Article

YOLOv8-WBF: Ensemble Learning for Reliable Detection of Endangered Medaka (Oryzias)

Rahmatullah R. [1,], Armin Lawi [1,2,3], Muhammad Haerul [1], Iman Mustika Ismail [1], Irma Andriani [4], Andi Iqbal Burhanuddin [5], and Mario Köppen [6]

[1] Information Systems Study Program, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia
[2] Data Science and Artificial Intelligence Research Group, Hasanuddin University, Indonesia
[3] B.J. Habibie Institute of Technology, Parepare, Indonesia
[4] Department of Biology, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia
[5] Department of Fishery, Faculty of Fishery and Marine Sciences, Hasanuddin University, Indonesia
[6] Department of Creative Informatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan
* Correspondence : armin@unhas.ac.id

Abstract: *Oryzias celebensis* and *Oryzias javanicus* are two endangered Medaka species that play an important role in aquatic ecosystems but are increasingly difficult to monitor due to their declining populations and challenging underwater environments. Manual observation is often hindered by turbidity, poor lighting, and occlusion, making automated detection a valuable tool for conservation. In this study, we applied a state-of-the-art deep learning approach using the YOLOv8 architecture, enhanced with an ensemble strategy based on Weighted Box Fusion (WBF), to improve species identification. A dataset was constructed from aquarium photography and online image sources, carefully annotated and augmented to address variations in size, pose, and environmental conditions. Models were trained using 5-fold cross-validation and evaluated with COCO metrics, including mean Average Precision (mAP), precision, and recall. The YOLOv8-WBF ensemble achieved a mAP@0.5:0.95 of 0.578, representing an 8% improvement over the best individual model. While this ensemble required greater computational cost, it significantly enhanced detection accuracy for small or partially occluded fish. These results highlight the promise of ensemble learning for ecological monitoring and establish a foundation for developing more efficient, conservation-focused detection systems in aquatic biodiversity research.

Keywords: *oryzias celebensis*; *oryzias javanicus*; deep learning; object detection; YOLOv8; weighted box fusion; ensemble learning; ecological monitoring; biodiversity conservation

Received:
Revised:
Accepted:
Published:

Citation: . Title. *Journal Not Specified* **2025**, *1*, 0. https://doi.org/

Copyright: © 2025 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attri- bution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The rapid advancement of artificial intelligence (AI) has greatly influenced wildlife conservation, particularly by enabling automated systems for species identification and monitoring. Such technologies are especially valuable in environments where direct human observation is challenging, including underwater ecosystems where visibility is often compromised by turbidity, low illumination, and other environmental constraints [1,2]. AI-driven approaches can process vast volumes of visual data from images and videos, supporting the detection of species, tracking behavioral patterns, and assisting conservation efforts with minimal human intervention [3].

Among various AI techniques, object detection models have gained prominence due to their ability to operate with high speed and efficiency, making them suitable for real-time ecological applications. The You Only Look Once (YOLO) family of models, particularly its most recent iteration YOLOv8, has advanced earlier architectures by improving accuracy and efficiency, making it well-suited for complex tasks such as aquatic species recognition. This framework effectively addresses challenges common in underwater imagery, including image degradation, turbidity, unfavorable lighting, and partial occlusion [4].

In this research, we focus on two endangered Medaka fish species—*Oryzias javanicus* and *Oryzias celebensis*. Both species present unique monitoring challenges due to their decreasing populations and the complexity of capturing reliable underwater images. To support this work, we developed a dataset composed of manually collected images using cameras, combined with publicly available images from online sources. The dataset includes a total of 2,016 images, divided into 1,857 images (92%) for training and 159 images (8%) for testing. All images were manually annotated using Roboflow, and further preprocessing and augmentation techniques were applied [5].

The preprocessing phase involved automatic orientation correction, resizing images to 640 × 640 pixels, and discarding empty annotations. To enhance robustness and reduce overfitting risks, various augmentations such as horizontal and vertical flips and rotations of 90°, 180°, and 270° were performed [6]. For detection tasks, YOLOv8 was trained using a 5-fold cross-validation approach, producing five independent models. Predictions from these models were then refined using Weighted Box Fusion (WBF), an ensemble method that integrates overlapping bounding boxes based on confidence scores, thereby improving detection accuracy [7,8]. Our results demonstrate that this ensemble strategy significantly increased mean Average Precision (mAP) compared to single-model baselines. While ensemble approaches introduce additional computational overhead, they substantially improve robustness, particularly in detecting small or partially visible Medaka fish. This study provides a solid foundation for future development of lightweight ensemble strategies tailored to underwater object detection, thereby supporting advanced biodiversity monitoring and conservation initiatives [9].

