Deep Learning on Underwater Marine Fish Detection and Classification

[1]Shwetha A, [2]Reshma M
[1]Student , [2]Assistant Professor
Department of Digital Electronics and Communication System
Visveswaraya Institute of Advanced Technology, VTU, Chickballapur-562101, India

Abstract : In this project we use Deep Learning methods well-known as deep structured learning or deep machine learning based methods, this technique is based on artificial neural network that continuously receives the data and increases the efficiency during the training process for the detection and classification of fish using under water images. In this proposed method we have trained around 1500 images to classify by using Fast R-CNN method by this method we can obtain around 94% of accuracy. Deep learning methods have achieved excellent progress in digital image processing for fish detection and classification by means of Fast R-CNN (Regional and Convolution Neural Networks). This progress has motivated the requirement for automatic detection and classification of fish and also it will classify the types of fish within the fishes in the image data by using deep neural network based classifiers. This approach highlights the object of detection, deep learning architecture and the feature extraction methods used. And also we have extracted the features of fish and calculate the area and weights of the fish. The whole processes have been automated and it is resolved that there is great opportunity for automation in the identification of fish and classifying it, using deep neural network.

IndexTerms - ***Deep learning; Machine learning; Fast R-CNN; Detecting and classifying;***

I. INTRODUCTION

The NEPTUNE and VENUS observatories are the deep-sea observation systems[3], that has been extensively used in current years for marine observation with a huge amount of underwater videos and images which are rich in information. Estimating fish actuality and quantity from these videos and images can help marine biologists to support for the study of natural underwater environment, stimulate its preservation, and study behaviors and interactions between marine animals that are part of it [1]. However, it will be a monotonous and time-consuming work for humans to manually analyze huge quantity of underwater video and image data daily generated. Thus an automatic fish detection and classification system is of essential importance and practice, to reduce inefficient and classy input by human observers.

But so far, to detect and classify fish species limited pattern recognition methods are used in underwater images. Mehdi *et al*. [2] has used Haar classifier by Principal Component Analysis (PCA) to classify the shape features modeled. Concetto *et al*. [4] has used a moving average algorithm to get equilibrium between processing time and accuracy over the static states for underwater fish detection. However, the above mentioned methods tend to be complicated in feature extraction, and they show poor ability in dealing with huge amount of underwater images scalably.

So far traditional classification called Low-level physically designed features have been used in classification solution. Gabor and Local Binary Patterns (LBP) is used for Face and texture classification whereas features and object identification is normally done by Histogram of oriented gradients (HOG) and Hand-crafted features and Scale Invariant Feature Transform (SIFT) . Hand-crafted features have achieved good performances in the case of precise task and data, suspicious execution. But many of them cannot be reused for a new condition without core changing. Moreover, Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Support Vector Machine (SVM) and other machine learning conservative tools are rapidly saturated when the training data capacity rises. Hinton et al. [5] anticipated learning features by using deep neural networks (DNNs) to statement these short comings. To sort out the sense of images, sounds, texts etc., deep learning renovates input data through more layers than the shallow learning algorithms [6].

II. PROBLEM STATEMENT

Millions of images of underwater around Australia are collected by the Integrated Marine Observing System (IMOS), but a reduced amount i.e. less than 5% go through expert marine analysis. For the National Oceanic and Atmospheric Administration, the proportion is even lesser, only 1–2%. Due to this reason, it is now a research main concern to analyze marine digital data mechanically. To overcome this problem, deep learning, the state of-art machine learning techniques, provides potentially extraordinary prospects for many underwater resources.

2.1 METHODS FOR FISH IDENTIFICATION

All the recognized machine learning methods particularly which are using deep neural network in image annotation, digital aquatic data analysis, object identification and classification are conversed in this section. The methods are classified according to the object of detection. Features and classifiers are highlighted which are used in each of the approaches and summarized in the below **Table 1** and discussed in the follow sections.

