Implementation of YOLOv8-Pose Model for Identification and Estimation of the Length of Skipjack and Tuna Like Species on the Website

Muhammad Fadhilah Tanhi*, Muhammad Iqbal

Department of Marine Science and Technology, Faculty of Fisheries and Marine Science, IPB University

Abstract. Skipjack and tuna species are crucial to Indonesia's capture fisheries sector as they are the largest export commodities. These groups are diverse, and the identification and measurement of their length can be time-consuming due to the abundance of caught fish and their morphological similarities. The objective of this study is to utilize an artificial intelligence algorithm to detect the species and estimate the length of four types of fish: bullet tuna (*Auxis rochei*), black skipjack (*Euthynnus lineatus*), mackerel tuna (*Euthynnus affinis*), and skipjack tuna (*Katsuwonus pelamis*). This algorithm will be implemented through a website. The YOLOv8-Pose deep learning model is employed to identify each fish species and estimate their length by determining keypoints. The dataset used consists of 148 images with rulers and 185 images of the four types of fish. Computer Vision Annotation Tool (CVAT) is used to assist in the labelling of the dataset, enabling the detection of boxes and keypoints. The labelled dataset is trained using Google Collaborator, resulting in the production of two model weights. Both models achieve an accuracy rate of 100%, as well as precision, recall, and an F1-score of 1. The coefficient value between the actual fish length and the detected fish length is 0.8649 or 86.5%, indicating a relationship between the two variables. To facilitate the identification, measurement, and storage of data in CSV format, a website is created using the Streamlit framework. In summary, the models accurately identify the limited number of datasets for *Auxis rochei*, *Euthynnus lineatus*, *Euthynnus affinis*, and *Katsuwonus pelamis*, and can provide estimates of fish length through the website.

1 Introduction

Marine and fisheries resources are of great importance in Indonesia and receive special attention. Fisheries, which are part of the marine sector, are relied upon by the Indonesian people due to their productivity and impact on various aspects of life. According to KKP data, as of 2019, domestic capture fisheries production reached 7.53 million tons, with capture fisheries accounting for 92.68% and the remaining 7.32% coming from land public waters.
The Central Statistics Agency reported that in 2020, tuna production reached 20,907 tons with a transaction value of Rp 457.34 billion, while skipjack fish had a production volume of 20,174 tons with a transaction value of Rp 316.19 billion. The significant production of cob and skipjack fisheries plays an important role in Indonesia's capture fisheries exports. Furthermore, cob and skipjack fisheries are utilized in the development of fresh fish products, processed canned fish, and frozen processed products, enabling the Indonesian capture fisheries sector, particularly tuna and skipjack, to meet international market demands.

Tuna like species in Indonesia has many species such as lisong tuna (Auxis rochei), black cobs (Euthynnus lineatus), white cobs (Euthynnus affinis), and skipjacks (Katsuwonus pelamis). Lisong Tuna (Auxis rochei) which has a small body characteristic compared to other types of cobs with an average length of 35 cm and rounded body shape like a cigar. Characteristics of the Lisong Tuna Body has a bluish back with a sloping pattern on the upper body. Black tuna (Euthynnus lineatus) has a darker color body characteristic than a white cob with a fine black spot above without a line, on the section of black tuna fish has a longer fin than a white cob. White tuna (Euthynnus affinis) has a long gray body and has a gray-colored dorsal fins with the upper edges of darkness with the characteristics of vertical lines on the upper back. Skipjack tuna (Katsuwonus pelamis) has an elongated and gray body. Skipjack tuna has a special characteristic, namely the horizontal line of three to four on the bottom body.

Morphological similarities that are almost the same of tuna or skipjack can cause many errors in the identification of these fish. The identification process carried out by fishermen, the community, and certain parties is carried out manually which is at risk of causing errors in identification and can affect statistical calculations and stock estimation of the species of tuna and skipjack. Fish identification certainly requires expertise from taxonomy, namely "Identification Technique". In practice, the process of identifying each fish species requires a method and supporting equipment and libraries regarding taxonomy. The number of fish taxonomists in Indonesia has a limited amount that causes a small amount of knowledge and related literature in determining the principles of fish identification in Indonesia.

