Comparative Analysis of YOLO Model Performance for Freshwater Fish Detection

Vrica Gede Penggalih[*1], Esi Putri Silmina[2]

[1]Department of Information Technology, Faculty of Science and Technology, Aisyiyah University Yogyakarta, Yogyakarta, Indonesia

[2]Department of Information Technology, Faculty of Science and Technology, Aisyiyah University Yogyakarta, Yogyakarta, Indonesia

E-mail: vricagedepenggalih@gmail.com[1], esiputrisilmina@unisayogya.ac.id [2]

Abstract. In 2024, the prevalence of stunting in Indonesia reached around 19.8%, still far from the target of 14.2%. Nutritional literacy in the community is one of the factors contributing to stunting. Freshwater fish is one solution for animal protein, but a lack of knowledge about fish is one of the obstacles to the utilization of freshwater fish. With digital image-based automatic object detection technology, YOLO offers great potential to support nutrition literacy and public education. This study aims to analyze and find the best model for detecting freshwater fish (Lele, Patin, Gurame, Nila, Bandeng) by comparing the performance of YOLOv11 and YOLOv12. The research method began with the collection and annotation of the dataset, followed by data pre-processing, model training, and analysis of the performance of both YOLO versions. Research conducted on the use of this technology is expected to be meaningful in reducing stunting rates, especially in Indonesia.

Keywords: yolov11, yolov12, freshwater fish detection, stunting, nutrition literacy.

1. Introduction

Indonesia still faces major challenges in terms of child nutrition [1]. Data from the 2024 Indonesian Nutrition Status Survey (SSGI) shows that the national prevalence of stunting reached around 19.8% in 2024 [2]. Although this represents a downward trend compared to the previous year, the figure is still far from the national target of 14.2% set in the 2024-2029 National Medium-Term Development Plan (RPJMN) [3]. One of the causes of stunting is low nutrition literacy [4]. This condition underscores that nutrition interventions are not only important in the short term, but also crucial for the long-term development of the nation.

Meanwhile, freshwater fish such as Lele, Patin, Nila, Bandeng, and Gurame are relatively affordable local commodities that are easily accessible to the Indonesian people [5], [6]. The use of freshwater fish as a source of animal protein can be an important strategy in the context of improving nutritional literacy and empowering local communities [7], [8]. However, in practice, there are obstacles in quickly and accurately identifying fish species by consumers or farmers, including in selecting fish that are highly nutritious and adequate for children's nutritional intake [9].

Object detection or automatic object detection based on digital images offers great potential to support nutrition literacy and public education [10]. In particular, the "You Only Look Once" framework of the YOLOv11 and YOLOv12 model families has shown significant improvements in terms of object detection accuracy and speed [11]. For example, a review study of the YOLO model series states that YOLOv12 is capable of providing a higher mAP with lower latency compared to previous versions [12]. This opens up the possibility that object detection technology can be applied to detect freshwater fish species from digital images, then integrated into educational applications or systems that help the public choose fish correctly.

Research on fish detection using YOLO-based models has been conducted extensively in recent years. The YOLO model was chosen because it provides better detection accuracy than other detection models, such as RF-DETR and Faster R-CNN [13], [14]. Zhang et al. [15] developed the YOLOv8 model, which is optimized for underwater fish detection, and the results showed a significant improvement in performance compared to YOLOv5 in low-light conditions and complex backgrounds. A similar study by Vijayalakshmi et al. [16] introduced AquaYOLO, a variant of YOLO adapted to detect fish in natural aquaculture environments, with an increase in mean Average Precision (mAP) of up to 8% compared to the basic YOLO version.

Meanwhile, local research by Zainuddin et al. [17] applied YOLOv5 on the Android platform to count catfish fry. The study showed that image-based detection technology can be used to assist fish farmers in the cultivation process, particularly in automatic seed counting. On the other hand, Al Muksit et al. [18] introduced YOLO-Fish, a YOLOv5-based model tailored to the morphological characteristics of marine fish, and achieved high precision in dynamic lighting conditions. These studies confirm that the YOLO model is effective in the field of fisheries, but most of them still focus on marine fish.