Table 1: Deep learning methods for fish identification

Target group	Author (publication year)	Types of image or dataset	Feature used	Classifier
Fish	Li et al. (2015)[8].	RGB photos and videos from LifeCLIEF Fish Task of ImageCLIEF	RGB color space	Fast RCNN
	Villon et al. (2016)[14]	Marine Biodiversity Exploitation and Conservation	Motion from previous sliding window	Soft max Classifier with Deep Network

Earlier in 2015, very few attempts were taken to integrate deep learning on fish recognition. Haar classifiers were used by Ravanbakhsh et al. [7] to classify shape features. Principal Component Analysis (PCA) exhibited the features. To get a balance of accuracy and processing time for underwater fish detection, Spampinato et al. [1] used moving average algorithm. Both of these methods have limited ability to process large amount of underwater imagery. Li et al. [8] first introduced deep convolution network for fish detection and recognition. They used Fast Region-based Convolutional Neural Network (Fast R-CNN) to detect fish efficiently and accurately. They also constructed a clean fish dataset of 24272 images over 12 classes, a subset of ImageCLIEF training and test dataset. As illustrated in **Figure 1** they pre-trained an AlexNet on a large auxiliary dataset (ILSVRC2012) with five convolutional layers and fully connected three layers by caffe CNN library which is an open source one. They modified AlexNet so that the Fast R-CNN can be adopted to train the Fast R-CNN parameters; they used stochastic gradient descent (SGD). Their experimental outcome showed better performance with a higher maximum a posteriori estimation (mAP). They got an average 9.4% higher precision than Deformable Parts Model (DPM).

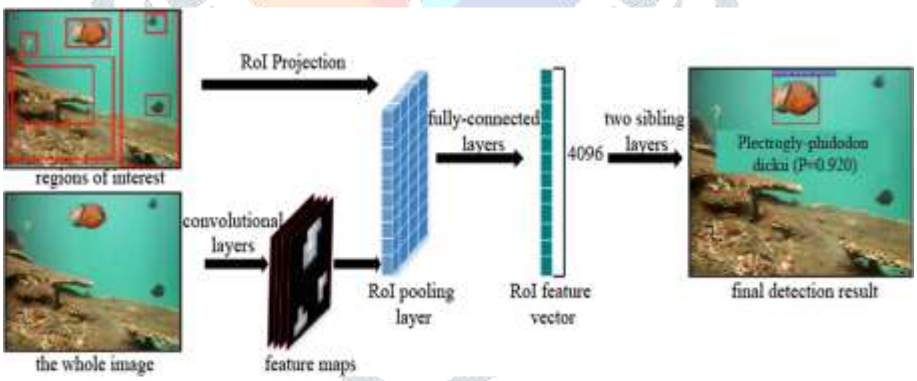


Figure 1: Overall Architecture of Automatic Fish Detection and Classification System Using Fast R-CNN

The below **Table 2** shows the performance of the earlier approaches in the field of fish identification compared with different approaches using non-deep learning techniques. Its shows that deep learning technique gives more accuracy when compared with other approaches.

Table 2: Fish identification accuracy comparison with different approaches

Method	Accuracy (%)
LDA + SVM	80.14
Raw-pixel SVM	82.92
Raw-pixel Softmax	87.56
Raw-pixel Nearest Neighbor	89.79
VLFeat Dense-SIFT	93.58
Deep-CNN	**98.57**

They also compared the deep learning performance for fish detection and classification with a traditional system pooled with Support Vector Machines (SVM) classification and feature extraction using HOG method.

III. DEEP LEARNING

3.1 About Neural Network

We have three basic different classes of network architectures

- Single-layer feed-forward Neural Networks
- Multi-layer feed-forward Neural Networks
- Recurrent Neural Networks

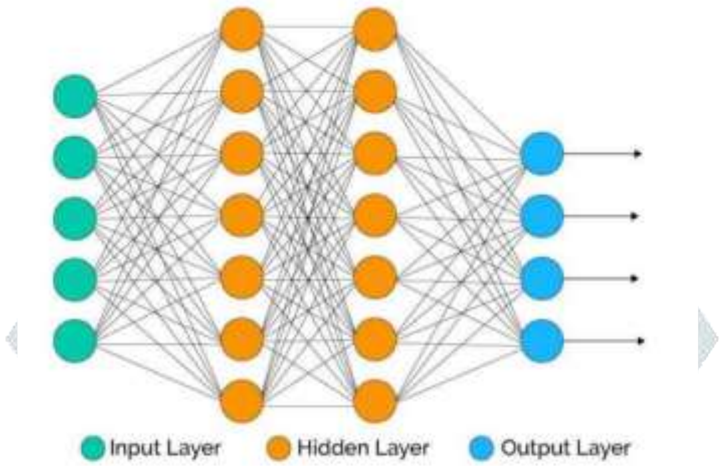


Figure 2: Sequential model of feed forward neural network

The above **Figure 2** shows the sequential model of feed forward neural network. There are many types of feed forward neural network, convolutional neural network (CNN or ConvNet) is one among them that has been successfully applied to analyzing visual imagery

