FishNet: Deep Neural Networks for Low-cost Fish Stock Estimation

No Author Given

No Institute Given

Abstract. Fish stock assessment typically involves a lot of manual counting that requires specialists, and so is both time-consuming and costly. We propose a method to perform both taxonomic classification and size estimation of fish, for the purpose of building an automated stock assessment system that uses a standard camera. We use a subset(single fish images), of a large dataset generously provided by the Nature Conservancy, and fisheries in Indonesia. We achieve a 95% top-1 classification accuracy, and 2.3cm mean error on size estimation, and propose a way to perform automated stock assessment on images with multiple fish. The resulting models are published on Hugging Face models.

Keywords: Computer Vision · Fish Stock Estimation · Classification · Size Estimation.

1 Introduction

Predictions that all of the world's stocks of commercially important fish could be collapsed by the year 2048 [24] have been tempered by recent evidence that fish stocks are recovering in many developed nations [25]. This turnaround is typically attributed to stringent catch limits backed by accurate scientific assessment of the number of fish in a particular stock. It is difficult to set a limit on catch appropriately unless scientists know how many fish are actually present in a wild population. Unfortunately, the majority of the world's fisheries remain "unassessed" and the prognosis for these fisheries is concerning [4]: populations are declining and are on a trajectory towards functional extinction. One of the major barriers to performing fish stock assessment in the developing world is the cost. The federal government in the US spends approximately $\$215$ million a year on fish stock assessment [17], which excludes spending by state governments on assessment of stocks within 3 nautical miles of the coast (state waters). As an example, the average cost of a fish stock assessment performed by NOAA's Pacific Islands Fisheries Science Center in Honolulu is $\$5.6$ million, which exceeds the total value of many developing country fisheries.

Costly stock assessment also prevents many well-managed fisheries from accessing lucrative markets for sustainable seafood. For example, the Marine Stewardship Council (MSC), which endorses seafood with their blue eco-label requires detailed stock assessment of the fishery for certification. In [21], the authors argue that the high costs of certification remain a barrier to many fisheries in Latin America and the Caribbean accessing the benefits of seafood ecolabels. For example, there is only one fishery certified as sustainable in Indonesia by the MSC, and this is, in part, because the species

Fig. 1: Fishers in Indonesia on their boats measuring their catch on a standard color coded measuring board.

being targeted (yellowfin tuna) occurs throughout the Pacific Ocean; the expensive assessment of this stock is performed by an intergovernmental body, funded by the US, Australia, France, Japan, and a number of other countries. Here we propose a methodology for drastically reducing the cost of fisheries stock assessment by combining citizen science with machine learning. We gave small-scale fishers as seen in Figure 1 in Indonesia digital cameras, and asked them to photograph everything they caught on a uniform background. A team of fish biologists then identified the species and length of fish from the photographs, resulting in over $300,000$ images, and over one million fish hand-labeled by species and length. We focus on the tropical snapper-grouper fishery in Indonesia, one of the most difficult fisheries in the world in which to perform a stock assessment. Fishers catch over a hundred different species of various sized fish using a variety of fishing gears spread across Indonesia's $17,000$ different islands. The complexity of the task and the limited available resources make technological innovations an attractive alternative to traditional methods. Furthermore, if novel artificial intelligence methods can be used to perform a low-cost and accurate assessment in this fishery, then the same approach should work in practically any other fishery in the world. In this work, we present a machine learning approach to predicting fish species and length from photographs taken by fishers. To do this, we collaborate with fishers willing to take photographs through The Nature Conservancy Indonesia, and field biologists who have annotated over 1 million fish in these photographs. By leveraging transfer learning, we train a segmentation model that feeds into species classification and size estimation models achieving performances worthy of practical deployment.