One method commonly employed for fish species identification is the Convolutional Neural Network (CNN/Convnet), a deep learning algorithm derived from Multilayer Perceptron (MLP) architecture. CNN operates by processing two-dimensional data directly from images, thus alleviating the programming burden. By analyzing visual images, CNN can detect and recognize objects within the image. The evolution of CNN methodology has led to the introduction of novel algorithms like YOLO (You Only Look Once), renowned for its rapid detection capabilities by inputting images directly into the network and promptly yielding detection results. The advancement of YOLO has been notable, with the latest iteration, YOLOv8, unveiled on January 10, 2023. YOLOv8 shares its architectural framework with YOLOv5, featuring backbone, head, and neck components, albeit with enhancements over its predecessors. Notably, YOLOv8 integrates a swifter and more accurate Darknet-53 backbone in comparison to the YOLOv7 algorithm. In summary, while both CNN and YOLO methodologies serve as potent tools for fish species identification, YOLO, particularly in its latest version, YOLOv8, stands out for its efficiency and precision, owing to its upgraded architecture and advanced algorithms.
2 Method

2.1. Data collection

The fish dataset was meticulously collected from two distinct locations: TPI Palabuhanratu in West Java and TPI Muara Baru in North Jakarta. Uniformity in image data acquisition was ensured using a smartphone camera with a resolution of 640px × 640px. During the data gathering process, a standardized fish survey board served as the backdrop, supplemented by an artificial ruler calibrated to measure precisely 10 cm. Fish specimens were methodically positioned facing left and perpendicular to the camera to maintain consistent orientation. In this study, two distinct objects were labeled: the four distinct types of fish (lisong tuna, black tuna, white cob, and skipjack) and the ruler. The inclusion of ruler labels was crucial for automatically determining the length of fish specimens, thus providing a calibration factor for translating pixel count to actual length measurements. Furthermore, an ancillary dataset comprising 148 meticulously labeled images of rulers was amassed for the secondary dataset model. Subsequently, both datasets were meticulously divided into distinct training and validation subsets, with 90% of the images allocated to the training dataset and the remaining 10% to the validation dataset. Consequently, the fish dataset comprised 168 training images and 17 validation images, while the ruler dataset contained 133 training images and 14 validation images.

2.2. Data labeling

Dataset labeling or labeling process that aims to give a sign to the image according to the type of fish or objects that you want to recognize in the training process. Dataset labeling is assisted using the Computer Vision Annotation Tool (CVAT) website by uploading all the datasets that you want to label the cvat website. Image labeling is done by creating a project and task on the CVAT website and starting by making points and a box (ground truth box). The next step gives the label name used for the training process, namely Ruler, Auxis Rochhei, Euthynnus lineatus, Euthynnus affinis, and Katsuwonus pelamis.

Fig. 1. Image labelling
2.3. Training and Evaluation

Artificial intelligence is an integrated science in a system to manage data in the form of text, sound, image, and video as a basis for achieving the desired goals [14]. In managing the data it takes training assisted by the Deep Learning YOLOv8-Pose algorithm. The YOLOv8-Pose algorithm performs the task by determining the location, classifying by displaying a detected object box, and providing a detected object keypoint.

Evaluation of the research model uses confusion matrix in seeing the correlation of matrix predictions with actual predictions detected by the model [15]. The confusion matrix method is useful as a measuring medium produced by the YOLOv8-Pose model after the Dataset training process. The confusion matrix method works using the comparison of the data tested and making it four categories, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Confusion Matrix is used to calculate matrix performance to measure the performance of the model that has been drained and produces matrices such as accuracy, precision, and recall and F1-Score.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)
\]

Accuracy (accuracy) is the ratio of true prediction (positive and negative) in the whole data.

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (2)
\]

Precision (precision) is a True Positive (TP) prediction ratio compared to all positive prediction results.

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (3)
\]

Recall is a true positive (TP) ratio compared to positive data as a whole.

\[
F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)
\]

F1-Score is a ratio of the precision value and the weighted average acquisition value [6].

