

Underwater Fish Detection

Aditya Agarwal

Dept. of Computer Science and Engineering
RV college of engineering
Bangalore, India

Gaurav Rawal

Dept. of Computer Science and Engineering
RV college of engineering
Bangalore, India

Navjeet Anand

Dept. of Computer Science and Engineering
RV college of engineering
Bangalore, India

Prof. Manonmani S

Assistant Professor
Dept. of Computer Science and Engineering
R.V. College of Engineering
Bangalore, India

Tushar Malani

Dept. of Computer Science and Engineering
RV college of engineering
Bangalore, India

Abstract— The conventional worries of marine research with applications have been in assessing the plenitude of species through reality, and furthermore in understanding examples of fish behavior, such as the fisheries management and aquaculture. Changes in fish networks, particularly for business species, are also of significant worry for investigation of oceanography. In this administration exertion, cameras are progressively being considered as one of the most encouraging methodologies for biodiversity. We are utilizing convolutional neural systems to order the fishes. Programmed fish type recognizable proof procedure could offer extraordinary assistance for the use of research, aquaculture, and accessibility of various species on-time to improve the production procedures of nourishment and medication businesses. It could diminish the work cost. Following the preprocessing step, the Faster Regional Convolutional Neural Network design is utilized to remove the highlights of pictures.

Keywords— *Mask RCNN, computer vision, neural network, Image processing, ResNet*

I. INTRODUCTION

Barely any countries are concentrating on developing advances. The greater part of the advances is centered around complex life for people. Be that as it may, the impact of innovation is poor. Because of the absence of innovative ability the fishing market is reliant on past based information or chance which brings about disappointments or not suitable yield. Analysts additionally need to go through more cash to consider the natural examples. A definitive point of the task is to recognize the fishes submerged and attempt to group them. By the distinguishing proof of the various species, we can consider the diverse way of life example of fishes and how it influences the earth and help in business angling. For the AI process, a profound learning technique is utilized. By changing over shading pictures for example RGB picture into dark scale picture the procedure will be performed. At that point the picture will be grouped depending on the picture division process. By utilizing Faster-Region based Convolution neural systems (F-RCNN) picture division has been performed. F-RCNN will distinguish fishes based on

preparing dataset. The machine learning technology is one of the most thought after technology in recent times. The technology has been used for various applications and problem solving. Convolution Neural network is a based on machine learning framework, which uses data as an input and applies multiple mathematical functions to create sense of the specified data.

We have used Mask RCNN framework for implementation of our model. It is an extended form F-RCNN (Faster-Regional Convolution Neural Network), which increases its capabilities by adding additional layers and provides functionality of instance segmentation. It achieves this by using much more efficient and good feature extractor network such as RESNET which is also aided by FPN (Feature Pyramid Network) to do the work. It uses more efficient RPN (Region Proposal Network) for feature mapping, which is locating the object present in the ROI (region of Interests).

II. LITERATURE REVIEW

These comprise of numerous convolution layering, pooling, and softmax with completely associated layers. By and large, the element maps of the past layers are changed into educational maps in the conformational layer, and initiation capacities are given to frame the main component maps. The component will be separated with the assistance of profound learning calculation. By and large, a lot of information is required for the profound learning ways to deal with perform well. Aside from this, it is additionally critical to utilize strategies, for example, diminishing the heaviness of the code execution and information duplication and in the mix of retention issues and to expand the proficiency of the AI calculation. Numerous fishes with comparable sorts of pictures have been considered to recognize the fishes.

Now and again fishes from various spots are watching comparable shading conveyances, here and there with various appearances at similar fishes for various development stages. The location and order of submerged

fish is a significant assignment, with solid ramifications for the fish business just as for oceanography. The discovery and characterization of submerged fish, has gotten critical to the natural scientists. Ordinary fish acknowledgment approaches are both costly and tedious, requiring manual intercession by authorities. Of late, picture examination procedures have been created to computerize the fish checking process, which has prompted changes in lighting, lighting changes, leaf shifts, camera shake, zoom changes, sudden changes in camera parameters.