To date, there have been few studies comparing the performance of the YOLOv11 and YOLOv12 models in the context of freshwater fish, especially Indonesian local species such as Lele, Patin, Nila, Bandeng, and Gurame. This is an important gap, considering that freshwater fish are an affordable source of protein that plays a strategic role in efforts to reduce stunting rates in Indonesia. Therefore, this study offers something new by conducting a comparative evaluation of the performance of the YOLOv11 and YOLOv12 models in detecting five types of local freshwater fish: Lele, Patin, Nila, Bandeng, and Gurame. Research conducted on the use of this technology is expected to be meaningful in reducing stunting rates, especially in Indonesia.

2. Research Method

This study uses a quantitative experimental approach with the aim of analyzing and comparing the performance of two YOLO-based object detection models, namely YOLOv11 and YOLOv12, in detecting five types of local Indonesian freshwater fish. This approach was chosen because it allows for measurable evaluation of the accuracy, computational efficiency, and generalization ability of the models on a dataset of fish images that are homogeneous in domain but diverse in visual appearance.

2.1. Research Design

The research design consists of six main stages, namely: (1) dataset collection, (2) dataset annotation, (3) dataset splitting, (4) pre-processing & augmentation (5) training of the YOLOv11 and YOLOv12 models, (6) analysis of results and interpretation. The overall flowchart of the research can be seen in Figure 1.
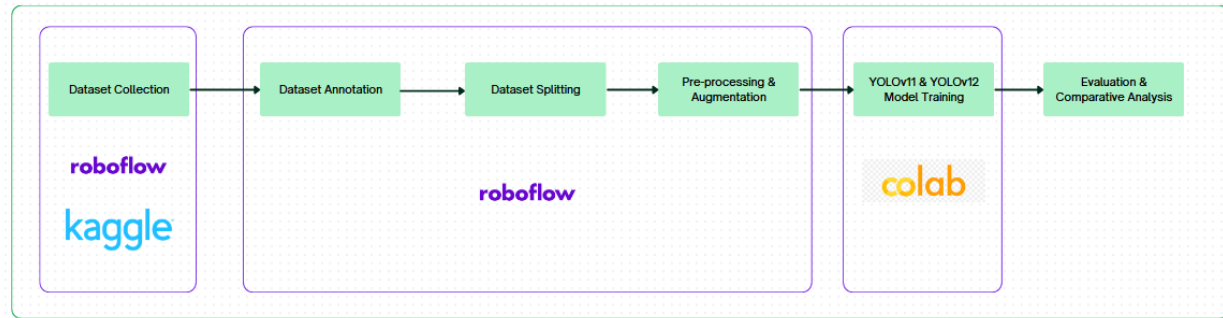
Figure 1. Research Stages

2.2. Dataset Collection

The dataset used in this study consisted of 500 freshwater fish images, covering five popular fish species in Indonesia, namely: Lele (*Clarias batrachus*), Patin (*Pangasius hypophthalmus*), Nila (*Oreochromis niloticus*), Bandeng (*Chanos chanos*), and Gurame (*Osphronemus goramy*) [19]. Each fish class was taken from various public sources (Roboflow, Kaggle) with consideration given to variations in lighting, shooting angle, background, and object position. This diversity was intended to create a dataset that was representative of real-world conditions in the field, such as in fish markets or aquaculture environments. All image data were converted to .jpg format to maintain model input consistency. Table 1 shows the distribution of the research dataset.

Table 1. Research Datasets

Class	Count
Lele	100
Patin	100
Gurame	100
Nila	100
Bandeng	100

2.3. Dataset Annotation

The entire object annotation process was carried out using the Roboflow platform, which provides a web-based interface for data labeling. Each image was manually given a bounding box according to the area of the fish's body and labeled according to its species. Manual annotation was chosen to ensure high label accuracy, which is an important factor in the quality of object detection model training. Annotation errors can cause the model to learn incorrect features and reduce detection performance. Figure 2 shows the dataset annotation process in Roboflow.

