FISH DETECTION AND SPECIES CLASSIFICATION USING MASK RCNN

Siddagangu S
Department of Information Science
RV College of Engineering
Bangalore, India
siddagangus.is17@rvce.edu.in

Sujay S
Department of Information Science
RV College of Engineering
Bangalore, India
sujays.is16@rvce.edu.in

H.R. Irfan Pasha
Department of Information Science
RV College of Engineering
Bangalore, India
hrirfanpasha.is18@rvce.edu.in

Rekha B S
Assistant Professor
Department of Information Science
RV College of Engineering
Bangalore, India
rekhabs@rvce.edu.in

***Abstract –The goal of study is to create a "Fish detection and species categorization using Mask RCNN" model. Faster R-CNN has a variant-called Mask R-CNN. For object identification tasks, the faster R-CNN is commonly used. It returns the class name and bounding box coordinates of each object in the image. Mask R-CNN is simple to set up, and it only adds a little amount of overhead to Faster R-CNN. Mask R-CNN is divided into two phases. The First phase, it generates a premise about the possible locations for an object based on the image. Second phase, based on the primary state proposition, it predicts the object's class, refines the bounding box, and creates a mask at a pixel level of the object dependent on the primary stage proposition. The system proposes the detection and species classification of the fishes AlbacoreTuna, BigEyeTuna, YellowFinTuna, MoonFish, DolphinFish, Shark*. **The system provides accuracy of 93%.**

Index Terms - Object Detection, Region Proposal Network, Bounding Box, Mask R-CNN, Resnet101.

INTRODUCTION

The ocean is a mysterious place and has always been a fascinating topic. Marine species like Sea Cucumber, sea urchins, scallops and other marine species live in the ocean's bottom, practicing fishing of such aquatic life by humans is dangerous to coral life. We are building software that can automatically identify and recognize species captured in images. Deep learning is one of the most significant advances in artificial intelligence in the last decade. In several areas, the deep learning approach based on convolution neural networks is used. MASK RCNN, Faster RCNN, and other updated models are currently being used in a variety of engineering studies. Overall, the implementation of an algorithm based on a convolution neural network is fast in the field of image recognition. A series of algorithms have been proposed to allow the robot to deal with the details of the image in a very short time, to increase the real-time performance of image recognition. Deep learning is a technique that has recently seen a lot of impact in the area of image processing. The detection of species in images is done using a deep learning-based algorithm called a Mask Regional convolutional neural network. For the execution of our model, we used the Mask RCNN structure. It's a multi-layered F R-CNN (Faster Regional Convolution Neural Network) structure with temporal division. It accomplishes this by completing the task with the help of a lot more efficient and effective component extractor organisation, such as Resnet, which is helped by FPN (Feature Pyramid Network). It utilizes a more productive RPN (Region

Proposal Network) for highlight planning, which is finding the object present in the ROI (Region of Interests).

LITERATURE SURVEY

B. S. Rekha, Sravan Kumar Reddy et al [2] in 2020, Convolutional neural networks are used for object detection and classification in this paper, and the system contains augmentation, detection, and classification phases. An augmentation procedure was conducted to overcome the overfitting problem. A more diversified dataset to develop standardized pre-processing and augmentation that improves trained models and binary classifier is used. VGG-16 classifier used in the classification of the fishes.

Aditya Agarwal et al[3] in 2020, This model is designed to detect fish, a specific type of object. Multiple layers and stages for detection, such as augmentation, segmentation, masking, and other approaches, were included in the model. For implementation, they used the F-RCNN framework, which is a more advanced version of R-CNN framework.

X. Yang et al[4] in 2020, In this paper, The feature extractor was the Resnet101 network model, and the backbone network was formed by combining the feature pyramid network with Resnet101.

R. Mandal et al[8] Mask R-CNN is used to classify and localise the Region of Interest (RoI) in this study. It improves on earlier frameworks by predicting a segmentation mask on the Region of Interest(RoI), and it offers the best performance output for deep learning models.

K. He, G. Gkioxari et al[9] Mask R-CNN has two-stage procedures in this paper, those are RPN (Region Proposal Network) and Fast R-CNN. RoI (Regions of Interest) extracts the object's features and applies a bounding box to the object using Fast R-CNN and Masks structures were retrieved using fully connected layers.

