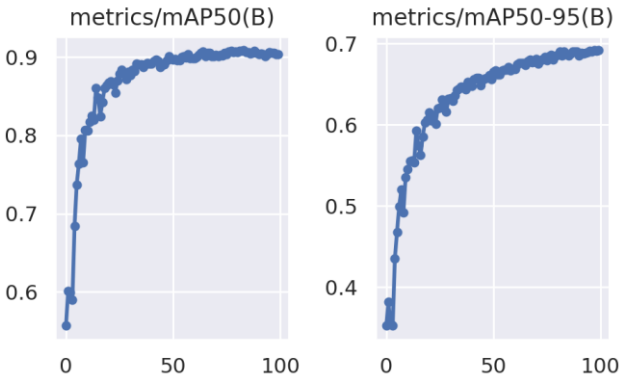

Fig. 3. mAP through epochs

Figure 3 visually depicts the progressive increase in mean Average Precision (mAP) across the training epochs, highlighting a notable upward trend. Notably, the mAP50-95 begins to plateau around the 90-epoch mark, suggesting diminishing returns in performance improvement. This observation guided the decision to cap the training duration at 100 epochs, striking a balance between optimizing the model's accuracy and mitigating the diminishing returns associated with prolonged training.

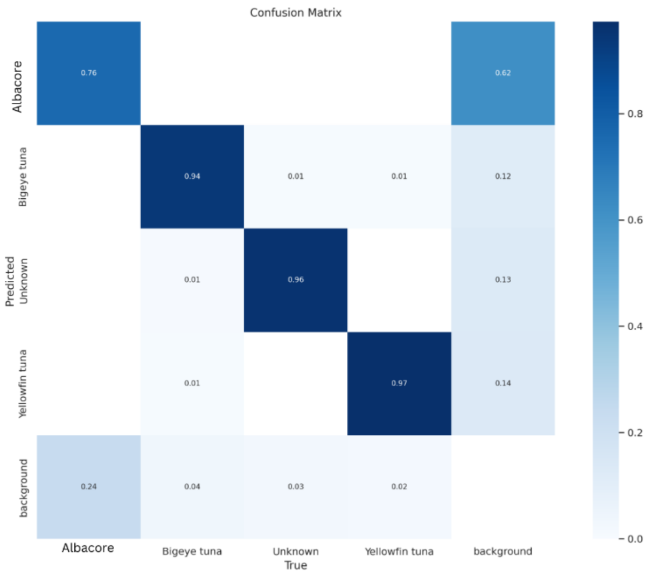

Fig. 4. Confusion Matrix

The generated confusion matrix provides a comprehensive visualization of the prediction outcomes for the four distinct classes, along with the background category. This matrix serves as a valuable tool for assessing the model's performance in terms of classification accuracy and potential areas for improvement.

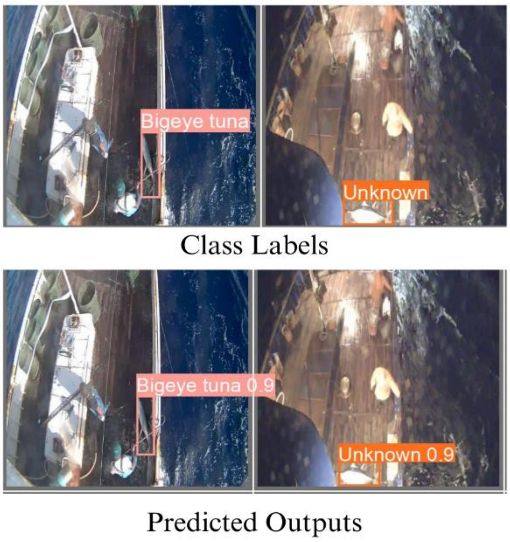

Fig. 5. Prediction Results

Figure 5 shows an actual example of model predictions. The higher photos clearly indicate class designations, providing a visual reference for the ground truth. Meanwhile, the lower images represent the model's predictions, which are shown as class names with matching confidence scores. The confidence scores range from 0 to 1, providing a quantitative assessment of the model's assurance in its predictions. This visual depiction clearly demonstrates the model's capacity to appropriately classify items while quantifying the level of confidence associated with each prediction.

