ILKOM Jurnal Ilmiah Vol. 15, No. 2, August 2023, pp. 317-325
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Combination of YOLOv3 Algorithm and Blob Detection Technique in Calculating Nile Tilapia Seeds

Diana Tri Susetianingtias [a,1]; Eka Patriya [a,2,*]; Rini Arianty [a,3]
[a] *Gunadarma University, Margonda Raya Street 100, Pondok Cina Depok, Indonesia*
[1] *diants@staff.gunadarma.ac.id;* [2] *ekapatriya@staff.gunadarma.ac.id;* [3] *rinia@staff.gunadarma.ac.id*
* *Corresponding author*

Article history: Received February 23, 2023; Revised June 09, 2023; Accepted June 27, 2023; Available online August 16, 2023

Abstract

Baby Fish must be counted accurately so it will not cause any loss, especially for fish seeds or fingerlings that have a small size. Generally, people still use conventional counting methods that produce low accuracy values. This research will make a Nila Baby Fish fingerlings counter program using the YOLOv3 algorithm and Blobb detection technique. The annotation data process will use labelImg, and the dataset training will use Google COLABoratory with the Darknet framework in an online environment. Images predicted in this program will be called and detected with an object detector. The object with a confidence score of more than 0.3 will be converted into a blob. The blob value will be forwarded to the output layer for scaling the bounding box objects. The output of this program is the predicted image, blob value, prediction time, and the number of Nila seeds. The model performance is evaluated using a confusion matrix and obtained a 98.87% for accuracy score.

Keywords: Baby fish; Blob; Nila; Bounding Box; YOLOv3.

Introduction

Indonesia is an agricultural country which means that most of population in Indonesian works in agriculture sector. The agricultural sector can be divided into 5 (five) sub-sectors including food crops, plantations, forestry, animal husbandry, and fisheries. One of the agricultural sub-sectors, the fisheries sector [1], has good business potential from increased national fish consumption [2] in the last decade. The fish consumption rate in 2021 reached 55.37 kg/capita, while in 2011, it was only 36.66 kg/capita. From these data, the national consumption rate has increased by around 69.17% over the last ten years [3]. One way to support increased fishery production is to improve good aquaculture technology by overcoming the obstacles experienced by fish cultivators, namely counting fish seeds that still use manual or measuring methods [4]. The method of counting fish seeds manually or using measurements takes a long time, and the results are inaccurate [5] when this method is used to count seeds in small quantities. Apart from the manual method using measurements, another method that is quite accurate in calculating fish seeds is measuring cups to calculate the number of large fish seeds [6]. This method is faster and more effective, but the level of accuracy is uncertain. The third method uses a scale similar to the measuring method using a glass, except that the measurement is converted to weight, where for speed and efficiency, this method is quite good, but the accuracy is still uncertain.

Barriers in calculating fish seeds can be overcome by carrying out automatic calculations by implementing the use of the object detection method [7] seeds in water using the You Only Look Once (YOLO) algorithm, which was developed to detect an object in real-time [8]. The YOLO algorithm is implemented [9] to detect an object by changing the image size to 448 × 448 pixels, then running the single-convolutional network in the seed image and setting the threshold value of the confidence model value obtained. Research [9] implemented the YOLOv3 algorithm to design an automatic counting program for shrimp larvae. The dataset consists of 325 images as training data and 99 test data. Research [10] performs sorting, resizing, and labeling the training data to improve the resulting accuracy. This study obtained an accuracy of 96.10% and an RMSE of 2.60 at a density of less than 50 larvae and an accuracy of 95.72% and an RMSE of 4.19 at a density of greater than 50 larvae. Research [11] applied the YOLO algorithm to design a program for calculating carp seeds using a webcam as a data input device. The dataset used is divided into three classes based on the size of the carp seeds, namely the first class of thumb-sized carp seeds (2-3 cm), the second class of gas-sized carp seeds (3-4 cm), and the third class is the finger-sized carp seeds (3-4 cm), and the third class is the carp seeds the size of a razor blade (4-5 cm). The accuracy obtained from dividing the three classes using 15 carp seeds was 85.33% for the first, 82.67% for the second, and 84% for the third. Two accuracy calculation techniques are used to count the number of carp seeds without differentiating the size of the carp seeds, namely the mode technique and the one detection technique. The results obtained using the mode technique were 97.62% for 29 seeds, 98.77% for 49 seeds, and 100%