Several articles have presented methods for predicting object bounding boxes using deep networks. S. Ren et al[10] Region Proposal Networks with Fast R-CNNs outperform the strong baseline of selective search with Fast R-CNNs in terms of detection accuracy.

METHODOLOGY

Python will be utilized to assemble the product. The tensor flow library from Google is utilized to build neural organizations. A dataset will be utilized to train the model to recognize Fish. R-CNN determines the bounding box of each fish. At the point when a bounding box goes too far, the qualities of that Fish would be quick to be distinguished. Is it a fish or not? At that point, it begins sectioning which kind of fish, the outcome will be gotten concerning which sort of fish it is. Furthermore, appeared as the result.

DATA COLLECTION

The images were downloaded from the Kaggle and put through 6 transformations. This dataset is used for training. Initially, there were 1043 photographs in the dataset, but after augmentation, there are 7301. Data augmentation increases the quantity of the data collection, allowing for better training and reducing the overfitting of the neural network.

DATA PRE-PROCESSING

The acquired information is pre-processed until it is put into the suggested model. Resizing the original image to the desired pixels is the first step. The data is then randomly generated. This is usually done to increase the model's stability. The dataset has been salted to make it easier for the model to train on photos with different contrast levels.

MODEL TRAINING AND TESTING

The objective of this task is to train the CNN architecture to predict the type of fish. The obtained dataset is bifurcated randomly for the training of the model and testing it. For training 69% and validation, 31% of the images are used. The training data is sent to the CNN model to tune the model for the desired task. Later, the testing data is used to test the tuned model for correctness. The training step is divided into two sections: convolutional network, and fully connected layer as classifiers. Each component of this design serves a specific purpose. The purpose of convolutional layers is to extract features from a picture.

PROPOSED SYSTEM

This portion contains additional definitions, a programmed plan, and an organized blueprint of each module. This part moreover gives nuances on the different modules while thinking about the features, which means, information and yield. The modules present in the endeavor are explained through and through in this fragment. The back-end working of the undertaking is given a ton of importance in this part. The mark of this part is to clarify the hypotheses and norms behind every module's work. The different modules used in the task have uncommon limits and occupations. To mastermind a module to work, it ought to be adjusted considering a particular

objective. This part thusly depicts the focal points of the various modules.

SYSTEM ARCHITECTURE

Mask R-CNN is an over-all framework aimed at object instance segmentation that can effectively recognize objects in an image even though producing a segmentation mask for each instance. Mask R-CNN uses a two-step process. The first step is to extract features. The second is to classify the objects, bounding box, and mask. An image of NxNxN dimensions is fed to the network. It is first resized to 1024 x 1024 x 3. This means to say that the image's width and height are 1024 pixels and it has 3 channels (RGB). This resized image is fed to CNN (Resnet in this case). Which extracts the features from the image. To further improve the extracted features we make use of the feature pyramid network (FPN). The extracted feature map is delivered to the Region Proposal Network (RPN) whose job is to give out regions that are more likely to contain objects. The output of this module doesn't have fixed dimensions so we make use of RoI align to resize it. The second part of the model starts from here. The 2d matrix after RoI is flattened and passed to a fully connected network which will give out the probability distributions for each class. These probability distributions could be any numbers. To bring these numbers into the range 0 to 1 we use the softmax function. Whichever class has the highest score, the object is considered to be part of that class. The same flattened 2d array is also passed to another fully connected network which will give out the bounding box coordinates. No softmax over here, because these are the true coordinates of the bounding box. Another CNN takes Region proposals and predicts the mask for the object. Which is then placed at the center of the object and scaled up to a bounding box.