2 Related Work

Taxonomic Classification: Deep learning algorithms (DL) have seen much success in image classification tasks post ImageNet [13]. ImageNet and other large image datasets such as MS COCO [16], demonstrated not only the versatility of such networks, but also their robustness to over-fitting on very large labelled datasets [23,15]. These models have been applied to various domains including, but not limited to work closest to ours: broadly segmentation and classification of fish species [22,6,9,20,2].

Fish Size Estimation: Estimating the size of a fish in centimeters from an image requires a reference object of known length in most cases. In our work, 10cm rectangular color areas situated on the sides of each measuring board that the fish is to be placed on are used. These act as fiduciary markers to guide the estimation of the fish size. In [1], a mask R-CNN similar to our choice is used to detect the head and rest of the fish's body. Each of these lengths is estimated separately. This method works well for a relatively small dataset containing fish of the same species that are all of the similar size. In our case, the sizes of fish range wildly, from as little as 32cm to more than 100cm, this necessitates the need for fiduciary markers for perspective. In [19], the authors place three fiduciary markers, one in the background, one in the forefront, and a laser marker for estimating fish length using standard smartphone cameras. In [18], the authors also place three ArUco fiduciary markers of different sizes on polypropylene sheets, along with the fish to be measured. They use mask R-CNN to detect and segment objects of interest. Their work is evaluated on a much smaller dataset (1000), of fewer fish species than we address, and does not address both species classification and size estimation, which are necessary for stock estimation.

3 Methods

In this section we outline implementation details for data collection, image segmentation, classification and fish length regression. The different models and experimental setup is explained.

3.1 Data Collection

Small-scale fishers in Indonesia were given digital cameras and asked to photograph everything they caught. To help standardize the photographs, the fishers were asked to photograph their catch on a 1-by-0.8 m plastic board provided to them. These are white, with multi-colored markings every 10 cm that serve as a fiduciary marker for scale. Fish are placed on these boards — sometimes many fish at a time — and photographs are taken from a variety of distances, angles, and lighting conditions (Figures 3 and 2). A team of fish biologists then examined each photograph, labeling the species and length of each fish. From over 300,000 photographs, approximately 1.2 Million fish are labeled in this way, with 163 different species represented. Massive data-labeling projects always contain some level of error. In this case, one source of error is that photographs containing multiple fish are not consistently labeled in the same orientation. In most images, fish are labeled from top-to-bottom, left to right, but many examples are found in which this is not the case, and the extent to which label noise on multi-fish images exists is unknown. It is important to have a standard way of assigning the provided labels to the fish on the board in cases of multiple fish so we can, in a later stage train models on the much larger dataset of $1,2$ Million fish, and make the dataset publicly available.

Fig. 2: An example of a typical image containing 2 fish on a marked board from the dataset

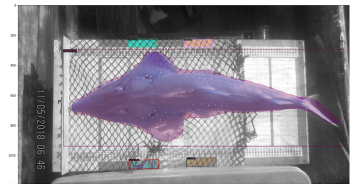


Fig. 3: Axample of an image segmented using Detectron2 to identify both the fish and the 10 cm colored rectangles used as fiduciary markers (bottom).

3.2 Image Segmentation

Since most images contain multiple fish, our system first applies an object detection model to identify each individual fish, then, we perform species classification and length regression. Separate machine learning models perform each of these three tasks. The object detection and segmentation model is pre-trained on MS COCO, and fine-tuned on a small dataset of images that we annotated. This is done by segmenting fish and a set of fiduciary markers: four 3-by-10 cm colored boxes (two yellow and two blue), that are located on the edges of the presentation board. These markers are chosen because at least one of the four is readily identifiable in most of the images. Using the VGG Image Annotation (VIA) tool [7], we sampled 700 random images containing single fish, and annotated these with polygon outlines of the three different types of object: (1) a fish; (2) yellow colored boxes; (3) blue-colored boxes. When these objects are occluded by humans, their shadows, or fishing nets, the shape of these objects is inferred by the annotator (Figure 4).

