DETECTION AND RECOGNITION OF BANGLADESHI FISHES USING SURF FEATURES AND CNN CLASSIFIER

This thesis was submitted in partial fulfillment of the requirement for the degree of

Bachelor of Computer Science and Engineering

Under the Supervision of

Dr. Jia Uddin

By

Syed Adiba Shuhin (14301080)

Jannatul Tajrin (14301084)

Afrin Akhter(14101046)

Chowdhury Israth(14101064)

School of Engineering and Computer Science

July 2018

BRAC University, Dhaka, Bangladesh

DECLARATION

We, hereby declare that the thesis "DETECTION AND RECOGNITION OF BANGLADESHI FISHES USING SURF FEATURES AND CNN CLASSIFIER" is solely based upon the results founded by ourselves under the supervision of Assistant Professor Dr. Jia Uddin. Material and related information that are used here from other resources are acknowledged by reference to our best ability. We have carried out this thesis for our degree of Bachelor of Science in Computer Science and Engineering and this thesis neither in whole or in part have been previously submitted for any degree.

Signature of Supervisor	**Signature of Authors**
Dr. Jia Uddin	Chowdhury Israth(14101064)
Assistant Professor	
Department of Computer Science and Engineering	Jannatul Tajrin (14301084)
BRAC University	Afrin Akhter (14101046)
	Adiba Shuhin (14301080)

ACKNOWLEDGMENT

We would like to start by expressing our deep gratitude to the Almighty Allah, the creator of everything, the most benevolent and the most gracious, for giving us the strength and determination to initiate and complete our research successfully.

After that, we are eternally grateful to our thesis Advisor Professor Dr Jia Uddin of the Department of Computer Science and Engineering at BRAC University. Whenever we were lost and feeling demotivated he guided us through the phase and steered us in the right direction. He was always there to solve the problems we faced when doing the thesis. Not only had that he has given us motivation and inspiration whenever we were feeling down with our progress. Not to mention his valuable constructive feedbacks during the whole research. We are grateful to have the opportunity to work under him and to learn from him.

And finally and greatly, we are indebted to our peers, friends, families and faculties for their unrelenting support in our thesis journey. It wouldn’t have been possible without them.

Thank you

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SURF – Speeded Up Robust Features

CNN- Convolutional Neural Network

RGB – Red Green Blue

ABSTRACT

This thesis proposes the detection and recognition of Bangladeshi local fishes using image processing. In the proposed model, we have successfully detected fishes using grass-fire algorithm along with other methods like Gaussian Elimination, noise margin for image pre-processing at first and binary masking, flood-filling of binary image later for detection. After that the species of the detected fish is recognized using Speeded up Robust System (SURF) algorithm and for further analysis it was also done by CNN classifier to check whether the accuracy rate is better than the previous one. This is to be mentioned that we have done the whole research on various species of Fishes available on Bangladesh which has not been done before. We implemented our custom Dataset consisting of 400 sample images in the proposed method to measure out it credibility. Our aim is to detect and recognize fish more accurately with less error percentage.

Keyword: Fish Detection, Edge Detection, Fish Recognition, SURF, Convolutional Neural Network, Binary masking, Keras

CHAPTER 01

Introduction

1.1 Motivation

It is surprising to learn how many types of fishes are available in this world and how very different they are from each other in aspect to their lifestyle, food value and appearance. In order to have a vague idea about known fish species first of all what is needed to know that more than half of vertebrate species are various types of fishes. The known extant species are around almost 28,000 and 27,000 of them are bony fishes, in the remaining 1,000 about 970 of them are different types of sharks, rays and chimeras and around 108 types of Hagfish and lampreys. One third of these species fall within nine largest families these families are Cyprinidae, Gobiidae, Cichlidae, Characidae, Loricariidae, Balitoridae, Serranidae, Labridae, and Scorpaenidae. Around 64 of these families are monotonic meaning they include only one species. The number of known extant species is still growing and may grow to exceed 32,500 [1].

Fishes are an incomprehensible part of human resource especially as food. A proper nutritious diet must have fish in menu as it's nutritious value is unbeatable but the taste and nutrition type differs from species to species and some species are not edible at all they are downright poisonous. Aside from food value Fish is a very important part of our eco-system as our planet earth consists of 70 percent water. Not to mention fishes can be great pets and their quality, price and the way of taking care of them also differs from species to species.

So all people should know species of the fishes when dealing with them but it is hard for many to differentiate between as some of them are quite similar and it is hard to remember so many types of fishes.So automatic detection and recognition of fish is needed. Not only for that it can become a great help for understanding marine ecosystem. Marine biologists will greatly benefit by an automatic fish detection and recognition system as then they wouldn't have to use manual annotation which tends to be very expensive and time consuming. Furthermore fish farming, meteorological monitoring and fish quota would be greatly benefited.

So in this paper we have researched fish detection and recognition methods. We have successfully detected the using grass-fire algorithm fish along with other methods like Gaussian Elimination and noise margin for image pre-processing at first. After that the species of the detected fish is recognized using Speeded Up Robust Features (SURF) algorithm and for further analysis it was also done by CNN classifier to compare the accuracy between.

In our thesis we focused on finding the species by using still images with the possibility of recognizing fish species in real time more specifically in underwater videos in future work

1.2 Contribution Summary

The summary of the main contributions is as follows:

- Image pre-processing using Gray Scaling, Noise Filtering and Gaussian elimination methods.
- Fish detection through Binary masking, Flood Filling of that Binary Image and Boundary extraction using Sequential Grass-Fire Algorithm
- After successfully detecting fish .with an accuracy of 95.23% we recognized Fish through SURF feature method.
- We also recognized Fish using CNN classifier to decide which one is better
- We have done the entire research on Bangladeshi Local Fishes which hasn't been done before.

1.3 Thesis Outline

- Chapter 2 outlines the previous works done in the field of Fish Recognition using Image processing and deep learning.
- Chapter 3 describes the proposed model including the implementation specifics.
- Chapter 4 experimental results based on a diverse dataset and their performance comparisons.
- Chapter 5 concludes the paper specifying its difficulties and limitations while stating possible future paradigms the system can be expanded to

CHAPTER 02

Literature Review

2.1 Neural Network

Neural network can defined by saying it is a parallel computer model which is created with a large number of adaptive processing units who are called artificial neurons and are loosely based on the model of animal neurons in brain and this units are interconnected between themselves and this connections are like synapsis in a biological brain and can transmit signal or data to one another through it [2, 3].

Other machine learning methods learn faster than neural networks but as it has one or more layers in other word neurons who enables the learning of complex task by progressively extracting features from an inputted image it predicts or give a more accurate result way faster than others also it has good models for nonlinear data [2].

Mathematically, the neural network was described with the following equation

$$y = \varphi(\sum i = \ln W_i * x_i + b) \tag{1}$$

Here, W_i : Weight vector,

x_i : Input vector,

b: bias activation function

For further understanding lets understand what perception is well in simple words it is an algorithm for learning a binary classifier or in other word a function that maps its input to an output.

$$f(x) = \begin{cases} 1 & [\text{if } w.x + b > 0] \\ 0 & [\text{ otherwise}] \end{cases} \quad (2)$$

Where

w = vector of real-valued weights

w.x= Is the dot product $\sum_{i=1}^{m} w_i \, . x_i$, m = number of inputs to the perceptron.

b = is the bias

The bias shifts the decision boundary away from the origin and does not depend on any input value.

After understanding perceptron which is a single-layer perceptron, multilayer perceptron is needed to understand because in neural network it is what is used. Now multilayer perceptron is a class of Artificial Neural Network (ANN) and much more complex it has at least three nodes except for input nodes (neuron). A supervised learning technique called back propagation has been utilized by it.