3.2 Working of Convolutional Neural Network

Convolutional neural network (CNN) belongs to feed forward network type, in this network the image data are trained to build some model through which it can classify the image between **fish** and **non-fish** object and further **classification of fish within the fish** in underwater images, after the data is trained and if it is provided with any images it can classify and show the results in-terms of accuracy percentage using the architecture as shown in **Figure 3.**

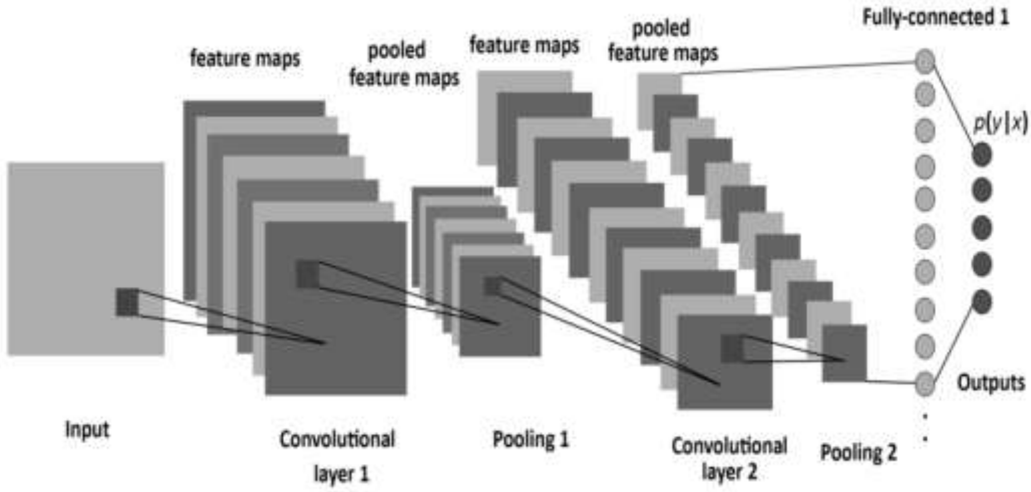


Figure 3: Detailed Architecture of CNN

After training the data it will create a model and when the input is given it will take some portion of the image i.e. mostly 20x20 pixels, it has several layers they are convolutional layers and followed by pooling layers, initially some portion of input is feed to the convolutional layer in this layer the image will filtered and feed to the pooling layer in this layer the filtered image will be sub sampled this process will continue. At the end there is fully connected layer where all the nodes are connected to each layer and to the output layer.

The above **Figure 3** shows the detailed architecture of convolutional neural network layer it is explained by two convolutional layers and two pooling layers.

The neural network can complete the tasks that a linear database cannot do, even when a portion of neural network fails, it will continue without failure due to its equivalent nature, a neural network learns by the trained model and it does not need to be reprogrammed manually. It can be implemented for multiple or any domains.