The annotations are used to fine-tune an object-recognition model: a Resnet-50 architecture [12] using Facebook's Detectron2 instance segmentation [26]. Detectron2, written in Pytorch, is a collection of re-implementations of state-of-the-art object-detection algorithms including Mask R-CNN [11] for detection and segmentation, that we use in this work. Mask R-CNN trains a single convolutional neural network for pixel-wise segmentation, classification, and bounding box regression. In mask R-CNN, a pre-defined number of regions of interest (ROIs) are proposed, of adjacent pixels similar in color, texture, or intensity from the features learned by a stack of convolutional layers. The network is then trained to minimize the classification and segmentation losses of the best ROIs.Once ROIs with very high probabilities of containing objects are found, a fully connected layer is added to predict the x, and y coordinates of a rectangular region that most tightly encloses the objects with high confidence. Given a new image, the trained model produces a list of detected objects, each with the following information: (1) the predicted class (fish, yellow color box, or blue color box); (2) a bounding box that represents the smallest rectangular region that completely contains the detected object; (3) a pixel-wise segmentation mask outlining the object; (4) an object-level confidence score. On the validation dataset, the system uses the predicted segmentation mask to crop out images of individual fish, and these cropped images are used as input to the classification and regression models.

3.3 Species Classification

The species classification model is an Inception-ResNet V3 (pre-trained on the ImageNet dataset [5]) fine-tuned to classify fish species from cropped images containing a single fish. This model is trained on the subset of $50,000$ images that only contain a single fish because it is for these images that we are most confident in the labels. Of this subset, $40,000$ images (80%) are used for training, and $10,000(20\%)$ are held out for validating the model. During training, a number of augmentation strategies are used to improve generalization; these are listed in Table 1. Since we eventually want to classify fish species even in cases where there are multiple fish in an image, we make use of the segmentation output of the Detectron2 model to crop out only the bounding box around each fish and use this as input in training the classifier.

Transformation	Range/Value
intensity normalization	[0,1]
Random Rotation	[0,30°]
width, length, channel shift and random zoom	[0, 0.2]
shear	[0, 0.3]
feature-wise normalization and horizontal flip	Yes
fill mode	nearest
re-scale	152×152

Table 1: Data augmentations applied during training of the species classification model.

The classifier is trained for 100 epochs with early stopping, and learning rate reduction on the plateau, based on the validation set performance at the end of each epoch, it produces the final prediction on the validation set that we evaluate the quality of. During testing, no random transformations are applied, but the images are first cropped to the size of the bounding box of each fish, then re-scaled to 152×152, and pixel values are divided by 255 to be in the range $[0, 1]$.

3.4 Length Regression

Length regression is performed in two steps. First, we use the outputs of Detectron2 to detect and locate fish and the colored markings on a board which are normally used by the biologist annotators to determine the fish length. The color boxes are of fixed size on every board in centimeters(cm), and thus provide scale information — our strategy is to detect these markings using object recognition to aid in the length prediction task. For each image with one or more fish, we restrict the problem to images with at least one visible Fiduciary color marker, this is necessary as the images are not captured

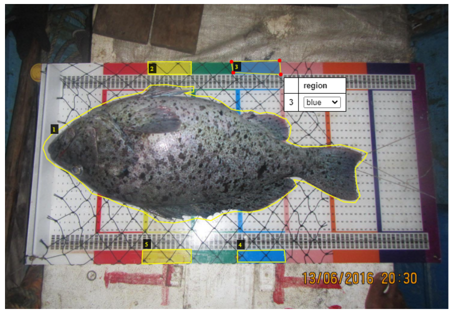


Fig. 4: The Detectron2 instance-segmentation model is trained to detect fish along with the yellow and blue color boxes on either side of the boards. In this example, a human annotator has outlined these objects using the VGG Image Annotation tool.