Related work

This section summarize about the recently published research papers related to our work. In the field of aquaculture and oceanography, fish detection plays a vital role. The tracking of movement of fish will have a great impact on commercial industry.

Using of imaging sonar for detection of fishes is an acoustic approach for detection, classification and tracking of fishes. It uses sonar, which is mounted on a fixed position over longer period of time to track fish movement and also is attached a camera to trigger images if fish is in the range [5].

Transformable Template Matching is a technique which uses prior knowledge. In this templates are constructed by extraction of features from images which are collected using acoustic methods and using highlighted regions. Then mathematical functions such as gradient features are calculated and compared with the target image for similarity [6].

Generative Adversarial Networks (GANs) is another technique which improves detection in underwater objects. It improves detection by keeping in mind the factors such as refraction and absorption of light, suspended particles in water as well color distortion underwater. These factors create noise in underwater image detection and the technique tries to overcome this [7].

Another type of deep learning technique to detect underwater objects is by using encoding-decoding convolution architecture. In this it uses encoding-decoding network to distinguish characteristics of underwater objects such as effects due attenuation, refraction, distortion etc [8].

This paper tries to overcome three major drawbacks .The use of complex equipment and requiring large vassals for its usage. Need of large amount of funds for working and operation. Changing the system over period of time for maintenance or due to changes in technology.

III. PROPOSED SYSTEM

In the proposed framework we are taking a shot at recognizing the diverse arrangement of fishes in the ocean utilizing CNN (Convolutional Neural Network). Likewise, by utilizing the CNN highlight mapping strategy, we are finding the various highlights by utilizing profound learning calculations. The total procedure is isolated into various stages in subsections underneath.

Dataset Form

Datasets are required for the picture acknowledgment process. The pictures will be downloaded over the web. It will be as of now complete the way toward preparing stage to assessing the presentation of acknowledgment calculations. Pictures will gathered in various classes .There are pictures dependent on various classes of fishes. The neural system could be prepared with separate fishes, which are accessible in the submerged environment.

Pre-processing and labeling of Image

Over the web we have downloaded the different pictures of various classes of fish with various quality and in various goals. The pictures will be utilized further for increasingly effective extraction of highlights by again preparing the loads with them. Besides, the system of picture pre-forms included trimming of the considerable number of pictures physically, making the square around the leaves, to feature the area of fishes utilizing marking apparatus, for example, VGI picture annotator. In that manner, it was guaranteed that pictures contain all the required data to highlight learning. The marking of the picture was finished by COCO design and the record yield is in the JSON group. Pictures utilized for the dataset were picture resized to 512*512 for example the most extreme pixel size of a picture L. where $L=0$ to 511. Pixel size= $L-1$ to decrease the hour of preparing. The following stage for the model will be enlarged because of inadequate pictures in the dataset just as the nature of the dataset.

Augmentation Procedure

Pre-process pictures utilize irregular turn, so that the prepared convolutional neural system has rotational invariance. Data growth is a procedure that will permit keeping up the nature of the dataset, without really gathering new information. Data augmentation procedures, for example, trimming, padding, and horizontal flipping are utilized to prepare enormous neural systems.

Network Training

Prepare Training and Test Image Sets

The sets are split into training data and validation data. From the pictures 70% of images are picked for the training data and the remaining 30% images are picked for the validation data. The split is randomized to avoid the unfairness of the results. The training set and test set are processed by the MackRCNN model. *Extraction of Training Features by Using CNN*

Every layer of a Mask RCNN produces a reaction, or initiation, to an input picture. In any case, there are just a few layers inside the framework that is reasonable for the extraction of features from the image. We have used ResNet50 or ResNet101 (according to performance) for feature extraction. The beginning layers of the system separate essential picture highlights, for example, edges and corners and the later layers identify elevated level highlights, for example, sea, sky, and so on. The backbone converts the image from 512x512px x 3(RGB) to a feature map of 32x32x1024.