So as it has been said Multilayer perceptron has a set of source nodes containing one or more input layers and also a set of hidden node inputs [4]. The signal propagates inside the network from layer to layer. In order to have a better understanding see the model below where the inputs in input layer, set of hidden nodes in hidden layer and the output had been shown figuratively. In Fig.1 below we see the Multilayer perceptron in Neural Network

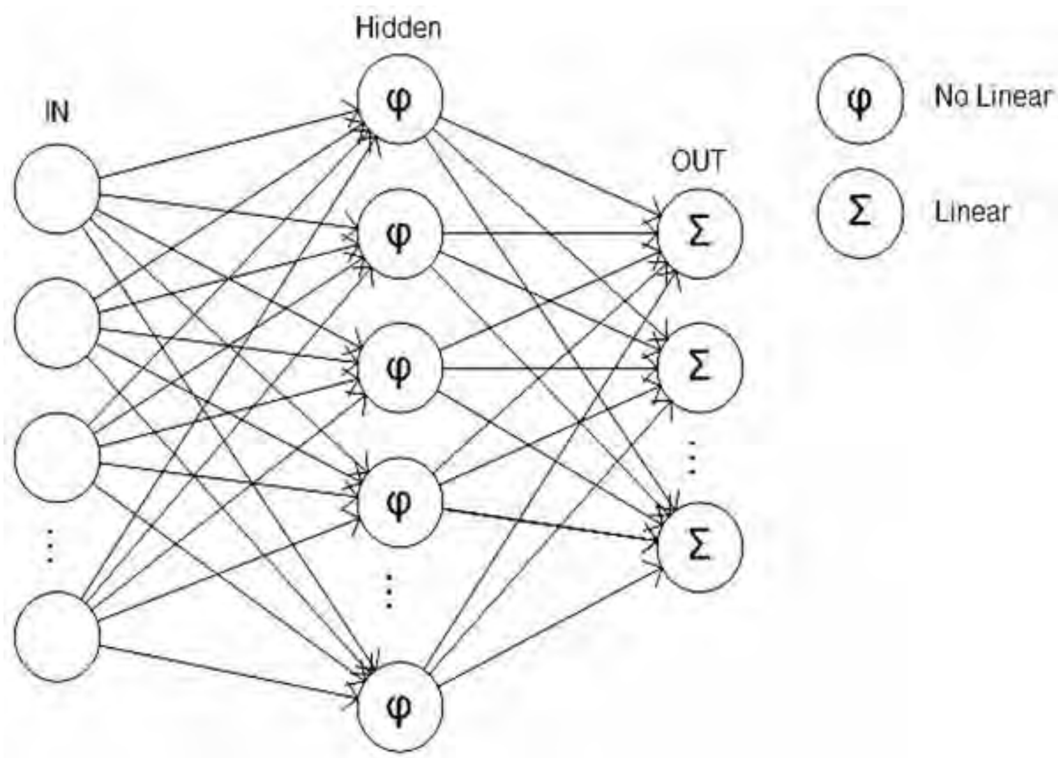


Fig .1.Multilayer perceptron.

So it can be seen from the figure the neural network structure is created by N inputs where N= $[N_1, N_2 \ldots N_n]$, a set of hidden nodes in hidden layer H and output vector $S = [S_1, S_2, \ldots, S_m]$. Binary signal [0, 1] has been found after each $S_{i \text{ was}}$ assessed by a single step. This step is a sigmoid activation step which is based on back propagation. In back propagation weights and biases were updated in the direction of the negative gradient of the performance and then updated in the opposite direction [1].

$$\mathrm{f}(x) = 11 + e - x \tag{3}$$

This equation is used for the sigmoid activation function for the hidden layer and for the determination of output layer.

2.2 Convolutional Neural Network

CNN is a multilayer neural network, it consists of the input layer, convolutional layer, pooling layer, fully-connected layer and output layer. The details are as follows. Input layer: The input layer works directly on the original input data, and for the input image, the input data is the pixel value of the image. Convolutional layer: It is also called the feature extraction layer .The main function of this layer is to extract features form input data. The features of the input images extracted by each convolutional kernel are different, and the more convolution[5]

2.3 Related Work

Not much work has been done on this area still Some of the ways that fish were be detected are by shape, color, texture and also by extracting feature also another way that has been explored is recognizing the species of the fish by Using CNN. Bai *et al.* has used image classification based on shapes and texture. It has been done by the feature extraction process. The purpose of feature extraction is to determine the most relevant and the least amount of data representation of the image characteristics in order to minimize the within-class pattern variability, whilst, enhancing the between-class pattern variability [6].Alsmandi has done Fish recognition based on the combination between robust features selection from the size and shape measurement using neural network[7].

Al-Omari *et al* used CNN (a deep neural network model which contains convolutional layers, subsampling layers and fully-connected layers) to recognize the features of fish. Its weight-sharing network structure for convolutional layers makes it more similar to a biosphere network and can reduce the number of weights and the complexity of the network, and improve the training efficiency [5].

F. Storbecka *et al* used a fish image which was easy to extract a fish region with a white background or uniform background for automatic processing. This research adapted an approach to give several feature points by manual operation by the user. The proposed approach is able to accept the fish image in the complicated background taken on the rocky place.to recognize fish species it has used computer vision and neural network [8].

Spampinato et al. proposed a vision system for detecting, tracking and counting fish from real-time videos, which consists of a series of video texture analysis, object detection and tracking procedures [9]. Hwang et al. reported an automatic segmentation algorithm for fish acquired by a trawl-based underwater camera system [10].

In this paper we investigated the efficient features for fish recognition, we defined various features, such as, image pre-processing, Noise Filtering, Gaussian blurring method and Sequential Grass-fire algorithm to detect the fish and Speeded-up Robust Features (SURF) algorithm to finally recognize the fish. Furthermore we used CNN classifier to recognize the fish to check whether the accuracy rate is better than the previous one. This is to be mentioned that we have done the whole research on various species of Fishes available on Bangladesh which has not been done before.

CHAPTER 3

Proposed Model

In this research we have detected and recognized fishes using our custom database consisted of 400 images. Fig.2. below shows the process of fish detection step by step.

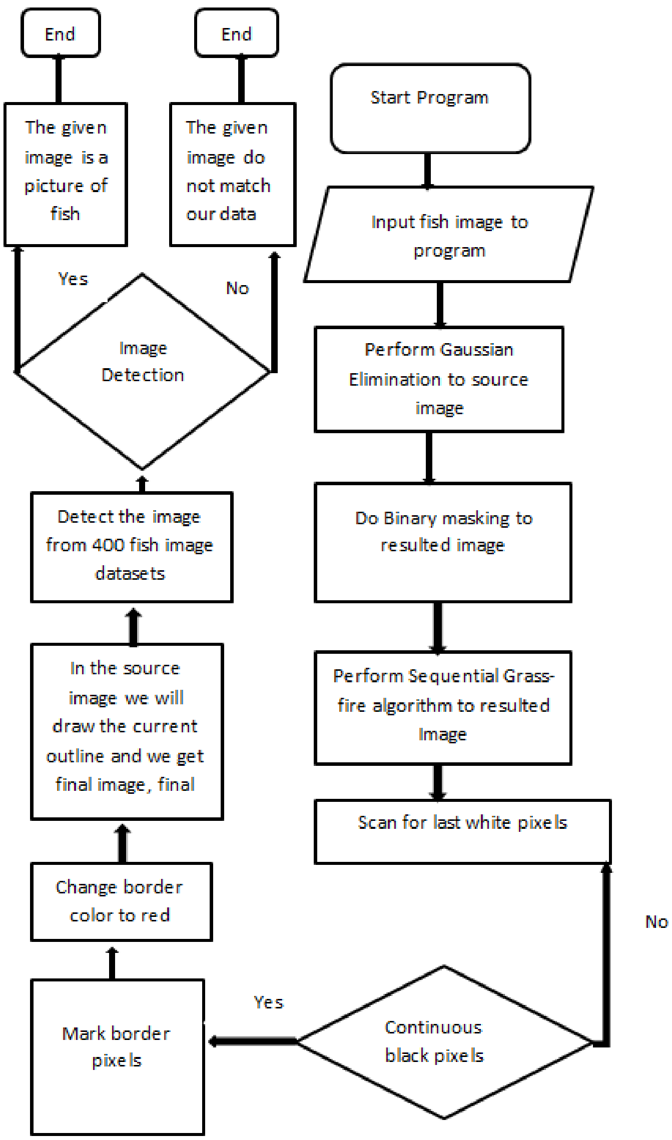


Fig.2. Workflow diagram

3.1 Image Pre-Processing

3.1.1 Data Acquisition

At first, we will acquire some of our selected data to begin the fish detection process. In this stage, we will take around 500 of our sample data. This sample data will be used for the fish detection in the later stages. We will take picture of different kinds of fish and those fish will be cropped from different angle and different location. Using this sample data, we will later on try to identify the fishes

3.1.2 Image Pre-Processing Segments

Image segmentation is a process of diving an image into multiple parts. This process is used to identification of any kind of object. So, at first we select a source image and divide it into different parts for the identification process.