Figure 2. Annotation Process in Roboflow

2.4. Dataset Splitting

After the annotation process is complete, the dataset is divided into three subsets with a ratio of 70% for training data (training set), 20% for validation data (validation set), and 10% for testing data (testing set). This proportion allows the model to have enough data to learn visual patterns during the training stage, while the validation data is used to assess the model's ability to generalize to new data during the training process. The testing data is used only at the final stage to measure the actual performance of the model after training is complete.

2.5. Pre-processing and Augmentation Dataset

The pre-processing stage is carried out in Roboflow and consists of two main steps: image standardization and normalization. All images are resized to 640×640 pixels, the standard size used by the YOLO model to maintain input dimension consistency. In addition, auto-orientation is performed to ensure uniform image orientation, avoiding differences in rotation angles that can affect the feature extraction process.

To expand data variation and improve model generalization, data augmentation is performed using several transformation techniques. Table 2 shows a list of augmentation techniques applied to the dataset and their purposes.

Table 2. Augmentation Technique Applied to Dataset

Augmentation Technique	Purpose
Horizontal flip	Creating variations in fish orientation
Rotasi -15° to +15°	Simulating variations in the angle of view
Saturation-15% to+15% dan brightness -15% to +15%	Anticipating variations in natural light intensity
Exposure -10% to +10%	Enriching lighting level variation
Blur up to 2 pixel	Training the model still enables it to detect objects in images with low sharpness

The augmentation techniques in Table 2 were selected based on common practices in object detection research to improve model robustness without causing excessive distortion to the shape of the fish [20] and to avoid model overfitting [21].

2.6. Model Training

The training process was conducted using Google Colab with a GPU runtime configuration (NVIDIA Tesla T4) to ensure efficient computation. The augmented dataset was imported into the Colab environment to

be trained using both models, YOLOv11 and YOLOv12, separately but with identical parameters. The training parameters set can be seen in Table 3.

Table 3. Training Parameter

PARAMETER
Epoch: 100
Batch Size: 16
Image Size (imgsz): 640
Optimizer: AdamW

During training, the model will iteratively update the network weights by minimizing the loss function, which includes three components: bounding box regression loss, objectness loss, and classification loss. This process produces a model capable of recognizing and classifying fish in images with high accuracy.

2.7. Evaluation and Comparative Analysis
The evaluation stage was conducted to assess and compare the performance of the YOLOv11 and YOLOv12 models in detecting freshwater fish based on digital images. The evaluation was carried out on a testing set subset (10% of the total data) that was never used during training or validation, so that the test results objectively represented the model's generalization ability to new data.

Performance was assessed using four main evaluation metrics commonly used in object detection research, namely Precision (P), Recall (R), mAP@0.5 (mAP@0.5), and F1-Score. These four metrics provide a comprehensive overview of how well the model is able to detect objects accurately and consistently.

a.   Precision (P)
     Precision measures the accuracy of a model in detecting objects, i.e., how many predictions are correct (true positives) compared to all detections made by the model, including false positives. The Precision formula is shown in equation (1).

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

     with:
     TP (True Positive) = number of correct detections, i.e., fish objects that are correctly detected according to class and position.
     FP (False Positive) = number of false detections, i.e., predictions that do not match the actual object (for example, the model detects "lele" when the object is actually "gurame").
     High precision indicates that the model rarely makes false detections or classification errors. In the context of this study, a high Precision value means that the model is able to identify fish with a low classification error rate.

b.   Recall (R)
     Recall measures the extent to which the model is able to find all objects that actually exist in the image. In other words, Recall measures the model's ability to find all fish that appear in the image. The Recall formula is expressed in equation (2).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

     with:
     FN (False Negative) = the number of objects that should have been detected but were not detected by the model.

A high Recall indicates that the model has good sensitivity to objects, meaning that almost all fish in the image are successfully detected. In the context of this research, high Recall is important to ensure that no fish are "missed" by the detection system, which is particularly relevant for applications such as monitoring catches or classifying fish in the market.

c. Mean Average Precision (mAP@0.5)

Mean Average Precision (mAP) is the main metric in object detection tasks because it combines Precision and Recall into a single overall performance measure. mAP is calculated based on the Average Precision (AP) value of each object class, then averaged.

