LONTAR KOMPUTER VOL. 11, NO. 3 DECEMBER 2020 p-ISSN 2088-1541
DOI : 10.24843/LKJITI.2020.v11.i03.p03 e-ISSN 2541-5832
Accredited B by RISTEKDIKTI Decree No. 51/E/KPT/2017

Fish Species Recognition with Faster R-CNN Inception-v2 using QUT FISH Dataset

Yonatan Adiwinata[a1], Akane Sasaoka[b2], I Putu Agung Bayupati[a3], Oka Sudana[a4]

[a]Department of Information Technology, Faculty of Engineering, Udayana University,
Jl. Raya Kampus Unud Bukit Jimbaran, Bali, Indonesia
[1]yonatanadiwinata@student.unud.ac.id (Corresponding author)
[3]bayuhelix@yahoo.com
[4]agungokas@unud.ac.id

[b]Electrical Engineering and Computer Science, Kanazawa University
Kanazawa, Ishikawa, Japan
[2]acannie0530@gmail.com

Abstract

Fish species conservation had a big impact on the natural ecosystems balanced. The existence of efficient technology in identifying fish species could help fish conservation. The most recent research related to was a classification of fish species using the Deep Learning method. Most of the deep learning methods used were Convolutional Layer or Convolutional Neural Network (CNN). This research experimented with using object detection method based on deep learning like Faster R-CNN, which possible to recognize the species of fish inside of the image without more image preprocessing. This research aimed to know the performance of the Faster R-CNN method against other object detection methods like SSD in fish species detection. The fish dataset used in the research reference was QUT FISH Dataset. The accuracy of the Faster R-CNN reached 80.4%, far above the accuracy of the Single Shot Detector (SSD) Model with an accuracy of 49.2%.

Keywords: *Fish Species Recognition, Object Detection, Faster R-CNN, QUT FISH Dataset, Deep Learning*

1. Introduction

Ocean makes up two-thirds of the earth's surface. Ocean ecosystems have an important role in the balance of nature, with a variety of living things that live in it, like fishes. More than 22,000 species of fishes make up nearly half of the total 55,000 species of vertebrates living on earth [1]. The development of technology related to the cultivation of fish species was very important for the preservation and protection of marine ecosystems because fish was an important factor in the marine ecosystem. The existence of efficient technology in fish species recognition could help the fish cultivation process because the cultivation method for each fish was not always the same. Fish species were identified through manual observation by humans in the past, which required humans to study various fish characteristics in order to recognize the fish species, and recently the fish species recognition could be done by utilizing artificial intelligence technology.

The latest research that has been done related to the fish species classification was using the Deep Learning method. The deep learning method that was used currently was the Convolutional Layer or Convolutional Neural Network (CNN) [1]–[3]. The classification method in the research references needs background removal preprocessing to recognize the fish species inside the image. This research experimented with using object detection method based on deep learning like Faster R-CNN, which possible to recognize the species of fish inside of the image without more image preprocessing (fish species detection).

The main method used in this research was the Faster R-CNN method with Inception-v2 architecture. Single Shot Detector (SSD) is also used in this research as a compliment. This research aimed to know the performance of the Faster R-CNN method against other object detection methods like SSD in fish species detection. Faster R-CNN was chosen because Faster R-CNN was a method that popular recently, and it had a great performance in object detection, which better than other basic object detection methods like SSD [4] and Yolo-V3 [5]. The result of this research was comparison performance in fish species recognition between Faster R-CNN and SSD object detection method.

The first research reference used was research from Praba Hridayami et al., which discussed the classification of fish species using the Convolutional Neural Network (CNN) with VGG-16 Architecture. The dataset used was the QUT FISH Dataset. The data used were 50 classes with ten training data and 5 test data for each class. The total data used was 750 cropped image data. Evaluation of test results was carried out using the Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR), and False Rejection Rate (FRR). The best test results obtained were with GAR 96.4%, FAR 3.6%, and FRR 3.6% [1].

The next research reference was about improving the performance of transfer learning in the Squeeze-and-Excitation networks + Bilinear CNN (SE+BCNN) method. This research was done by Chenchen Qiu et al. The highest accuracy achieved on the QUT FISH dataset was 71.80% [2].

The next research reference was researched by M. Sarigül and M. Avci about the comparison of test results from three different custom Convolutional Layers architectures. The dataset used was the QUT FISH Dataset. This research used 93 species from this dataset. The highest accuracy that was obtained in this study was 46.02% [3].

The next research reference was about traffic light detection research from Janahiraman and Subhan. This research was comparing the results of traffic light detection between SSD-MobileNet-V2 and Faster R-CNN Inception-v2 Architecture. The results of the test accuracy that have been obtained by the Faster R-CNN Inception-v2 method was 97.02%, and SSD-Mobilenet-v2 was 58.21% [4].

The next research reference was about livestock detection, which was also carried out by comparing several object detection methods by Han et al. The dataset used was an image containing livestock with a resolution of 4000 pixels x 3000 pixels taken from the air. The methods compared in the journal were Faster R-CNN, YOLOv3, and the Unet + Inception Method. On the Faster R-CNN, the accuracy obtained was 89.1%, Yolo-V3 gets 83% accuracy, and Unet + Inception gets 89.3% accuracy [5].