Segmentation is a set of pixels. It is possible that an area contains almost similar pixels due to homogeneity criteria such as color, intensity and texture [11].

There are total 4 types of segmentations and those are:

i) Region based segmentation

ii) Pixel based segmentation

iii) Edge based segmentation

iv) Model based segmentation

Here edges based segmentation is being used to get the image edges properly. The different parts of the image are more focused with this approach and by having a clear look of the images from the data set due to high pixel value on the edges. Edge segmentation technique avoid a bias in the size of segmented object and in logical term it uses first order derivatives by that the position of edge is given extreme. For more preprocessing and to reduce the level of gray scale Gaussian Elimination Method and Noise Filtering are being applied.

3.1.3 Gray scale

Firstly the source image are converted into gray scaling to reduce the complexity of the color. It is reduce the build-up noise and signal distortion during processing.

The rules for gray scale Formula R×0.118 + B×0.385+ G ×0.51 is being used to reduce the complexity of color image [11, 13].In Fig.3. Below image before Gray Scaling and after Gray Scaling is shown.

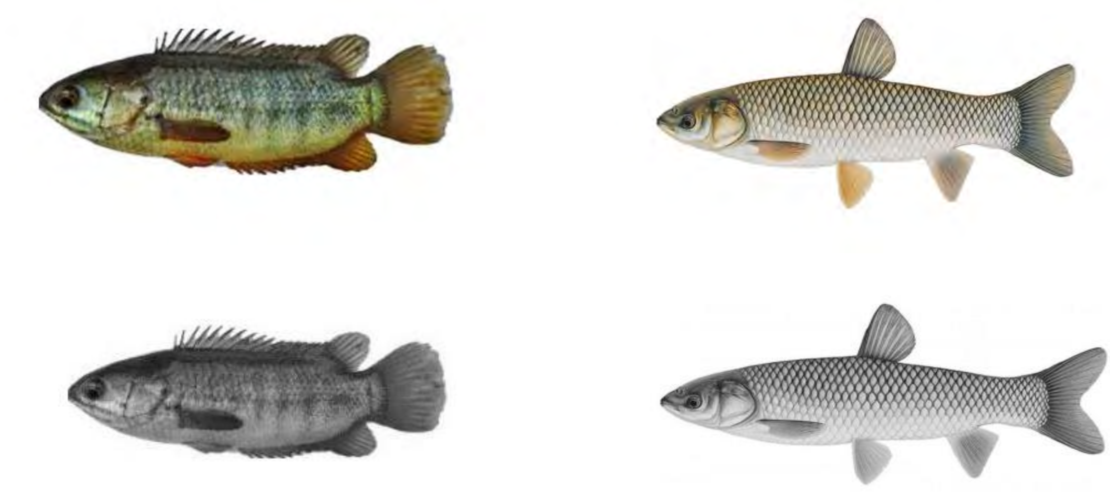


Fig.3.Gray Scale before and after

3.1.4 Noise Filtering

Since the image is going back to preprocessing, it should ensure that the images are without noise. Specially salt and pepper noises as it will hamper in segmentation and feature extraction part and also the quality of the image processing or analysis result as well [14]. The whole process of reducing noise is done using mean filtering.

Mean filtering is a non-linear approach to deduct noises from image. It is a very effective approach to remove the noises from image. It works by moving through pixel by pixel in the image, as it replaces every value by median value of neighbor pixels. The pattern of neighbor is called window, which slides, pixel by pixel entire image [14].

The equation for median filter is,

$$y_{i,j} = \begin{cases} y_{i,j} & if\ y_{i,j} \in \left(s_{i,j}^{min,w},\ s_{i,j}^{max,w} \right) \\ s_{i,j}^{med,w}\ , & else \end{cases} \tag{4}$$

Here Y is median filter and S is impulse noise.

3.1.5 Gaussian Blur Method for smoothing

Gaussian blur method is an effective approach to blur an image. It is widely used to reduce to image noise and details. In the source image we have to perform Gaussian blur method to reduce noises of the image. It is also known as Gaussian smoothing. Gaussian elimination depends on standard deviation. Our preferred standard deviation (σ) for this project is σ = 0.5 [15].

In the Fig. 4 below, one image's both RGB and Grey Scaled Image are shown

(a) (b)

Fig 4. (a) RGB image, (b) Gray Scaled image.

The function that is used for Gaussian blur for 2d image is,

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

In the following equation, x is the value of distance from origin to horizontal axis and y is the value of distance from origin to vertical axis and σ is the value of standard deviation [15].

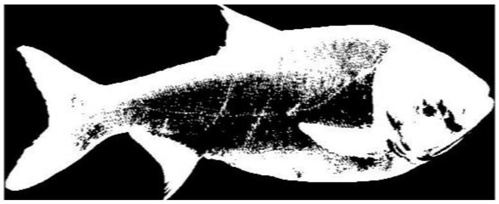
3.2 Image Segmentation

3.2.1 Binary Masking of the image

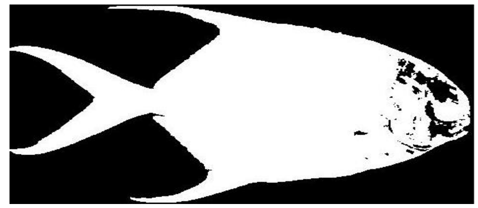
The binary masking of an image is the process of converting an image to a black and white image which can be misleading sometimes [17] .Grey scale and binary image, both, can be black and white image. Grey scale image is an image which consists color range of [0,255] and in binary image the color range is [0, 1]. Thus the image of binary and greyscale both can be black and white but the binary image contains only two color which is 0 and 255(0 and 1) [17,18].

So we have to be careful not to convert our image to grayscale and convert our image to binary for the next approach [17].

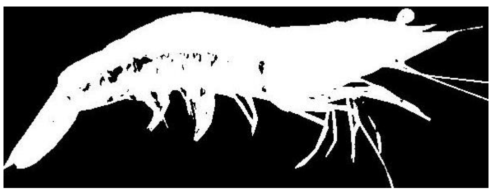
In Fig.5 it shows Images after Binary masking.



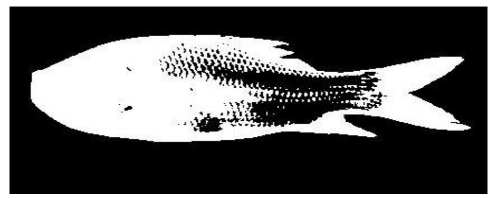
(a)



(b)



(c)



(d)

Fig.5. (a), (b), (c), (d) Source images after binary masking

The image matrix between a normal image, grey scale and binary image differs in its value. Example:

$$\begin{matrix} 94 & 12 & 4 \\ 5 & 1 & 5 \\ 5 & 12 & 5 \end{matrix} \qquad \begin{matrix} 6223 & 5124 & 6223 \\ 5124 & 5521 & 441 \\ 5531 & 4441 & 441 \end{matrix} \qquad \begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{matrix}$$

(a) (b) (c)

Fig .6. (a)Grayscale, (b) Color (c) Binary image array

3.2.2 Flood Filling of Binary Image

Flood fill is a process to fill the image of connected area with similar color. It is also known as a seed fill. In flood filling algorithm we start with some seed color and examine the neighbor colors, but in this case the pixels will check for a specific color bit and that specific color will be replaced by boundary color with this algorithm [19]. In this process, we will fill the holes with similar color for detection purpose. So, in our approach we are changing the background pixels (0s) to foreground pixels (1s) stopping when it reaches the boundary for our binary image. We will approach to fill the boundary as hole. By this approach we can have a more clear vision of the object that being represented in the array as shape.

Fig.7. shows Flood filling of the binary images.