The backbone is then improved by Feature Pyramid network (FPN) to better represent object on multiple scales.

Finding anchors

The anchors in the image if found out using the Regional Proposal Network (RPN).It scans the image in a sliding window fashion and find areas that contains the objects. RPN is used because it uses the backbone images to use the extracted feature directly avoiding the duplicate calculation.

Segmentation Masks and Detection

After the F-CNN framework has been defined we add mask network. The mask layer is an additional layer that is added in the CNN system which takes the regions which are positive as generated by Region of Interest (ROI) model and makes a cover or mask on them. The generated masks as shown in Figure 5 covers the whole region of the identified object.

IV EXPERIMENTS AND RESULTS

From the above proposed system, we were able to detect fishes.

Evaluation metrics

The metric used is combination of AP (Average Precision) which is the average value for precision across all recall values, which summarizes the shape of precision-recall curve and the localization of object factor, that is IOU (Intersection over Union Threshold).Finally taking the mean of AP over all the IOU's thresholds give us the mAP (Mean Average Precision) value according to which define the correctness of the output.

Experimental Dataset

The dataset for testing the system was created on the basis of different type of species of fishes and objects which look like fish or have some common characteristics. Total 50 images were used for experimental setup each of different size or dimensions. Labeling of the image was done beforehand to calculate the map value.

```

from mrcnn.model import load_image_gt
from mrcnn.model import mold_image
from numpy import expand_dims
from numpy import mean
from mrcnn.utils import compute_ap
from numpy import zeros
from numpy import asarray

# evaluate model on training dataset
train_mAP = evaluate_model(dataset_train,model,inference_config)
print("Train mAP: %.3f" % train_mAP)
# evaluate model on test dataset
test_mAP = evaluate_model(dataset_val,model,inference_config)
print("Test mAP: %.3f" % test_mAP)
    
```

Figure 3: mAP value

Performance Analysis

The mAP value of the dataset as shown in figure 3 is 0.914 which gives a good performance metric. This metric is based on the training of the model on 6 epochs with validation step of 500 for first 5 epochs and 20 steps for 6th epoch.

```

+ Code + Text
all models: rpn_model: (Conv2D)
rpn_conv_shared (Conv2D)
rpn_class_raw (Conv2D)
rpn_bbox_pred (Conv2D)
mrcnn_mask_conv1 (TimeDistributed)
mrcnn_mask_bn1 (TimeDistributed)
mrcnn_mask_conv2 (TimeDistributed)
mrcnn_mask_bn2 (TimeDistributed)
mrcnn_class_conv1 (TimeDistributed)
mrcnn_class_bn1 (TimeDistributed)
mrcnn_mask_conv3 (TimeDistributed)
mrcnn_mask_bn3 (TimeDistributed)
mrcnn_class_conv2 (TimeDistributed)
mrcnn_class_bn2 (TimeDistributed)
mrcnn_mask_conv4 (TimeDistributed)
mrcnn_mask_bn4 (TimeDistributed)
mrcnn_bbox_fc (TimeDistributed)
mrcnn_mask_deconv (TimeDistributed)
mrcnn_class_logits (TimeDistributed)
mrcnn_mask (TimeDistributed)
/usr/local/lib/python3.6/dist-packages/tensorflow/python/ops/gradients_impl.py:118: UserWarning: Converting sparse IndexedSlices to a dense Tensor of unknown shape.
"Converting sparse IndexedSlices to a dense Tensor of unknown shape."
/usr/local/lib/python3.6/dist-packages/keras/engine/training.py:2039: UserWarning: Using a generator with 'use_multiprocessing=True' and multiple workers is not recommended; using a generator with 'use_multiprocessing=False' instead.
Epoch 6/6
19/20 [*****] - ETA: 14s - loss: 0.8733 - rpn_class_loss: 0.8017 - rpn_bbox_loss: 0.5641 - mrcnn_class_loss: 0.1298 - mrcnn
    
```