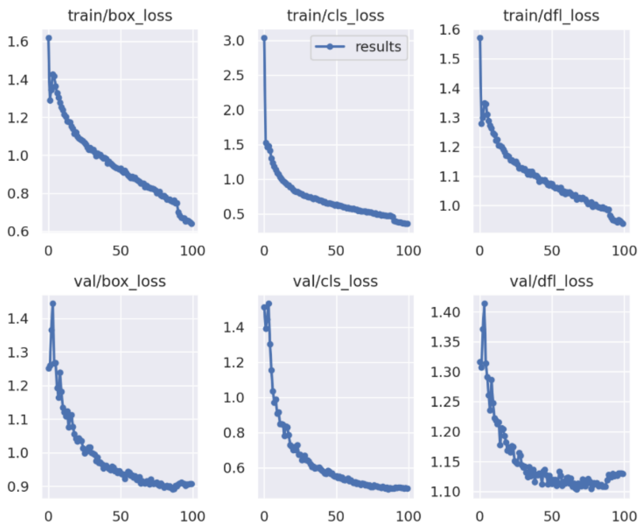

Fig. 6. Loss Metrics during training

Figure 6 depicts a range of loss metrics, where lower values indicate greater model performance. Notably, during the 90th epoch, the majority of the loss metrics show no discernible variation. This result is consistent with the concept of diminishing returns, implying that more training beyond this point may produce minimal improvements in performance. The thorough examination of these loss indicators yields useful insights into training dynamics and aids in establishing the best training duration to balance model development and computing efficiency.

VI. CONCLUSION

To conclude, our study presents a thorough examination of an upgraded YOLO (You Only Look Once) model applied to the fishnet dataset, revealing promising outcomes that carry significant implications for fisheries management practices. As we navigate the dynamic waters of technological innovation, our findings illuminate a promising path toward a more refined and efficient system for monitoring fisheries activities.

Looking ahead, there exists ample room for future research to enhance the YOLO model's practical impact by integrating advanced features like fish tracking and automated counting. This expansion into real-time video feeds not only bolsters accuracy but also holds the potential to revolutionize the enforcement of fishing quotas, addressing challenges associated with manual reporting and mitigating the risks of underreporting.

In essence, the proposed enhancements position the YOLO model as a transformative tool for sustainable fisheries practices, elevating its technical capabilities beyond its current scope. This research not only lays the foundation for future investigations into automated fish tracking but also presents innovative solutions with far-reaching implications for global fisheries conservation and management. As we continue to navigate the confluence of technology and environmental stewardship, our study envisions a future where automated systems contribute significantly to the preservation and sustainable management of our marine resources.

REFERENCES

[1] Khokher, M. R., Little, L. R., Tuck, G. N., Smith, D. V., Qiao, M., Devine, C., O'neill, H., Pogonoski, J. J., Arangio, R., & Wang, D. (2022). Early lessons in deploying cameras and artificial intelligence technology for fisheries catch monitoring: where machine learning meets commercial fishing. Canadian Journal of Fisheries and Aquatic Sciences, 79(2), 257–266. https://doi.org/10.1139/cjfas-2020-0446

[2] Zhang, J., Wang, S., Zhang, S., Tang, F., Fan, W., Yang, S., Sun, Y., Liu, Y., Zhu, W., Zhang, S., & He, R. (2022). Research on target detection of Engraulis japonicus purse seine based on an improved model of YOLOv5. Frontiers in Marine Science, 9. https://doi.org/10.3389/fmars.2022.933735

[3] Kay, J., & Merrifield, M. (2021). The Fishnet Open Images Database: A Dataset for Fish Detection and Fine-Grained Categorization in Fisheries. http://arxiv.org/abs/2106.09178

[4] Qiao, M., Wang, D., Tuck, G. N., Little, L. R., Punt, A. E., & Gerner, M. (2021). Deep learning methods applied to electronic monitoring data: Automated catch event detection for longline fishing. ICES Journal of Marine Science, 78(1), 25–35. https://doi.org/10.1093/icesjms/fsaa158