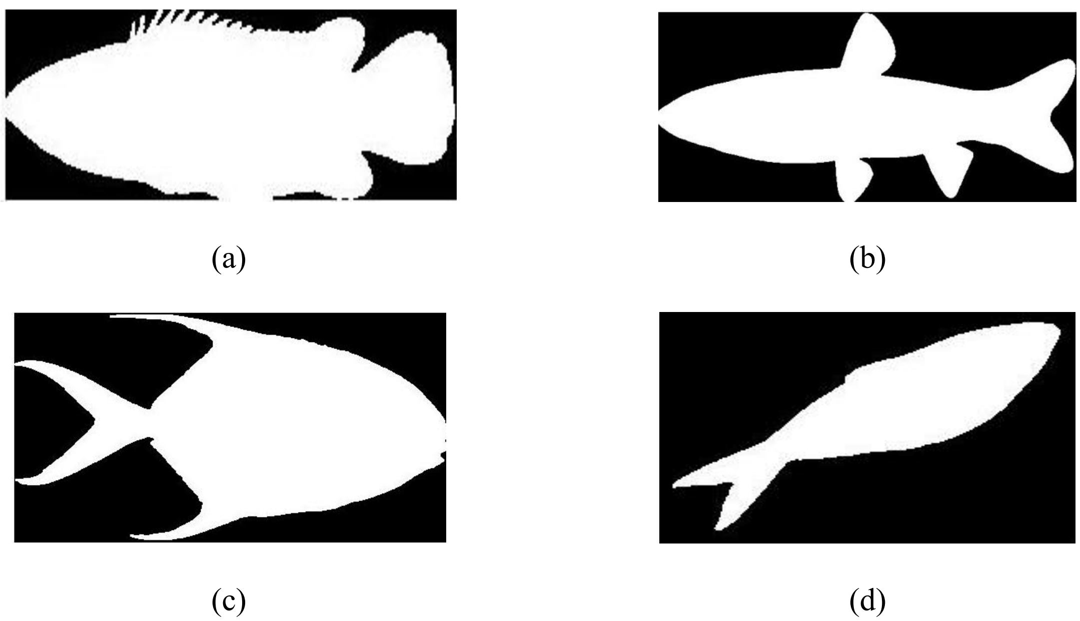


Fig .7. (a), (b), (c), (d) Flood filling of the binary image

3.2.3 Boundary extraction using Sequential Grass-Fire Algorithm

The grass-fire algorithm can be implemented in many ways. But among them sequential grass-fire algorithm has less shortcomings and that is the reason why we are using sequential grass-fire algorithm for our purpose.

The sequential grass-fire algorithm starts from the top left and ends with scanning all the way to bottom right. After reaching its object pixel it will do two tasks. Firstly, it will create a label for the object pixel in the output image and secondly it will remove its object pixel from the input image. Then it will check for its neighbor object pixels. The neighbor object pixels are also identified as label and then deleted from the input image and also they are put down in a list. The next stage, we take the first pixel from the list and neighbor and if we find any object pixel are in output we set to zero in input and set it to list. The procedure is then repeated until the all object pixels are investigated. Then it continues the scanning until it finds the next object pixel and then start the next sequential grass-fire algorithm [21].

In our objective, the scanning of the binary image after filling starts from top left and ends in bottom right. And after running the algorithm we will get black and white image as well. In the white part it is the fish that we are detecting and in the black side it is the background.

3.2.4 Marking Border Pixel

We have to scan again. In this stage, we have to scan for white pixels. A group of white pixels that is staying close to each other or continuous white pixel that ends with black pixels. Now we have to mark this group of white pixels as border pixels. Border pixels are very useful to compute approximate output value of border-dependent pixels. Border pixels help us to get identical border with the source image to output image.

3.2.5 Changing Border Pixel Color

In the current output image we have to change the border pixel colors. The color will be changed to red. It is done by changing the RGB color of the output image parameter. Now, it will help us even more to detect the fish.

In Fig.8. below we can see a example of an output image after changing the border color

Fig .8. Changing the border color of the output image

3.2.6 Outline the source

Now we will get a copy of the source image. This image doesn't have any effect on the changes we have made for detection. We will outline the source image with the help of our previous steps. The border pixels will be replaced in the source image and the source image will be updated [22].

In Fig.9. we can see examples of Boundary Tracing.

(a)

(b)

(c)

(d)

Fig .9. Boundary Tracing of Fishes

3.2.7 Shape Detection

After doing all this procedures, we will be able to detect shape almost 95% without fail. We will have around 400 sample images to detect the shape and it will be quite efficient [23, 24].

3.2.8 Determining Height and width

So in this process we will focus on the pixel. In our output image we will count the white pixels. The white pixel is efficient to measure the size of the fish. Thus, we will be able to know the if the fish is small, medium or large. And from our dataset we can approximately measure its weight from the size of the fish by learning the current dataset we will shortlist our fish from 400 fish image datasets.

3.3 Image Classification

3.3.1 Convolutional Neural Network

Convolutional Neural Network (CNN) has developed rapidly in recent years for its high accuracy in image classification and recognition. It is highly used for its weight sharing network structure and the ability to reduce the number of weights and the complexity of the network. CNN is a multi-layer, fully trainable model which can capture highly nonlinear mappings between inputs and outputs [25].

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. A CNN consists of an input and an output layer and multiple hidden layers. The hidden layers are consisting of convolutional layers, pooling layers, fully connected layers [26].

Input layer

The input layer works on the original Input data and input image, the input data is the pixel value of the image.

Convolutional layer

Convolutional layer is the core layer of CNN. It is also called the feature extraction layer. The aim of this layer is to extract features form input data. The features of the input images extracted by each convolutional kernel are different. The more convolution kernels of the convolutional layer, the more features of the input data can be extracted.

Pooling layer

Pooling layer is another important concept of CNN. It is also known as the subsampling layer, the main purpose of this layer is to reduce the number of data while retaining useful information and to speed up the training speed. The more layers of convolution and pooling, the more abstract features can be extracted. In general, the convolutional layer and the pooling layer appear alternately.

Fully-connected layer

Finally, several convolutional and max pooling layers, the high-level reasoning in the neural network is done via fully connected layers. In fact, it is the hidden layer of multi-layer Perception machine.

In Fig.10. Below we can see a typical CNN architecture

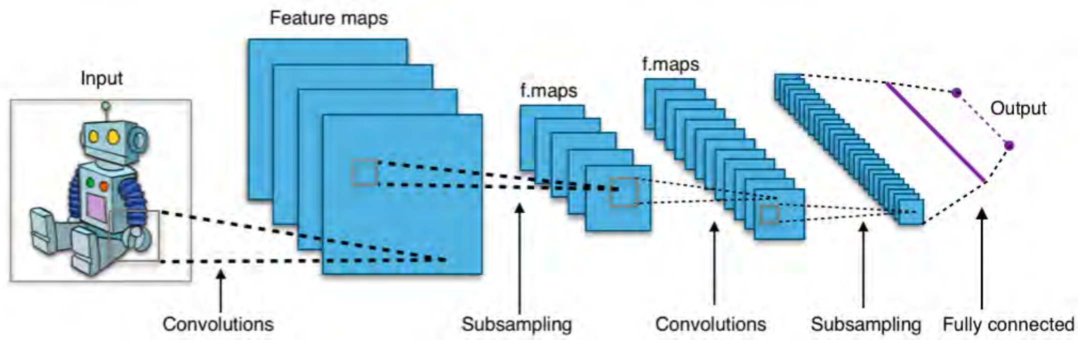


Fig.10. Typical CNN architecture

3.3.2 Keras

Keras is an open source neural network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit or Theano. It is designed to enable fast experimentation with deep neural networks; it focuses on being user-friendly, modular, and extensible. It supports convolutional networks and recurrent networks, as well as combinations of the two. It also runs seamlessly on CPU and GPU [27].

3.4 Fish Recognition

After successfully detecting fish in order to fulfill the main purpose of these paper that is recognizing fish the Object Recognition method Speeded-Up Robust Features have been used. Speeded-Up Robust Features (SURF) algorithm is a scale and rotation invariant robust features detector and descriptor, It was firstly presented by Herbert Bay, et al., at the 2006 European Conference on Computer Vision and published in 2008 [28]. IT is widely used for object recognition, image registration, traffic monitoring in image sequences, 3D reconstruction or augmented reality applications etc.