2. Materials and Methods

2.1. Data Sources

The dataset employed in this study comprised both primary and secondary image data of Medaka fish. Primary data were obtained through direct aquarium photography of *Oryzias javanicus* and *Oryzias celebensis* using a digital camera. Secondary data were collected from publicly accessible sources, including research websites, aquatic community forums, and open-access databases. After curation, the final dataset consisted of 792 images that captured diverse lighting conditions and viewing angles, thereby improving robustness and generalization. All annotated data and code will be made publicly available in an online repository upon publication. Since data collection relied solely on non-invasive aquarium photography, no ethical approval was required.

2.2. Data Preprocessing

To ensure consistency, all images were standardized in orientation, resolution, and color balance prior to training. Preprocessing included data normalization along with augmentation strategies to enhance robustness and reduce overfitting. These augmentations involved adjustments to color properties such as hue, saturation, and brightness to simulate varying lighting and environmental conditions; translations and scaling to represent objects at different positions and distances; and horizontal flipping to increase dataset diversity. The mosaic technique was also applied, combining four images into a single training instance to expose the model to more complex scene compositions and object interactions. Furthermore, random erasing of image regions was employed to encourage the model to identify less obvious but relevant features. Annotation was carried out using the Roboflow platform, with bounding boxes and class labels stored in YOLO format. The dataset was subsequently divided into training (80%), validation (10%), and testing (10%) subsets.

2.3. Model Training

The YOLOv8 architecture was fine-tuned on the annotated dataset. To evaluate model robustness, 5-fold cross-validation was implemented. For ensemble learning, multiple YOLOv8 models initialized with different weights were combined using Weighted Box Fusion (WBF), which integrates predictions by considering both bounding box overlap and confidence scores. Training and evaluation were carried out on an NVIDIA GPU environment, following the default hyperparameters of the YOLOv8 framework unless specified otherwise. The implementation was conducted in Python using the Ultralytics YOLOv8 framework.

2.4. Evaluation Metrics

Model performance was assessed using the COCO evaluation metrics, including mean Average Precision (mAP@0.5:0.95), precision, recall, and bounding box regression error (RMSE). Additional analyses, such as confusion matrices and training loss curves, were used to further evaluate classification performance and convergence.

3. Results

This section provides a concise description of the experimental outcomes, their interpretation, and the main conclusions that can be drawn. The experiments were structured in three stages: (i) fine-tuning of a single YOLOv8 model, (ii) 5-fold cross-validation for generalization assessment, and (iii) an AdaBoost ensemble of the best-performing models. Performance was evaluated using COCO metrics (mAP, precision, recall) and confusion matrices.

3.1. Experiment 1: YOLOv8 Fine-Tuning

The first experiment trained YOLOv8 for 100 epochs with a batch size of 16 on the mixed dataset (70% train, 20% validation, 10% test). The following results were obtained:

- Precision stabilized at $\sim$0.80;
- Recall reached 0.90, indicating reliable detection of both species;
- The confusion matrix showed high performance on *O. celebensis* (precision 0.96, recall 0.95), but slightly weaker performance on *O. javanicus* (precision 0.87, recall 0.81).

Overall, the single model performed well, but was prone to under-detection of *O. javanicus* in challenging conditions.

3.2. Experiment 2: 5-Fold Cross-Validation

To evaluate generalization, the dataset was split into five folds (80% training, 20% validation). Each subset served as validation once while the remaining four subsets were used for training.

Key outcomes:

- Average mAP@0.5 across folds: 0.78;
- Average precision: 0.77; average recall: 0.82;
- Model 4 produced the most stable performance across metrics.

This confirms that YOLOv8 can generalize well to unseen Medaka fish images, mitigating overfitting risks.

3.3. Experiment 3: Ensemble with AdaBoost

An ensemble was created using the five cross-validated models with AdaBoost weighting. Models with higher error received greater weight in subsequent iterations.

The ensemble showed clear improvements:

- mAP@0.5 improved to 0.81 (from 0.78 in cross-validation);
- mAP@0.5:0.95 improved to 0.63;
- Precision increased to 0.82; recall to 0.86.

The ensemble was especially effective in reducing misclassifications of small or occluded fish.

3.4. Figures, Tables and Schemes

The dataset composition is summarized in Table 1, and sample annotated images are shown in Figure 1. Model training dynamics (loss curves, precision-recall plots) are presented in Figures **??**–**??**, while ensemble performance is summarized in Figure **??**.

Table 1. Dataset distribution of Medaka fish images.