2.4. Extraction of fish length estimation data and ruler

Fish length estimation data is obtained from the key point pinpoint on the dataset. In this study there are 4 key points determined to measure the estimated length of fish, namely points A, B, C, and D. These points are connected to one another to determine the estimated length of fish morphometric. Estimated Morphometric Length of Fish measured is the fork length by measuring the body starting from the front end of the head to the outer tip of the tail fin branch. The calculation of the length from the point to point is also carried out for the ruler detection model, but the ruler only has two coordinates at the left end and the right end of the ruler with the actual length of 10 cm.

Calculating the morphometric length must first determine the length of each connected point. Point A will be connected to point B, point B will be connected to point C, and point C will be connected to point D.
Fig. 2. The relationship between key point points to determine the length of the fish.

The key point is obtained from the prediction of the model that has been trained with YOLOV8-POSE. Each point has the coordinates of X and the coordinate value of Y. To get the length from point A to point B and so on, the formula used is the Pythagorean formula obtained by quadratic the difference between X coordinates and coordinates Y which is then preferred (5).

\[ AB = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} \]  

(5)

Where,

- \( AB = \) Pixel length of point A to point B
- \( x_B = \) coordinating point \( x_B \) axis
- \( x_A = \) coordinating point \( x_A \) axis
- \( y_B = \) coordinating point \( y_B \) axis
- \( y_A = \) coordinating point \( y_A \) axis

Subsequently, after determining the length between each point relative to the fish and ruler, the next step involves aggregating these point lengths to compute the overall length of the fish in pixel values. Notably, the ruler utilized in the images is standardized to a length of 10 cm. Therefore, to convert pixel values to centimeters, the number of pixels representing the length of the fish is compared to the number of pixels representing the length of the ruler, and the result is multiplied by 10 cm (6).

\[ l = \frac{a}{b} \times 10cm \]  

(6)

Where,

- \( l = \) Actual Fish Prediction length (cm)
- \( a = \) Length of Fish Pixel Prediction
- \( b = \) Ruler Pixel Prediction Length

The correlation analysis test was carried out to see the relationship between fish length detected with the length of the actual fish by measuring the level of strength of the relationship between two variables by looking at the correlation coefficient [16]. The correlation and linear regression coefficient equation used is as follows:

\[ y = ax + b \]  

(7)

Where,

- \( x = \) value on the x axis (real data)
- \( y = \) value on the y axis (prediction results)

2.5. Website creation

The final stage involves the development of a website serving as a user interface for interacting with the deep learning model. The website is created utilizing Python 3.10.11 programming language, supported by a range of modules, libraries, and Python frameworks including Pathlib, Pills, Math, Datetime, Pandas, Ultralytics, and Streamlit. These tools are leveraged to streamline and expedite the website development process. Notably, Streamlit, a Python-based module, simplifies the creation of sections, features, and widgets on the website. The website adopts a one-page configuration, featuring a sidebar for adjusting the model confidence level, a button for image upload, and another button for initiating image identification. This layout ensures user-friendly navigation and seamless interaction with the deep learning model.
On the main page displayed the equipment needed, imagery instructions, uploaded photos, photos of the identification of ruler, photos of fish identification, and tables that record fish species and the length of fish predicted in cm. The data will then be stored in CSV format files containing the clock date, fish name, and also estimated fish length.

The website's backend comprises computations derived from the two deep learning models developed: the ruler detection model and the fish detection model. The ruler detection model serves as the fundamental reference for converting pixel lengths to actual lengths. This involves comparing the pixel length obtained from the fish detection model with the known pixel length of the ruler, thereby yielding the actual length value in centimeters (cm). These calculations are executed using Python, with assistance from the math module inherent in Python 3.10.

3 Result and discussion

3.1. Training result

The training process with the YOLOv8-Pose algorithm produces two weight or weight model values which are weights to detect ruler and weight to detect tuna and skipjack fish. The results of the training process produced an evaluation matrix including precision, recall, training loss, map50 and map50-95 from each box or pose of the YOLOv8-Pose algorithm. Precision is a positive prediction part compared to the whole predicted positive results, recall or other designations Sensitivity is a part of the actual positive prediction that is correctly identified by the weight of the model [17].