The AP calculation is based on the area under the Precision–Recall (PR) curve. The value of mAP@0.5 indicates that the evaluation was performed with an Intersection over Union (IoU) threshold of 0.5, meaning that predictions are considered correct if the overlap between the predicted bounding box and the ground truth is greater than or equal to 50%. The general formula for mAP is shown in equation (3).

$$mAP = \frac{1}{N} \sum_{i=1}^{N} APi \tag{3}$$

with:
N = number of classes (in this study N = 5 types of fish),
AP_i = Average Precision value for class i.

A high mAP@0.5 value indicates that the model has a good balance between Precision and Recall across all fish classes. This metric is often used as a key indicator of the success of an object detection model because it considers both the accuracy and completeness of detection.

d. F1-Score

F1-Score is a measure of the balance between Precision and Recall. This value is used to evaluate the overall performance of a model in a single number. The F1-Score formula is shown in equation (4).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \tag{4}$$

The F1-Score is high if both Precision and Recall are high, and low if either one is low. In this study, the F1-Score is used to provide an overview of the balance between the model's ability to detect fish correctly (without classification errors) and ensuring that all fish are successfully detected.

3. References

[1]   F. Ernawati, A. Syauqy, A. Y. Arifin, M. Y. E. Soekatri, and S. Sandjaja, "Micronutrient deficiencies and stunting were associated with socioeconomic status in indonesian children aged 6–59 months," *Nutrients*, vol. 13, no. 6, 2021, doi: 10.3390/nu13061802.

[2]   Kemenkes, *Survei Status Gizi Indonesia (SSGI) 2024*. 2025. Accessed: Oct. 25, 2025. [Online]. Available: https://www.badankebijakan.kemkes.go.id/survei-status-gizi-indonesia-ssgi-2024/

[3]   P. P. Indonesia, *Peraturan Presiden (Perpres) Nomor 12 Tahun 2025 tentang Rencana Pembangunan Jangka Menengah Nasional Tahun 2025 - 2029* . 2025. Accessed: Oct. 25, 2025. [Online]. Available: https://peraturan.bpk.go.id/Details/314638/perpres-no-12-tahun-2025

[4]   F. C. Wahyuni, U. Karomah, R. W. Basrowi, N. L. Sitorus, and L. A. Lestari, "The Relationship between Nutrition Literacy and Nutrition Knowledge with the Incidence of Stunting: A Scoping Review," 2023, *Airlangga University Faculty of Public Health*. doi: 10.20473/amnt.v7i3SP.2023.71-85.

[5]   G. H. Martono, A. Azhari, and K. Mustofa, "An extended approach of weight collective influence graph for detection influence actor," *International Journal of Advances in Intelligent Informatics*, vol. 8, no. 1, pp. 1–11, Mar. 2022, doi: 10.26555/ijain.v8i1.800.

[6]   M. Suparta, M. D. Pradana, and K. S. Jit Singh, "Financial Analysis of Catfish Cultivation Businesses in Sedenganmijen Village, Krian District, Sidoarjo Regency," *International Journal of Economics, Business and Management Research*, vol. 07, no. 12, pp. 54–66, 2023, doi: 10.51505/ijebmr.2023.71204.

[7]   Candra Wahyuni and Retno Palupi Yonni Siwi, "Increasing Public Awareness of the Importance of 'Gemari' (Love to Eat Fish) Through Mentoring Catfish Cultivation and Processing," *Journal of Community Engagement in Health*, vol. 7, no. 1, pp. 84–88, Mar. 2024, doi: 10.30994/jceh.v7i1.578.

[8]   D. S. Sudaryanti, A. Helmi Munawar, N. Sahroni, and T. Badriatin, "COMMUNITY EMPOWERMENT THROUGH DIVERSIFICATION OF PROCESSED FISH TOWARDS FOOD SECURITY BASED ON LOCAL WISDOM IN THE PANDEMIC COVID-19," vol. 5, no. 1, pp. 29–40, 2022, doi: 10.31764/jces.v3i1.5982.