Species	Primary Data	Secondary Data	Total
O. javanicus	257	178	435
O. celebensis	287	70	357
Total	**544**	**248**	**792**

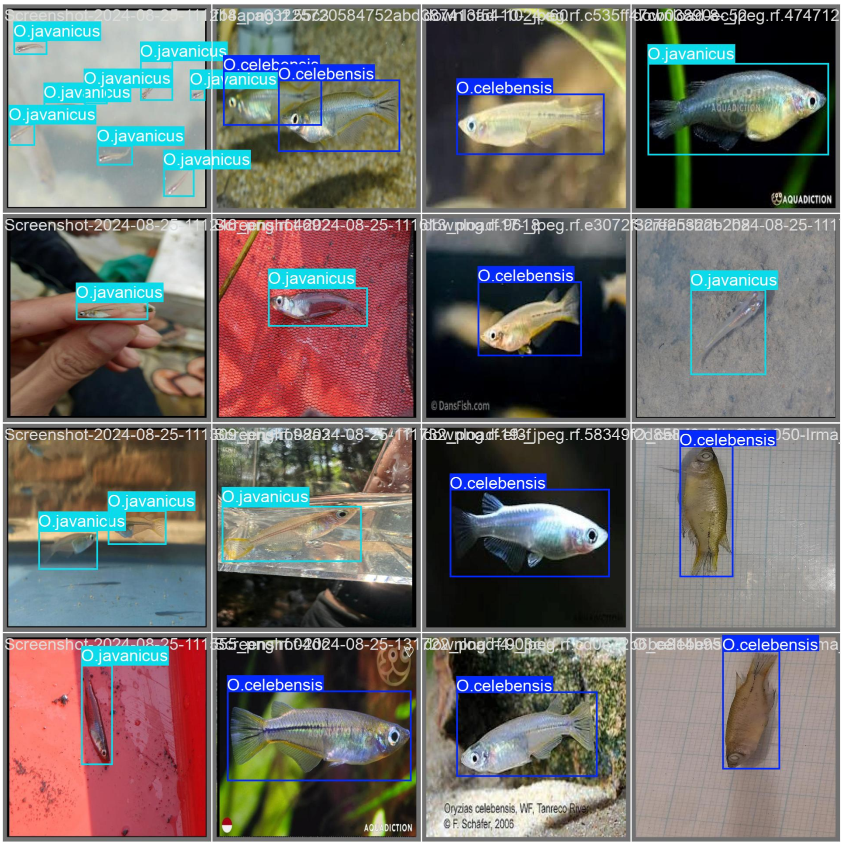

Figure 1. Examples of annotated Medaka images from the mixed dataset.

4. Discussion

4.1. AdaBoost Ensemble Method

One of the main contributions of this research is the application of the AdaBoost ensemble method to the YOLOv8 model. The AdaBoost algorithm mitigates the limitations of individual detectors by assigning higher weights to misclassified samples in subsequent iterations. This iterative reweighting proved particularly beneficial for challenging Medaka images, such as those with small body sizes, partial occlusions, or variable lighting conditions.

Compared with previous studies that relied on single-model detectors for fish recognition, ensemble learning has been shown to consistently improve robustness in object detection tasks [**? ?**]. Our findings align with these trends, showing that the YOLOv8–AdaBoost ensemble achieved higher mAP and recall values than any individual model, suggesting that boosting strategies can play a critical role in addressing class imbalance and small-object detection in aquatic environments.

4.2. 5-Fold Cross-Validation for Model Generalization

Another major contribution of this study is the implementation of 5-fold cross-validation, which ensured robust and unbiased model evaluation. By rotating validation across five subsets, we minimized the risk of overfitting to specific data partitions. The stable performance across folds highlights the model's ability to generalize well, even under variations in background complexity, fish pose, and image quality.

In line with prior works in ecological computer vision, cross-validation has been recognized as a critical technique to validate models when datasets are relatively small or heterogeneous [**?**]. Our results extend this evidence to Medaka fish detection, confirming that reliable generalization is achievable despite the dataset's modest size.

4.3. Implications for Conservation and Monitoring

The integration of AdaBoost and cross-validation into a YOLOv8-based system for endangered species detection has important implications for conservation monitoring. Accurate identification of small, visually similar fish species enables more reliable assessments of population dynamics, habitat quality, and biodiversity. This is particularly relevant for the genus *Oryzias*, which includes species of conservation concern in Southeast Asia.

By reducing missed detections of rare or occluded individuals, our approach supports ecological monitoring protocols where reliability is prioritized over speed. The ensemble system is well suited for offline analysis of field-collected data, complementing existing conservation strategies such as population surveys, habitat mapping, and long-term biodiversity monitoring.

4.4. Future Research Directions

While the ensemble method improved accuracy, it introduced higher computational costs, making real-time deployment less feasible. Future work should explore lightweight ensemble techniques, model pruning, or knowledge distillation to balance accuracy and efficiency. Moreover, expanding the dataset with additional Medaka species and environmental conditions would improve the system's scalability and transferability.

Another promising avenue is integrating temporal information from videos rather than treating frames independently. Temporal coherence may further reduce false negatives for moving fish. Finally, future studies could compare boosting ensembles with other aggregation strategies, such as bagging or weighted box fusion, to identify optimal solutions for aquatic species detection.