http://dx.doi.org/10.33096/ilkom.v15i2.1634.317-325

for 100 seeds. The single detection technique results were 95.83% for 29 seedlings, 95.68% for 49 seedlings, and 97.60% for 100 seedlings. Research [12] used a combination of the DIRT Olive Fruit Fly dataset and the YOLOv4 algorithm as the basis for modifying the Original YOLOv4 with the DIRT method with a precision of 97%. Research [13] compared the YOLOv2 and YOLOv3 algorithms to find out the difference in the performance of the two versions using a video of a human being in an elevator as an input object. The result of the comparison between the YOLOv2 and YOLOv3 algorithms is that the confidence value obtained by YOLOv3 is 0.9, while YOLOv2 only gets a value of 0.6 with a confidence threshold set at 0.3.

This study implements the YOLOv3 algorithm and the BLOB detection technique in automatic calculations of Nile Tilapia seeds as a freshwater fish species, one of the most common and easily cultivated fish species. Freshwater fish can be cultivated using simple technology, lower costs [14], and the availability of easy seeds to obtain when compared to sea fish [15]. The dataset of Nile Tilapia seeds in this study results from taking seed images directly using a smartphone camera with various sizes of seeds. The resulting dataset consists of 87 data which are categorized into one class. The dataset has been labeled automatically at the training stage using the Darknet framework on Google COLAB. The blob detection technique in this study is implemented using a forward pass using the blob obtained through the output layer to calculate the time needed to detect one seed image. Implementing this algorithm is expected to speed up the time of counting fish seeds and be accurate in providing optimal results when counting fish seeds to increase fresh fish production in particular.

Method

This study uses the YOLOv3 algorithm with blob detection techniques to calculate Nile Tilapia seeds automatically. The research began with data acquisition to obtain images of Nile Tilapia seeds. The data acquisition results are annotated and trained to get the Nile Tilapia dataset ready to be used for predictions. The data training and test data results are called into the program for detecting and counting Nile Tilapia seeds.

A. Image Acquisition of Nile Tilapia Seeds

The first stage is acquiring digital image data for Nile Tilapia seeds, as seen in **Figure 1**. The data is an image of Nile Tilapia seeds taken using a smartphone camera with 48MP specifications at approximately 10 cm from the container. Nile Tilapia seeds were placed in a 21 × 36 cm container with a water level of approximately 4 cm, and pictures were taken with variations in the position and number of seeds. Taking pictures using a flashlight on a smartphone camera so that there are no shadows in the image.

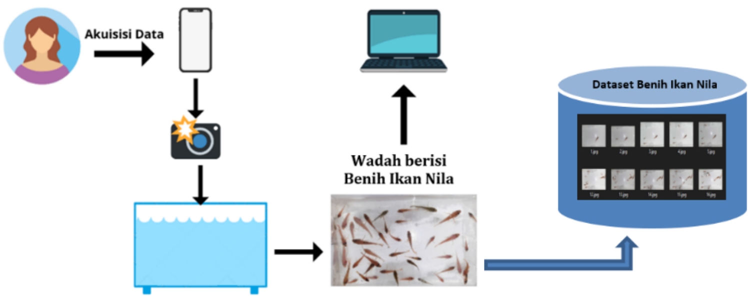


Figure 1. Data Acquisition Process

At this stage, data annotation and training will be carried out so that the dataset used for prediction has good quality. Annotation of seed images is done using the labelImg program. The image will be divided into four parts to make the image annotation process clearer and more detailed, especially for images with many objects.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 2. Darknet-53 Network Layer Model [18]