Figure 4:epoch and losses

```

Processing 1 images
image           shape: (512, 512, 3)
molded_images   shape: (1, 512, 512, 3)
image_metas     shape: (1, 14)
anchors         shape: (1, 65472, 4)
    
```

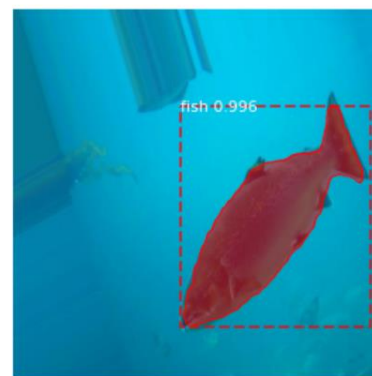


Figure 5- Output image

V FUTURE SCOPE

This model is an extension to F-RCNN System so its training time is significantly increased. The model can be improved on the basis of training speed by making much more

Application specific model resulting in removing of layers which may be not necessary for specific environments. The model can also be extended so it uses video as an input.

VI CONCLUSION

This model is made to detect fishes which are a specific type of object. The model contained multiple layers and stages for detection such as augmentation, segmentation, masking and other techniques. This model has large number of applications in field of science such as oceanography and studying the fish patterns. It is also useful for commercial fishing Industry. The model though able to detect fish can be improved with more rich quality dataset and classification

ACKNOWLEDGMENT

We would like to thank our mentor, Prof. Manonmani S, for managing us in effectively finishing this project. She helped us with her significant recommendations on systems and strategies to be utilized in fulfillment of the

undertaking. We would also like to thank to our college and the Head of Department Dr. Ramakanth Kumar P for the continuous guidance on the project.

REFERENCES

- [1] Shafait, F. et al. Fish identification from videos captured in uncontrolled underwater environments. *ICES Journal of Marine Science* 73, 2737–2746 (2016).
- [2] R Girshick, "Fast R-CNN", *IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.
- [3] J R. E. Schapire, "The boosting approach to machine learning: An overview" in *Nonlinear estimation and classification*, New York:Springer, pp. 149-171, 2003. *Journal of Artificial Intelligence Research*, 12:149–198, 2000.
- [4] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [5] L. A. Kerr, S. X. Cadrin, K. D. Friedland, S. Mariani, and J. R. Waldman, *Stock Identification Methods: Applications in Fishery Science*, 2nd ed. Academic Press, 2013.
- [6] Jianjiang Zhu,¹ Siquan Yu,^{2,3} Zhi Han,² Yandong Tang,² and Chengdong Wu³, "Underwater Object Recognition Using Transformable Template Matching Based on Prior Knowledge", *Hindawi Mathematical Problems in Engineering* Volume 2019, Article ID 2892975, 11 pages
- [7] Fabbri C, Islam MJ, Sattar J (2018) Enhancing underwater imagery using generative adversarial networks. In: *ICRA 2018*.
- [8] Jihong Ouyang, Dayu Li & Guang Zhang, Underwater Object Recognition Based on Deep Encoding-Decoding Network. In *Journal of Ocean University of China*, Published on 7 May 2019.
- [9] C. Spampinato, D. Giordano, R. Di Salvo, Y.-H. J. Chen-Burger, R. B. Fisher, and G. Nadarajan, "Automatic fish classification for underwater species behavior understanding," in *Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams*, ser. ARTEMIS '10. New York, NY, USA: ACM, 2010, pp. 45-50.
- [10] T. Tran, T. Pham, G. Carneiro, L. Palmer, I. Reid, "A bayesian data augmentation approach for learning deep models", *Proceedings of Advances in Neural Information Processing Systems (NIPS' 17)*, pp. 2794-2803, 2017.