The next research reference was about investigating fruit species detection with Faster R-CNN from Basri et al. This research used object class images of mango and dragon fruit as image data. The object detection model moves with the help of the Tensorflow library. The results in this research were reached accuracy, up to 70.6% [6].

2. Research Methods

There was 4 phase in this research. These phases were Data Collecting Phase, Data Processing Phase, Data Training Phase, and Testing Phase.

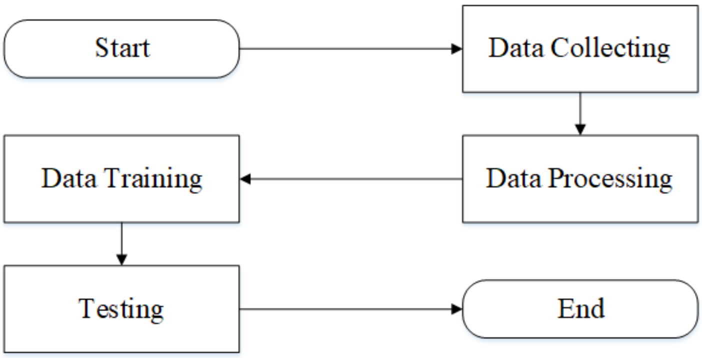

Figure 1. Research Flowchart

The data collection phase was the phase of collecting data needed in this research. The data that must be collected was fish dataset. The fish dataset used in this research was the QUT FISH dataset [7].

The data processing phase was the phase of adjusting the data from the dataset obtained for use in the Data Training Phase. This research used 50 fish classes with ten training images and 5 test images for each class. The total image data used was 750 data. The reason for using this amount of data was so that the results obtained could be compared with current research references [1] because of similar data usage conditions. The 50 names of fish data classes from QUT FISH used in this research could be seen in Table 1.

The data training phase was the phase of training the object detection model with the Faster R-CNN Method Inception V-2 Architecture using training data that has been prepared in the previous phase. The data training process was done with Google Colab Cloud service.

The testing phase was the phase to test the performance of the object detection model that has been trained and evaluating the test results. Evaluation of test results would be compared with the results obtained from previous related research [1]–[3].

The 50 classes of fish used in this research were selected based on the consideration that each class must have a minimum of 15 data from this research reference [1]. These 15 data would be used in the training and testing phase.

2.1. Faster R-CNN

The popularity of machine learning was increasing following the popularity of Artificial Neural Networks (ANN). ANN was a non-linear complex learning system that occurs in a network of neurons [8]. Convolutional Neural Network (CNN) was one of the most developed ANN derivatives currently [1]. CNN was a deep learning algorithm that uses a convolutional layer for feature extraction and a fully connected layer for classification [9]. CNN could be applied in image and text classification [10], [11]. The method used in this research was Faster R-CNN.

Faster R-CNN was a deep learning algorithm developed from CNN that could be used in object detection systems [12]. The object detection system was a system that has a function to localize objects in the image, so the classification process would get better results [13].

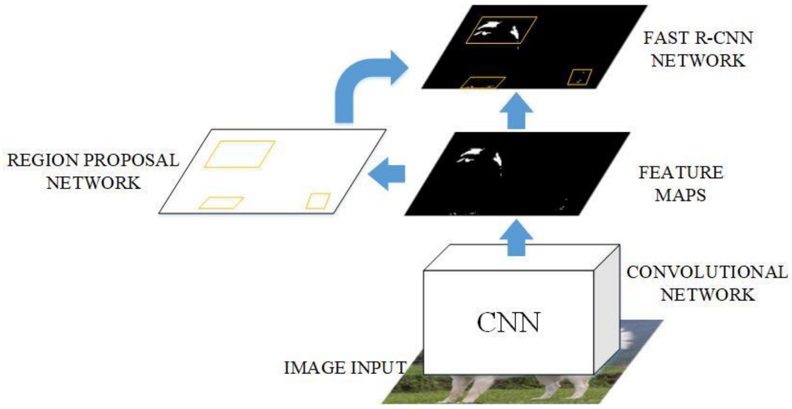

Figure 2. Illustration of the Faster R-CNN Method Workflow [16]

Faster R-CNN was the development of Fast R-CNN. Fast R-CNN was an object detection method that used the selective search method in the region proposal search process [14]. The region Proposal module task was to find regions or areas that may contain objects in it [15]. Shaoqing Ren, in his research on the implementation of Faster R-CNN as Real-Time Object Detection, revealed that this method generally consists of two modules, namely Region Proposal Network (RPN) and Fast R-CNN. [16]. Figure 2 was an illustration of the Faster R-CNN method workflow. The input that was entered into the system will be processed in the Convolutional Network first to get the feature of the object in the image, named Feature Maps. Then Feature Maps from the Convolutional Network will be forwarded to the Region Proposal Network (RPN) module and the Fast R-CNN module. The region Proposal function was to find regions or areas that may contain objects in it (Region Proposal) [17]. The Fast R-CNN module function was refining the region proposals of the RPN and classifying the objects in it [16].