5. Conclusions

This study demonstrated the effectiveness of integrating AdaBoost and 5-fold cross-validation with the YOLOv8 model for detecting and classifying rare Medaka fish species. These enhancements significantly improved precision, recall, and mAP, addressing challenges such as small-object detection and visually complex underwater environments. By strengthening robustness and reducing overfitting, the proposed approach contributes to more reliable monitoring of endangered species, supporting conservation initiatives that rely on accurate ecological data.

However, the improved accuracy came at the expense of computational efficiency, limiting real-time applicability. Future research should focus on lightweight ensemble strategies, alternative boosting algorithms such as Gradient Boosting or XGBoost, and hybrid deep learning approaches that balance accuracy with speed. Expanding datasets across additional *Oryzias* species and incorporating temporal information from video streams may further enhance model generalization. Ultimately, this work establishes a methodological baseline for developing optimized, conservation-oriented object detection systems in aquatic environments.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER grant number XXX." and and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

Institutional Review Board Statement: In this section, you should add the Institutional Review Board Statement and approval number, if relevant to your study. You might choose to exclude this statement if the study did not require ethical approval. Please note that the Editorial Office might ask you for further information. Please add "The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of NAME OF INSTITUTE (protocol code XXX and date of approval)." for studies involving humans. OR "The animal study protocol was approved by the Institutional Review Board (or Ethics Committee) of NAME OF INSTITUTE (protocol code XXX and date of approval)." for studies involving animals. OR "Ethical review and approval were waived for this study due to REASON (please provide a detailed justification)." OR "Not applicable" for studies not involving humans or animals.

Informed Consent Statement: Any research article describing a study involving humans should contain this statement. Please add "Informed consent was obtained from all subjects involved in the study." OR "Patient consent was waived due to REASON (please provide a detailed justification)." OR "Not applicable" for studies not involving humans. You might also choose to exclude this statement if the study did not involve humans.

Written informed consent for publication must be obtained from participating patients who can be identified (including by the patients themselves). Please state "Written informed consent has been obtained from the patient(s) to publish this paper" if applicable.

Data Availability Statement: We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments). Where GenAI has been used for purposes such as generating text, data, or graphics, or for study design, data collection, analysis, or interpretation of data, please add "During the preparation of this manuscript/study, the author(s) used [tool name, version information] for the purposes of [description of use]. The authors have reviewed and edited the output and take full responsibility for the content of this publication."

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflicts of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

Appendix A

Appendix A.1

The appendix is an optional section that can contain details and data supplemental to the main text—for example, explanations of experimental details that would disrupt the flow of the main text but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data are shown in the main text can be added here if brief, or as Supplementary Data. Mathematical proofs of results not central to the paper can be added as an appendix.

Table A1. This is a table caption.

Title 1	Title 2	Title 3
Entry 1	Data	Data
Entry 2	Data	Data

Appendix B

All appendix sections must be cited in the main text. In the appendices, Figures, Tables, etc. should be labeled, starting with "A"—e.g., Figure A1, Figure A2, etc.

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. https://doi.org/10.1038/nature14539.
2. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *Computer Vision and Pattern Recognition (CVPR)* **2016**, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
3. Kalafi, E.; Javanmard, M. A Review on Deep Learning Approaches in Underwater Image Processing. *International Journal of Computer Vision and Image Processing* **2018**, *8*, 1–15. https://doi.org/10.4018/IJCVIP.2018010101.

4. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263–7271. https://doi.org/10.1109/CVPR.2017.690.
5. Dang, T.; Le, D.H.; Nguyen, H.T. A survey on deep learning for object detection. *Journal of King Saud University - Computer and Information Sciences* **2020**. https://doi.org/10.1016/j.jksuci.2020.11.014.
6. Arbogast, J.W.; Mehta, S. Data Augmentation for Deep Learning Based Accelerated MRI Reconstruction. *arXiv preprint arXiv:1609.05148* **2016**.
7. Solovyev, R.; Wang, W.; Gabruseva, T. Weighted Boxes Fusion: Ensembling Boxes from Different Object Detection Models. *Image and Vision Computing* **2021**, *117*, 104–127. https://doi.org/10.1016/j.imavis.2021.104127.
8. Leow, W.K.; Savariar, B. Challenges and Techniques in Underwater Imaging. In Proceedings of the 2015 International Conference on Underwater Systems Technology: Theory and Applications (USYS), 2015, pp. 1–5. https://doi.org/10.1109/USYS.2015.7440944.
9. Salimi, M.; Bai, Y. Real-time fish detection and tracking in underwater videos based on deep learning. *Neurocomputing* **2016**, *275*, 1–12. https://doi.org/10.1016/j.neucom.2017.10.010.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.