The divided or split image will be opened in the labelImg program, then annotated by giving a box to each object [16]. The training stage in this study utilizes configuration files from Darknet53 [17] and uses transfer learning techniques. The transfer learning technique is a technique that uses a pre-trained model to classify a new dataset without having to do data training from scratch. This transfer learning process utilizes configuration files from Darknet, namely cfg files and pre-trained weights [18]. Darknet 53 in this study uses a convolution layer size of 3×3 in a row, as seen in **Figure 2.**

B. Detection of Nile Tilapia Seeds using YOLOv3 and Blob Detection

Following the annotation process and data training on the seed dataset is to implement and recognize Nile Tilapia seed objects using YOLOv3 and Blob Detection by:

1. Read the input image in the form of a Nile Tilapia seed image.
2. Divide the input image into $S \times S$ grids, where each grid cell has only one object with a fixed number of bounding boxes using the YOLO mathematical equation [19]. Each bounding box has one confidence value and x, y, w, and h values which represent the coordinates and bounding box size of the objects.

$$S \times S \times (B \times 5 + C) \quad (1)$$

The grid cell size value is represented by $S \times S$, B is used to represent the number of bounding boxes that will be created in each cell, and C represents the number of object classes that must be detected and predicted. The B value is multiplied by 5 because a bounding box has 5 values that need to be stored, namely x, y, w, h, and confidence values. The x and y values are the midpoints of the resulting bounding box, while the w or width and h or height values are the width and height of the bounding box. Each bounding box has an objectness score which is predicted using logistic regression as can be seen in **Figure 2**. The objectness score will have a value of 1 if there is overlap between the bounding box prior and the ground truth object more than other bounding box priors (having the highest IOU value). The system will only set one bounding box prior for each ground truth object, if there are no bounding boxes, then there will be no loss for coordinate and class predictions, only objectness [20].

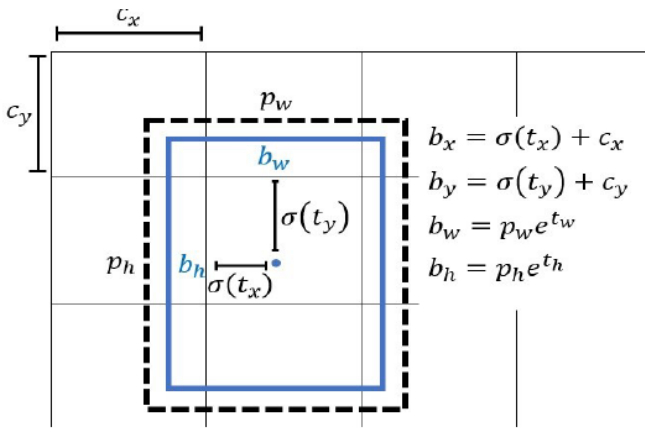
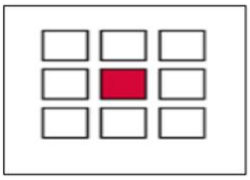


Figure 3. Bounding Box with Objectness Score

The process of getting the bounding box for each object to get all the detected seed object information will be extracted and displayed on the screen. Repeat for each output layer and loop for each detection function in the output by extracting the classID and confidence. The function of confidence > CONFIDENCE_THRESHOLD is to ensure that objects that enter the loop or will be detected are objects that have a confidence value of more than 0.3. Coordinates and dimensions are extracted from the bounding box by returning the bounding box coordinates in the form centerX, centerY, width, and height. Store back all the values that have been obtained in each variable.

3. Determine the confidence threshold to detect objects that have a low prediction probability and load object detectors from the Darknet. The confidence threshold value used in this study is 0.3, which means that objects with a confidence value of less than 0.3 will be missed. Determination of confidence threshold.
4. Detect Nile Tilapia seed objects by implementing the Blob Detection Technique. This technique is done by comparing the same color from a set of pixels with the object's background color. The blob counting process can be done by performing a pixel adjacency analysis. Pixels are called neighboring if a pixel has a distance of 1 with the original pixel. The blob calculation utilizes the 8-neighbors pixel relation as can be seen in **Figure 4**. The object mapping process will trace each pixel in each existing row and label pixels that have a color value other than black (RGB = 0, 0, 0). Every pixel that has an 8-neighbors relationship will be given the same label.