2.2. Single Shot Detector (SSD)

SSD was a single-shot detector for multiple class objects that was faster than YOLO. The SSD method was based on a feed-forward convolutional network that produces a fixed-size collection of bounding boxes and scores for the presence of object class instances in those boxes, followed by a non-maximum suppression step to produce the final detections. SSD only needs an input image and ground truth boxes for each object during training. SSD object detection method was designed to create a deep learning object detection method with a lighter process than other object detection methods based on deep learning processes like YOLO and Faster R-CNN [18].

2.3. Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR), False Rejection Rate (FRR) and Accuracy (ACC)

Genuine Acceptance Rate (GAR) was the percentage of the number of objects that were correctly recognized [19]. The results of the object classification must get the correct class with a probability above the threshold value used. The formula of GAR showed in (1) [20].

$$\mathrm{GAR} = 1 - \mathrm{FRR} \tag{1}$$

False Acceptance Rate (FAR) was the percentage of the number of objects received, but the class classification results were wrong [21]. False Acceptance Rate could also be said as *False Positive*. The formula of FAR showed in (2) [20].

$$\mathrm{FAR} = \frac{\text{Total number of fish species identified with another fish species}}{\text{Total number of test data}} \tag{2}$$

False Rejection Rate (FRR) was the percentage of the number of objects that do not get a single classification result. The False Rejection Rate was also commonly referred to as the False Negative. The formula of FRR showed in (3) [20].

$$\mathrm{FRR} = \frac{\text{Total number genuine of fish species rejected}}{\text{Total number of test data}} \tag{3}$$

Accuracy (ACC) was calculated as the number of all correct predictions divided by the total number of the test data. The formula of ACC showed in (4).

$$\mathrm{ACC} = \frac{\text{Total number of fish species identified correctly}}{\text{Total number of test data}} \tag{4}$$

2.4 Evaluation Protocol

The process of evaluating test results was calculating the values of GAR, FAR, FRR, and Accuracy from both of Faster R-CNN and SSD models. GAR, FAR, and FRR was used based on this research references [1]. Accuracy is used to complement the evaluation of test results. The formula of GAR, FAR, FRR, and Accuracy could be seen in section 2.3. The detection result used was the recognition result with the highest confidence percentage.

3. Result and Discussion

This section describes the results and discussion of this research about fish species recognition using the R-CNN Faster and SSD Method with Inception-v2 Architecture with the QUT FISH Dataset.

3.1. Preparing Training Data and Testing Data

Training data was the data used in the model training process. The training data used were 10 data for each class. Total fish classes used in this research were 50 fish classes. Total training data used were 500 data from the QUT FISH Dataset. Examples of training data used in this research could be seen in Figure 3.

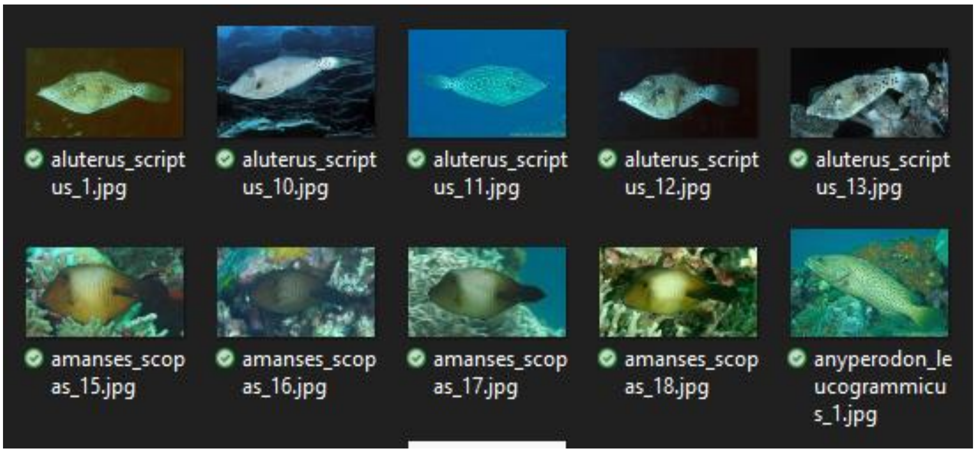

Figure 3. Training Data Samples

Test data was the data used in the testing data phase. The test data used were 5 data for each class. The total fish class used in this research were 50 fish classes. Total test data used were 250 data from the QUT FISH Dataset. Examples of test data used in this research could be seen in Figure 4.

3.2. Implementation

This subsection contained the implementation of the testing phase. The testing phase was done by running the detection process upon the test image using the object detection model of the training result. The optimum threshold used in Faster R-CNN model testing was 72%, which was the optimum threshold of the FAR and FRR values. The optimum threshold used in Single Shot Detector (SSD) model testing was 54%, which was the optimum threshold of the FAR and FRR values. Figure 5 was an example of a detection result with one correct detection result.