This Algorithm has three main steps. They are -1. Feature Extraction or Detection

2. Feature Description

3. Feature Matching

3.4.1 Feature extraction:

Feature extraction is a must step in almost all object detection algorithm. Now the question is what exactly is feature extraction, well feature extraction is the process of extraction useful information, identifiable attributes referred as features from an input image.

Step i: When an input image is given the output returns with an array of extracted feature points.

Step ii: Convert the image in a 8 bit gray scale image. When doing the detection part input image was already converted so no need for this step in this case.

Step iii: Calculating the integral image in order to find the sum of pixels enclosed within any rectangular region of the image the function MyIntegralImage.m is used to do it.

Step iv: Hessian based blob detector is used in Speeded-Up Robust Features to find interest points the function is called FastHessian.m. The determinant of a hessian matrix expresses the extent of the response and is an expression of the local change around the area [29].

$$H(x,\sigma) = \begin{matrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{matrix} \quad (6)$$

Where,

$$L_{xx}(x, \sigma) = I(x) * \frac{\partial^2}{\partial x^2} g(\partial) \tag{7}$$

L_{xx} (x, σ) in equations are the convolution of the image with the second derivative of the Gaussian. The heart of the Speeded-Up Robust Features detection is non-maximal-suppression of the determinants of the hessian matrices.

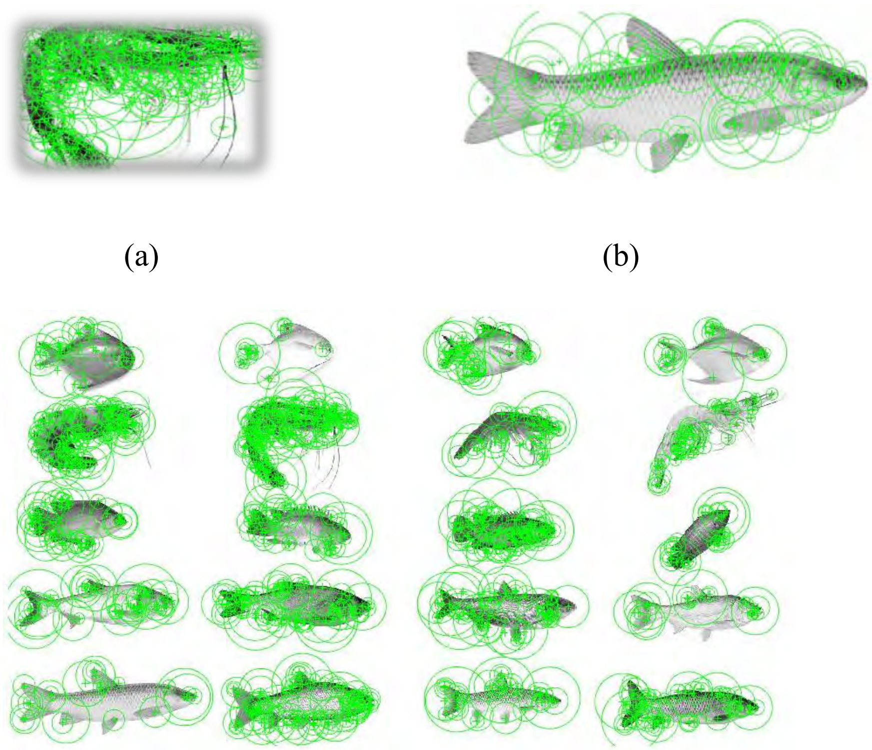
Step v: Filtering out the only useful interest points from the output gotten from FastHessian.m function based on some factors. It can be done by NonMaxSuppression_gpu.m function.

In order to find the exact location of interest points are interpolated by fitting a 3D quadratic in scale space.[4]An interest point is located in scale space by (x, y, s) where x, y are relative coordinates $(x, y, \in [0; 1])$ and s is a scale.

Step vi: Calculating and assigning orientation to the interest points located in the Step v by OrientationCalc.m Function. The final Result that is gotten from feature extraction is an array of interest points and is a structure with following fields.

x, y (coordinates), scale, orientation, lapl

Below in the Fig .11, the 1000 feature keypoints are visualized in training and testing dataset.



(a) (b)

(c)

Fig. 11. (a), (b) SURF strongest feature detection all possible points (left) 100 points (right) and

(c) For all fishes.

3.4.2 Feature Description

The interest points founded each should have a have a unique description that does not depend on the features scale and rotation. The Speeded-Up Robust Features descriptor is based on Haar wavelet responses and can be calculated efficiently with integral images. The descriptors are robust to rotations and an upright version. . The interest area are weighted with a Gaussian centered at the interest point to give some robustness for deformations and translations [24]. The Calculation

$$v = \{\sum dx \sum |dx| \sum dy \sum |dy|\} \tag{8}$$

For each subarea a vector v is calculated, based on 5×5 samples. The descriptor for an interest point is the 16 vectors for the subareas concatenated.

3.4.3 Feature Matching

Recognizing the object and a possibly transforming it from the given input image based on based on predetermined interest points is what done on matching.

After calculating the descriptors we can find a match between them by testing given the distance between the two vectors is sufficiently small. This is called putative point matching/ The SURF algorithm does add another detail to speed up matching that is the sign of Laplacian.

$$\nabla^2 L = tr(H) = L_{xx}(x,\sigma) + L_{yy}(x,\sigma) \tag{9}$$

The laplacian is the trace of the hessian matrix and when calculating the determinant of the hessian matrix these values are available, it is a matter of storing the sign. The sign of laplacian should be stored because it distinguishes between bright blobs on dark backgrounds and vice versa. Only on the cases with same sign the full descriptors vectors are needed to be compared this significantly lowers the cost of matching. It is only necessary to compare the full descriptor vectors if they have the same sign, which can lower the computational cost of matching. In Fig.12. Result of putative points matching is shown.

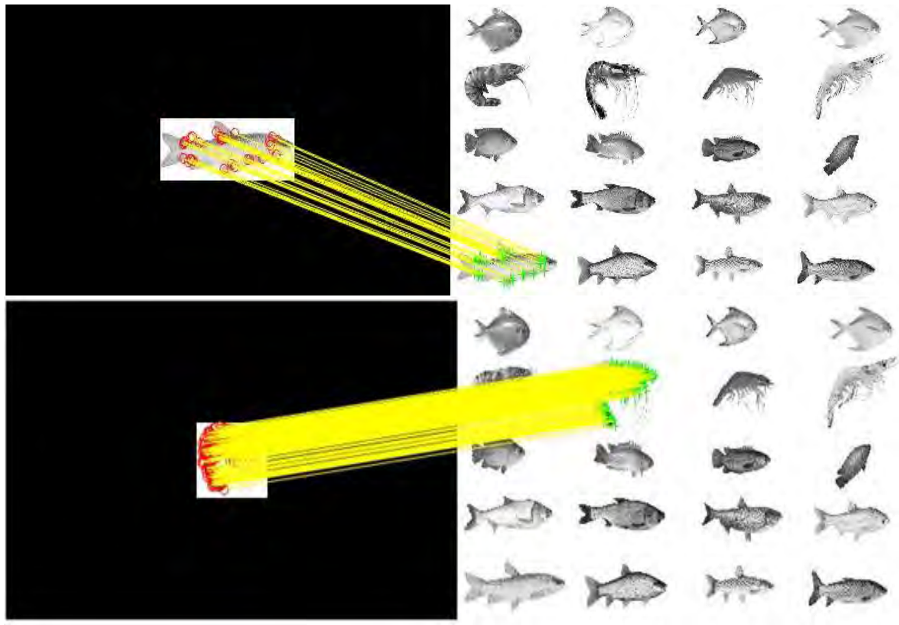


Fig.12. Putative matching points

In order to understand how it works with images let's see the following flowchart. In Fig.13. It can be seen how a sample image is inputted and then processed after than Feature is extracted the information is sent to database and by taking help from the trained pattern it predicts the result based on pattern and give a result.