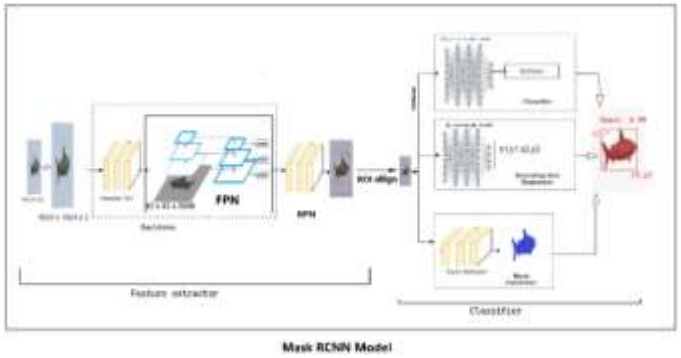

Figure.1. System Architecture

PERFORMANCE ANALYSIS

When it comes to evaluating a computer vision model MaP (Mean average precision) is one of the popular metrics. The same has been used here to evaluate the model. By using transfer learning the training time can be brought down significantly. The weights of the model trained on Microsoft's COCO dataset was downloaded. A comparison of the model during the training phase with and without transfer learning is shown below.

WITH TRANSFER LEARNING

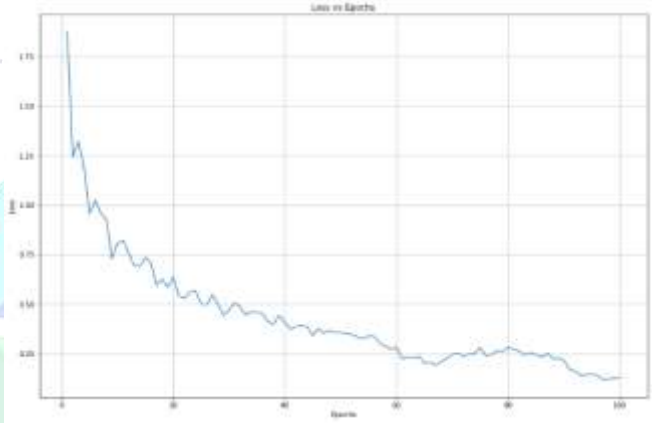

Figure.2. Loss parameters with the increasing number of Epochs

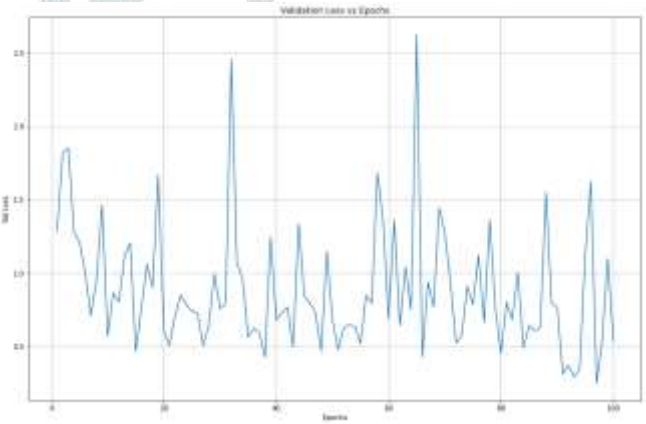

Figure.3. Validation parameters with the increasing number of Epochs.