from a single view distance and angle. We then extract features from the output of the segmentation model based on the current image and use these to train two ensemble regression models to predict the size of the fish in centimeters. In the case where there is only one fish on the board, the vertical and horizontal pixel distances of the bounding box and segmentation mask are used together with the confidence scores for the color boxes. These aggregate features extracted from the color boxes are especially useful in cases where some color boxes are occluded by the fish or have bad lighting for the segmentation model to predict the highest intersection over union(IOU) percentage. Estimation of fish length requires the use of fiduciary markers since the photos are taken from a variety of distances and angles. Thus, we restricted our analysis to images for which at least one colored box is detected. For each fish and colored box object detected, we calculate a set of features, then a fish's length is predicted from a set of summary statistics:

1. Colour Boxes: Color box count, median color box length, mean color box length, median color box segment length, maximum color box segment length.
2. Detected fish: The confidence score for each fish object, length of the bounding box around the fish, and length of the segment mask of the fish, we use both since the segmentation mask can sometimes be slightly shorter than the actual fish object as we do not achieve a 100% IoU on fish segmentation.

These features are used as input to a random forest regression model. We train two ensemble regression models, a random forest regressor with 300 trees, and a gradient boosting regressor with 300 trees as well. Both models optimize a Mean Squared Error loss. The model is trained and evaluated using 5-fold cross-validation on the same subset of $50,000$ images containing one fish since we are most confident in those labels. We then pick the model that performs best on the 5-fold cross-validation as the final length regression model.

4 Results

In this section, test results for the different components are presented starting with performance on single fish $n_i = 1$ images, followed by images for $n_i > 1$ fish on a board.

4.1 Single Fish Detection and segmentation

We evaluated the segmentation model in terms of accuracy and the intersection over union (IoU) score on the held-out validation set images. These metrics inform us of how often we are detecting and correctly assigning objects, as well as how well the segmentation mask covers the object since these outputs directly feed into the length estimation algorithm. A high IoU score (near 100%) means the bounding boxes fully and tightly enclose the fish, and this will result in better data for the downstream classification and regression tasks

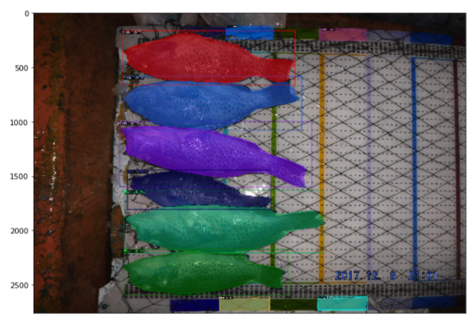


Fig. 5: Prediction of the Detectron model showing 6 segmented fish as well as 2 yellow and 2 blue Fiduciary color markers, all enclosed in their bounding boxes

The trained model is able to detect fish in the image 99% of time while detecting the fiduciary markers correctly 97% of the time. The IoU of the predicted segmentation masks has mean 92% (median 94%) for fish and mean 86% (median 88%) for the fiduciary markers (see Figure 6). Figure 5 is an example of predictions on an image where high IoU is achieved for all the fiduciary markers and fish on the board. The detection rate and IoU impact the performance of the classification and length regression models. A good IoU score means the bounding boxes fully and tightly enclose the fish, and this will result in better training data for length estimation in the case of single fish images. In the case of multi-fish images, both classification and regression are affected by a bad IoU from the Detectron2 in that having bounding boxes that do not tightly enclose the fish will result in an overestimation of the lengths, and lead to parts of one fish appearing in the crop of another fish that could be of a different species.

4.2 Single Fish Species Classification

For species classification, We fine-tune a pre-trained ResNet-50 model on images of single fish for 100 epochs using a dataset of $50,000$ images. The baseline model is trained on the whole image, including all the background surrounding the fish. The second model is trained only on the area inside the bounding box containing the detected fish from the instance segmentation model.