[9]   H. Mustafidah, R. Setiawan, and A. Karim, "Image-Based Classification of Freshwater Fish Species to Support Feed Recommendation Using Random Forest," 2025.

[10]   R. Agil, Y. Arjun, and E. P. Silmina, "Deteksi Bahan Pangan Tinggi Protein Menggunakan Model You Only Look Once (YOLO)," *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6889.

[11]   M. Chaman, A. El Maliki, H. El Yanboiy, H. Dahou, H. Laâmari, and A. Hadjoudja, "Comparative Analysis of Deep Neural Networks YOLOv11 and YOLOv12 for Real-Time Vehicle Detection in Autonomous Vehicles," *International Journal of Transport Development and Integration*, vol. 9, no. 1, pp. 39–48, Mar. 2025, doi: 10.18280/ijtdi.090104.

[12]   R. Sapkota *et al.*, "YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series," *Artif Intell Rev*, vol. 58, no. 9, Sep. 2025, doi: 10.1007/s10462-025-11253-3.

[13]   R. Sapkota, R. H. Cheppally, A. Sharda, and M. Karkee, "RF-DETR Object Detection vs YOLOv12 : A Study of Transformer-based and CNN-based Architectures for Single-Class and Multi-Class Greenfruit Detection in Complex Orchard Environments Under Label Ambiguity," Apr. 2025, [Online]. Available: http://arxiv.org/abs/2504.13099

[14]   J. Justam, A. Malik, E. Erlita, D. Mangellak, and Y. Yuyun, "Perbandingan Kinerja YOLO vs Faster R-CNN untuk Deteksi & Estimasi Berat Ikan," *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI)*, vol. 7, no. 2, pp. 363–376, Oct. 2024, doi: 10.57093/jisti.v7i2.273.

[15]   Z. Zhang *et al.*, "An Improved YOLOv8n Used for Fish Detection in Natural Water Environments," *Animals*, vol. 14, no. 14, Jul. 2024, doi: 10.3390/ani14142022.

[16]   M. Vijayalakshmi and A. Sasithradevi, "AquaYOLO: Advanced YOLO-based fish detection for optimized aquaculture pond monitoring," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-89611-y.

[17]   M. Zainuddin and M. S. Zuhri, "Implementasi algoritma YOLOv5 pada platform Android untuk penghitungan bibit ikan lele (clarias sp.)," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 19, no. 2, pp. 114–121, Sep. 2025, doi: 10.32815/jitika.v19i2.1184.

[18]   A. A. M. Tamim, F. Hasan, F. Emon, M. Haque, M. A. Anwary, and S. Shatabda, "YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment," *Ecol Inform*, p. 101847, Oct. 2022, doi: 10.1016/j.ecoinf.2022.101847.

[19]   Pusaran-KP Foundation, "LAPORAN HASIL SURVEY PERKEMBANGAN USAHA BUDIDAYA IKAN AIR TAWAR DI INDONESIA," Oct. 2025. Accessed: Nov. 20, 2025. [Online].   Available: https://pusarankp.org/wp-content/uploads/2025/10/LAPORAN-HASIL-SURVEY-PERKEMBANGAN-BUDIDAYA-AIR-TAWAR-FINAL.pdf

[20]   E. Putri Silmina, S. Sunardi, and A. Yudhana, "COMPARATIVE ANALYSIS OF YOLO DEEP LEARNING MODEL FOR IMAGE-BASED BEEF FRESHNESS DETECTION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 11, no. 1, pp. 250–265, Aug. 2025, doi: 10.33480/jitk.v11i1.6784.

[21]   A. Yudhana, E. P. Silmina, and Sunardi, "Deteksi Kesegaran Daging Sapi Menggunakan Augmentasi Data Mosaic pada Model YOLOv5sM," *JRST (Jurnal Riset Sains dan Teknologi)*, pp. 63–71, Apr. 2025, doi: 10.30595/jrst.v9i1.24990.