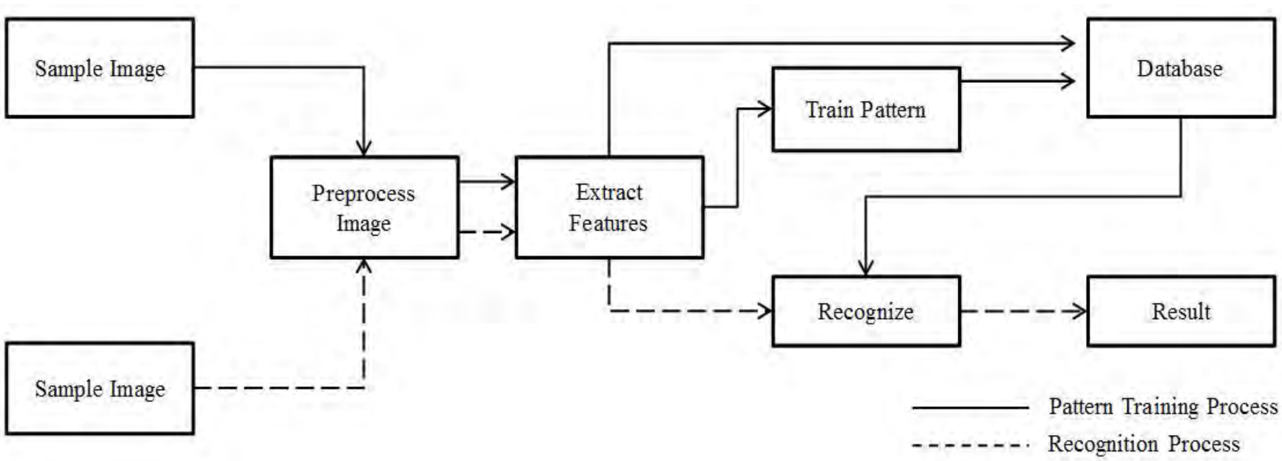


Fig .13. Flowchart for Fish Recognition

Chapter 4

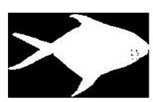
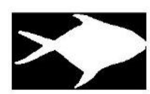
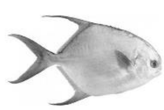
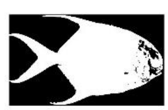
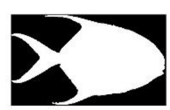
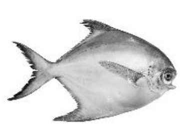
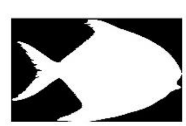
Result and Analysis

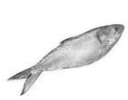
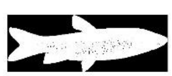
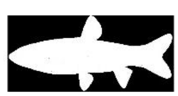
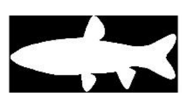
The proposed system is designed to detect and recognize regional fishes available in Bangladesh. It contains both image processing and convolutional neural network**.** Image processing algorithm like grass-fire algorithm is used for outlining contour of fish's body where the inputted is pre-processed, then has been through binary masking. The accuracy of contour marking depends on the threshold values of binary masking. Recognizing is at first done by speeded up robust feature (SURF) algorithm, where a set of fishes are made putting image in row by row based on species . SURF algorithm creates keynote points and connecting the featured keynote point of testing image to dataset image it denotes the recognized species of fish. The accuracy is gained by trial and error method that how many featured point are points wrong row of fishes' species. In the convolutional neural network part, the accuracy of recognition is gained from the cross validation of training and testing dataset.

4.1 Detection of Fish Using Image Processing

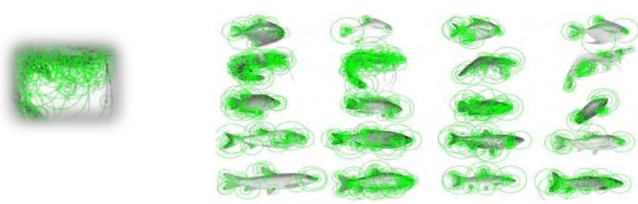
Source image are converted into gray scaling to reduce the complexity of the color. It is reduce the build-up noise and signal distortion during processing. Following that, binary making is applied so that only white and black pixels can exit which is needed for binary tracing. Now before going to be boundary detecting part, the images need to be reconstructed by morphological reconstruction flood-fill algorithm. As a result, if black pixels exist in a white pixel bounded region, then it is converted into white pixels and finally median filter is used for reducing the rest noises. Grass-fire algorithm which has less shortcomings and that is the reason why we are using sequential grass-fire algorithm for tracing the boundary of the binary image. Finally the contour array is stored and the value is converted into (255, 0,) color space. Table 1 shows each step of fish detection and accuracy rate based threshold value during binarization.

Table 1. Fish detection with accuracy score

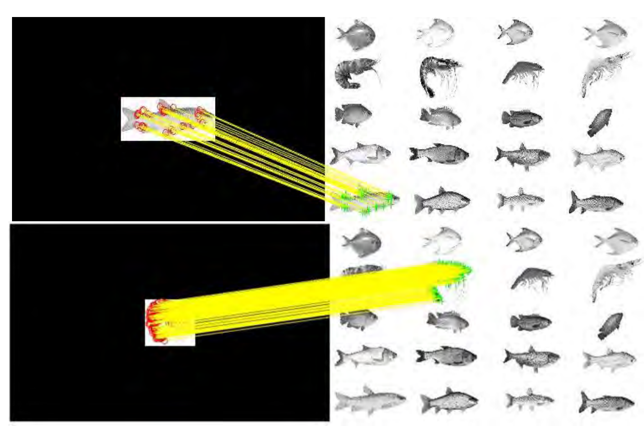
Acquised Image	Gray Scaled	Binary Masked	Flood Fill and Median Filtered	Detection of Fish	Accuracy
					97.23%
					97.64%
					89.58%
					91.87%
					87.93%

					89.72%
					94.68%
					94.97%
					98.67%
					98.55%
					98.79%
					83.12%
					81.08%

SURF algorithm creates keynote points and connecting the featured keynote point of testing image to dataset image it denotes the recognized species of fish. Here 1000 featured points are selected for testing and dataset image. The accuracy is gained by trial and error method that how many featured point are points wrong row of fishes' species. In the convolutional neural network part, the accuracy of recognition is gained from the cross validation of training and testing dataset. When it is processed, the error connected parts are gained when testing image's one of 1000 featured points is connected to the wrong specifies. In Fig.15. Below we can see result of feature point marking.



(a)



(b)

Fig.14. (a) 1000 key feature point marking of training and testing image, (b) Recognizing using Surf algorithm

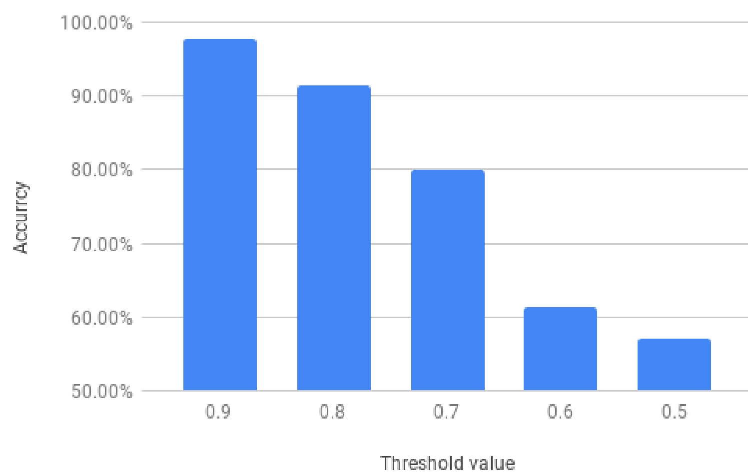


Fig .15. Accuracy vs Threshold Value

The threshold value is very important for image binarization for fish detection part. It works like edge detecting process . The whole figure will be detected if it the threshold value is .814 to 0.9 which is gained by trail and error method , manually testing again and again. When the whole fish is detected, then area is gained and compared with other threshold values' areas as it can't fully get the whole fish, the area will less than the actual one, and finally, deference is presented in parentage. while the threshold is 0.5 then the accuracy of tracing the fish region decreases less than 60%, when threshold value is 0.6, accuracy is between 60 to 63 percent. When the threshold value is 0.7, accuracy becomes 80%. But when threshold value is something between 0.8 and 0.9 and have Gaussian blurring applied with it, increases the accuracy of trace region and contour of fish to 90 to 97% accuracy.