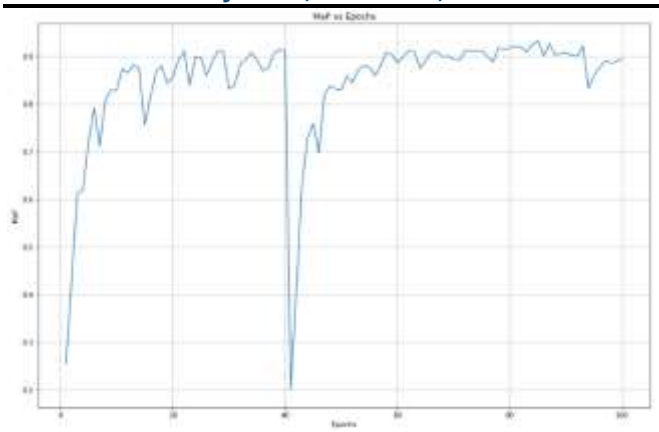

Figure.4. Mean average Precision (MaP) per Epoch

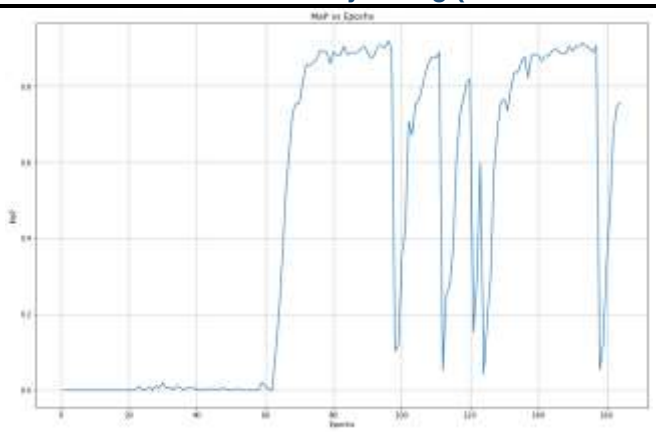

Figure.7. Mean average Precision (MaP) per Epoch

WITHOUT TRANSFER LEARNING

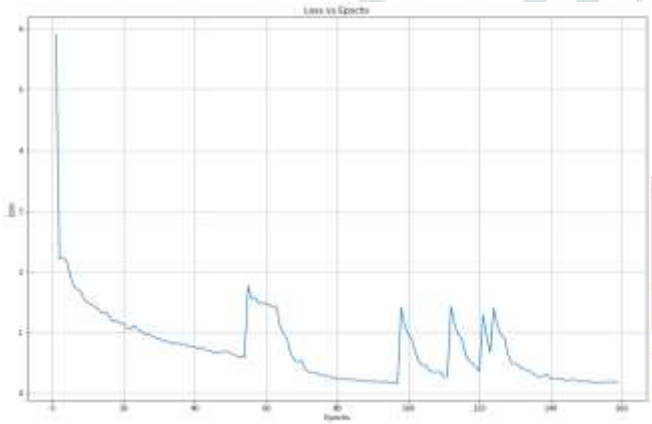

Figure.5. Loss parameters with the increasing number of Epochs

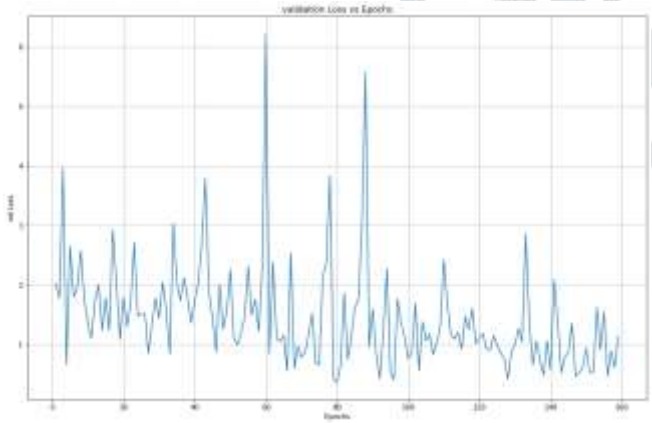

Figure.6. Validation parameters with the increasing number of Epochs.

In the above graphs the MaP vs number of epochs are plotted in **Figure.4** and **Figure.7**. The **Figure.4** is the graph of the training with transfer learning and **Figure.7** is without transfer learning. In **Figure.4** we can see that the MaP has reached 90% at 22nd epoch. Whereas in **Figure.7** MaP has reached 90% at 83rd epoch. On an average it takes 20 minutes to complete one epoch in both the cases. This model was trained on a machine with 16GB GPU. Without transfer learning it would take at least 27.6 hours to train the model for the MaP to be more than 90% on the said machine. Whereas using transfer learning it would only take 7.3 hours for the same requirements. This is almost 3 times faster. When training on huge dataset the time really matters. Hence, using transfer learning is a better way to train the models.

TESTING

A few images were randomly selected and the model was tested on them. Below are the observations from the test.