Figure 5 was an example of test results with one correct detection result, which belonged to the Genuine Acceptance Rate (GAR). The class object contained in the image was Aluterus Scriptus, and the detection results obtained were the Aluterus Scriptus class with 99% confidence. The confidence value was the percentage of object similarity in the image to the object recognized according to the object detection model or object classification model. In the detection results of the object detection method, there might be images that had more than one detection result which had confidence above the threshold value. An example of this case could be seen in Figure 5.

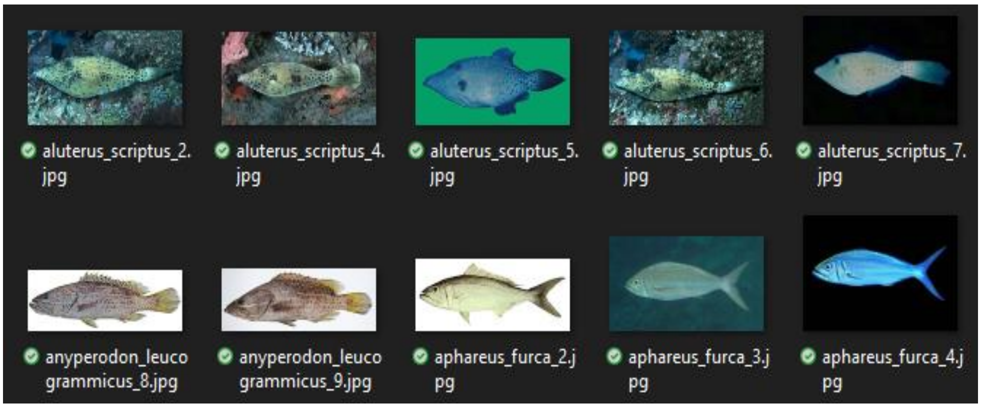

Figure 4. Testing Data Samples

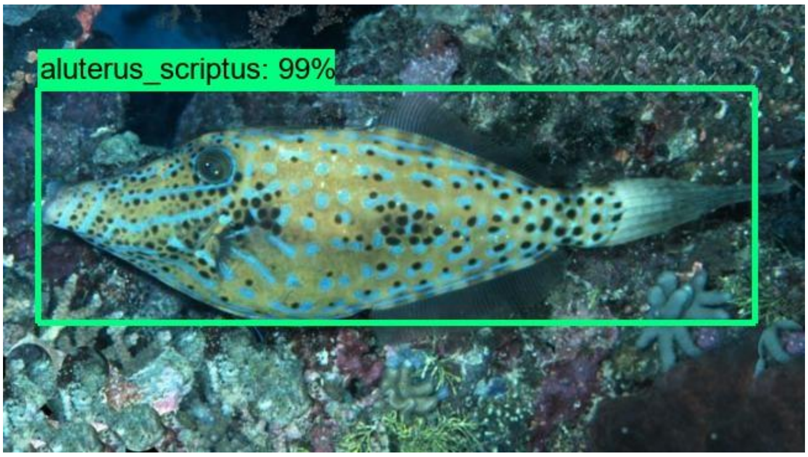

Figure 5. The Test Results with One Correct Detection Result

Figure 6 was an example of test results that get more than one detection result. The image used was Bodianus Diana class test image. The detection results obtained were the Cirrhilabrus Cyanopleura class with 86% confidence and Diana Bodianus class with 77% confidence. In this test image, the detection results used were Cirrhilabrus Cyanopleura class because it had the highest confidence percentage. This test data result belonged to the False Acceptance Rate (FAR) because it had wrong recognition. Figure 7 was an example of test results that did not get