IV. METHODOLOGY

4.1 Flow Chart of Neural Network

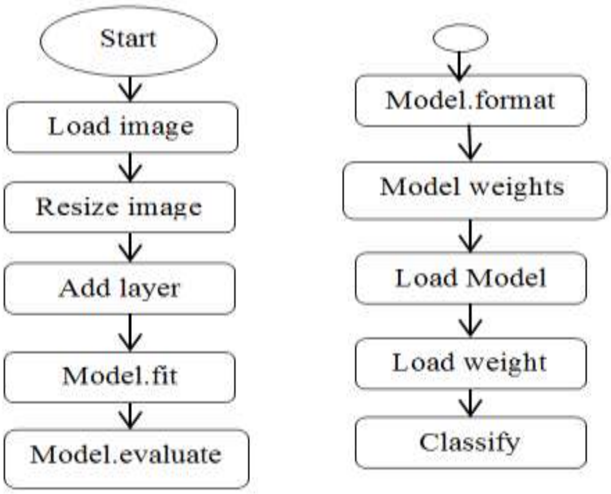


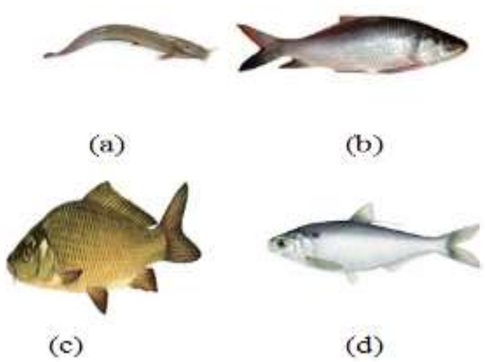
Figure 4: Flowchart of Neural Network

The above **Figure 4** shows the flowchart of neural netowork, obtained image is resized to 32*32 using open CV, which is library of Python designed for bindings to solve computer vision problems and also import the Numpy package for performing the numerical operations such as Resize(), Add(), sum(),etc. Hence from this the image is resized to 32*32 with three images i.e. RGB and are flattened. The number of input pixel nodes obtained is 32*32*3=3072 nodes as input. The input arguments are provided to the pixels and each node is labeled. This will measure pixels of input image to be in the range of [0, 1], and then alter the vector labels in the range [0, num_classes]. For each label it will generate the vector where the index of the label is set to '1' and all other entries to '0'. The complete vector is divided by 255 pixels.

The labels are provided for differentiating the categories of image. It uses 75% of dataset for training the model and remaining 25% for the testing purpose. This provides the test data, train data and list of both. The model used is Sequential model. The model created generates 2 files namely, "model.json" and "filename.hd5". The Json file consists of complete information of the model and store the complete data of Fishes and Not-fishes. The hd5 file store the weights and bias provided to the image in binary data format and it let store large amounts of numerical data, and it can simply handle that data from the data obtained.

This model is worked for new images for testing by using batch size=2 and epoch=60. The weights are stored in json and hd5 file in binary data format.

After training when input is given by using 25% of data which is kept for testing, if the image is with fish then it display as "0" and if the image is non-fish then it is displayed as "1". Further when it is classified within the fish then it shows for cat fish as '0', for catla fish as '1', for common carp as '2' and for silver carp it is displayed as '3'.



The above shown figures are the four types of fishes which we are going to classify
(a) Cat fish, (b) Catla fish, (c) Common carp, (d) Silver carp.

.4.4 FEATURE EXTRACTION

Feature extraction plays a very important role in image processing technique, here we mainly focus on the interested features of the image. Before extracting the features different image-pre-processing techniques like binaries, normalization, resizing, thresholding etc. are applied on the input image. The main aim of feature extraction is to extract the useful information from the original data. Feature extraction is done after the pre-processing of image, this is an important step in construction of any object recognition and classification, and aims for the extraction of useful information that characterizes each type it will become very easy for the classifier to classify between different classes by considering the features that are extracted, and also it distinguish between objects very easily.

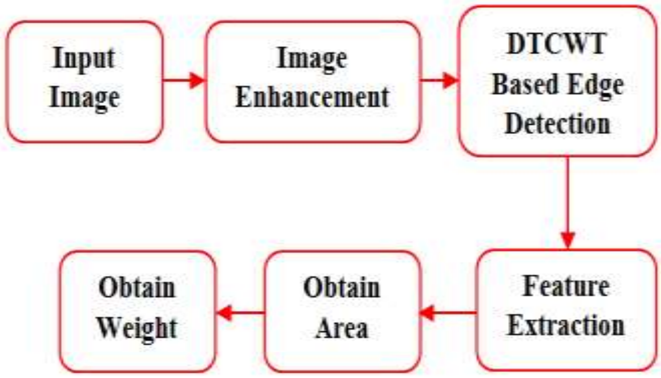


Figure 5: Block Diagram of Feature Extraction

The **Figure 5** shows the block diagram of feature extraction, it consist of five main blocks i.e. input image block followed by image enhancement block then DTCWT based edge detection block, then useful information is extracted by using feature extraction block, then we use obtain area and obtain weight blocks for obtaining area and weights of the objects. This complete feature extraction action is done by using matlab program and later called by python using callmatlab function.