P (x-1, y-1)	P(x, y-1)	P (x+1, y-1)
P (x-1, y)	P (x,y)	P(x+1, y)
P (x-1, y+1)	P(x, y+1)	P (x+1, y+1)

Figure 4. 8-neighbor relations [19]

5. Implement a forward pass using the blob that has been obtained through the output layer then calculate the time needed to detect one image.
6. Minimizing stacked bounding boxes on an object and selecting the best bounding box for an object using non-maxima suppression (NMS). NMS will consider two things, namely the objectivity score given by the model and the IoU from the bounding box. The following pseudocode minimizes the use of bounding boxes by using the parameters boxes, confidences, and CONFIDENCE_THRESHOLD.
7. Provide an information label containing the number of Nile Tilapia seeds, confidence values, and bounding boxes.

C. Darknet file configuration for Nile Tilapia Seed Training Dataset

At the training stage using Google COLAB, there is a Darknet file configuration to suit research needs. Data training utilizes the GPU on Google COLAB so that the training process can be carried out faster, therefore the GPU and CUDNN functions are changed to 1 so that they can be used as shown in **Table 1.**

Table 1. Configuration on Darknet53

Configuration Type	Information
Load Model	Darknet53
Load Weight	YOLOv3
OPENCV	1
GPU	1
CUDNN	1

Configuration is also done in the custom.cfg file to determine the maximum batch value, steps value, number of classes, and number of filters. The maximum value of the batch or max_batches is the iteration limit for training. Equation (2) is used to calculate the max_batches value, which means that each class can iterate at least 2000 times. In this study the max_batches value used was 4000 so that the training results obtained could be better.

$$max\_batches = \text{ number of classes } \times 2000 \qquad (2)$$

Equation (3) is the formula for determining the value of steps

$$steps = (80\%max\_batches), (90\%max\_batches) \qquad (3)$$

Equation (4) is used to determine the number of filters

$$filter = (j\text{number of classes } + 5) \times 3 \qquad (4)$$

Table 2 is a summary of the configuration results in the custom.cfg file. The batch value is used to determine the number of images to be processed before the network weight is updated. The subdivision value is used to determine the number of batches to be processed.

Table 2. Configuration in the custom.cfg file

Configuration Type	Information
batch	64
subdivision	16
Max_batches	4000
steps	3200, 3600
classes	1
filters	18

D. YOLOv3 Performance Parameters and Blob Detection Techniques in Counting Nile Tilapia Seeds

This stage is used to determine whether the implemented model is successful in detecting seeds. The computation is performed by comparing the actual data from manual calculations to the predicted outcomes. Manual computations are performed by drawing dots on the seed image, as shown in **Figure 5**.

Figure 5. Example of Manual Calculation Results for Nile Tilapia Seeds

The calculation of confusion matrix is performed on each test data by comparing the manually calculated seed images with the predicted seed images, then analyzing detected and undetected objects. **Figure 6** compares the manual calculation image and the expected result image. The difference in the number of seeds between manual calculation and prediction can be seen in **Figure 6** (a) 16 seeds, and (b) 15 seeds respectively.

(a) (b)

Figure 6. Example of Comparison of Manual Calculations and Predictions
(a) Manual Calculation, (b) Prediction Results

In **Figure 7**, two different bounding box colors are blue and red. The blue bounding box is considered a True Positive (TP), which means that the Nile Tilapia seed object (actual positive) is detected as an object (positive prediction) by the system. The red bounding box is considered a False Negative (FN), which means that the fish seed object (positive actual) is not detected as an object (negative prediction). From the analysis of **Figure 7**, the confusion matrix values are TP = 15, FN = 1, FP = 0, and TN = 0. This calculation is performed on all test data to obtain the overall confusion matrix value. The confusion matrix values obtained from all test data can be used to calculate accuracy, precision, recall, and RMSE.

