A Novel Approach to Coral Fish Detection And Classification in Underwater Footage Based on Convolutional Neural Network

Zhijian Sun
The Governor’s Academy, Byfield, MA, USA

Abstract. Tracking and identifying fish species is crucial to understanding marine ecosystem and its role in the world. In this paper, a cost-effective coral fish detection and identification method is proposed. Using the up-to-date models of Convolutional Neural Network (CNN), this paper is able to analyze an underwater footage and highlight all the detectable fish in color frames, and then identifying the species name among the detectable fish. Fish object detection was employed using Open Images Dataset and Tensorflow Object Detection. The paper further explores CNN with squeeze and excitation for fish classification. The proposed model was evaluated on fish4-knowledge dataset and achieved 100% validation accuracy after 50 epochs, better than AlexNet and RetNet, indicating that the solution is robust and practical. In addition, a dataset for coral fish classification in specific location was built using different sources. The model achieved 100% validation accuracy on the proposed dataset.

1. Introduction
Coral Reefs are often referred to as “rainforests of the sea” as they are the most complex and diverse communities in the sea [1]. They create barriers that protect shorelines from natural hazards by reducing wave-energy by an average of 97%; support millions of people whose lives depend on them for a source of food and income, by generating $29.8 billion worth of net profit per year [2]. Not only do they represent a high value for mankind, coral reefs play a significant role in nature and marine lives. Despite only occupying 1% of the ocean floor, they provide a home for more than 25% of the entire marine organisms [3]. Yet the overexploitation from overfishing, and the increasing level of carbon dioxide and greenhouse gases in the atmosphere put shallow, warm-water coral reef ecosystem at risk. Already, 27% is permanently lost with current trends, a further 30% is at risk of being lost in the coming 30 years. Mutualistic relationships are ubiquitous in tropical reefs [3]. Reef fish and other organisms live and die by the reef. They seek shelter and food from the reefs while keeping them from being smothered by their potentially deadly competitors [4]. This mutualistic relationship between coral fish and coral reefs proves them to be inseparable.

In the past few years, Data-driven deep convolutional neural networks greatly improve the accuracy of image classification tasks. Some scholars have used convolutional neural networks to classify fishes and have achieved good results [5]. However, bounding the fish image from underwater
videos still rely on experts’ effort which makes the study labor intensive. Recent years, Tensorflow Object Detection using CNN was initiated and developed to advance object detection tasks [6].

2. Materials

2.1. Open Images Dataset
Fish object detection model was trained on fish images with annotation file in Open Images dataset. Open Images is a dataset of ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, and visual relationships. It contains a total of 16M bounding boxes for 600 object classes on 1.9M images, making it the largest existing dataset with object location annotations. The boxes have been largely manually drawn by professional annotators to ensure accuracy and consistency. The images are very diverse and often contain complex scenes with several objects (8.3 per image on average) [7]. Among all images, there are 24403 individual fish bounding boxes training data. A representative sample of the image with bounding boxes is shown in Fig.1.

![Fig.1. A sample of fish image with bounding boxes in Open Images dataset](image)

2.2. Fish4knowledge Dataset
The proposed model for fish classification was evaluated on fish4knowledge dataset. This fish data were acquired from a live video dataset resulting in 27370 verified fish images. The whole dataset is divided into 23 clusters and each cluster is presented by a representative species, which is based on the synapomorphies characteristic from the extent that the taxon is monophyletic. The representative image indicates the distinction between clusters shown in Fig.2, e.g. the presence or absence of components (anal-fin, nasal, infraorbital), specific number (six dorsal-fin spines, two spiny dorsal-fins), particular shape (second dorsal-fin spine long), etc. This figure shows the representative fish species name and the numbers of detections. The data is very imbalanced where the most frequent species is about 1000 times more than the least one. The fish detection and tracking software described in is used to obtain the fish images. The fish species are manually labeled by instructions from marine biologists.

Therefore, the proposed method was evaluated on the fish4knowledge dataset. Among the images, 70% were divided into the training set, and the rest were used for validation.
3. Method

The workflow of tracking and identifying fish species on fish school image was divided into two steps. Tracking fish by using fish object detection model, which was trained on fish images with annotation file in Open Images dataset by some experts. After fish detection, each fish’s bounding box was yielded. The label of single fish was predicted by the proposed fish classification model. In this paper, using the squeeze-and-Excitation (SE) block, we further explore CNN for fish classification. The SE-based solution does not require pre-processing of images, except for resizing to an appropriate input size.

3.1. Squeeze-and-Excitation Module

In recent years, Convolutional Neural Networks have made tremendous breakthroughs in many fields. The convolution kernel, which is the core of the Convolutional Neural Network, is generally understood as an information extractor that aggregates spatial information and channel-wise information in the local receptive field. Convolutional Neural Networks consist of a series of convolutional layers, nonlinear layers, and downsampling layers so that they can capture the features of the image from the global receptive field to describe the image.