[5] Gilman, E., Legorburu, G., Fedoruk, A., Heberer, C., Zimring, M., & Barkai, A. (2019). Increasing the functionalities and accuracy of fisheries electronic monitoring systems. Aquatic Conservation: Marine and Freshwater Ecosystems, 29(6), 901–926. https://doi.org/10.1002/aqc.3086

[6] Li, Y., Li, S., Du, H., Chen, L., Zhang, D., & Li, Y. (2020). YOLO-ACN: Focusing on small target and occluded object detection. IEEE Access. https://doi.org/10.1109/ACCESS.2020.3046515

[7] Hartill, B. W., Taylor, S. M., Keller, K., & Weltersbach, M. S. (2020). Digital camera monitoring of recreational fishing effort: Applications and challenges. Fish and Fisheries, 21(1), 204–215. https://doi.org/10.1111/faf.12413

[8] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. http://arxiv.org/abs/1506.02640

[9] Hartill, B. W., Taylor, S. M., Keller, K., & Weltersbach, M. S. (2020). Digital camera monitoring of recreational fishing effort: Applications and challenges. Fish and Fisheries, 21(1), 204–215. https://doi.org/10.1111/faf.12413

[10] Mean Average Precision (mAP) in Object Detection. (n.d.). Retrieved January 28, 2024, from https://blog.roboflow.com/mean-average-precision/

[11] What is YOLOv8? The Ultimate Guide. (n.d.). Retrieved January 28, 2024, from https://blog.roboflow.com/whats-new-in-yolov8/

[12] Gilman, E., Legorburu, G., Fedoruk, A., Heberer, C., Zimring, M., & Barkai, A. (2019). Increasing the functionalities and accuracy of fisheries electronic monitoring systems. Aquatic Conservation: Marine and Freshwater Ecosystems, 29(6), 901–926. https://doi.org/10.1002/aqc.3086

[13] Wageeh, Y., Mohamed, H., Fadl, A., et al. (2021). YOLO fish detection with Euclidean tracking in fish farms. Journal of Ambient Intelligence and Humanized Computing, 12(1). https://doi.org/10.1007/s12652-020-02847-6

[14] Muksit, A., Hasan, F., Hasan Bhuiyan, E., Haque, M., Anwary, A., & Shatabda, S. (2022). YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment. Ecological Informatics, 72. https://doi.org/10.1016/j.ecoinf.2022.101847

[15] Shafait, F., Mian, A., Shortis, M., Ghanem, B., Culverhouse, P., Edgington, D., Cline, D., Ravanbakhsh, M., Seager, J., & Harvey, E. (2016). Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science: Journal du Conseil, 73(10). https://doi.org/10.1093/icesjms/fsw106

[16] Bekkozhayeva, D., & Cisar, P. (2022). Image-Based Automatic Individual Identification of Fish without Obvious Patterns on the Body (Scale Pattern). Applied Sciences (Switzerland), 12(11). https://doi.org/10.3390/app12115401

[17] Sandford, M., Castillo, G., & Hung, T. (2020). A review of fish identification methods applied on small fish. Reviews in Aquaculture, 12(2). https://doi.org/10.1111/raq.12339

[18] Ovalle, J., Vilas, C., & Antelo, L. (2022). On the use of deep learning for fish species recognition and quantification on board fishing vessels. Marine Policy, 139. https://doi.org/10.1016/j.marpol.2022.105015

[19] Prasetyo, E., Suciati, N., & Fatichah, C. (2020). A Comparison of YOLO and Mask R-CNN for Segmenting Head and Tail of Fish. In ICICoS 2020 - Proceeding: 4th International Conference on Informatics and Computational Sciences. https://doi.org/10.1109/ICICoS51170.2020.9299024

[20] Kuswantori, A., Suesut, T., Tangsrirat, W., & Nunak, N. (2022). Development of object detection and classification with YOLOv4 for similar and structurally deformed fish. EUREKA, Physics and Engineering, 2022(2). https://doi.org/10.21303/2461-4262.2022.002345