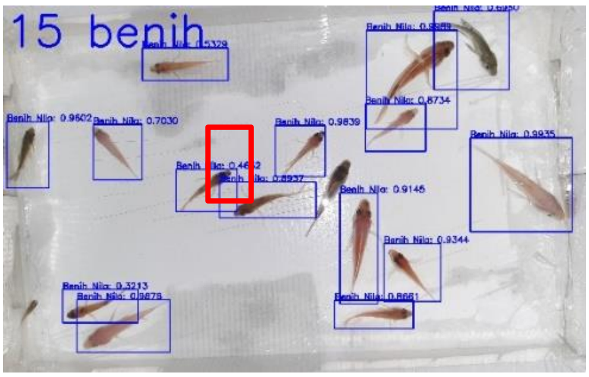


Figure 7. Example of Confusion Matrix Value Analysis on Detected Nile Tilapia Seeds

Results and Discussion

The study used a dataset of Nile Tilapia seeds which were acquired directly using a smartphone camera. Figure 8 is a display of objects that are being annotated by giving a RectBox to each object which will later be detected as a bounding box.

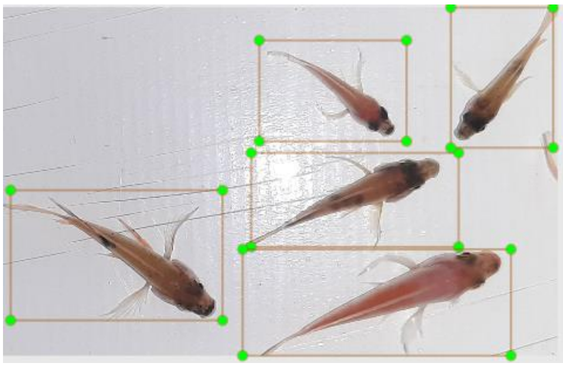


Figure 8. Example of a seed image being annotated

Figure 9 is the result of object annotations performed using the labeling program. The result of this object annotation has two formats, namely JPG and txt.

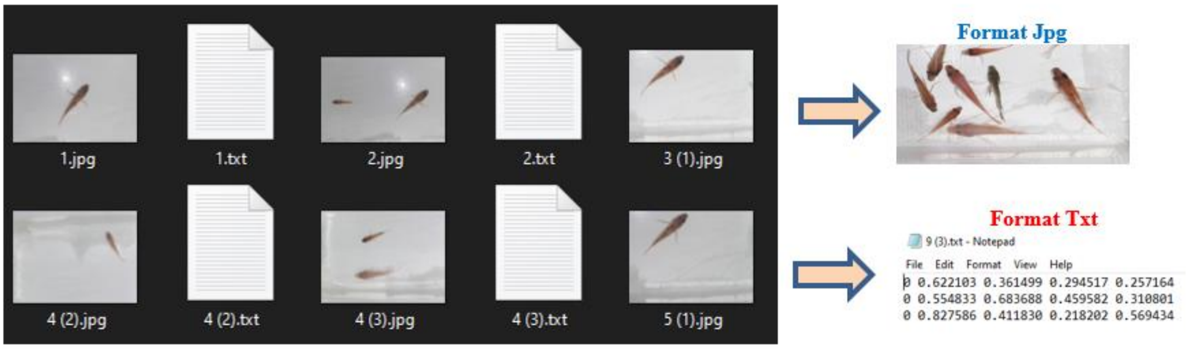


Figure 9. Example of Dataset Annotation Results

JPG images are the results of annotations that have previously been divided or split so that images can be annotated more clearly and in depth in the form of images, and txt files contain object classID, object coordinates, bounding box width, and bounding box height. All classIDs are 0 because there is only 1 class to predict. The maximum batch in this training has been set at 4000, thus the software stores the results of the training data until the last weight every 1000 iterations. Figure 10 shows the identified seeds' results, including the number of seeds detected, the bounding box, and the confidence value.