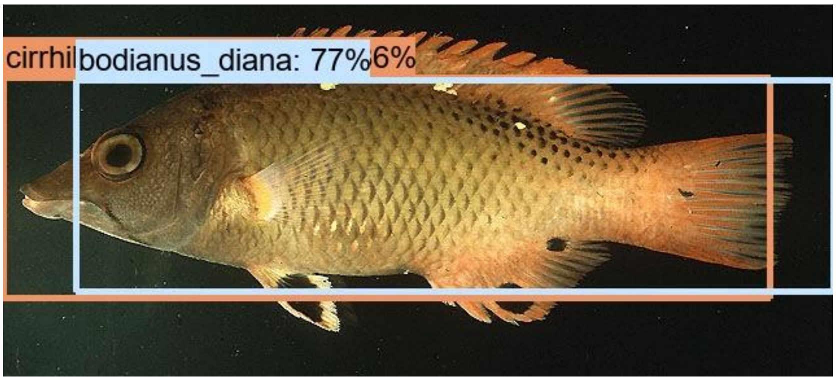

Figure 6. The Test Results with more than One Detection Result

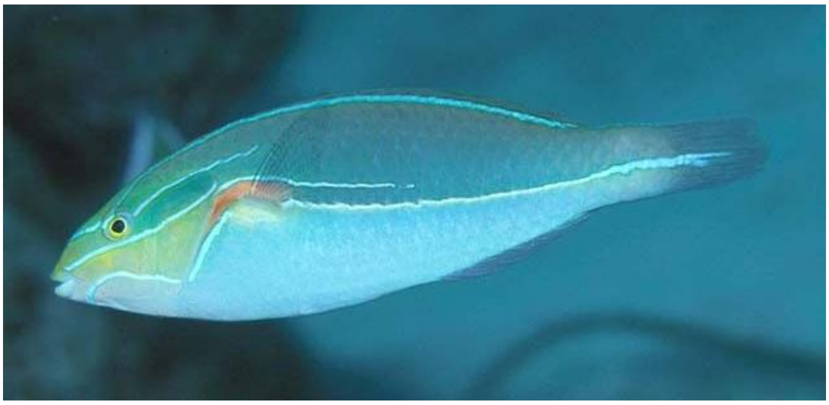

Figure 7. The Test Results without any Detection Result

any detection results. The test results from Figure 7 belonged to the False Rejection Rate (FRR). The test image used was the Stethojulis Bandanensis class test image.

3.3. Testing Result

This section contained the testing result of the Faster R-CNN and SSD model in fish species detection. Fish species detection was recognized as the fish species inside a raw fish image. The raw fish image was an image of fish that not has been preprocessed. Table 3 contained a comparison of the testing result between Faster R-CNN and SSD. Evaluation of the testing result used was GAR, FAR, FRR, and Accuracy.

Table 3. Comparison of GAR, FAR, FRR, and ACC between Faster R-CNN and SSD

Method	GAR	FAR	FRR	ACC
SSD Inception-v2	74%	24.8%	26%	49.2%
Faster R-CNN Inception-v2	90.4%	10%	9,6%	80.4%

The performance of each Faster R-CNN and SSD model could be seen in Table 3. The Faster R-CNN model had much better performance than the SSD model. Faster R-CNN accuracy was 80.4%, much better than SSD accuracy that was 49.2%. SSD model made a more wrong prediction of up to 24.8% (FAR) and more no detection result up to 26% (FRR). More wrong predictions and no detection result cause the SSD model to have low accuracy, although already using the optimum threshold in the testing phase. Faster R-CNN had higher performance than the SSD model proved that Faster R-CNN was more reliable for fish species detection.

Test data that had the most failed prediction in Faster R-CNN were from four class fish, such as Anyperodon Leucogrammicus, Bodianus Diana, Cephalopholis Sexmaculata, and Pseudocheilinus Hexataenia.

All test data from the Anyperodon Leucogrammicus class got a failed prediction. Three test data got the wrong prediction, and two test data got no detection result.

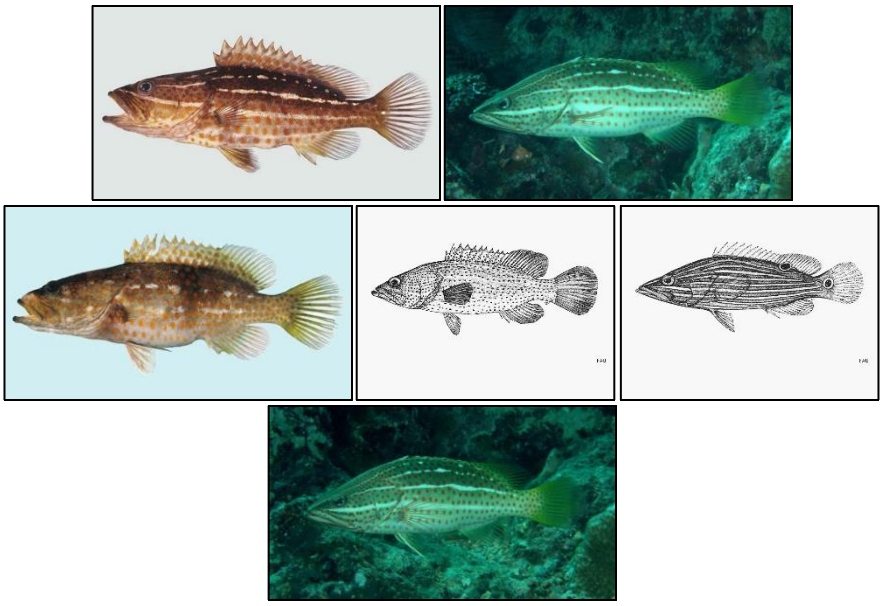

Figure 8. Some Samples of Anyperodon Leucogrammicus Training Data

Figure 8 showed some samples of Anyperodon Leucogrammicus training data. Those Anyperodon Leucogrammicus training data had inconsistent features that made the model difficult to recognize the fish species in Anyperodon Leucogrammicus tests data. There were two training data with binary color mode. In Faster R-CNN, object color was an important feature of an object. Those binary colored images interfered with the training of the model. Then there were two training data with greeny color that the fish did not have a dorsal fin. That inconsistency shape in training data caused the Faster R-CNN model could not detect the fish in test data correctly.

Four test data from Bodianus Diana's class got failed prediction. Two test data got the wrong prediction, and two test data got no detection result.