4.2 Recognition of Fish using CNN with Keras

In convolution neural network different dataset need to be split into Train, Test and Validation dataset. During each epoch data is trained over and over to learn the feature of data. Later it can accurately predict the data which is not seen before. It takes the decision based on what it learns from Training data. Validation data is used to validate our data during training the dataset. Validation data is separate from Train data. Validation data is needed to ensure that model is not over feeding which means model become really good to classify the data from train set. Now the

Table 2. Loss and Accuracy result for Training and Validation

Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
0.4375	0.8877	0.5783	0.8696
0.4483	0.8822	0.7073	0.8696
0.4262	0.8877	0.4877	0.8696
0.4592	0.8849	0.5979	0.8696
0.3789	0.8795	0.4542	0.8696
0.4015	0.8822	0.5898	0.8696
0.3734	0.8767	0.5002	0.8478
0.3423	0.8959	0.4692	0.8696
0.3381	0.8822	0.5756	0.8478

test set is the set which predict the fish after train and validation .Difference between test set and other two set is it must be unlabeled. Keras is expecting training dataset in numpy or list of numpy array. To create validation set need to pass the parameter to split and give a fraction that instruct keras to use as validation set from training set.

In Fig.16. below Feature of image during training epoch is shown

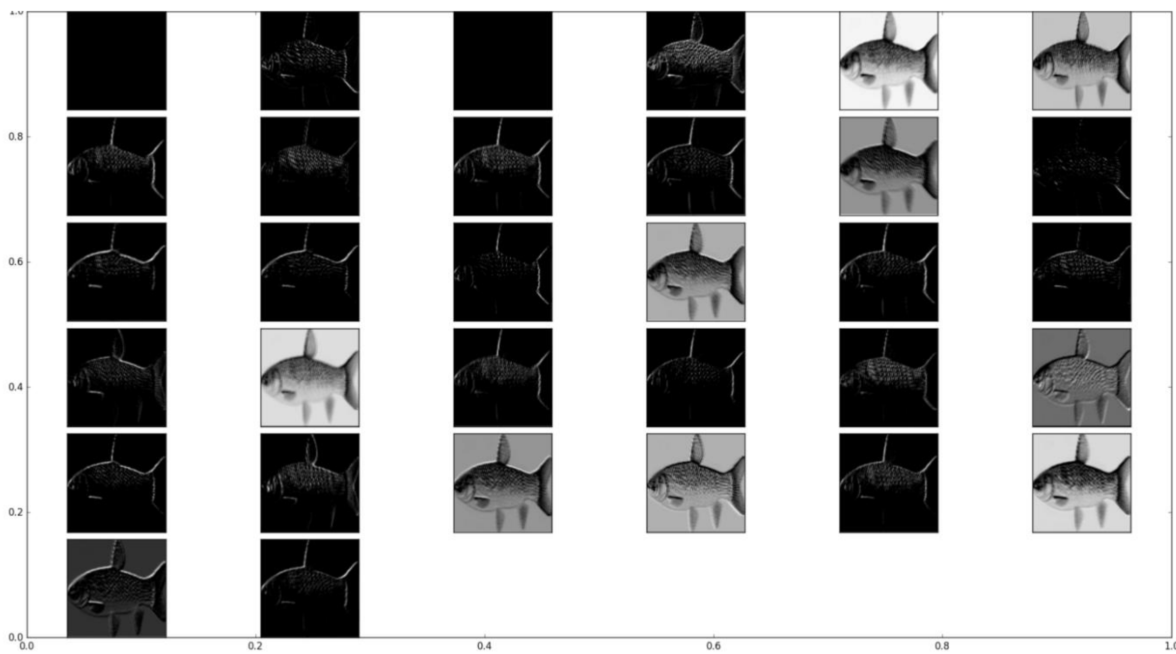


Fig.16.Feature of image during training epoch

Below Fig.17. describes the loss of training dataset

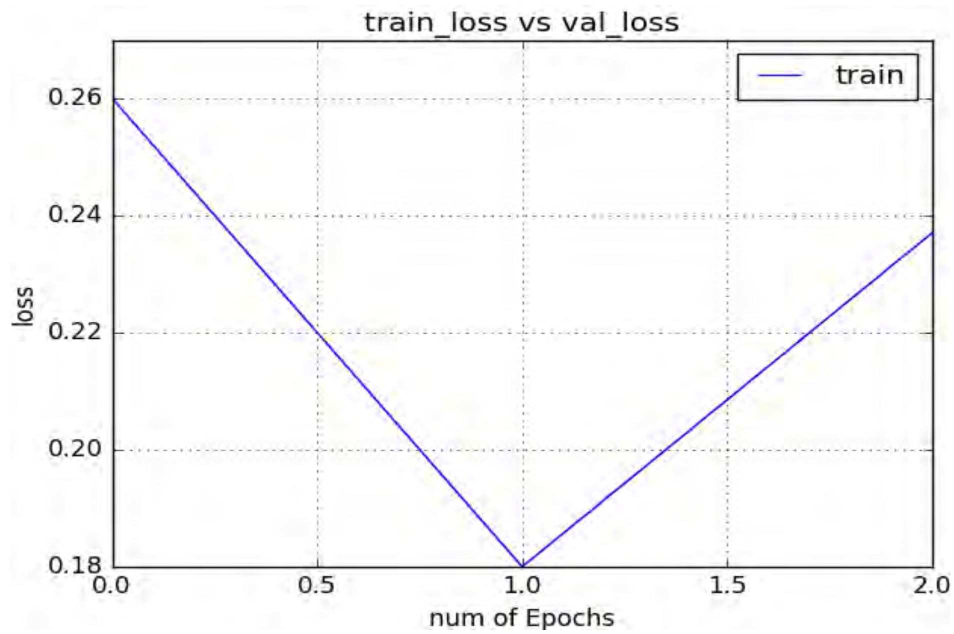


Fig.17. Loss of training Dataset

Below in Fig.18 describes the result of accuracy of train dataset against validation Dataset

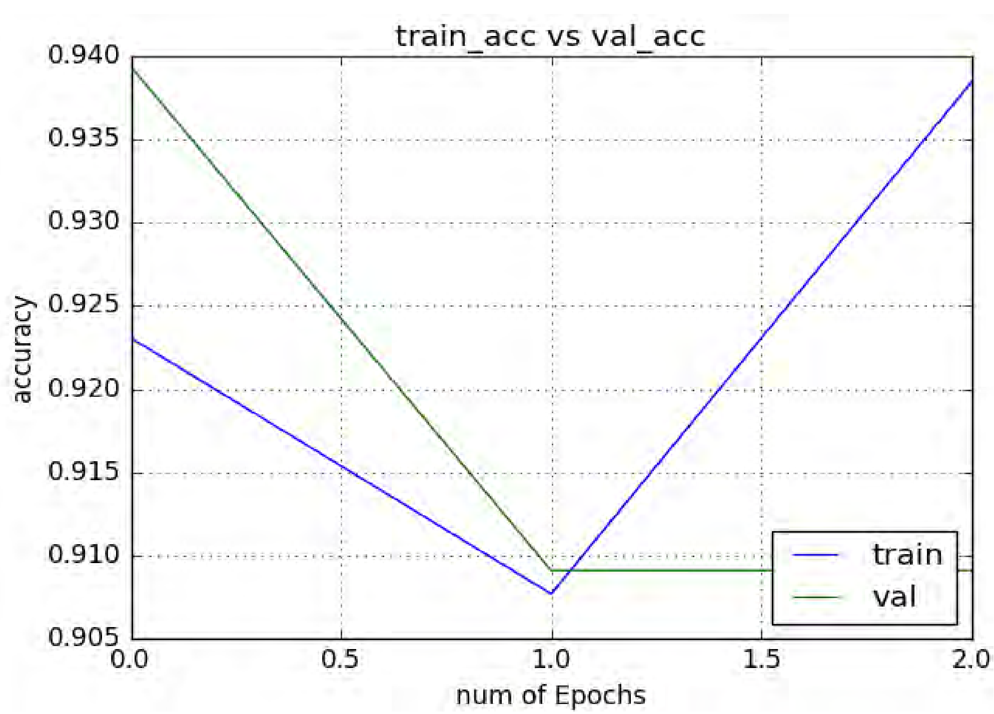


Fig.18. Train accuracy vs. Validation accuracy

Chapter 5

Conclusions and Future Works

5.1 Conclusion

We have given our best in this thesis to detect and recognize native Bangladeshi fishes using SURF and CNN (Convolutional Neural Network) to do so we had classified our custom dataset of 400 images. Methods like Gaussian Elimination, noise margin were used for image pre-processing and then binary masking, flood-filling of binary image and grass-fire algorithm was used for the detection of fishes. After that for recognition we used SURF algorithm at first and observed the accuracy later on we used CNN and noted its accuracy too. The accuracy of SURF is from 87.08 to 98.67% where grass crap fish species has the highest and shrimp has the lowest accuracy . At the period of CNN image classifier it gains 90.9% training accuracy and 23.7% loss. When a fish is known or gained from dataset SURF algorithm will work better, but when a testing image acquisition is inserted from a unknown source CNN image classifier will work better as it compares layer by layer.