Loss is a value that represents the sum of errors in the model, the loss value sees the performance of how good the performance of the model has been made [18]. The Training Loss graph will show the process of improving the error value during the training process. Training loss is used to maximize the performance of the algorithm rather than Machine Learning by looking at the performance of the model at the training stage and the validation of the model. Training loss can indicate how bad or how well a model after iteration of optimizing the model [17]. To find out changes in the value of training loss can be seen through changes in values in each iteration. Iteration is the repetition of tasks performed to achieve an outcome [19].

![Fig. 3. (a) Train/Box_loss weight ruler, (b) train/pose_loss weight ruler](image_url)
The weight of ruler model obtained from the Yolov8-Pose training validation test on epoch 300 produced precision values of 0.99662 and the recall 1 value on the task box or on the task pose. There are two charts of training loss namely training loss on the task box for identification of bounding boxes and training loss on the task pose for determining the key point. Both graphs tend to decrease close to point 0 as epoch increases. In the training loss produced by the Box Training Loss Value quickly is below the value of 1 when iteration or epoch 2 and when epoch 300 loss value is at a value of 0.18429 while the training loss graph in the pose tends to experience the value of the value and is only at the loss value and is only at the loss value Under 1 when Epoch 14. When Epoch 300 Loss value is valued at 0.00869.

![Fig. 4. (a) Train/Box_loss weight fish, (b) train/pose_loss weight fish](image)

The weight of the fish model obtained from the Yolov8-Pose training validation test on epoch 1000 produced precision values of 0.99403 and the recall 1 value on the task box while on the task pose with epoch 1000 produced precision values of 0.80534 and recall of 0.81042. This shows that the fish weight model for task poses has no better capabilities than task boxes in predicting results based on precision and recall values. In Figure 4 it is known that the two charts of training loss decreased during the addition of epoch/iteration. In Loss Training Task Box loss value decreases below the number 1 during the 6th loop and at the last repetition of loss values of 0.22289 while in the task pose, the value of the loss training with Epoch 1000 gets a loss value of 0.70513. A large enough loss pose value can affect the performance of the model including the results of precision and recall. Small loss pose can easily determine the key point and determination of the estimated fish length using the key point will also be increasingly in accordance with the actual. However, if the loss key point value is greater, it can cause the detection of the key point to occur an error and does not match the desired point position. The magnitude of the loss key point value of Yolov8-post training can be caused by the GPU capacity on a lack of Google collaborator and inappropriate training configuration.

### 3.2. Model evaluation

Model evaluation is used to see the performance of the two models that have been made. Model performance measurements using confusion matrix. The results of the value of confusion matrix will be used as a basis for calculating the accuracy of the model based on precision, recall, accuracy, and F1-score.
Table 1. The results of the confusion matrix value in the weight of the ruler and fish model

<table>
<thead>
<tr>
<th>Weight</th>
<th>True Positive (TP)</th>
<th>False Positive (FP)</th>
<th>False Negative (FN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruler</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fish</td>
<td>75</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Based on the validation using a ruler and dataset of the fish produced the confusion matrix value as in Table 5. The weighting weight of the model can identify all data properly worth 15 rulers, as well as the weight of the model in fish that succeeded in identifying all species well without an identification of 75 species. This value is used as a basis for calculating the model evaluation, namely accuracy, precision, recall, and F1-Score (Table 2).

Table 2. Percentage performance ruler model weight and fish model weight

<table>
<thead>
<tr>
<th>Model Weight</th>
<th>Accuracy (%)</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruler</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fish</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

3.3. Estimated fish length

Determination of the estimation of the length of the fish using 17 dataset photos that are the same as the dataset photo of fish validation. The dataset consists of 3 datasets of lisong tuna (*Auxis rochei*) with 8 fish in each dataset, 4 black datasets (*Euthynnus lineatus*) with a total of 4 fish in each dataset, 5 datasets white tuna (*Euthynnus affinis*) with a total of 3 fish in each dataset, and 4 dataset skipjack (*Katsuwonus pelamis*) with a total of 4 fish in each dataset. Then the total number of fish used amounted to 75.
Figure 5 illustrates the ratio between the length estimates generated by the actual model and the lengths of Lisong Tuna (*Auxis rochei*), Black Cob (*Euthynnus lineatus*), White Tuna (*Euthynnus affinis*), and Skipjack (*Katsuwonus pelamis*). Lisong Tuna exhibits a substantial variation in length estimates, ranging from 5.07 to 8.14 cm. In contrast, the length difference for Black Tuna (*Euthynnus lineatus*) is relatively minimal, ranging from 0.00 to 4.32 cm. For White Tuna (*Euthynnus affinis*), the difference between estimated and actual lengths ranges from 2.86 to 5.95 cm. Finally, Skipjack fish (*Katsuwonus pelamis*) demonstrate a slight discrepancy, with differences ranging from 0.00 to 2.36 cm. These variations in length among species may stem from errors in weight assignment during key point identification, leading to inaccuracies in fish length estimation. Model weight discrepancies in key point determination may arise from differences in dataset sizes for each species, resulting in unequal weight distribution and subsequently affecting length measurements performed by the model.