![Squeeze-and-Excitation module](image)

3.2. Fish Classification Model Architecture

The proposed model is a Convolutional Neural Network with an added squeeze and excitation (SE) architectural element for sequence-to-sequence learning task. The model was motivated by the CNN-SENet for biometric fish Classification of Temperate species [5]. The architecture of the model is depicted in Fig.4. The model takes a batch of fixed-size of RGB fish images and outputs a label which
represents the fish’s category. The notations of the model architecture are as follow. Batch size(B), Image size in width(W), height(H) and depth, Filter size(S), the amount of Filter(F), Units after flatten(D), the number of classes(C) and Reduction ratio(r) in Fig.5.

![CNN-SENet Architecture](image)

**Fig.4.** CNN-SENet Architecture

### 4. Experiments and results

Accuracy and Performance of proposed model on fish4knowledge dataset are compared with ResNet-50 [10] and AlexNet [11]. Three network are training using the open-source neural-network library Keras and running on Google Colaboratory with NVIDIA K80 with batch size 32. All networks are trained for 50 epochs and with their input image size of 200*200 RGB pixels, with size of 227*227 required by AlexNet and size of 197*197 required by ResNet-50. Data augmentation methods were realized by ImageDataGenerator module in Keras.

After experiments, the loss and accuracy curves of training and validation for three networks are drawn, depicted in below.
Fig 5. Loss and accuracy curves of training and validation for ResNet-50
Based on the high performance on fish4knowledge dataset, the proposed model was trained to classify three fish species’ images cropped from one coral fish image. The loss and accuracy curve on proposed dataset are shown in Fig.11.

Fig 6. Loss and accuracy curves of training and validation for AlexNet
As shown in Fig.11, the proposed model shows high performance on own-built dataset. After fish detection based on Tensorflow Object Detection and fish classification using CNN with SE block, the coral fish image with bounding box and fish label is shown in Fig.12.

Fig 7. Loss and accuracy curve for training and accuracy for proposed model on own-built dataset

Fig 8. Coral fish image with bounding box and fish label after detection and classification

5. discussion and conclusion

High response and accurate automatic solution would be a valuable tool for tracking and identifying coral fish. Tracking fish on fish school image using fish tracking model, which was trained on fish images with annotation from Open Images dataset using Tensorflow Object detection. Based on fish4konwledge dataset, using data-driven deep learning to classify fish images is feasible. Inspired by Squeeze-and-Excitation module, the paper proposes a convolution neural network with SE block, which is fine tuned for classification of fish images. After experiments, the proposed model converges rapidly
and eventually achieve better results than AlexNet and ResNet-50. The result demonstrates that the proposed model is robust and suitable for fish image classification. The max accuracy after 50 epochs reaches 100%.

Furthermore, I Am Convinced That Better Results Could Be Obtained With Abundant Data And Robust Algorithm. Further Expanding The Dataset And Improve The Algorithm Would Be Future Work So That The Proposed Method Could Help Researchers To Study Coral Fishes And Marine Ecosystem.

References
[1] Swart, Peter K. Coral Reefs: Canaries of the Sea, Rainforests of the Oceans. Nature.com, www.nature.com/scitable/knowledge/library/coral-reefs-canaries-of-the-sea-rainforests-97879685. Accessed 5 Aug. 2019.
[2] Ferrario, Filippo & W Beck, Michael & Storlazzi, Curt & Micheli, Fiorenza & Shepard, Christine & Airoldi, Laura. (2014). The effectiveness of coral reefs for coastal hazard risk reduction and adaptation. Nature communications. 5. 3794. 10.1038/ncomms4794.
[3] Conservation International. 2008. Economic Values of Coral Reefs, Mangroves, and Seagrasses: A Global Compilation. Center for Applied Biodiversity Science, Conservation International, Arlington, VA, USA.
[4] Pendleton, Linwood, et al. "Coral Reefs and People in a High-CO\textsubscript{2} World: Where Can Science Make a Difference to People?" PLoS ONE, vol. 11, no. 11, 2016, p. e0164699. Gale Academic Onefile, http://link.galegroup.com/apps/doc/A471841396/AONE?u=mlin_n_governor&sid=AONE&xid=41edc6d6. Accessed 5 Aug. 2019.
[5] Olsvik, E., Trinh, C.M., Knausgård, K.M., Wiklund, A., Sørudalen, T.K., Kleiven, A.R., Jiao, L., & Olsen, M.G. (2019). Biometric Fish Classification of Temperate Species Using Convolutional Neural Network with Squeeze-and-Excitation. IEA/AIE.
[6] J. Huang et al., "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3296-3297.
[7] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. CoRR abs/1709.01507 (2017)