Figure 10. Detected Seed Yield

The model performance results were obtained from the use of the average formula, confusion matrix, accuracy, precision, recall, and RMSE, as can be seen in Table 3, which is the average result of the time required to perform one image detection of Nile Tilapia seed, which is 0.57 seconds.

Table 3. Seed Prediction Time

Image	Number of Seeds	Prediction Time
1	1	0.710100
2	2	0.739972
...	…	…
89	44	0.561458
90	41	0.528547
Average		**0.571058**

The accuracy value is obtained from the value of the confusion matrix per record, which then calculates the average overall accuracy. **Table 4** is the result of calculating the accuracy per record, where in the last row of the table, it can be seen that the average accuracy of the Nile Tilapia seed counter program reaches an accuracy of 90.07%.

Table 4. Results of Nile Tilapia Seed Detection Accuracy

Image	TP	FN	FP	TN	Accuracy (%)
1	1	0	0	0	100.00
2	2	0	0	0	100.00
…	…	…	…	…	…
89	44	12	0	0	78.57
90	41	12	0	0	77.36
Average Accuracy					**90.07**

Visualization of Nile Tilapia seed detection accuracy can be seen in the blue line graph of **Figure 11**. The blue line represents the accuracy value of each prediction, while the green line represents the actual data or the actual number of seeds. From **Figure 11** it can be seen that the accuracy value reaches 100% for the number of seeds 1 to 11, then the accuracy value drops irregularly. The decrease in accuracy is caused by objects that overlap or are too close together so that the program is not able to detect these objects or detect them as one object.

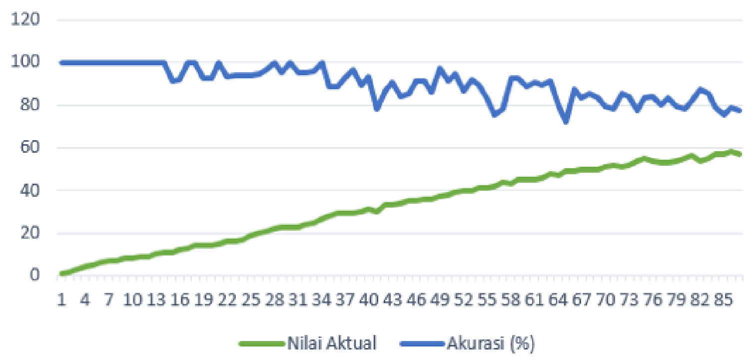


Figure 11. 9 Graph of Nile Tilapia Seed Detection Accuracy

The average time needed by the program to make predictions is 0.57 seconds, the prediction accuracy obtained is 90.07%, the precision value obtained is 1, the recall value obtained is 0.87, and the RMSE value obtained is 7.42 as can be seen in **Table 5**.

Table 5. Summary Results of Model Performance Calculations in Detecting Nile Tilapia Seeds

Calculation Type	Results
Average prediction time	0.57 second
Accuracy	90.07 %
Precision	1
Recall	0.87
RMSE	7.42

Conclusion

Based on the trial results of calculating the number of Nile Tilapia seeds using the YOLOv3 algorithm with the Blob detection technique, several conclusions can be drawn, including the data annotation process using the labelImg program produces several JPG format images and txt files that are ready to be trained. Data training is done online using the Darknet53 framework, producing 6 custom weight files. The custom weight used for prediction is final_custom_weight. The program for calculating the Nile Tilapia seeds using the YOLOv3 algorithm and the blob detection technique works well. It can produce predicted images, blob values, predicted time, and the number of detected seeds. Calculating model performance using the confusion matrix produces an accuracy value of 90.07%, a precision of 1, a recall of 0.87, and an RMSE of 7.42.

In future research, it is expected to add a dataset pre-processing stage with the pixel neighbor technique so that seeds that accumulate or are close together can be detected and can improve the accuracy of the program. The data annotation process in advanced research can use the polygonal segmentation annotation type so that seed objects can be detected in more detail and more accurately, especially in several seeds where one seed object is stacked with other seeds.