The input image is given to the image enhancement block the output obtained from the image processing may be either an image or a set of characteristics or constraints related to the image, the RGB Image is resized to 512*512 RGB image.

The RGB image is converted to YCbCr image. Y indicates brightness, Cb indicates Chrominance of blue, and Cr indicates Chrominance of red. The RGB image is converted to Y Cb Cr because it is easy to convert back to the RGB whenever necessary. The histogram values are considered for the RGB image. The values are enhanced. All the histogram values are 1% saturated at the start and at the peak of the histogram values. The Y component i.e. the brightness component is extracted and the Cb and Cr components are also extracted separately and they are color mapped using the coefficients. Dual-Tree Complex 2D Discrete Wavelet Transform functions are considered taking the coefficients of analysis filter and synthesis filter by taking random values. The coefficients are considered for calculating the sub bands.

The feature extraction at first the pre-processed and enhanced image is taken and then the edges are detected by using DTCWT coefficients. The edge-detected image is then threshold to evade low intensity pixels. Later the boundaries of the object are found are extracted by using some suitable algorithm.

V. RESULTS AND DISCUSSION

In our proposed method, current approaches for identifying and classifying underwater fish using deep learning methods are discussed and also we have implemented our method to classify the fishes within the fish. Features and deep learning architectures used are summarized. In this proposed method we have trained around 1500 images to classify by using Fast R-CNN method by this method we can obtain around 94% of accuracy. Deep learning methods have accomplished excellent achievement in digital image processing for fish identification and classification by means of Fast R-CNN (Regional and Convolution Neural Networks). This approach highlights the object of detection, deep learning architecture and the feature extraction methods used. And also we have extracted the features of fish and calculate the area and weights of the fish..

For the proposed fish detection and classification method the images are collected from Indian fishery department around 800 fish and 800 non-fish images are collected and trained to a build model using deep learning method.
The overall result i.e. identification of fish, classification of fish within fish, estimation of area and weight are show in the below **Figure 6** using python script.

The below **Figure 7** shows the estimation of area and weight of the fish by applying the bounding box to the image, it returns the boundary values by using this value we can calculate the percentage of area and weight with respect to area.

Figure 6: Result in Python Script

Figure 7: Estimation of Area and Weight

VI. ACKNOWLEDGMENT

This work is deemed incomplete without acknowledging the various individuals immensely instrumental in ushering in a great deal of effort, time and valuable guidance. I would like to thank my guide Assistent Prof. Reshma.M, Dr. sarika raga, HOD Dept. of DECS , VIAT and my family for providing valuable suggestions, relentless support, help and guidance throughout the course of our project.

REFERENCES

[1] C. Spampinato, R B. Fisher, and B. Boom. Image Retrieval in CLETFish task. http://www.imageclef.org/2014/lifeclef/fish, 2014.

[2] R. Mehdi, et al. Automated Fish Detection in Underwater Images Using Shape-Based Level Sets. The Photogrammetric Record 30.149 (2015): 46-62.

[3] H. Qin, Y. Peng, and X. L. Foreground Extraction of Underwater Videos via Sparse and Low-Rank Matrix Decomposition. Computer Vision for Analysis of Underwater Imagery (CVAUI), 2014 ICPR Workshop on. IEEE, 2014.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV, 2014.

[5]. Hinton, G.E., Osindero, S.: A fast learning algorithm for deep belief nets. Neural Comput. 18 (7), 1527–1554 (2006)

[6]. Deng, L., Yu, D.: Methods and applications. Found. Trends. Sig. Process. 7, 197–387 (2013)

[7]. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25 (NIPS) (2012)

[8]. Lavery, P.S., McMahon, K., Weyers, J., Boyce, M.C., Oldham, C.E.: Release of dissolved organic carbon from seagrass wrack and its implications for trophic connectivity. Mar. Ecol.
Prog. Ser. 494, 121–133 (2013)

[9]. Liu, M.: AU-aware deep networks for facial expression recognition. In: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1–6. IEEE Press, Shanghai (2013)

[10]. Ciresan, D., Meier, U., Schmidhuber, J: Multi-column deep neural networks for image classification. In: Computer Vision and Pattern Recognition, pp. 3642–3649 (2012)