5.2 Future Work

In our current system, we are detecting the object from a still image. We want to customize it a more and making it workable for real time image. In the real time image we can detect the fish when is it moving and making it a more viable option for marine scientist and researchers. It will be able to solve a lot of frustrating issue regarding this matter.

There are a lot of fishes around the world. They have lots of genus and species. The family of fishes is called shoal or school of fishes and it has so many species. In our current project we have detected 7 of our native available fishes in our country. It can be done by firstly including the all native fishes and then we will move on to including fishes around the world.

We are using around 500 sample data to detect fish. If we increase the sample data more we may able to detect the fishes more accurately and moreover if we change the algorithm of our current process to faster one we may get better outcome from it.

The current detection process is fast enough considering our data set of 500 images but when we will get more dataset and the data table will have thousands of value than there is a chance of being the process slower and that is the reason why there is a lot of work to be done to make the detection process faster.

Current approach is a fundamental approach to establish our target but there is a lot of chances to present this solution to the people in need by making a mobile application or computer software that can be run with laptop for portability.

References:

[1] Wikipedia contributors. (2018, July 23). Fish. In Wikipedia, The Free Encyclopedia. Retrieved from: https://en.wikipedia.org/w/index.php?title=Fish&oldid=851610198

[2] Andrés Hernández-Serna, Luz Fernanda Jiménez-Segura((Nov,2014)),Automatic identification of species with neural networks Chadimova, L. (2015). Creation of 3D models of chosen historical buildings for supporting knowledge transfer. 2015 Digital Heritage. doi:10.1109/digitalheritage.2015.7419617

[3] G. Singh and M. Sachan, "Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition," 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2014, pp. 1-5. doi: 10.1109/ICCIC.2014.7238334

[4] K. L. Goh, K. H. Lim, A. A. Gopalai and Y. Z. Chong, "Multilayer perceptron neural network classification for human vertical ground reaction forces," 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, 2014, pp. 536-540. doi: 10.1109/IECBES.2014.7047559

[5] Al-Omari, S.A.K., P. Sumari, S.A. Al-Taweel and A.J.N. Husain, 2009. Digital recognition using neural network. J. Comput. Sci., 5: 427-434.

[6] Bai, X., X. Yang and J.L. Latecki, 2008. Detection and recognition of contour parts based on shape similarity. Patt. Recog., 41: 2189-2199.

[7] Alsmadi, M.K.S., K.B. Omar, S.A. Noah and I.Almarashdeh, 2009. Fish recognition based on the combination between robust features selection, image segmentation and geometrical parameters techniques using artificial neural network and decision tree. Int. J. Comput. Sci. Inform. Secur.,6: 215-221.

[8] F. Storbecka and B. Daan, “Fish species recognition using computer vision and a neural network,” Fisheries Research, vol. 51, pp. 11–15, 2001.

[9] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher, "Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos," VISAPP (2), vol. 2008, pp. 514-519, 2008

[10] M.-C. Chuang, J.-N. Hwang, K. Williams, and R. Towler, "Automatic fish segmentation via double local thresholding for trawl-based

[11] M. Sonka., V. Hlavac. and R. Boyle., "Image pre-processing," in Image Processing, Analysis and Machine Vision., Boston, Springer, 1993, pp. 56-111

[12] EngineersGarage, "Introduction to Image Processing," 2012. [Online]. Available: https://www.engineersgarage.com/articles/image-processing-tutorial-applications.

[13] A.Kumar, "Image pre processing," 06 12 2012. [Online]. Retrieved From https://www.slideshare.net/ASHI14march/image-pre-processing. [Accessed 21 07 2018].

[14] D. Guo, X. Qu, X. Du, K. Wu and X. Chen, "Salt and Pepper Noise Removal with Noise Detection and a Patch-Based Sparse Representation," Advances in Multimedia, vol. 2014, pp. 1-14, 2014.

[15] I.Kutskir, "Fastest Gaussian Blur (in linear time)," [Online]. Available: http://blog.ivank.net/fastest-gaussian-blur.html. [Accessed 21 07 2018].

[16] Jason Corso, "Linear Filters and Image Processing," 24 09 2014. [Online]. Available: https://web.eecs.umich.edu/~jjcorso/t/598F14/files/lecture_0924_filtering.pdf. [Accessed 21 07 2018]

[17] N. Roman and J. Woodruff, "Ideal binary masking in reverberation," 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, 2012, pp. 629-633.

[18] S. Mischie, "On using time-frequency binary masking for dereverberation," International Symposium on Signals, Circuits and Systems ISSCS2013, Iasi, 2013, pp. 1-4. doi: 10.1109/ISSCS.2013.6651226

[19] E. M. Nosal, "Flood-fill algorithms used for passive acoustic detection and tracking," 2008 New Trends for Environmental Monitoring Using Passive Systems, Hyeres, French Riviera, 2008, pp. 1-5.doi: 10.1109/PASSIVE.2008.4786975

[20] R. Khudeev, "A new flood-fill algorithm for closed contour," 2005 Siberian Conference on Control and Communications, 2005, pp. 172-176. doi: 10.1109/SIBCON.2005.1611214

[21] F. Leymarie and M. D. Levine, "Simulating the grassfire transform using an active contour model," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 1, pp. 56-75, Jan 1992. doi: 10.1109/34.107013

[22] Xiaoyin Xu, Jie Cheng, R. M. Witt, B. L. Sabatini and S. T. C. Wong, "A shape analysis method to detect dendritic spine in 3D optical microscopy image," 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006., Arlington, VA, 2006, pp. 554-557.doi: 10.1109/ISBI.2006.1624976

[23] K. Nasreddine and A. Benzinou, "Shape-based fish recognition via shape space,"2015 23rd European Signal Processing Conference (EUSIPCO), Nice, 2015, pp. 145-149. doi: 10.1109/EUSIPCO.2015.7362362

[24] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013.

[25] L. Kang, J. Kumar, P. Ye, Y. Li and D. Doermann, "Convolutional Neural Networks for Document Image Classification," 2014 22nd International Conference on Pattern Recognition, Stockholm, 2014, pp. 3168-3172. doi: 10.1109/ICPR.2014.546

[26] D. P. Sudharshan and S. Raj, "Object recognition in images using convolutional neural network," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 2018, pp. 718-821.doi: 10.1109/ICISC.2018.8398912

[27] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013.

[28] L. Kang, J. Kumar, P. Ye, Y. Li and D. Doermann, "Convolutional Neural Networks for Document Image Classification," 2014 22nd International Conference on Pattern Recognition, Stockholm, 2014, pp. 3168-3172. doi: 10.1109/ICPR.2014.546

[29] M. Muthugnanambika and S. Padmavathi, "Feature detection for color images using SURF," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017, pp. 1-4. doi: 10.1109/ICACCS.2017.8014572

[30] Pedersen, J.T. (2011). Study group SURF: Feature detection & description. Retrieved from: http://cs.au.dk/~jtp/SURF/report.pdf

[31] Wikipedia, “Blob detection — Wikipedia, the free encyclopedia,” 2011, [Online; accessed 14-July2010]. [Online].Available: https://secure.wikimedia.org/wikipedia/en/wiki/Blob detection

[32] M. B. David Lowe, “Invariant features from interest point groups,” BMVC, 2002.

[33] T. Sledevič and A. Serackis, "SURF algorithm implementation on FPGA," 2012 13th Biennial Baltic Electronics Conference, Tallinn, 2012, pp. 291-294. doi: 10.1109/BEC.2012.637