While the first model can capitalize on surrounding objects to better assign species classes, it is the latter that is better suited for application on images of more than one fish. We achieve a 91% top-1 and 98% top-5 accuracy on the full image validation dataset, and a 89% top-1, 97% top-5 accuracy on the model trained on cropped single fish images.

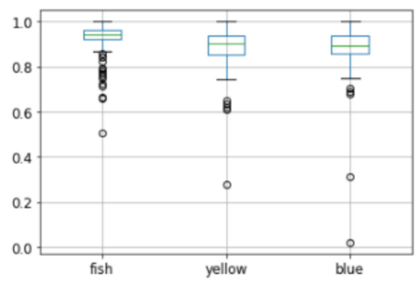
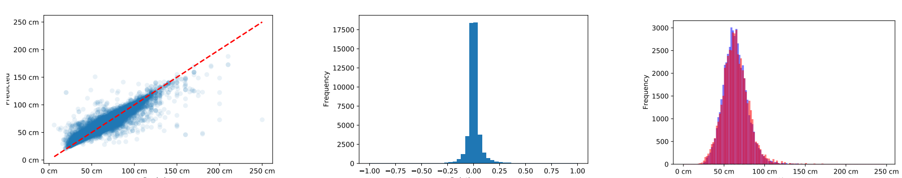


Fig. 6: Box and whiskers plot showing five key statistics of IoU results on the validation set

4.3 Single Fish Length Regression

Fig. 7: Size estimation results.



(a) Scatter plot of length predictions with Random Forest.

(b) Relative errors of the Random Forest predictions.

(c) Restored population distribution.

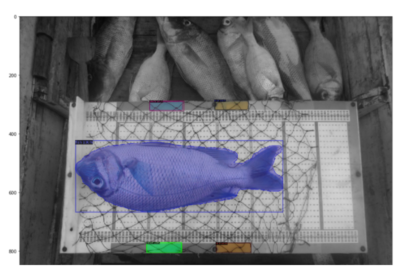
Five-fold cross-validation produced a Mean Absolute Error of 2.3cm and a coefficient of determination (R^2), of 79%. Figure 7 (a) shows a scatter plot of the predicted lengths of the fish against the true labels on the validation set. In the figure, predictions are accurate, especially for mid-sized fish $(60 - 80cm)$, which form the majority of the validation set as shown in the distribution of size in Figure 7 (c). On the longer end of the spectrum, the model tends to underestimate sizes, and this could largely be due to a lack of enough large fish training examples for the model to learn from. This problem can be solved to a large extent by having more large fish in the training dataset or applying oversampling techniques on large fish images. A look at Figure 7 (b) shows the relative errors are centered around zero with a very small deviation. We can expect similar results on a larger dataset drawn from a reasonably similar distribution.

4.4 Multiple Fish

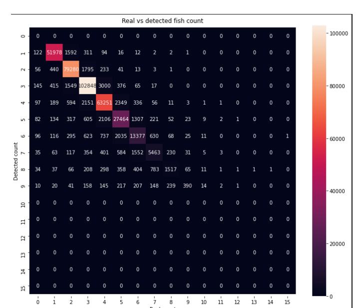
At this point, we only apply models trained on the single fish images to the complete dataset as a baseline with future improvements in mind that will be discussed in Section 4.6. In Figure 8(b), Out of the 300000 images with multiple fish, we detect the exact number of fish in 92% of prediction instances, predicting over the true count in 3%, and predicted count less than the actual count in 5% of images.

Multi-fish Classification After running the fine-tuned Detectron2 model on the 1.2M dataset and achieving a fish detection rate of 92%, We looked at the confusion matrix of the number of fish objects the model detects versus the actual number of objects in each image. This analysis is important in understanding the margin of error in the cases where we do not detect fish. It can also aid in discovering whether the model focuses only on fish on a board as trained, and not any other fish in the background outside the board as depicted in Figure 8 (a).