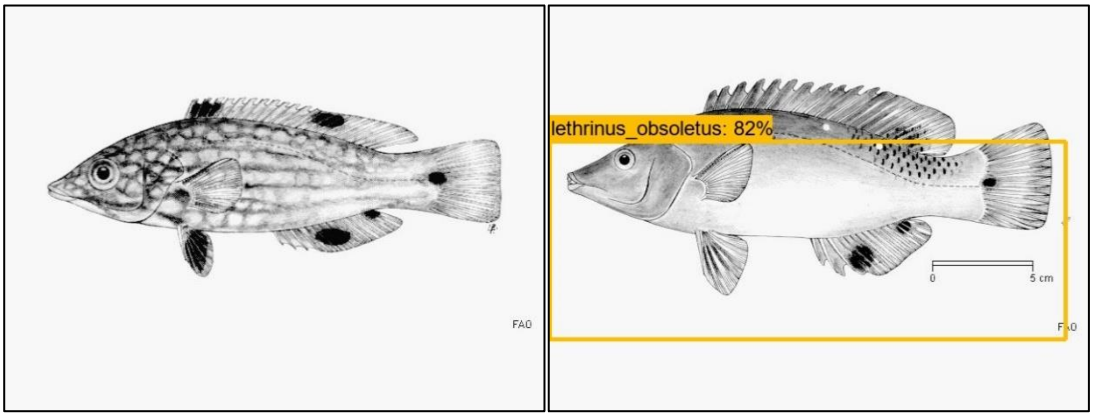

Figure 9. Some Samples of Bodianus Diana Testing Data

Figure 9 showed some samples of not good Bodianus Diana training data. These two binary colored images caused one wrong prediction and one test data with no detecting result. Another two test data with failed prediction was in good quality images so that two another failed prediction caused by the failure of Faster R-CNN model.

Four test data from the Cephalopholis Sexmaculata class got failed prediction. Those four test data got got no detection result.

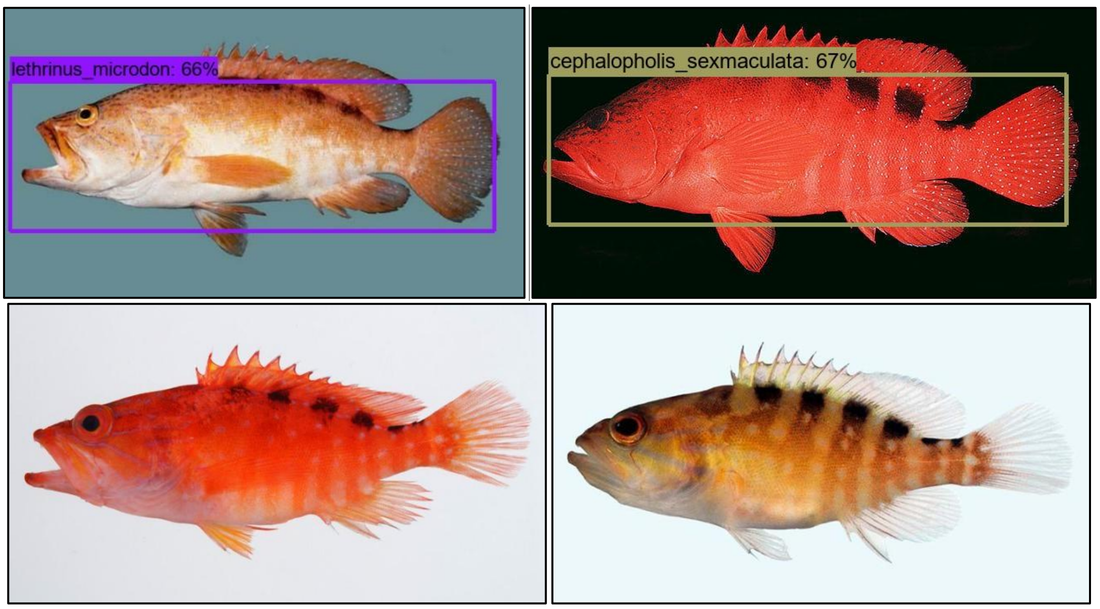

Figure 10. Cephalopholis Sexmaculata Testing Data that Got Failed Prediction

Figure 10 showed Cephalopholis Sexmaculata testing data that got failed prediction. There were two data that got any prediction result, but the confidence level below the optimum threshold used (72%). One of those two data got the correct prediction result, so one failed prediction result was caused by the threshold used to high. Another three test data with failed prediction was in good quality images, so that three another failed prediction caused by the failure of Faster R-CNN model. Three test data from Pseudocheilinus Hexataenia got failed prediction. Those three test data got got wrong prediction result.