Reference

[1] L. Grace and R. Van Anrooy, *Guidelines for micro-finance and credit services in support of small-scale fisheries in Asia. A handbook for finance and fisheries stakeholders. In support of the implementation of the Voluntary Guidelines for Securing Sustainable Small-Scale Fisheries in*. 2019.

[2] K. K. dan Perikanan, *KKP (2019) Laporan Tahunan*. 2019.

[3] K. K. B. Kemaritiman, *Kemenko Maritim (2018) Laporan Kinerja*. .

[4] E. A. Awalludin, W. N. A. Wan Muhammad, T. N. T. Arsad, and W. N. J. H. Wan Yussof, "Fish Larvae Counting System Using Image Processing Techniques," *J. Phys. Conf. Ser.*, vol. 1529, no. 5, p. 052040, May 2020, doi: 10.1088/1742-6596/1529/5/052040.

[5] S. M. D. Lainez and D. B. Gonzales, "Automated Fingerlings Counting Using Convolutional Neural Network," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, Feb. 2019, pp. 67–72, doi: 10.1109/CCOMS.2019.8821746.

[6] M. Solahudin, W. Slamet, and A. S. Dwi, "Vaname (Litopenaeus vannamei) Shrimp Fry Counting Based on Image Processing Method," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 147, p. 012014, May 2018, doi: 10.1088/1755-1315/147/1/012014.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[8] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, doi: www.fao.org/3/ca5128en/CA5128EN.pdf.

[9] C. Geraldy and C. Lubis, "Pendeteksian dan Pengenalan Jenis Mobil Menggunakan Algoritma You Only Look Once dan Convolutional Neural Network," *J. Ilmu Komput. dan Sist. Inf.*, vol. 8, no. 2, p. 197, Aug. 2020, doi: 10.24912/jiksi.v8i2.11495.

[10] S. Armalivia, "Perhitungan Otomatis Larva Udang Menggunakan Metode YOLO," *Undergrad. Thesis*, 2021.

[11] A. R. Arganata, S. T. Rasmana, and W. I. Kusumawati, "Analisis Perhitungan Bibit Ikan Gurame Menggunakan Webcam Dengan Metode YOLO(You Only Look Once)," *J. Control Netw. Syst.*, vol. 5, no. 1, pp. 119–125, 2016.

[12] N. Mamdouh and A. Khattab, "YOLO-Based Deep Learning Framework for Olive Fruit Fly Detection and Counting," *IEEE Access*, vol. 9, pp. 84252–84262, 2021, doi: 10.1109/ACCESS.2021.3088075.

[13] B. P. . Pamungkas, N. Nugroho, and F. T. Anggraeny, "Deteksi Dan Menghitung Manusia Menggunakan YOLO-CNN," *J. Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 67–75, 2021.

[14] E. K. Galappaththi, S. T. Ichien, A. A. Hyman, C. J. Aubrac, and J. D. Ford, "Climate change adaptation in aquaculture," *Rev. Aquac.*, vol. 12, no. 4, pp. 2160–2176, Nov. 2020, doi: 10.1111/raq.12427.

[15] O. O. Ahmed and O. O. Solomon, "Effects of Catfish (Clarias gariepinus) Brood-stocks Egg Combination on Hatchability and Survival of Fish Larvae," *J. Aquac. Res. Dev.*, vol. s2, 2017, doi: 10.4172/2155-9546.S2-014.

[16] L. Ambalina, *What is Image Annotation? – An Intro to 5 Image Annotation Services.* 2019.

[17] J. Redmon and A. Farhadi, "Sequence Classification Using Decision Trees," 2007, pp. 187–213.

[18] M. F. İnkaya and H. Gurkan, "A YOLOv3 Based Smart City Application For Children's Playgrounds," *J. Innov. Sci. Eng.*, Dec. 2020, doi: 10.38088/jise.813664.

[19] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[20] M. Guerrieri and G. Parla, “Deep Learning and YOLOv3 Systems for Automatic Traffic Data Measurement by Moving Car Observer Technique,” *Infrastructures*, vol. 6, no. 9, p. 134, Sep. 2021, doi: 10.3390/infrastructures6090134.