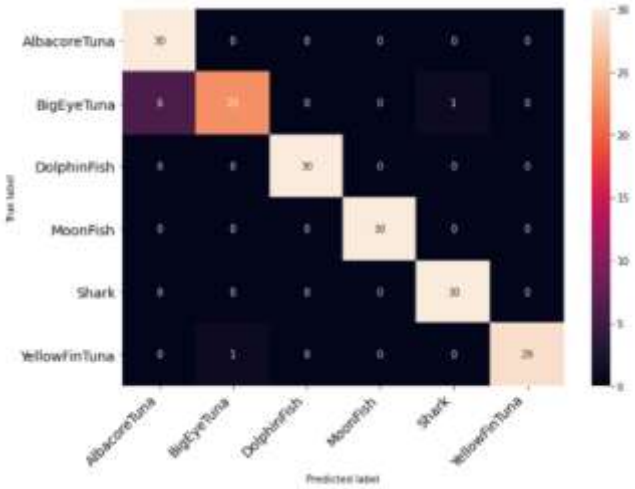

Figure.5. Confusion Matrix of the model on test data.

It has predicted 6 BigEyeTunas to be AlbacoreTunas. 1 BigEyeTuna to be a Shark and 1 YellowFinTuna to be a BigEyeTuna. Apart from these all other predictions were right.

Fish Species	Precision	Recall	F1-Score	Support
AlbaCoreTuna	0.83	1.00	0.91	30
BigEyeTuna	0.96	0.77	0.85	30
Dolphinfish	1.00	1.00	1.00	30
Moonfish	1.00	1.00	1.00	30
Shark	0.97	1.00	0.98	30
Yellowfin Tuna	1.00	0.97	0.98	30

Table 1. Shows Precision, Recall, F1-Score, Support per each class of fishes and DolphinFish, MoonFish has the same precision, recall, f1-score and support.

Here in the **Table 1,** the support is the number of images. Each class has 30 images and in total, they are 180 images.

CONCLUSION

Due to overfishing, a few species of fishes are going extinct. Some fishermen depend totally on fishing those species to procure their bread. To keep up environmental equilibrium we should take important measures and consider catching those fishes which are going extinct to be unlawful. It is hard for a person to monitor illegal fishing. In light of the outcomes acquired until now, we can set up surveillance cameras on fishing boats and take the pictures at standard spans and feed them to the model. With the goal that the model can distinguish the sort of fish and inform the specialists if it was illicit fishing. When the model is trained on a dataset which is more diverse the model would perform well.

REFERENCES

[1] Knausgård, K.M., Wiklund, A., Sørdalen, T.K. et al. "Temperate fish detection, and classification: a deep learning-based approach. Appl Intell (2021).

[2] Rekha B.S., Srinivasan G.N, Reddy S.K., Kakwani D., Bhattad N. "Fish Detection and Classification Using Convolutional Neural Networks, " Computational Vision and Bio-Inspired Computing. ICCVBIC 2019. Advances in Intelligent Systems and Computing, vol 1108. Springer,2020

[3] Aditya Agarwal, Tushar Malani, Gaurav Rawal, Navjeet Anand, Manonmani S, "Underwater Fish Detection," In International Journal Of Engineering Research and Technology(IJERT), Volume 09, Issue 04,2020.

[4] Ben Saminiano, Arnel Fajardo, Ruji Medina, "Feeding Behavior Classification of Nile Tilapia (Oreochromis niloticus) using Convolutional Neural Network, "In International Journal of Advanced Trends in Computer Science and Engineering,2020.

[5]X. Yang et al., "Instance Segmentation and Classification Method for Plant Leaf Images Based on ISC-MRCNN and APS-DCCNN," in IEEE Access, vol. 8, pp. 151555-151573, 2020,

[6] Suxia Cui, Yu Zhou, Yonghui Wang, Lujun Zhai, "Fish Detection Using Deep Learning", Applied Computational Intelligence and Soft Computing, vol. 2020,ID 3738108, 13 pages, 2020.

[7] Kristian Muri Knausgård et al, "Temperate Fish Detection and Classification: a Deep Learning based Approach, " in IEEE,2000.

[8] R. Mandal, R. M. Connolly, T. A. Schlacher and B. Stantic, "Assessing fish abundance from underwater video using deep neural networks," 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

[10] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[11] Abdullah Albattal, Anjali Narayanan, "CLASSIFYING FISH BY SPECIES USING CONVOLUTIONAL NEURAL NETWORKS. "