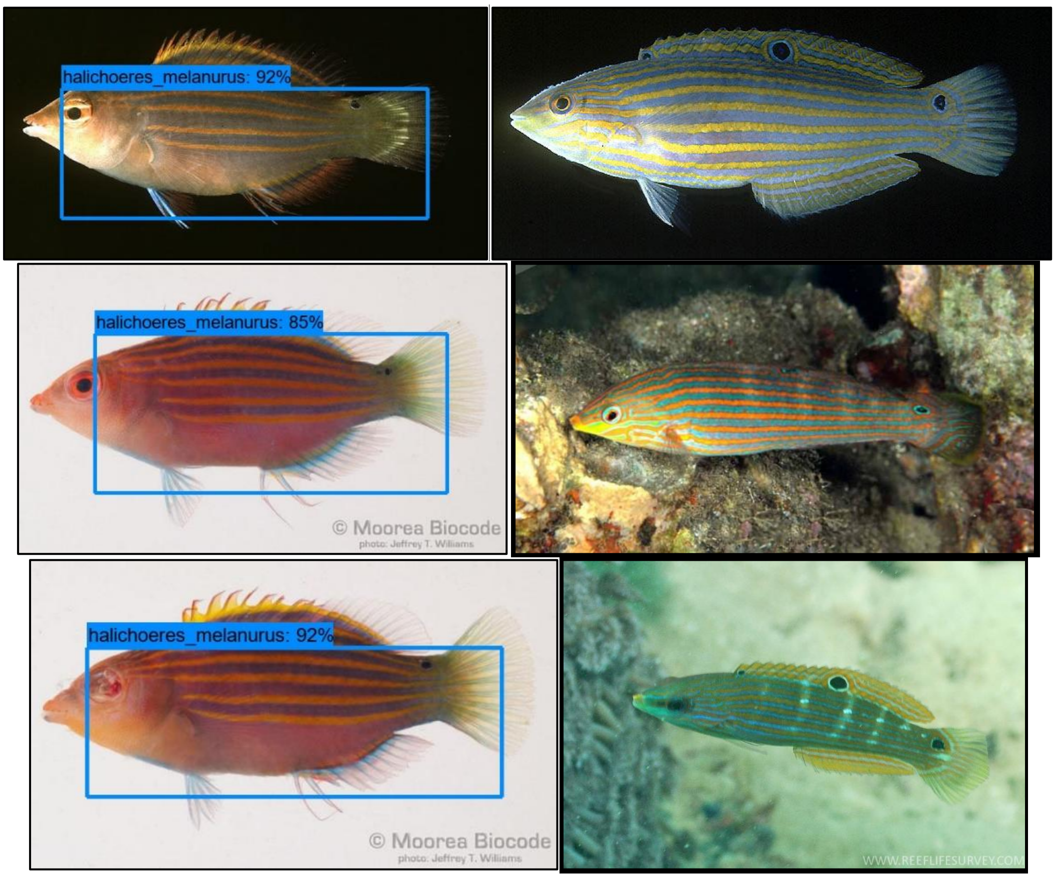
halichoeres_melanurus: 92%
halichoeres_melanurus: 85%
© Moorea Biocode
photo: Jeffrey T. Williams
halichoeres_melanurus: 92%
© Moorea Biocode
photo: Jeffrey T. Williams
WWW.REEFLIFESURVEY.COM

Figure 11. Three Pseudocheilinus Hexataenia Testing Data (Left) and Three Samples of Halichoeres Melanurus Training Data (Right)

Figure 11 showed three Pseudocheilinus Hexataenia testing data on the left side and three samples of Halichoeres Melanurus training data on the right side. Pseudocheilinus Hexataenia had a similar pattern with Halichoeres Melanurus that was horizontal lines. Faster R-CNN failed to extract more features from Pseudocheilinus Hexataenia like head shape and the fish fin, so the model probably made the wrong prediction in Pseudocheilinus Hexataenia test data.

Overall Faster R-CNN model had a good performance on fish species detection with 80.4% accuracy than SSD with 49.2% accuracy. Faster R-CNN probably could get better accuracy in fish species detection if using other architecture that more suitable for extracting fish features. Need more research to got that more suitable architecture for extracting fish features in Faster R-CNN.

4. Conclusion

Overall Faster R-CNN model had a good performance on fish species detection with 80.4% accuracy than SSD with 49.2% accuracy. Faster R-CNN got worse prediction result upon test data on Anyperodon Leucogrammicus, Bodianus Diana, Cephalopholis Sexmaculata, and Pseudocheilinus Hexataenia class object. Faster R-CNN probably could get better accuracy in fish species detection if using other architecture that more suitable for extracting fish features. Need more research to get more suitable architecture for extracting fish features in Faster R-CNN.

References

[1] P. Hridayami, I. K. G. D. Putra, and K. S. Wibawa, "Fish species recognition using VGG16 deep convolutional neural network," *The Journal of Computer Science and Engineering*, vol. 13, no. 3, pp. 124–130, 2019, doi: 10.5626/JCSE.2019.13.3.124.

[2] C. Qiu, S. Zhang, C. Wang, Z. Yu, H. Zheng, and B. Zheng, "Improving transfer learning and squeeze- and-excitation networks for small-scale fine-grained fish image classification," *IEEE Access*, vol. 6, pp. 78503–78512, 2018, doi: 10.1109/ACCESS.2018.2885055.