Fig. 8: Detectron2 segmentation prediction example and detected fish count heat-map.



(a) Example image containing multiple fish with the model successfully detecting only fish placed on the board.



(b) Detected vs ground truth fish count heat-map in multi-fish images. The Darker the diagonal entry color, the better detection rate by the model.

The confusion matrix heat-map in Figure 8 (b), illustrates that the detection model is mostly off by 1 or 2 fish when it under or over-predicts fish instances. It also shows that: the more fish you have on a board the more likely we are to predict the wrong

Number of Fish	1	2	3	4	5	6	7
Top-1 Accuracy %	95	91	81	72	68	68	66

Table 2: Multi-fish classification accuracy across the number of fish objects in an image, using the model trained on single fish images. This is from a total of 300000 images containing 1.2 Million fish objects

count of fish objects. This normally happens more often with smaller-sized fish than it does with larger fish. The single fish dataset is only a small subset of the larger dataset of 300000 images containing 1.2 Million fish. While the whole dataset is labeled and represents the largest annotated collection for the tasks of fish species classification and length regression. The labels to images with a large number of fish on one board are not straightforward to assign to the placement of fish on the board, so some level of analysis is required in assigning the labels correctly. With further fine-tuning of the segmentation model, we are able to run species classification and length regression on each fish instance in a picture. This allows us to train both the classifier and length estimator on a much larger dataset of 1.2 Million fish. Table 2 shows classification accuracy results on the complete 1.2M dataset. Performance on images with one or two fish is consistent and slightly better than that on the validation split of the 50000 single fish images discussed in Section 4.2. Performance drops significantly from three fish upwards. This can be explained by the observation that it is smaller fish that are packed onto the board in groups of three or more, and these smaller-sized fish are not very well represented in the single fish images training data, as can be seen in the size distribution in Figure 7 (c). To combat this problem, we seek to compile a dataset of all 1.2M fish images, extracted using the trained segmentation model, and cropped out as individual fish. Once this dataset is compiled, it will form the largest labeled dataset of fish images with more than 150 species.

4.5 Multi-fish image dataset

The construction of this dataset depends heavily on the accuracies of two prior processes: (1) The process of placing and labeling the fish on the boards done by the human experts and (2) How well the trained segmentation model generalizes to images with multiple fish. As mentioned before, assigning labels from a list, onto each fish on multi-fish images, turns out to have not always been done consistently during labeling. Some images are labeled top to bottom, while others bottom up or right to left depending on the orientation of the board and the position of the photographer when the picture was taken. Some engineering is required to be able to match fish to labels to a reasonable level of confidence though more work is still required.

Fish to label Matching For images with one or two fish, matching the labels to the fish is trivial as there are only 1 or two possible combinations. In cases where the detection and segmentation model detect fewer fish than we have on a board, such images are excluded from our target dataset at this time. For images with three fish and above on a board, we use a trained single fish species classifier's log-likelihood on each image and take the combination with the highest likelihood as the correct orientation: label match. Once the dataset of 1.2M fish is labeled following this procedure, we then train a species classifier on it to achieve an 85% top-1 accuracy. A likely reason for the drop in performance is the error inherent in our matching process. With these promising results on multi-fish images, we seek better ways of getting to a complete 1.2M fish dataset with little to no uncertainty in the labels.

4.6 Discussion and Future Research

This paper presented results on methodology to predict both fish species and size using widely open source deep learning (DL) architectures, transfer learning, and a large collection of fish images generously provided by the Indonesia fisheries and the nature conservancy. It also introduces work being done towards the ImageNet of fish: FishNet, a large collection of labeled fish images. This will greatly improve the predictive capabilities of DL models for marine life that will eventually be deployed to help in fish stock assessments, greatly reducing both the time and money required to perform assessments. We make the trained segmentation and classification models publicly available on Hugging Face. Based on our findings, an avenue for future research involves considering deep learning-based algorithms in the presence of label noise, as well as deep active learning methods for image segmentation and classification with a noisy labeling oracle [10,3,14]. It would also be useful to consider self-supervised pre-training [8] to combat label noise.