Based on Figure 6, the graph of the actual length of the fish and the length of the fish detection gets a coefficient value of 0.8649 or 86.5%. Relationships between two variables can be said to be perfect if the relationship has a correlation coefficient value = 1 or = -1 [20]. This means that there is a relationship between the actual data results and the estimated data obtained.

### 3.4. Website implementation

Website uses streamlit framework provided on Github. Streamlit runs a website on localhost on a laptop and can classify species, determine keypoint points, estimate the length of the fish, and record the date of detection and can store the data into the CSV (Comma-Separated Value) file that will be stored in local storage. The websites created in this study have one main page that has titles, equipment needed, shooting instructions, and the presence of a menu sidebar used to configure the Deep Learning model, upload the file to be detected, and to do the detection process by pressing the 'detect image'.
In the Deep Learning Model configuration, it can later be used to increase the number of models apart from the 'cob and skipjack' model by adding options to the deep learning model that will be used. The current Deep Learning model 'cob and skipjack' refers to the previous training files that have been obtained. In the configuration settings of the training model the results of confidence can be adjusted by shifting on the slidebul that is available with the starting point on confidence 70 and can be arranged from 25 to 100. For image configuration there is only one source/source option that is image. When only the source of the image is chosen, a place to upload an image with a limit of 200 MB which is a limit of streamlit with an acceptable format in the form of JPG, JPEG, PNG, BMP, and Webp. When the website has not received the best image, the website will display examples of shooting and examples of identification results carried out by the website (Figure 16 and Figure 17). Then there is the 'detect image' button to carry out the process of identifying and estimating the length measurement when the website has received an image upload.

When the website receives an uploaded image file, the website will display the image by changing the 'example of shooting' into 'the best image' with the image uploaded by the user but the identification results will remain empty because the user has not pressed the 'detect image' button for Perform the process of identifying and measurement of estimated fish length. When the user presses the 'detect image', the website will issue images of identification and classification of species, determining the key point, and measurement of fish length estimation. Data dates, fish names, and the length of the estimated fish will be stored in CSV format files.
The detect image process occurs on the backend rather than a website. After pressing the detect image, the website will process the uploaded image by detecting a ruler using a ruler model by determining the two key point points and looking for the length of the two key point points using equation 5. The length will be a reference of the image pixel with the original 10 cm. Furthermore, the classification of fish and key point of each fish in the photo uses a fish model so that the pixel value is obtained from the length of each fish. The next step, the backend on the website will make a comparison using equation 6 and the estimated value of fish is detected in each fish. These results are displayed in the table on the website and are directly stored to CSV format files.

4 Conclusion

The website based on the Deep Learning model successfully detects each fish species and provides accurate estimations of their lengths, including Lisong (Auxis Rochhei), Black Cob (Euthynnus lineatus), White Tuna (Euthynnus affinis), and Skipjack (Katsuwonus pelamis). Both Deep Learning models achieve a remarkable accuracy rate of 100%, with precision, recall, and F1-Score values all reaching 1. Additionally, the coefficient value between the actual fish length and the detected fish length is calculated at 0.8649, indicating a strong correlation of 86.5% between the two variables.
5 Suggestion

Suggestions for further research are to multiply the dataset of other species not only limited to cobs and skipjacks. As well as the addition of features on the website to identify and measure the estimated fish length in different forms such as video or real-time from the camera.

References

10. Iahiyah, S, Nilogiri, A, Jurnal Sistem Dan Teknologi Informasi Indonesia, 3(2), 49-56 (2018)