[3] M. Sarigül and M. Avci, "Comparison of Different Deep Structures for Fish Classification," *International Journal of Computer Theory and Engineering*, vol. 9, no. 5, pp. 362–366, 2017, doi: 10.7763/ijcte.2017.v9.1167.

[4] T. V. Janahiraman and M. S. M. Subuhan, "Traffic light detection using tensorflow object detection framework," *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) 2019 - Proceeding*, no. October, pp. 108–113, 2019, doi: 10.1109/ICSEngT.2019.8906486.

[5] L. Han, P. Tao, and R. R. Martin, "Livestock detection in aerial images using a fully convolutional network," *Computational Visual Media*, vol. 5, no. 2, pp. 221–228, 2019, doi: 10.1007/s41095-019-0132-5.

[6] H. Basri, I. Syarif, and S. Sukaridhoto, "Faster R-CNN implementation method for multi-fruit detection using tensorflow platform," *International* Electronics *Symposium* on *Knowledge* Creation and Intelligent *Computing* (*IES-KCIC*) *2018 Proceedings*, pp. 337–340, 2019, doi: 10.1109/KCIC.2018.8628566.

[7] A. Karad, G. Padhar, R. Agarwal, and S. Kumar, "Fish species detection using computer vision," vol. 4, no. 6, pp. 2–6, 2020.

[8] D. Kristianto, C. Fatichah, B. Amaliah, and K. Sambodho, "Prediction of Wave-induced Liquefaction using Artificial Neural Network and Wide Genetic Algorithm," *Lontar Komputer Jurnal Ilmiah Teknologi Informasi*, vol. 8, no. 1, p. 1, 2017, doi: 10.24843/lkjiti.2017.v08.i01.p01.

[9] O. Sudana, I. W. Gunaya, and I. K. G. D. Putra, "Handwriting identification using deep convolutional neural network method," *Telkomnika (Telecommunication, Computing, Electronics and Control*, vol. 18, no. 4, pp. 1934–1941, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14864.

[10] I. M. Mika Parwita and D. Siahaan, "Classification of Mobile Application Reviews using Word Embedding and Convolutional Neural Network," *Lontar Komputer Jurnal Ilmiah Teknologi Informasi*, vol. 10, no. 1, p. 1, 2019, doi: 10.24843/lkjiti.2019.v10.i01.p01.

[11] S. Wang, M. Huang, and Z. Deng, "Densely connected CNN with multi-scale feature attention for text classification," *International Joint Conference on Artificial Intelligence.*, vol. 2018-July, pp. 4468–4474, 2018, doi: 10.24963/ijcai.2018/621.

[12] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," *Proc. - 12th IEEE International Conference on Automatic Face Gesture Recognition, FG 2017 - 1st Int. Work. Adapt. Shot Learn. Gesture Underst. Prod. ASL4GUP 2017, Biometrics Wild, Bwild 2017, Heteroge*, pp. 650–657, 2017, doi: 10.1109/FG.2017.82.

[13] P. Garg, D. R. Chowdhury, and V. N. More, "Traffic Sign Recognition and Classification Using YOLOv2, Faster RCNN and SSD," *2019 10th Int. Conf. Comput. Commun. Netw. Technol.*, pp. 1–5, 2019.

[14] K. Wang, Y. Dong, H. Bai, Y. Zhao, and K. Hu, "Use fast R-CNN and cascade structure for face detection," *VCIP 2016 - 30th ANNIVERSARY OF VISUAL COMMUNICATION AND IMAGE PROCESSING*, pp. 4–7, 2017, doi: 10.1109/VCIP.2016.7805472.

[15] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster R-CNN doing well for pedestrian detection?," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9906 LNCS, pp. 443–457, 2016, doi: 10.1007/978-3-319-46475-6_28.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016, doi: 10.1109/TPAMI.2016.2577031.

[17] Y. Nagaoka, T. Miyazaki, Y. Sugaya, and S. Omachi, "Text Detection by Faster R-CNN with Multiple Region Proposal Networks," *Proc. International Conference on Document Analysis and Recognition*, pp. 15–20, 2017, doi: 10.1109/ICDAR.2017.343.

[18] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector Wei," *European Conference on Computer*

Vision, vol. 1, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0 2.

[19] T. S. Indi and Y. A. Gunge, "Early Stage Disease Diagnosis System Using Human Nail Image Processing," *International Journal of Information Technology and Computer Science.*, vol. 8, no. 7, pp. 30–35, 2016, doi: 10.5815/ijitcs.2016.07.05.

[20] Z. Waheed, A. Waheed, and M. U. Akram, "A robust non-vascular retina recognition system using structural features of retinal image," *Proc. 2016 13th International Bhurban Conference on Applied Sciences & Technology Technol. IBCAST 2016*, pp. 101–105, 2016, doi: 10.1109/IBCAST.2016.7429862.

[21] S. Bharathi and R. Sudhakar, "Biometric recognition using finger and palm vein images," *Soft Computing,* vol. 23, no. 6, pp. 1843–1855, 2019, doi: 10.1007/s00500-018-3295-6.