References

1. Alvarez-Ellacurı´a1, A., Palmer1, M., Catala´n, I., Lisani, J.: Image-based, unsupervised estimation of fish size from commercial landings using deep learning. ICES Journal of Marine Science (2020) **77**(4), 1330–1339 (2019)
2. Amanullah, B., Mushstaq, A., Faqir, G., Sadia, B., Ibrar, A.: Fish image segmentation algorithm (fisa) for improving the performance of image retrieval system. International Journal of Advanced Computer Science and Applications **8** (01 2017)
3. Chicheng, Z., Kamalika, C.: Active learning from weak and strong labelers. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015)
4. Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., Lester, S.E.: Status and solutions for the world's unassessed fisheries. Science **338**(6106), 517–520 (2012)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. IEEE (2009)
6. Dhruv, R., Sushant, J., Indu, S.: Underwater fish species classification using convolutional neural network and deep learning. In: 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR). pp. 1–6 (2017)
7. Dutta, A., Zisserman, A.: The VIA annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia (2019)
8. Evgenii, Z., Chaim, B., Avi, M., Alex, B., Or, L.: Contrast to divide: Self-supervised pretraining for learning with noisy labels. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 387–397 (2022)

9. Guang, C., Peng, S., Yi, S.: Automatic fish classification system using deep learning. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). pp. 24–29 (2017)
10. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., Sugiyama, M.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. In: 32nd Conference on Neural Information Processing Systems. pp. 8536–8546. NIPS'18, Curran Associates Inc. (2018)
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988 (2017)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
13. Jia, D., Wei, D., Richard, S., Li-Jia, L., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
14. Jun, D., X, L.: Active learning with human-like noisy oracle. In: 2010 IEEE International Conference on Data Mining. pp. 797–802 (2010)
15. Lange, S., Riedmiller, M.: Deep learning of visual control policies. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (09 2010)
16. Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, L.: Microsoft coco: Common objects in context. In: Computer Vision – ECCV 2014. pp. 740–755 (2014)
17. Merrick, R., Methot, R.: Noaa's cost of fish stock assessments. National Oceanic and Atmospheric Administration (2016)
18. Monkman, G., Hyder, K., Kaiser, M., Vidal, F.: Using machine vision to estimate fish length from images using regional convolutional neural networks. Methods in Ecology and Evolution) **10**(12), 2045–2056 (2019)
19. Monkman, G., Hyder, K., Kaiser, M., Vidal, F.: Accurate estimation of fish length in single camera photogrammetry with a fiducial marker. ICES Journal of Marine Science (2020) **77**(6), 2245–2254 (2020)
20. Oguzhan, U., Diclehan, K., Mehmet, T.: A large-scale dataset for fish segmentation and classification. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). pp. 1–5 (2020)
21. Pérez-Ramírez, M., Castrejón, M., Gutiérrez, N.L., Defeo, O.: The marine stewardship council certification in latin america and the caribbean: A review of experiences, potentials and pitfalls. Fisheries Research **182**, 50–58 (2016)
22. Prasetyo, E., Suciati, N., Fatichah, C.: Multi-level residual network vggnet for fish species classification. Journal of King Saud University - Computer and Information Sciences **34**(8, Part A), 5286–5295 (2022)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations **abs/1409.1556** (2015)
24. Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B., Lotze, H.K., Micheli, F., Palumbi, S.R., et al.: Impacts of biodiversity loss on ocean ecosystem services. Science **314**(5800), 787–790 (2006)
25. Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., et al.: Rebuilding global fisheries. Science **325**(5940), 578–585 (2009)
26. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2. https://github.com/facebookresearch/detectron2 (2019)